# Exploring the Diversity of Dwarf Galaxy Populations with Hydrodynamical Simulations of Targeted Cosmological Environments

by

Jemima Mae Briggs

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

February 2025

# Declaration

The work presented in this thesis was carried out at the Astrophysics Research Institute, Liverpool John Moores University. Unless otherwise stated, it is the original work of the author.

While registered as a candidate for the degree of Doctor of Philosophy, for which submission is now made, the author has not been registered as a candidate for any other award. This thesis has not been submitted in whole, or in part, for any other degree.

Jemima Mae Briggs
Astrophysics Research Institute
Liverpool John Moores University
IC2, Liverpool Science Park
146 Brownlow Hill
Liverpool
L3 5RF
UK

February 2025

# Abstract

The standard Λ-Cold Dark Matter (CDM) model of cosmology has provided accurate predictions and explanations for the large-scale structure of the Universe and its evolution with time, however, on length and halo mass scales below $\sim$1Mpc and $\sim 10^{11}\mathrm{M}_\odot$ discrepancies between theory and observation have propelled the study of the low mass regime to the forefront of cosmology and astrophysics (see e.g. Bullock & Boylan-Kolchin, 2017a). Dwarf galaxies provide unique insight to galaxy formation and evolution: their early formation times and shallow gravitational potentials make them equally sensitive to internal astrophysical processes, such as feedback from young stars and supernovae (SNe), and external environmental influences, such as the ram pressure stripping of gas and the tidal stripping of stars. Furthermore, isolated dwarf galaxies are shielded from the extreme environmental effects that occur in high density environments and provide interesting 'laboratories' for examining galaxy growth in the ΛCDM framework, in particular, how low mass galaxies occupy haloes and how their large scale environment influences key properties.

I introduce the Columba suite of simulations of galaxy formation in the ΛCDM cosmogony, which are designed primarily to examine the influence of large-scale cosmic environment on the evolution of isolated dwarf galaxies, i.e. those not in close proximity to larger structures at the present day. I focus on this subset of dwarf galaxies in order to explore the influence of large-scale overdensity and the intrinsic properties of their host DM haloes on galaxy evolution in the low-mass regime, in the absence of direct environmental influences driven by tidal and ram pressure forces exerted by near neighbours. I describe the selection of the zoom-in regions of the Columba suite which are selected from a $L = 400$Mpc parent simulation and span a range of underdense environments, from voids to filamentary structures, and host a diverse population of dwarf galaxies. The Columba simulations incorporate a galaxy formation model that represents an evolution of that used for EAGLE which explicitly follows the cold, dense, neutral gas of the ISM, adopts a star formation model that accounts for turbulence, a treatment of pre-supernova feedback from massive stars, more sophisticated methods of distributing the energy liberated by SNe and active galactic nuclei (AGN), and updated treatments of BH growth and dynamics.

The Columba simulations give an insight into the formation and evolution of isolated dwarf galaxies in diverse cosmic environments. I utilise the simulations to examine the effect of large scale environment on key properties of field dwarf galaxies quantified by their spherical overdensity in $r = 5$Mpc. While the amplitudes of the halo mass functions (HMF) and stellar mass functions (SMF) exhibit a dependence on the region's density, their slopes, when normalized to the cosmic mean, do not show an environmental dependence. Similarly, the stellar mass halo mass (SMHM) relation does not depend on the region's environment, maintaining consistent slopes despite density variations. The luminous fraction of galaxies is not significantly affected by their large scale environment but show a correlation with halo concentration. Halo concentration has been shown to be strongly indicative of a halo's formation time and studies have shown that the scatter in the SMHM relation is strongly correlated with the concentration of their DM halo. I further investigate the influence of halo concentration on properties of isolated field dwarf galaxies. At a fixed mass, high concentration haloes host larger dwarf galaxies, quantified by a peak Spearman rank correlation coefficient of $\rho = 0.85$. This effect propagates through their cumulative star formation histories which show that high concentration, and hence earlier forming, haloes assemble their stellar mass earlier across the full mass range examined, and is reflected by shorter $t_{90}$, the time it takes for 90% of the total stellar mass to assemble. The central galaxies in the Columba suite show a diversity in morphology from disc to elliptical structures, and though halo concentration shows a strong influence on the evolution of isolated dwarf galaxies, it does not demonstrate a strong correlation with their kinematic morphology.

Jemima Mae Briggs February 2025

# Acknowledgements

First and foremost, I would like to thank my supervisors, Rob Crain and Azi Fattahi. You have both supported me through every up and down of my PhD journey and I never would have reached this point without your encouragement and guidance. Your kindness, patience, and dedication have been unwavering. Thank you for being such wonderful mentors and for always believing in me, it has been a privilege to work with you!

The Astrophysics Research Institute has been an incredible place to work with such a welcoming and friendly atmosphere. I am immensely grateful to everyone here, particularly the HPC and cosmology group, for the engaging journal clubs and thought provoking discussions. To Rob's unofficial gossip group: Ryan, Chris, and Anna - I wish you all success, good luck! A huge thanks to the SWIFT and Leiden teams for their help, Matthieu Schaller, Joop Schaye, and Evgenii Chaikin, this project would not have been possible without your expertise. Thank you also to everyone on the Prospero team for your technical support.

Office 3.18 has undoubtedly been the best office to work in over the past couple of years. After the challenges of lockdown, sharing this space with all of you - past, present, and honorary - has been a true joy. Flo, Phoebe, Tutku, Andrew, Ryan, Em, Sara, Elena, Kyle, Adrian, Sarah, Shobhit, and Bethan, you are all incredible people. I will dearly miss our coven horoscope readings, weekly pub quizzes, and our coffee trips. Flo, you've become one of my dearest friends and these past few years wouldn't have been the same without you, we've grown so much together and you are the sister I never had.

Erin, Fran, and Philly, you are my found family and the constants in my life. I cannot imagine having done this without you all. You have lifted me out of my darkest times and I am so grateful to have you in my life. Chloe and Nona, thank you for making my undergraduate degree so enjoyable, I would never have made it to this point without you and I miss you both in Liverpool dearly!

Lastly, to my brother, my parents, and of course my cats and dogs who have made the occasional Zoom cameo. Max, your big sister is incredibly proud of you, thank you for all the unofficial tech support and your candid advice, I will do my best to avoid any more coffee spills on my laptop! Mum, you are the most inspiring person I know. Your resilience and strength have shaped who I am today, we have weathered so much together in these past few years and I couldn't have done this without you. Thank you.

*“Strange about learning; the farther I go the more I see that I never knew even existed. A short while ago I foolishly thought I could learn everything - all the knowledge in the world. Now I hope only to be able to know of its existence, and to understand one grain of it. Is there time?”*

Daniel Keyes, Flowers for Algernon

# Contents

# List of Figures

# List of Tables

# 1

# Introduction

## 1.1 Cosmological Background

### 1.1.1 Expansion of the Universe

A key finding for cosmology arose from observations of distant galaxies by Hubble (1929). These observations showed that the light from distant galaxies (originally identified as extra-galactic nebulae) is shifted to longer, redder, wavelengths, commonly known as red-shifted, meaning the objects are moving away from the observer. The red-shift $z$ of light is quantified by the relative difference between the observed $\lambda_0$ and emitted $\lambda_e$ wavelengths,

$$z = \frac{\lambda_0 - \lambda_\mathrm{e}}{\lambda_\mathrm{e}} \tag{1.1}$$

Specifically, Hubble found that the recessional velocity, $v$, of a galaxy is proportional to the proper distance, $D$, such that galaxies further away are receding at higher velocities. This had significant implications for our understanding of the Universe, presenting evidence against the commonly held theory of the time that on large scales the Universe is static (Hoyle, 1948). In particular, Hubble showed that the Universe is expanding and dynamic, and opened up the idea that our Universe has evolved from an inception point and will potentially reach its end. Hubble's law,

$$v = H_0 D \tag{1.2}$$

quantifies the current rate of expansion of the Universe at present day, z=0, by the Hubble constant $H_0$.

In an expanding Universe the physical distance to an object increases over time, it is therefore useful to define a coordinate system in which the distance between two points is independent of time. These are known as comoving coordinates, $\vec{x}(t)$, and are related to the proper, physical distance, $\vec{r}$, through a dimensionless quantity known as the cosmic scale factor, $a(t)$,

$$\vec{r}(t) = a(t)\vec{x}(t) \tag{1.3}$$

The cosmic scale factor is related to the red-shift, $z$, following,

$$a = \frac{1}{1+z} \tag{1.4}$$

This scale factor increases as the Universe expands and is normalised to unity at present day. Within this paradigm, the Hubble constant is given the general form with respect to the scale factor $a$, and can be applied to any given cosmology,

$$H = \frac{\dot{a}}{a} \tag{1.5}$$

The evolution of the Hubble constant with the expansion of the Universe can also be found through Einstein's theory of relativity. The Friedmann–Lemaître- Robertson–Walker (FLRW) metric is a metric based on the exact solution of the Einstein field equations of general relativity (Friedmann, 1922; Lemaître, 1927). The FLRW metric invokes the Cosmological principle which assumes that on large scales, the Universe is homogeneous and isotropic. A homogeneous Universe means that there is no preferred location throughout space and an isotropic Universe means that there is no dependence on direction for the distribution of matter and radiation.

Friedmann found that Einstein's field equation could be reduced to the following solutions: Friedmann's force Equation which defines the change in velocity with the expansion of the Universe,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) - \frac{\Lambda c^2}{3} \tag{1.6}$$

where $p$ and $\rho$ are the density and pressure of the matter-radiation fluid comprising the Universe. The constants $G$ and $c$ denote the gravitational constant and the speed of light respectively and $\Lambda$ is the cosmological constant.

The equation of state of the fluid can be written as $p = w\rho^2$, where $w$=1/3 for a fluid that is radiation dominated and $w$=1 for a matter dominated fluid. For an isotropic Universe, as instated by the Cosmological principle, where there are no heat inflows or outflows, the expansion is adiabatic and the energy Friedmann equation can be derived,

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{1.7}$$

The curvature of spacetime is represented by $k$ where $k$ = +1, 0, or -1 depending on whether the shape of the universe is assumed to be spherical, flat (Euclidean space) or hyperbolic and thus the energy Friedmann equation yields the time evolution, geometry, and hence fate of the Universe as a function of the fluid density. The term $\Lambda$ was added to the field equations by Einstein to achieve a static Universe.

Under the assumption that the shape of the Universe is flat, with zero curvature ($k$=0), it follows that there is a critical density for a given expansion rate that will bring the expansion asymptotically to a halt. This critical density, $\rho_{\rm crit}$ is expressed as,

$$\rho_{\rm crit} = \frac{3H^2}{8\pi G} \tag{1.8}$$

and from this it is standard to define a density parameter, $\Omega$, given by the ratio between the actual density and the critical density of the Universe,

$$\Omega = \frac{\rho}{\rho_{\rm crit}} \tag{1.9}$$

### 1.1.2 The Big Bang and Inflation

One of the most significant implications of the Friedmann equations is the origin of the Universe; a point in time when $a \to 0$ with an initial state of high density and temperature, now commonly known as the Big Bang singularity. The Big Bang model explains an extensive range of observed phenomena, including the existence and extremely smooth thermal spectrum of the Cosmic Microwave Background (CMB) radiation (Penzias & Wilson, 1965), the abundance of light elements (Olive et al., 2000), and the expansion of the Universe implied by Hubble's law (Hubble, 1929).

Despite its major successes, the BB model by itself has several outstanding problems. The flatness problem points out the fine-tuning required to maintain a flat Universe with a density close to the critical density ($\Omega = 1$). Small deviations from this required energy density will propagate through time and the Universe would rapidly evolve into a different state at present time. At the Planck time, the density is required to be within 1 part in $10^{57}$ of the critical density and the BB model has no explanation for this fine-tuning of the initial conditions of the Universe (Guth, 1981).

Another significant challenge is known as the horizon problem. Observations show that the CMB is isotropic and this uniformity implies that at the surface of last scattering, radiation was isotropic, homogeneous, and in thermal equilibrium. However, the Universe consists of regions that were not in causal contact before the last scattering surface; light had not yet had time to travel between regions and hence no information could be exchanged. Therefore, the BB model cannot explain the isotropy of the CMB on wider scales without a means of violating causality.

A possible solution to these problems was proposed by Guth (1981) by invoking a period in the early Universe when the Universe experienced rapid expansion or 'inflation'. If the Cosmological constant is sufficiently large to dominate the dynamics of the early Universe, it can expand exponentially, rapidly inflating small sizes ($R \propto e^{\sqrt{\frac{\Lambda}{3}}t}$). The inflationary model is able to solve the flatness problem as the rapid expansion of space makes any curvature negligible ($k \rightarrow 0$) on finite scales. This also increases the size of the causal horizon beyond the observable Universe meaning that regions were in causal contact prior to inflation and could attain uniform temperature.

The BB model paired with inflation is now the most widely accepted theory for the early evolution of the Universe. Immediately after the BB, the Universe existed at an extremely high temperature and density with baryons strongly coupled with photons, under these conditions all matter was ionised. After inflation the Universe continued to expand with decreasing density and temperature, after approximately $380,000$ years ($z \approx 1100$) the Universe was cool enough that electrons and nuclei were able to form neutral atoms, known as the era of recombination. Recombination decoupled photons and allowed them to propagate freely through space which we can observe as the Cosmic Microwave Background (CMB) radiation (Penzias & Wilson, 1965).

### 1.1.3 Dark Matter

There are several strong pieces of evidence for the contribution to the matter budget from an invisible and, as of yet, undetected, dark matter. One such piece of evidence comes from the disparity in inferred total matter and observed luminous matter of objects. For example, the gas and stars in the disk of spiral galaxies orbit around the centre of the galaxy, assuming Newtonian gravity the rotational velocity $v(r)$ can be approximated by,

$$v(r) = \sqrt{\frac{GM(r)}{r}} \tag{1.10}$$

where $M(r)$ is the total mass within radius $r$. For the observed stars and gas in spiral galaxies, the rotation curve is expected to decrease as it approaches larger radii, however, measured rotation curves exhibit an increase in circular velocity at small radii which plateaus to a constant velocity at larger radii (Rubin & Ford, 1970; Rubin et al., 1980). The rotation curve at larger radii is higher than accounted for by the luminous, observed matter, and hence implies the presence of an unseen mass.

The mass of an object can also be inferred through a phenomenon called gravitational lensing. General relativity dictates that the gravitational potential of massive objects, like galaxy clusters, produces a curvature in spacetime which bends the path of incoming light with the massive object acting as a lens. Strong gravitational lensing can be observed where light from distant objects is distorted, often into arcs or 'Einstein rings' (King et al., 1998). Weak gravitational lensing describes smaller distortions of the background source that can only be detected through statistical analysis of large galaxy surveys. Observations of gravitational lensing allow us to probe the matter distribution, and quantify the mass, of the intervening object. Using this method, the mass-to-light ratio of galaxy clusters can be measured and high ratios have indicated the presence of a mass that does not contribute to the observed light and is hence non-luminous and 'dark'.

Compelling evidence can also be found from observations of the Bullet cluster, a collision of two clusters of galaxies (Clowe et al., 2004). X-ray imaging shows the hot gas component which is slowed by drag forces during the collision, and weak gravitational lensing maps reveal the gravitational mass distribution, which aligns with the galaxies rather than the gas. This causes a significant offset between the centre of mass of the total system and the baryonic centre of mass which is dominated by the gas, suggesting that the dominant matter component is unseen and collisionless. These observations pose a significant challenge to alternative arguments to dark matter, like modifying gravity (e.g. Modified Newtonian Dynamics (MOND) (Milgrom, 1983)), but these observations can be explained in a ΛCDM framework where the dark matter haloes of the colliding clusters pass through each other without interacting whilst the hot gas in each cluster is slowed by the collision.

Evidence for the existence of dark matter led to the first numerical simulations following the non-linear gravitational evolution of the structure of the Universe. These early models were critical in constraining the nature of DM; comparing theoretical predictions to the observed large scale clustering of galaxies and anisotropies of the cosmic microwave background led to the leading theory that dark matter must be 'cold' instead of 'hot' (Peebles, 1982; Blumenthal et al., 1984; Davis et al., 1985).

### 1.1.4 ΛCDM

The current standard cosmological model of the Universe is the ΛCDM (Cold Dark Matter) model which can be defined by three major components: the cosmological constant of dark energy, cold dark matter and baryons. Λ denotes a positive Cosmological constant linked to the accelerating expansion of the Universe often referred to as 'dark energy' and cold dark matter refers to a form of dark matter that is non-relativistic, dissipationless, collisionless, and whose interactions with baryonic matter are dominated by gravity. This framework is defined by a set of independent cosmological parameters, with four related to the matter content of the Universe. The total density parameter, $\Omega_0$, is defined as the ratio of the present day mean density and the critical density of the Universe, where the critical density is defined as the density required for the Universe to halt expansion after an infinite amount of time, below which the Universe will continue expanding forever. In a flat Universe ($\Omega_\mathrm{k}$=0), $\Omega_0$, defined by the sum of contributions from matter, radiation, and energy is equal to 1,

$$\Omega_0 = \Omega_{\mathrm{m},0} + \Omega_{\Lambda,0} + \Omega_{\mathrm{rel},0} = 1 \tag{1.11}$$

At present day $\Omega_{\mathrm{rel},0}$, the parameter that accounts for the cosmic microwave and neutrino background that dominated the early Universe, forms a negligible contribution. The matter density parameter can be further defined by its contributions from cold dark matter $\Omega_{\mathrm{c},0}$ and baryons $\Omega_{\mathrm{b},0}$ as,

$$\Omega_{\mathrm{m},0} = \Omega_{\mathrm{b},0} + \Omega_{\mathrm{c},0} \tag{1.12}$$

## 1.2 Galaxy Formation

### 1.2.1 Dark Matter Haloes

In ΛCDM the structure of the Universe is formed from the gravitational collapse of DM seeded by the density fluctuations from inflation. DM gravitationally collapses in overdensities and bound structures known as DM haloes; haloes are able to grow by accreting dark matter and merging with other structures. This bottom-up structure growth is known as hierarchical clustering (Peebles, 1980) and produces the clustering of more massive objects, like galaxies and galaxy clusters, that we observe today. Our galaxy, the Milky-Way, resides in the Local Group which in turn resides in the Virgo Supercluster comprised of over 1000 galaxies. Hierarchical clustering shapes the Universe on the largest scales and gives rise to the cosmic web (Bond et al., 1996): a vast network of interconnecting walls and filaments of primarily dark matter and intergalactic gas, nodes of galaxies and clusters where filaments intersect, separated by empty voids. The structure of the cosmic web is reproduced in cosmological simulations under ΛCDM cosmology, shown in 1.1 is the surface density of CDM in the FLAMINGO simulation (Schaye et al., 2023).

The density profile of a DM halo, $\rho(r)$, can be well described by a double power law known as the Navarro-Frenk-White (NFW) profile (Navarro et al., 1996b),

$$\rho(r) = \rho_{\mathrm{crit}} \frac{\delta_c}{\frac{r}{R_s}\left(1 + \frac{r}{R_s}^2\right)} \tag{1.13}$$

which is characterised by a dimensionless characteristic density $\delta_c$ and a scale radius $R_s$. The scale radius can be defined by $R_s = r_{200}/c$, where $c$ is a dimensionless halo concentration parameter which depends strongly on halo mass and redshift (Navarro et al., 1996b; Bullock et al., 2001). As a direct consequence of the hierarchical structure formation in CDM, the concentration of DM haloes is anti-correlated with their mass

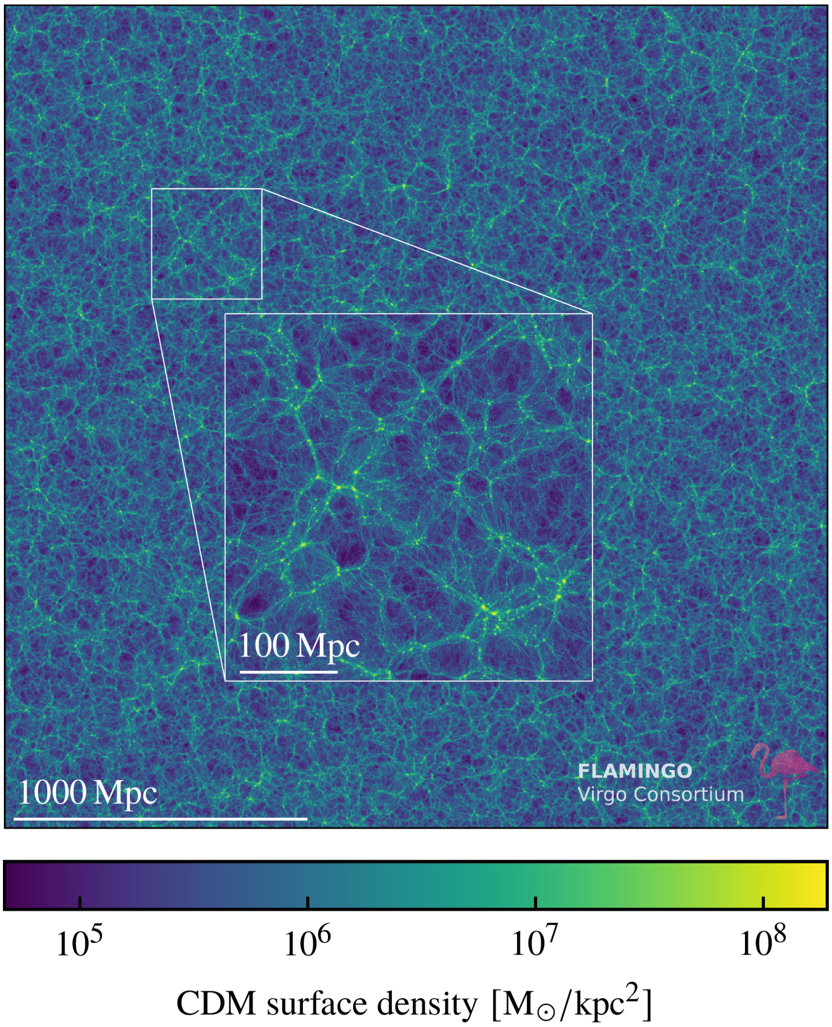


Figure 1.1: A projection of the large scale structure of the cosmic web coloured by the CDM surface density in a 20Mpc thick slice through the fiducial FLAMINGO ($L$ =2.8Gpc) run at z=0 from Figure 4 by Schaye et al. (2023). Filaments intersect at dense nodes and are separated by low density voids.

(e.g. Ludlow et al., 2014) with low mass haloes having higher concentration. The characteristic density of a halo is proportional to the density of the universe at its initial time of collapse time and hence the higher concentrations of low mass haloes reflect their collapse at higher densities and higher redshift. This strong link has shown that concentration can be used to infer the formation time of a halo (Neto et al., 2007). Though the NFW profile is often considered a universal density profile, it has been shown that halo profiles are not entirely self-similar and other profiles have been preferred such as the Einasto (Einasto, 1965) or Burkert (Burkert, 1995) profiles.

Dark matter haloes are fundamental in the formation and evolution of galaxies; it is within the early gravitational potentials of DM haloes that baryonic matter accumulates and can produce the conditions for star formation. The filaments of the cosmic web that arise from hierarchical clustering provide channels for the accretion of new gas, further star formation, and enable the growth of galaxies.

### 1.2.2 Gas and Stars

#### 1.2.2.1 Star Formation and Evolution

Though CDM dominates the matter density of the Universe, baryons make up the gas and stars, and hence the luminous component of the Universe that we see. Unlike CDM, gas is able to dissipate energy through processes such as radiative cooling (Dalgarno & McCray, 1972; Sutherland & Dopita, 1993). These radiative processes include thermal Bremsstrahlung (free-free emission), recombination, and atomic line emission.

The dissipative nature of gas is essential for star formation which occurs in cold, dense gas. Gas entering the potential well of a DM halo can be shocked to the virial temperature of the halo, is able to dissipate energy through various cooling mechanisms, and condense to the centre of the halo reaching the high density required for star formation. The dominant cooling mechanism in the CGM depends upon the density and temperature of the gas, Bremsstrahlung cooling operates at temperatures above $T \gtrsim 10^7$K, metal-line cooling occurs due to the collisional excitation of metals and subsequent cooling at temperatures between $10^7 \gtrsim T \gtrsim 10^5$K, and Hydrogen cooling at $T \sim 10^4$K. When clouds of gas cool and lose their thermal pressure support, they collapse under their own gravity, enabling the formation of stars. Mathematically, gravitational collapse occurs when the mass of the cloud exceeds the Jeans mass $M_\mathrm{J}$,

$$M_\mathrm{J} = \left(\frac{5k_\mathrm{B}T}{Gm}\right)^{3/2}\left(\frac{3}{4\pi\rho}\right)^{1/2} \tag{1.14}$$

where $T$ is the average temperature, $m$ is the mean particle mass, $k_\mathrm{B}$ is the Boltzmann constant, and $G$ is the gravitational constant. The dependence of the Jeans mass on $\rho^{1/2}$ means that as the cloud collapses and its density increases, the Jeans mass decreases and fragmentation, the process by which larger gas clouds break into smaller, denser clumps, occurs. Once the fragments reach a high enough density and exceed the temperature threshold for hydrogen fusion ($T > 10^7$K), a zero-age main sequence (MS) star is born.

The distribution of masses in a stellar population at the time of their formation is described by the stellar initial mass function (IMF) and in general is dominated by a high number of low mass ($\sim 1\mathrm{M}_\odot$) stars and few massive ($\sim 100\mathrm{M}_\odot$) stars. One of the earliest and simplest forms, the Salpeter (Salpeter, 1955) IMF ($\xi(M)$), is described by a single power-law,

$$\xi(M) = \xi_0 M^{-\alpha} \tag{1.15}$$

with a characteristic slope of $\alpha = -2.35$ and normalisation constant, $\xi_0$, and is valid for masses $M > 0.5\mathrm{M}_\odot$. Other forms of the IMF adopt broken power law with a shallower slope at the low mass end below $\sim 1\mathrm{M}_\odot$ (Kroupa, 2001; Chabrier, 2003) and it is debated whether the IMF can be considered universal (Bastian et al., 2010).

All stars begin their life burning Hydrogen on the main sequence, and their time spent in a stable state of hydrostatic and thermal equilibrium is determined by their mass, as more massive stars burn through their Hydrogen more quickly. These massive stars evolve rapidly, reaching temperatures that allow further fusion and the production of heavier elements up to iron, and subsequently lose mass through strong stellar winds and core-collapse supernovae at the end of their life. These mass-loss mechanisms disperse these elements into, and chemically enrich, their surrounding gas. Lower mass stars are longer lived; they stay on the main sequence longer until they reach the Asymptotic Giant Branch (AGB) evolutionary stage where they enrich their surroundings through stellar winds. The white dwarf that remains after the AGB phase can undergo thermonuclear explosion as a type Ia supernova following the mass accretion from a companion star in a binary system or a merger with another white dwarf. Type Ia supernovae are a significant source of iron enrichment to the interstellar medium (ISM) (Nomoto et al., 2013).

#### 1.2.2.2 Reionisation

After recombination, the Universe was comprised of neutral Hydrogen and Helium, however, observations of the present-day Universe show that the intergalactic medium (IGM) is mostly ionised suggesting that there was a second phase transition of the early Universe. This is known as the epoch of reionisation; as the first luminous objects formed, the energy emitted by these sources was able to ionise their surrounding IGM and Hydrogen transitioned from neutral to ionised. The sources responsible for reionisation and the precise time it took place are uncertain with ongoing debates around the contribution from quasars, bright galaxies, and faint galaxies (Robertson et al., 2010). Dwarf galaxies have been proposed as a significant source of ionising photons during the epoch of reionisation, however constraining the number of these faint sources in the early Universe is challenging.

### 1.2.3 Galaxies

As established, the gas fuelling star formation exists in clouds of molecular gas in the interstellar medium (ISM) of galaxies. The ISM is accreted from the intergalactic medium (IGM) between galaxies that is primarily composed of Hydrogen and Helium. Galaxies accrete gas from the IGM for star formation via two modes: the cold mode provides cool, dense gas from the IGM typically through inflows along filaments and the hot mode involves the accretion of gas that has been shock-heated to high temperatures before cooling through hydrostatic equlibrium (Katz et al., 2003; Kereš et al., 2005). The cold mode dominates accretion at high redshift and the hot mode accretion is more prominent at late time in more massive galaxies ($M_{\rm halo} \sim 10^{13}{\rm M}_{\odot}$) (van de Voort et al., 2011).

The star formation rate (SFR) of a galaxy can be directly inferred from the surface density of gas in the ISM from the Kennicutt-Schmidt relation (Kennicutt, 1998). The average galaxy SFR density in the Universe peaks at z∼2 (Madau & Dickinson, 2014) and present day observations show both galaxies with ongoing star formation and galaxies with no young stars. Galaxies that are undergoing active star formation are characterised by their blue colour from the presence of young, hot, massive stars. Galaxies that show no recent star formation are said to be quenched and are characterised by their red colour from older stellar populations. This dichotomy can be seen by the red sequence and blue cloud galaxy populations on the $u - r$ colour - magnitude plane (Baldry et al., 2004). The colour of the galaxy is commonly associated with their morphology; spiral galaxies are typically blue, with an abundance of gas and young stars, whereas elliptical galaxies are older, have little to no recent star formation, and are hence redder.

These characterisations, and the evolution of a galaxy, can also be linked to their environment. Galaxies in dense environments like galaxy clusters are subject to interactions that can strip their gas content through, for example, ram pressure stripping (Gunn

& Gott, 1972) and tidal stripping (Read et al., 2006), and subsequently quench star formation. Isolated galaxies in less dense environments can typically retain their gas content for longer and continue forming stars (Geha et al., 2012).

Galaxies require mechanisms to regulate their growth and balance the accretion of gas and star formation. The stellar mass - halo mass relation shown in Figure 1.2 (reproduced from Behroozi et al., 2019) demonstrates that the efficiency of galaxy formation peaks at halo masses of $M_{\rm halo} \sim 10^{12}$ with declining efficiency towards lower and higher stellar masses. In these regimes galaxy growth is regulated by feedback: in the low mass ($M_{\rm halo} < 10^{12}{\rm M}_{\odot}$) regime stellar feedback is the dominant mechanism and in the high mass regime ($M_{\rm halo} > 10^{12}{\rm M}_{\odot}$) feedback from Active Galactic Nuclei (AGN) dominates.

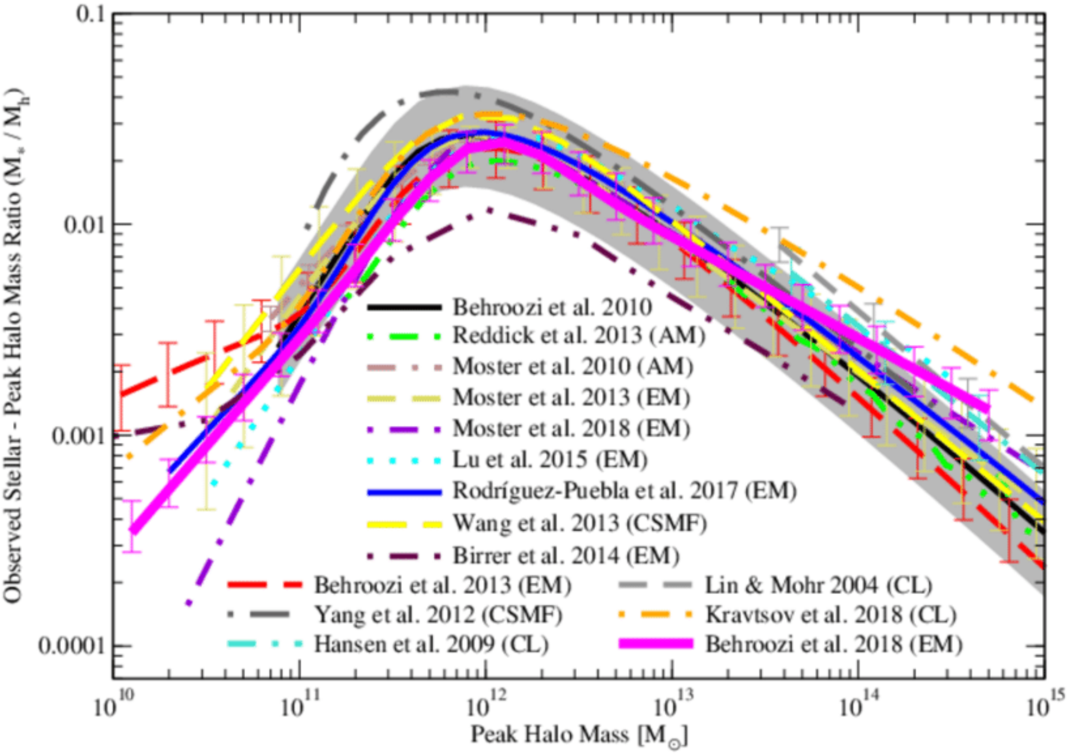


FIGURE 1.2: The present-day stellar mass - halo mass relation from a compilation of results from empirical modelling (EM), abundance matching (AM), Conditional Stellar Mass Function (CSMF) modelling, and cluster X-ray mass measurements (CL). The galaxy formation efficiency exhibits a peak at $M_{\rm halo} \sim 10^{12}{\rm M}_{\odot}$ with declining efficiency towards lower and higher halo mass.

Stellar feedback refers to the various processes through which stars influence the surrounding ISM, these include supernovae (SN), radiation pressure, stellar winds, and photoionising radiation. SN are the dominant form of stellar feedback whilst the latter three mechanisms are significant before the final stages of stellar evolution. Radiation pressure refers to the force exerted on matter by the radiation emitted by stars. Stars emit photons which inject momentum into the surrounding gas and dust through direct radiation pressure which can disrupt molecular clouds and disperse gas. Stellar winds are most important for the evolution of massive stars, these winds carry energy, mass, and momentum (Cassinelli, 1979) that is able to drive turbulence in the ISM and disperse and heat the surrounding gas. Photoionising radiation refers to the high energy photons emitted by massive stars that form HII regions; these photons exceed the energy required to ionise neutral Hydrogen and form bubbles of hot, ionised gas from the ISM surrounding massive stars. Finally, core collapse supernovae (CCSNe) occur when massive stars reach the end of their life and release $\sim 10^{51}$erg of energy, they impact star formation through a combination of gas expulsion, heating the ISM, driving turbulence, and creating outflows. These processes disrupt the conditions necessary for gas to cool and collapse into new stars, effectively halting star formation in the regions affected by the supernova explosions.

While feedback from star formation is sufficient to decrease the galaxy formation efficiency at the low mass end of the SMHM relation, it is unable to regulate the growth of more massive galaxies. Following the hierarchical framework, it is logical to assume that the most massive galaxies experience prolonged, and significant late time, stellar mass assembly, however, observations indicate that massive galaxies form the majority of their stars at high redshift ($z > 1$), this phenomenon is known as 'down-sizing' (Cowie et al., 1996) and requires a mechanism to quench the most massive galaxies at early times, which is now commonly accepted as AGN feedback. Supermassive black holes (SMBHs) exist in the centres of galaxies spanning a wide range of masses. The mass of the SMBH closely follows scaling laws with properties of their host galaxies, including the bulge luminosity and mass (Kormendy & Ho, 2013). SMBHs can grow through mergers and the accretion of material, and the radiated energy from this is capable of driving large outflows (Fabian, 2012), which are observed at both high redshift and present day. In addition to the removal of gas from the galaxy via outflows, AGN prevent gas in and around galaxies from cooling, suppressing star formation and playing a key role in galaxy evolution.

## 1.3 The Low Mass Regime

### 1.3.1 Challenges to LCDM

The currently preferred ΛCDM cosmological model has provided successful predictions for, and explanations of, the properties, structure, and evolution of the large-scale Universe. However, on length and halo mass scales below ∼1Mpc and $\sim 10^{11}\mathrm{M}_{\odot}$, this cosmological framework has encountered challenges that have arisen from discrepancies between theory and observations (see e.g. Bullock & Boylan-Kolchin (2017a)). The classic challenges to ΛCDM identified from small-scale numerical predictions are as follows:

**Core-Cusp Problem:** The core-cusp problem arises from the discrepancy between the density profiles of haloes inferred from observations of low mass galaxies and those measured in collisionless ΛCDM numerical simulations (Moore, 1994; Flores & Primack, 1994). The DM distribution of haloes within N-body simulations can be quantified by their density profile and commonly modelled using an NFW (Navarro et al., 1996c) profile with an inner slope following $\rho \propto r^{-1}$. Such haloes were termed to have 'cuspy' profiles owing to their steep rise in density towards their centres. However, this prediction is found to be in conflict with the density profiles inferred from the observed rotation curves of some dwarf galaxies. Many low mass galaxies have observed rotation curves that exhibit a shallower rotation curve that is more consistent with a 'cored' or constant inner density profile, for example, a Burkert profile (Burkert, 1995).

**Missing Satellite Problem:** The missing satellite problem refers to the overabundance of sub-haloes in collisionless ΛCDM numerical simulations compared to the observed number of satellite galaxies in the Local Group. ΛCDM DM-only simulations showed that through hierarchical structure formation, haloes should host their own satellite haloes, with original predictions citing that up to 500 subhaloes of mass $> 10^{8}\mathrm{M}_{\odot}$ should be found within the Virial radii of MW-mass haloes (Moore et al., 1999; Klypin et al., 1999). This prediction was in contest with observations that show a significantly lower number of dwarf galaxy satellites orbiting the MW or Andromeda (M31).

**Too-Big-To-Fail Problem:** The too-big-to-fail problem refers to ΛCDM simulations that predict that many of the most massive subhaloes of the MW are too dense to host any of its observed bright satellites with luminosities of $L_{\mathrm{V}} > 10^{5}\mathrm{L}_{\odot}$ (Boylan-Kolchin et al., 2011). The observed stellar kinematics of stars within a satellite's half-light radius can be used to infer their dark matter mass, these measurements were found to be in tension with simulations as the most massive observed satellites of the MW typically have a smaller dark matter mass than their analogues (the most massive haloes found) in DM only simulations. A possible solution is to require a number of the most massive haloes ($V_{\text{peak}} \geq 30\text{kms}^{-1}$) from the simulations to remain dark, however, such subhaloes are massive enough to form stars and are hence too big to fail at hosting galaxies (e.g. Sales et al., 2022).

These problems arose from the use of collisionless DM only simulations compared to observations, however, more recent studies have shown that many of these tensions can be resolved with the inclusion of a baryonic component in simulations. The missing satellite tension can be reconciled by a combination of improved observational completeness and improved simulations. The existence of 'dark' satellites means that a significant number of haloes in the dwarf mass regime do not host a stellar component and therefore cannot be observed, these dark subhaloes arise from a combination of reionisation and environmental processes that suppress galaxy formation (Sawala et al., 2016b). Alternatively, satellites may host a very small and hence faint stellar component, known as Ultra-Faint galaxies, that have only recently been discovered by improved observations around the MW and M31 (Simon & Geha, 2007). The uncertainty of the MW halo mass may also contribute to the missing satellite problem given that the number of massive satellite subhaloes is sensitive to the mass of the host halo. The MW-analogue haloes examined from the Aquarius Project simulations may be towards the high mass end of the predicted MW halo mass range, and therefore a lower halo mass within the range $5 \times 10^{11}\text{M}_{\odot} > \text{M}_{\text{halo}} > 1 \times 10^{12}\text{M}_{\odot}$ would have fewer satellites, consistent with observations, and would further alleviate the missing satellite tension (Wang et al., 2012). The inclusion of baryonic processes in simulations has also been shown to alter the DM distributions within dwarf galaxies. Stellar feedback drives gas out of the galaxies creating a variable gravitational potential that transfers energy to the orbits of dark matter particles and effectively lowers the central density of dark matter, flattening the cusp into a cored density profile (Navarro et al., 1996a; Pontzen & Governato, 2012). In the context of the too-big-to-fail problem, this means that the DM mass in the inner regions of dwarf galaxies is reduced without requiring them to reside in lower mass haloes.

The tensions surrounding the small scale regime have demonstrated how important baryonic effects are in reproducing observations where we are only able to directly observe luminous objects and the effects of DM on their structure and distribution. Baryons comprise $\sim 17\%$ of the total mass budget of the Universe and the non-linear physical processes that dominate them are essential in simulations. That being said, with the advancement of galaxy formation simulations there still remains tensions in this regime (e.g. Sales et al., 2022), including the planes of satellites (Boylan-Kolchin, 2021), diversity of rotation curves (Oman et al., 2015), and diversity of sizes (McConnachie, 2012) of dwarf galaxies, which will likely require refinements in numerical modelling and further observational constraints to authoritatively address.

### 1.3.2 Dwarf galaxies

Dwarf galaxies are typically classified as galaxies with a stellar mass $M_\star < 10^9 \mathrm{M}_\odot$ and include some of the oldest and most dark matter dominated systems. Dwarf galaxies are of particular interest as 'laboratories' for studying the astrophysics of galaxies as their early formation times and shallow gravitational potentials render them particularly sensitive to the physical processes that regulate galaxy growth. These include both internal mechanisms such as radiative and mechanical feedback from young stars and supernovae (SNe), and external environmental influences such as the ram pressure stripping of gas and the tidal stripping of gas and stars.

Since the advent of panoramic, digital sky surveys like the Sloan Digital Sky Survey (SDSS) in 2005, the discovery of new dwarf galaxies has rapidly accelerated, revealing ever-fainter systems (Simon, 2019). The lowest-luminosity galaxies, referred to as Ultra Faint Dwarfs (UFDs) represent the extreme lower limit of galaxy formation which are hypothesised to be relics of the early Universe that formed during the early cosmic epochs and were subsequently quenched during the epoch of reionisation (Bovill & Ricotti, 2009).

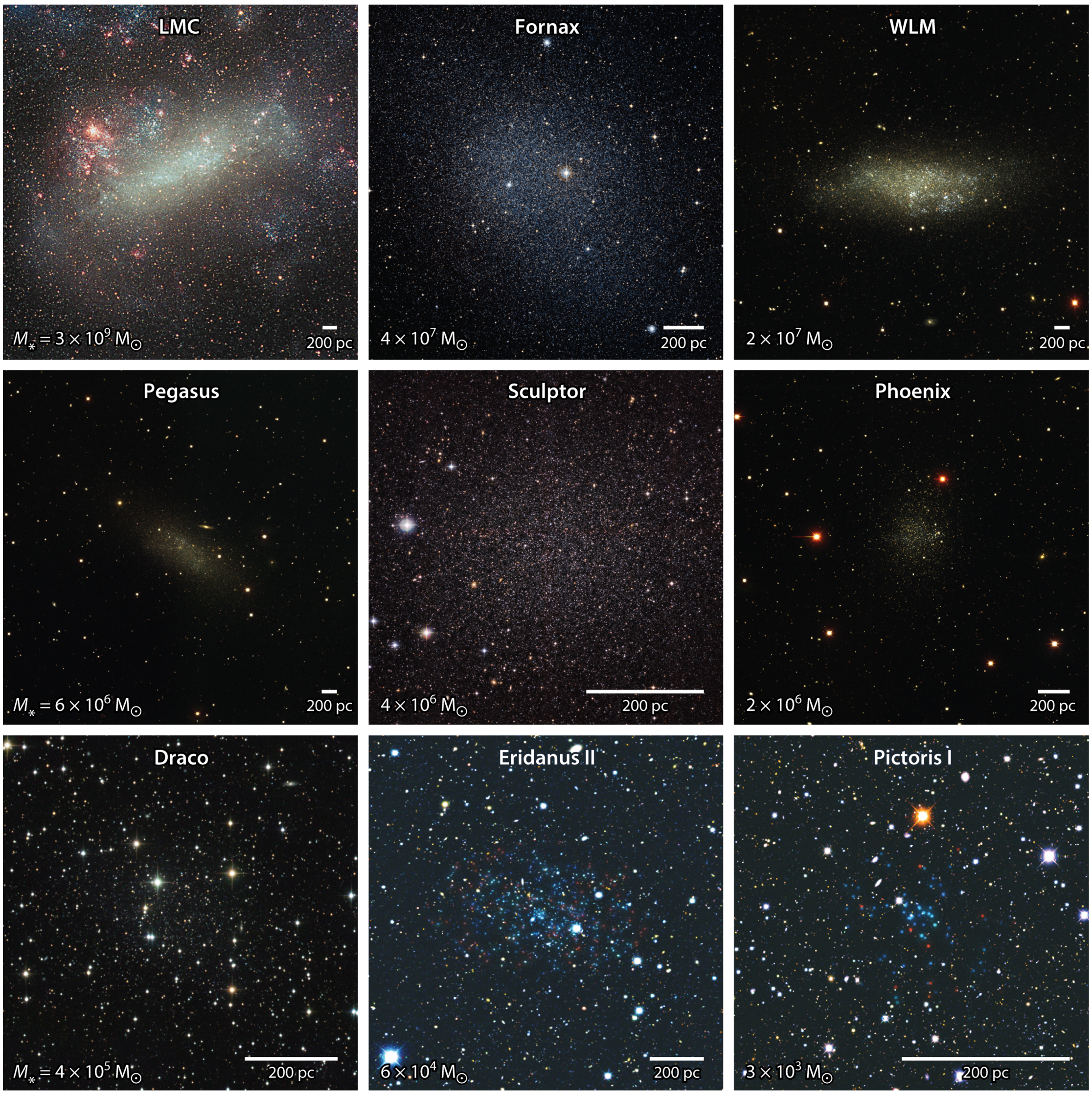
Dwarf galaxies are commonly subdivided by their morphology into dwarf spheroidals (dSph) and dwarf irregulars (dIrrs), classifications which often reflect the dwarf's local environment. Figure 1.3 shows the diversity in dwarf galaxies across a range of stellar mass ($10^3 < M_\star < 10^9 \mathrm{M}_\odot$) and morphology (dIrr to dSph). Dwarf spheroidals are characterised by their lack of gas and recent star formation activity, with high mass–to–light ratios of 10–1000 making them ideal probes of the underlying dark matter halo. Observations of the Local Group show that most satellites of the Milky-Way and M31 are dSph galaxies lacking atomic hydrogen (HI, Spekkens et al., 2014) and on-going star formation (Geha et al., 2012) which is often attributed to interactions with their nearby host that strips the satellites of their gas (Teyssier et al., 2012). Dwarf irregulars are typically gas rich with ongoing star formation; with the notable exception of the Large Magellanic Cloud (LMC), the Local Group's dIrrs are typically isolated field galaxies that are able to retain their gas content and active star formation by evolving without more massive and disruptive neighbours.

The sensitivity of dwarf galaxies to their environment is apparent in the dichotomy in gas content and current SFRs between satellite galaxies and field galaxies in the Local Group. Isolated galaxies are therefore ideal tests of galaxy evolution as they are insulated from extreme external influences like tidal interactions and mergers, and retain much of their primordial characteristics. They are expected to exhibit prolonged star formation and observations have suggested that quenched dwarf galaxies are extremely rare in isolation (Geha et al., 2012) with a possible stellar mass threshold of $M_\star = 10^9 \mathrm{M}_\odot$, below which a massive neighbour is required to halt star formation. Quenched star formation is however observed in isolated field dwarf galaxies, for example Cetus and Tucana in the Local Group (although it has been suggested that they were quenched by their interaction with the MW or M31 Teyssier et al., 2012), and farther afield KKR 25 ($d_{\mathrm{MW}} \simeq 1.93\,\mathrm{Mpc}$, Karachentsev et al., 2001; Makarov et al., 2012) and KKs 3 ($d_{\mathrm{MW}} \simeq 2.12\mathrm{Mpc}$, Karachentsev et al., 2015).

### 1.3.3 Linking Dwarf Galaxies to their Dark Matter Haloes

Understanding galaxy formation involves understanding the luminous component of the Universe, the collapse of dark matter haloes under ΛCDM cosmology and the link between these processes. This galaxy-halo connection can be quantified by the stellar mass - halo mass relation that probes which mass galaxies occupy which mass haloes. Dark matter halo masses are challenging to measure observationally, however, various modelling approaches including halo occupation models, abundance matching, and hydrodynamical simulations have helped to quantify this relation with good constraints above a halo mass of $M_{200,\mathrm{crit}} \gtrapprox 10^{11}\mathrm{M}_\odot$ (e.g. Wechsler & Tinker, 2018).

Abundance matching techniques match the stellar masses or luminosities of galaxies at a given abundance, to haloes of the same abundance derived from a DM only simulation.



Bullock JS, Boylan-Kolchin M. 2017.
*Annu. Rev. Astron. Astrophys.* 55:343–87

FIGURE 1.3: Observations of dwarfs galaxies compiled by Bullock & Boylan-Kolchin (2017b) that span six orders of magnitude in stellar mass. The LMC, WLM, and Pegasus are dwarf irregular (dIrr) galaxies that have gas and ongoing star formation. The remaining six galaxies shown are gas-free dwarf spheroidal (dSph) galaxies that have no current star formation. The faintest dwarfs shown are only detectable in limited volumes around the Milky Way (Bullock & Boylan-Kolchin, 2017b).

The relationship is generally assumed to be monotonic and assumes that every dark matter halo hosts a galaxy. Though abundance matching techniques and different galaxy formation simulations yield good agreement in the SMHM relation for haloes $\gtrsim 10^{11}\mathrm{M}_{\odot}$, the same assumptions lead to discrepancies in the low mass regime. Theoretical predictions show a high degree of scatter in the low mass SMHM relation: different simulations loosely follow the extrapolations of abundance matching models, however, in detail the slope and scatter varies, and how low mass galaxies occupy haloes is generally uncertain. Figure 1.4 shows a compilation of the SMHM relation predictions for field dwarfs from various models and environments compiled by Sales et al. (2022). The assumption that every halo must host a galaxy is also contrary to predictions from simulations. In low mass haloes, reionisation suppresses, or may completely shut down, star formation. Simulations find low mass haloes that are entirely devoid of stars due to reionisation (e.g. Gnedin, 2000), however, the mass scale at which this occurs and the environmental dependence are not well understood.

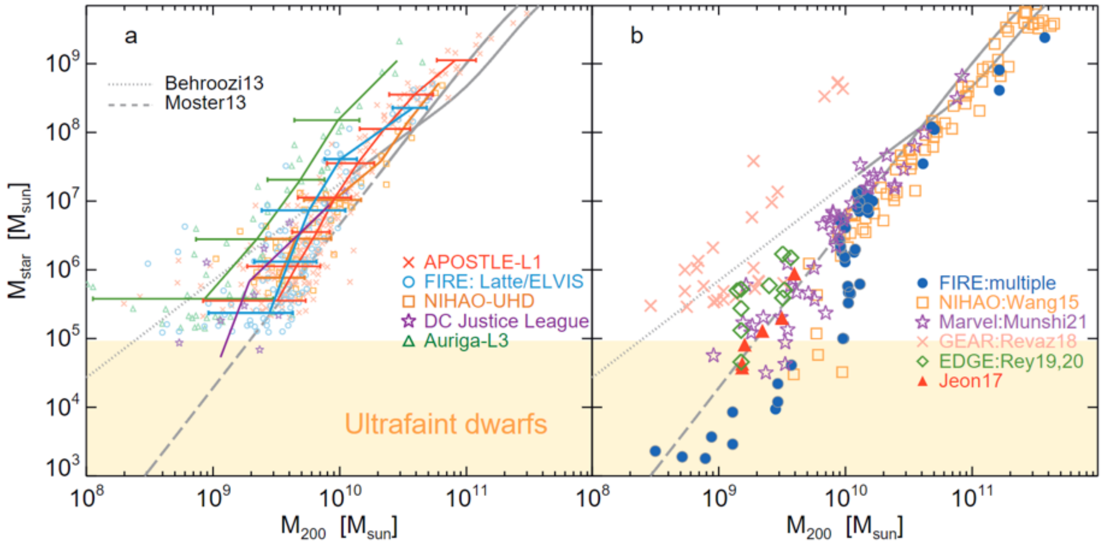


FIGURE 1.4: A compilation of the present-day stellar mass - halo mass relation of central dwarf galaxies by Sales et al. (2022) comparing the slope and scatter from abundance matching and different simulations. *Left:* Predictions from zoom-in simulations in a Local Group-like or MW-like environment using various models. *Right:* Predictions from zoom-in simulations of individual dwarf galaxies using various models.

## 1.4 This Thesis

Hydrodynamical simulations of representative cosmic volumes (side length $L \sim 100$ comoving Mpc, hereafter cMpc) have, over the last decade, matured such that they yield galaxy populations with properties that broadly correspond to those of the observed population (see e.g. Crain & van de Voort, 2023). Such models reveal the formation and assembly histories of galaxies that reside at the present epoch in diverse environments. However, the study of low-mass galaxies in such simulations is hindered by the large volume they follow, which requires sufficiently high resolution to adequately sample the structure and star formation history of the galaxies, as volume and resolution are competing 'uses' of the total number of a simulation's resolution elements. The significant difference in the scale of individual galaxies ($\sim 10^{3-4}\,\mathrm{pc}$) and their large-scale environment ($\sim 10^{6-7}\,\mathrm{pc}$) compounds the dynamic range challenge posed by the difference in scale between galaxies and the individual stars ($\sim 10^{-7}\,\mathrm{pc}$) and black hole (BH) accretion discs ($\sim 10^{-3}\,\mathrm{pc}$) that are the origin of the feedback processes that regulate galaxy growth.

The most common method to achieve the necessary resolution and sampling of dwarf galaxies at a reasonable cost is to adopt the zoom-in technique (e.g. Katz & White, 1993), wherein typically the immediate environment of an individual galaxy is followed with full gas dynamics at high resolution, and the remaining cosmological volume is evolved with reduced resolution to ensure the galaxy experiences the correct tidal forces. This method has been used to resimulate Milky-Way and Local Group-like analogues, including populations of field and satellite dwarf galaxies in these environments, in projects such as (but not limited to) Eris (Guedes et al., 2011), APOSTLE (Sawala et al., 2016a), Auriga (Grand et al., 2017), the D.C. Justice League Simulations (Applebaum et al., 2021), NIHAO-UHD (Buck et al., 2020), Latte (Wetzel et al., 2016) and ELVIS (Garrison-Kimmel et al., 2014). Zoom simulations have also been widely used to follow individual galaxies, enabling ultra-faint dwarf galaxies to be simulated e.g. the EDGE project (Rey et al., 2019, 2020), LYRA (Gutcke et al., 2021), and dwarf galaxies from FIRE (Wheeler et al., 2019). However, by construction such studies cannot explore the role played by diverse cosmic environments.

By targeting regions significantly larger than individual galaxies, the zoom technique also enables diverse environments to be simulated at a reasonably high resolution without incurring an infeasible computational cost and memory footprint. For example, the GIMIC simulations (Crain et al., 2009) follow the evolution of galaxies within five roughly spherical regions of comoving radius $r = 18 - 25\,\mathrm{cMpc}$ within the $L = 500\,\mathrm{cMpc}$ Millennium Simulation volume. The regions were chosen to have overdensities of $(-2, -1, 0, +1, +2)$ times the root-mean-square deviation from the mean overdensity, and thus span $\simeq 95$ percent of the overdensities in the parent volume whilst simulating only 0.13 percent of its volume. A similar approach was adopted by the FLARES project (Lovell et al., 2021) to examine diverse environments in simulations of the epoch of reionisation, and by the MARVEL-ous project to study dwarf galaxies in the environment defined by a cosmic ‘sheet’ (Bellovary et al., 2019; Munshi et al., 2021).

This thesis aims to probe small-scale galaxy formation and evolution and address the science questions surrounding dwarf galaxy populations in diverse environments. This thesis is built around the Columba suite of simulations of galaxy formation in the ΛCDM cosmogony, which are designed primarily to examine the influence of large-scale cosmic environment on the evolution of isolated dwarf galaxies, i.e. those not in close proximity to larger structures at the present day. The Columba simulation suite will explore the influence of large-scale overdensity and the intrinsic properties of their host DM haloes on galaxy evolution in the low-mass regime, in the absence of direct environmental influences driven by tidal and ram pressure forces exerted by near neighbours.

In Chapter 3 I present the development of the Columba suite of zoom-in simulations, including the strategy for selecting the simulated regions of interest and the generation of the multi-resolution zoom-in initial conditions (ICs). The final suite of simulations comprises 25 isolated regions that are selected to be devoid of any objects more massive than the MW and have a diverse population of low mass and LMC-like haloes. These regions are systematically selected to span the diversity in the overdensity which naturally tends towards low density void and filament environments. I describe the numerical and subgrid components of the model used to evolve the ICs to the present day, including the hydrodynamics and gravity code SWIFT and the galaxy formation model. Finally, I discuss the techniques for identifying galaxies and their parent haloes.

In Chapter 4, I investigate the environmental effect on global properties of low mass field haloes and galaxies in the Columba suite of simulations. Firstly, I compare basic properties of the simulated galaxies with observational measurements to validate the model and test the simulations for convergence across numerical resolution. I find that the simulations successfully reproduce the observed GSMF, however, are inconsistent with observed galaxy sizes. I then examine the influence of large scale overdensity variations on key population diagnostics such as the halo mass function, galaxy stellar mass function, SMHM relation, and luminous fraction. Over the full range of environments sampled in the simulation suite, large-scale overdensity does not have a significant influence on these properties and the SMHM relation does not exhibit significantly different slope between the environments sampled. I then find that the luminous fraction of galaxies shows a higher dependence on halo concentration than their large scale environment, indicating the important role of formation time on the ability of haloes to form and assemble their stellar mass.

In Chapter 5 I further investigate the influence of halo concentration, and hence formation time, on key properties of isolated dwarf galaxies in the Columba simulations. I explore the origin of scatter in the SMHM relation of field galaxies in the Columba suite and find a strong positive correlation with halo concentration such that high concentration, and earlier forming, haloes host higher stellar mass galaxies at a fixed halo mass, extending previously found correlations (e.g. Matthee et al., 2016) to lower masses. I find that high concentration haloes assemble their stellar mass earlier than low concentration haloes, reflected strongly in the time it takes for haloes to assemble 90% of their stellar mass. I find a broad range in morphology of isolated dwarf galaxies in the Columba suite and though halo concentration influences the stellar assembly of these dwarf galaxies, their morphology does not show a similar correlation.

Finally, in Chapter 6 I summarise the findings of the work in this thesis and discuss avenues for future research and extensions to the Columba simulations.

# 2

# Computational Galaxy Formation

Cosmological simulations have been instrumental in advancing our knowledge of the Universe. Galaxy evolution involves non-linear processes that span extreme time and length scales which are increasingly challenging to capture, and understand the physics of, observationally. Simulations provide an alternative method to confront the questions at the forefront of galaxy formation and allow us to trace in detail how galaxies form and evolve over cosmic time. Comparing the outcomes of simulations to observations allows us to test and constrain the underlying cosmological framework, our theoretical understanding of galaxy formation and the physical mechanisms that have led to the present day galaxy population in the observable Universe. In recent years simulations have been able to produce remarkably realistic galaxy populations and make a wide range of detailed predictions. In this Section I discuss how we evolve the Universe from specific initial conditions, follow the evolution of dark matter and baryons, and incorporate models of astrophysical processes like the cooling of gas, formation of stars, and feedback from massive stars and SMBHs.

## 2.1 Gravity

Within the ΛCDM paradigm CDM forms the dominant matter contribution and its evolution underpins the formation of galaxies, which are assumed to form within the centres of dark matter overdensities, or haloes. Within cosmological simulations CDM is modelled as a non-relativistic collisionless fluid that only interacts via gravitational forces and evolves as described by the collisionless Boltzmann equation under the influence of the Newtonian gravitational potential given by Poisson's equation. Poisson's equation,

$$\nabla^2\Phi = 4\pi G\rho \tag{2.1}$$

relates the gravitational potential $\Phi$ to the mass density $\rho$ where $G$ is the gravitational constant.

Instead of solving the highly dimensional Collisionless Boltzmann Equation (CBE), many simulations adopt an $N$-body technique whereby the continuous distribution of collisionless matter is sampled by a finite set of particles and the system is advanced in discrete timesteps. In an $N$-body system, one can compute the acceleration felt by a particle and update its motion and position. Owing to the long range nature of gravity, calculating the acceleration is an intensive operation since the gravitational influence of each particle must be taken into account:

$$\vec{F}(\vec{r}_\mathrm{i}) = \sum_{j=1,j\neq i}^{N-1} \frac{Gm_\mathrm{i}m_\mathrm{j}}{(|\vec{r}_\mathrm{i}-\vec{r}_\mathrm{j}|)^2} \tag{2.2}$$

This results in an increasing number of calculations needing to be performed at each timestep, scaling as $\mathcal{O}(N(N-1)) \sim N^2$ for a simulation containing $N$ particles. The computational expense of this approach rapidly increases with particle number and is unfeasible for modern simulations that can contain in excess of $10^9$ particles. In modern simulations algorithms are employed to improve the efficiency of the derivation of the gravitational potential, these typically involve using a multipole expansion and/or mapping the tracer distribution to a mesh and solving in Fourier space using fast transform methods.

### 2.1.1 Hierarchical Multipole Expansion

One means of reducing the computational expense of gravitational calculations in N-body simulations is by employing a hierarchical multipole expansion. Gravity is a long range force that is inversely proportional to the distance between masses, therefore, nearby particles can dominate the net gravitational potential at a given location compared to more distant particles. In contrast, there are more particles at large distances than close and hence gravitational calculations for particles with large separations will dominate the computational expense despite their lesser gravitational influence. This motivates sacrificing the precision of gravitational interactions between distant particles, and is exploited by the multipole expansion technique.

Tree codes use multipole expansion by grouping particles at long distances in a given location by one term and approximating their interaction collectively. The algorithm hierarchically divides the simulation volume into cells, grouping particles at each hierarchical level into decreasing numbers until reaching a predetermined minimum number of particles per cell (as small as a single particle). The resulting cell structure resembles a tree: the simulation volume representing the root node, each recursion level splits into a branch and the final, finest level of cells represent the leaf nodes (Barnes & Hut, 1986; Dehnen, 2000). So-called ‘oct-trees’ are commonly implemented whereby the volume is subdivided into 8 new subcells at each level by dividing each dimension by 2.

In the gravity operations, the tree allows distant particles to be grouped together and their gravitational forces to be approximated. The gravitational force felt by each particle is computed by a simple recursive calculation or ‘walking the tree’; starting at the root of the tree, containing the entire system, the current level of cells is assessed against an ‘opening criterion’ of accuracy. If the criterion is met, the gravitational force of that cell is accounted for by means of a single multipole force, included in the accumulated total and the walk is terminated. Otherwise, the current cell is subdivided and each cell is recursively examined until the criterion is met or the leaf node is reached. The opening criterion controls the accuracy of the approximation of the gravitational forces and many codes, e.g. GADGET-2 and 3, vary this criterion with a dependence on the distance to a node in order to achieve higher force accuracy for a given computational cost (Springel, 2005). Overall, this method effectively reduces the number of gravitational force calculations from $\mathcal{O}(N^2)$ to $\mathcal{O}(NlogN)$.

### 2.1.2 Particle Mesh

An alternative technique is the particle-mesh (PM) method. This method involves overlaying an equally spaced Cartesian grid onto the particle distribution and estimating the density field at each grid point. Poisson’s equation is solved for the gravitational potential on the mesh using a Fast Fourier Transform (FFT) and the gravitational force on a given particle is computed by interpolating from the force on the grid to the particle coordinates. For a simulation of $N$ particles and $N_{\mathrm{grid}}$ grid points the number of operations scales as $\mathcal{O}(N_{\mathrm{grid}} \log N_{\mathrm{grid}})$ achieving impressive computational speedup (Bertschinger, 1998). Despite their advantages, PM methods can incur a large memory cost. For interactions on the scale of the cell size, the inverse square law is poorly approximated and therefore high resolution can only be achieved by reducing the mesh spacing, i.e. using a finer grid, due to the fixed spatial resolution of the mesh. Refining the grid to achieve sufficiently high spatial resolution involves very large FFT meshes that incur an extremely high memory cost.

To balance the computational speedup, memory demands, and desired resolution, PM methods are often combined with other gravity solvers. The particle-particle particle-mesh method ($\mathrm{P^3M}$) combines the PM method and direct summation for nearby particles to mitigate the effects of the fixed spatial grid on small scales. Other gravity solvers, for example SWIFT (Schaller et al., 2023), combine the Fast Multipole Method (FMM) for nearby particles with a PM for long-range gravitational forces.

## 2.2 Hydrodynamics

Whilst dark matter dominates the matter budget of the Universe, for simulations to reproduce the observable Universe, the luminous baryonic matter, i.e. gas and stars, must be taken into account. The baryonic component is subject to hydrodynamical forces alongside gravity and hence requires additional modelling. Simulations make the assumption that cosmic gas is ideal, collisional, inviscid, and non-conducting, enabling the dynamics to be solved by the Euler equations. Within computational hydrodynamics the solutions to the Euler equations are typically done by discretising the fluid by volume or by mass. The former, or the Eulerian approach, is concerned with the fluid properties at a specific space-time point and is often used by mesh-based schemes. The latter, or Lagrangian approach, tracks a discretised fluid element or particle as it moves through space and is used by particle-based schemes.

Eulerian techniques are popular for computational fluid hydrodynamics and use fixed uniform meshes to calculate the fluid properties at points in space-time where fluid passes through the mesh faces. The extreme dynamics ranges in galaxy formation require the use of more complex adaptive refinement meshes (AMR, e.g. Abel et al., 2002)) that change the spacing of the grid to provide higher resolution where necessary, these have been used extensively in the field, for example RAMSES (Teyssier, 2002) and ENZO (Bryan et al., 2014), but comprehensive comparisons between AMR and Lagrangian Smooth Particle Hydrodynamics (SPH) suggest outstanding numerical issues like unphysical heating at early times in preshock regions (O'Shea et al., 2005).

Lagrangian techniques are appealing for cosmological simulations due to their inherent adaptive resolution which adapts to local conditions by increasing the resolution in overdense regions, their ability to manage the extreme dynamic ranges in galaxy formation, and the full available history of all particles.

### 2.2.1 SPH

Amongst Lagrangian techniques, Smoothed Particle Hydrodynamics (SPH) (Lucy, 1977; Gingold & Monaghan, 1977) is a popular numerical method for solving the fluid equations by discretising the fluid into mesh-free tracer particles (e.g. Springel, 2010), and is the method used in the simulations in this thesis. For any field $F(r)$ attributed to a discrete fluid element, or particle, a smoothed interpolated version, $F_s(r)$, can be defined through a convolution with a smoothing kernel $W(r, h)$, with an extent defined by the smoothing length, $h$, at a given coordinate:

$$F_s(r) = \int F(r)W(r - r', h)dr' \tag{2.3}$$

The most fundamental property for SPH is density: a particle carrying a given mass, $m$, is spread out by the smoothing kernel, $W(r, h)$, with a radius of smoothing length, $h$. The interpolation can be approximated by the particle summation and the density can be computed as,

$$\rho_i \approx \sum_j^N m_j W(r_i - r_j, h_i) \tag{2.4}$$

Given the density, any smoothed quantity can be approximated by the summation over the particles,

$$F_s(r) \approx \sum_i^N \frac{m_i}{\rho_i} F_i W(r_i - r_j, h_i) \tag{2.5}$$

To maintain conservation of mass, the smoothing kernel must be normalised to unity, and to conserve angular momentum must be spherically symmetrical. The simplest example of a kernel function is Gaussian, $W(r,h) = 1/(\sqrt{\pi}h)^3 exp(-r^2/h^2)$, characterised by the size of the kernel, or smoothing length, $h$. The smoothing length is generally taken to vary over time so that a fixed number of neighbour particles, or fixed mass, is enclosed and included in the kernel sum. The variable smoothing length means the resolution is automatically adapted depending on local conditions of the simulation, with a finer smoothing kernel in higher density regions where there are more clustered particles.

## 2.3 Resolution

As discussed, Lagrangian techniques adopt a particle based method of modelling the Universe; representing mass elements of different species of matter, e.g. dark matter, gas or stars, as particles. Each particle represents a given amount of mass, e.g. $m_{\rm p} = 10^6 \rm M_{\odot}$, and this value sets the mass resolution of the simulation. Methods for simulating physical processes that take place on smaller scales than this resolution which cannot be modelled directly are known as subgrid routines and are discussed in Section 2.5. In order to reliably analyse a galaxy and its properties it is necessary for the object to be well-resolved by a sufficient number of particles. This number is dependent on the mass resolution of the simulation and the focus of the analysis. For example, investigating the dynamics of an individual galaxy may require $\sim 10^2 - 10^3$ particles to be considered well-resolved and converged across resolution: Power et al. (2003) demonstrate a DM halo mass profile can require 100-3000 particles to be considered converged, Crain et al. (2015) report 600 star particles to obtain well-converged and resolved stellar sizes, and Hopkins et al. (2018) demonstrate that global galaxy properties, such as rotation curves and sizes, converge at moderate resolutions (typically requiring $\sim 10^3$ particles per galaxy component).

The spatial resolution of a simulation is limited by the softening factor ($\epsilon$) that is implemented into the calculations of Newtonian gravity described in eq. 2.6. To prevent unphysical scattering when a pair of particles have a close encounter, the gravitational force is softened on small scales, for $r_{\rm ij} \gg \epsilon$ the softened force approaches the regular

Newtonian force, and when $r \ll \epsilon$, the forces reaches a constant maximum:

$$\vec{F}_{\mathrm{soft}} = \frac{G m_{\mathrm{i}} m_{\mathrm{j}}}{\vec{r}_{\mathrm{ij}} + \epsilon} \tag{2.6}$$

The softening length limits the spatial scales that are resolved in the simulation and in order to maintain its desired effect whilst not unduly forfeiting resolution it is typical to adopt a softening scale that is fixed in comoving space, of the order of a fraction of the mean interparticle separation, and limited to a maximum proper size to ensure that the internal structure of dark matter haloes can be resolved (Power et al., 2003). This limit prevents the softening from becoming excessively large in physical space at low redhsift which can compromise the resolved halo structure, causing the inner halo density profile to be artifically smooth and form artificially low density cores.

## 2.4 Initial Conditions and Zoom Simulations

Before running a cosmological simulation, the initial conditions of the dark matter and gas content must be constructed that specify the fluctuations in the density field at high-z. Inflation predicts that the density fluctuations in the early Universe are Gaussian and described by the primordial matter power spectrum. Post-recombination, the density field is given by the linear convolution of the primordial fluctuation field (predicted by inflation) with a transfer function, that accounts for various physical processes that modify the initial density perturbations, in Fourier space. For cosmological simulations, the generation of Gaussian random fields with a specified power spectrum is implemented by displacing the positions and velocities of the particles from an unperturbed, uniform particle distribution.

Using a random distribution of particles to represent the unperturbed Universe introduces a Poisson noise power spectrum, if the simulation is run from such initial conditions the fluctuations will evolve into non linear structure without introducing any other fluctuations. The most commonly used solution is a uniform grid of particles, evenly spaced in each dimension. This procedure, however, can introduce strongly preferred directions into the ICs and the initial grid pattern can persist in evolved low density regions that may affect statistical properties, e.g. void statistics (White, 1994). The ideal solution, an unperturbed particle load that has no preferred direction, can be achieved using a glass distribution. A glass distribution is generated by randomly placing particles and evolving them in an expanding Einstein-de Sitter Universe with reversed gravity such that the forces become repulsive. Running this evolution over many expansion factors, the particles will settle into a configuration with negligible forces on each particle and no preferred direction (White, 1994). The displacements of the unperturbed particle distribution are calculated with linear theory (Zel'dovich, 1970) or low order Lagrangian perturbation theory. Standard cosmological simulations are typically constructed as 3-dimensional cubes of side-length $L$ with imposed periodic boundaries. The periodicity reproduces the homogeneity and isotropy of the Universe and ensures that no artificial boundaries are imposed.

As discussed in Section 2.3 increasing the resolution of large periodic simulations incurs increasing computational demands, a useful technique to mitigate expenses are zoom-in simulations. The zoom-in technique typically involves selecting a smaller volume within a cosmologically representative, dark-matter only parent simulation of coarser resolution, reconstructing the initial conditions with a multi-resolution approach, and resimulating the targeted region with full hydrodynamics at a higher resolution. The region of interest is simulated at high resolution, enabling more detailed analysis without incurring the cost of simulating the whole periodic volume in high resolution. The resolution of the surrounding volume is degraded whilst retaining the long-range gravitational forces and hence large-scale cosmological context (Katz & White, 1993).

## 2.5 Subgrid Physics

The volume of the simulation constrains the resolution that can be achieved and thus limits the physical scales that can be explicitly modelled. To simulate statistically significant galaxy populations, large representative volumes are required which necessitate the use of a higher number of particles to populate the full volume, therefore, improving the mass resolution incurs more calculations and increases the run time and memory requirements of the simulation. The extreme dynamic range of the length scales of astrophysical processes that influence galaxy formation and evolution is illustrated in Figure 2.1 which emphasises how macroscopic properties of galaxies and their populations are governed by microscopic processes beyond the resolution limitations of cosmological simulations. This limitation means that scales below the mass or spatial resolution cannot be explicitly modelled and 'subgrid' routines are used to approximate the small scale processes. The primary goal of subgrid methods is to reproduce the effects of unresolved processes on larger, resolved scales of the simulation and they have become a critical component in cosmological simulations for modelling processes such as radiative cooling, star formation, stellar feedback, and AGN feedback.

Atop the challenge of modelling unresolved processes, the choice of subgrid routine, and related parameters, can affect the outcome of the simulation. Theory and observation can be used to inform the numerical values of the subgrid parameters, however, these values are often not known a priori and may be resolution dependent. It is now common practise to calibrate the subgrid schemes, often for each resolution, to produce a realistic galaxy population. This was the case in, for example, the EAGLE simulations where the efficiency of the stellar feedback and BH accretion were calibrated to match the z=0 GSMF and produce sensible galaxy sizes, and the AGN feedback efficiency was calibrated to the observed stellar mass - BH relation (Crain et al., 2015). With the advancements in machine-learning techniques in recent years calibration has been made more efficient with simulations like FLAMINGO utilising Gaussian process emulators to reproduce the low redshift GSMF and cluster gas fractions (Kugel et al., 2023). Though the predictive power for the calibrated observables are forfeited, other emergent results from the simulation can be considered predictions of the model and used to guide our understanding of galaxy formation and evolution.

### 2.5.1 Radiative Cooling and Heating

Radiative gas cooling is a fundamental process in the formation of stars, and hence galaxy evolution as it allows cosmic gas to dissipate its internal energy. Cosmic gas dissipates energy due to cooling processes like inverse Compton scattering, recombination, and free-free emission, and experiences heating from processes like collisional excitation and ionisation. The gas is typically assumed to be in ionisation equilibrium with a redshift dependent spatially uniform X-ray/UV cosmic background expected from star-forming galaxies and quasars. In modern simulations it is common to compute the cooling and heating rates based on individual elements' abundances. Simulations typically trace the abundances of individual elements with significant contributions to the cooling rates, including H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe, which are pre-computed using a

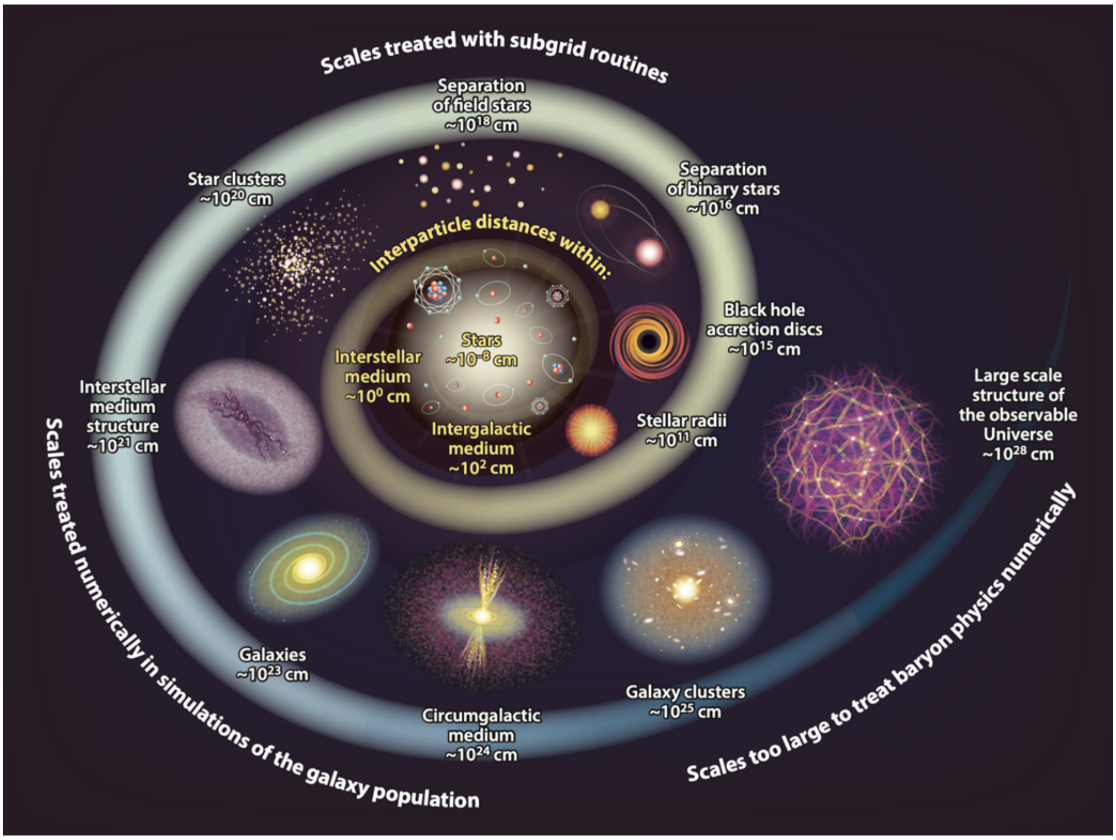


FIGURE 2.1: An illustration of the extreme dynamic range of the length scales of astrophysical processes that influence galaxy formation and evolution from Crain & van de Voort (2023). Cosmological hydrodynamical simulations are limited by their mass and spatial resolution and require subgrid models to reproduce the macroscopic effects from small scale processes.

photoionisation code. One of the most widely known photoionisation codes is CLOUDY (Ferland et al., 2017), which given the gas density, composition, and, if present, the spectrum of the radiation background can be used to tabulate the cooling and heating rates of gas as a function of density, temperature, and redshift (e.g. Wiersma et al., 2009a; Ploeckinger & Schaye, 2020).

### 2.5.2 Star Formation

The scales on which star formation occur following the collapse of molecular clouds in the cold, dense ISM, are not resolved within cosmological simulations and require subgrid modelling to implement on resolved scales. For star formation, this means reproducing the stellar distribution throughout the Universe without directly modelling the fragmentation of molecular clouds on sub-pc scales.

Stars can only be formed in gas that is sufficiently cold and dense, therefore, simulations typically use star formation criteria that allows stars to form in regions that meet specific physical conditions. This can be implemented using a temperature or density condition where a gas resolution element must be below a temperature threshold or above a critical density, or using the local Jeans criterion to determine if gas is gravitationally unstable. Various implementations are used in the field, a notable example is the EAGLE simulation (Schaye et al., 2015) that uses a metallicity-dependent density threshold.

When a gas element satisfies the star formation criteria, it is assigned a star formation rate based on empirical relations that are governed on scales resolved in the simulation. Commonly, the Schmidt and Kennicut laws that specify the SFR density as a function of the volume averaged gas density and surface densities, respectively, are used. Combined, these can be used to assign the resolution element a SFR. The final conversion of gas particle to star particle is typically implemented stochastically. For a given star formation rate and time-step, the probability that a gas element will be converted into a star particle is given by:

$$P(SF) = max\left(\frac{\dot{m}_{\mathrm{SF}}}{m_{\mathrm{gas}}}\Delta t, 1\right) \tag{2.7}$$

A random number is drawn from a uniform distribution between 0 and 1, and if less than $P(SF)$ then the gas particle (or cell) is replaced by a single stellar particle with an initial mass equal to $m_{\mathrm{gas}}$ with the same position, velocity and elemental abundances as the parent gas particle (or cell).

### 2.5.3 Stellar Evolution and Mass Loss

Models for chemical enrichment determine the mass of various chemical elements produced by stars and the fraction of this mass returned to the ISM. In simulations this requires defining how the mass ejected by stellar particles depends on their internal properties and then how the ejecta is distributed to the surrounding gas particles (or cells). This is typically implemented by tabulating the net stellar yields per unit stellar mass for all individual elements traced in the simulation as a function of stellar age and metallicity. The tables for galaxy simulations typically include chemical enrichment due to stellar mass loss in the AGB phase, the elements produced and ejected by massive stars via stellar winds and CCSNe, and the elements produced via type Ia SNe.

### 2.5.4 Feedback Processes

The ratio between galaxy stellar mass and halo mass as a function of halo mass exhibits a peak at halo mass of $\sim 10^{12}\mathrm{M}_\odot$ that is often associated with maximum galaxy formation efficiency and declining efficiency towards lower and higher halo mass. The suppression of galaxy growth in these regimes is attributed to distinct feedback mechanisms: stellar feedback in the low mass regime and AGN feedback in the high mass regime. Furthermore, the implementation of feedback using subgrid models in simulations has been pivotal in improving their agreement with observations. The injection of energy by these mechanisms occurs on scales that are not resolved within simulations and require subgrid implementation to approximate their macroscopic effects.

Through various mechanisms including stellar winds, radiation, and supernovae, stars inject energy into the ISM and if sufficiently energetic this can disrupt star forming molecular clouds and drive large scale galactic outflows, effectively suppressing or quenching star formation. Implementing these in simulations involves injecting the liberated energy into the stellar particle's neighbours. Each stellar particle represents a simple stellar population (SSP) with its stellar content described by an IMF, $\Phi(M)$, therefore, following Dalla Vecchia & Schaye (2012) the number of stars per unit stellar mass that end their life as a SNII is given by the integral of the IMF over the mass range $[M_{\mathrm{min}}, M_{\mathrm{max}}]$ which represent the minimum and maximum initial mass of stars that will explode as core-collapse SNe,

$$n_{\mathrm{SNII}} = \int_{m_{\mathrm{min}}}^{m_{\mathrm{max}}} \Phi(m) dm \tag{2.8}$$

The total available energy per unit stellar mass provided by SNII is hence given by $\epsilon_{\mathrm{SNII}} = n_{\mathrm{SNII}} E_{\mathrm{SNII}}$, where $E_{\mathrm{SNII}}$ is the available energy from a single SNII event. The amount of energy from SNII available in an SSP particle is then $m_* \epsilon_{\mathrm{SNII}}$ for a particle of initial mass $m_*$. Assuming $E_{\mathrm{SNII}} = 10^{51}\mathrm{erg}$, the total energy can be written as,

$$E_{\mathrm{SN,tot}} = 10^{51}\mathrm{erg}\; m_* \int_{m_{\mathrm{min}}}^{m_{\mathrm{max}}} \Phi(m) dm \tag{2.9}$$

Each stellar particle injects this energy into its surrounding neighbours. The simplest implementation involves injecting the energy thermally by increasing the temperature of the surrounding fluid elements. However, these early implementations of stellar feedback were plagued by the ‘overcooling’ problem (Katz et al., 1996) whereby thermal injected energy was quickly radiated away instead of driving winds. At the typical resolution scales of simulations, the fluid elements are orders of magnitude more massive than the SN ejecta and consequently the energy is distributed across a much larger mass, as a result the injected energy is only able to heat each element to $\sim 10^5$K. In this regime, the cooling time of the gas is much shorter than the sound crossing time, losing the energy through radiative cooling and making the SN feedback ineffective. Popular solutions to the overcooling problem involve delayed cooling (Stinson et al., 2006) which mitigates overcooling by temporarily disabling the cooling of gas, heating gas particles stochastically by large temperature increments $\Delta T >> 10^5$K (Dalla Vecchia & Schaye, 2012), or utilising kinetic feedback models (Dubois & Teyssier, 2008) that modify the peculiar velocities of the surrounding fluid elements. These approaches allow the feedback to be effective, enabling low mass galaxies to drive outflows and achieve self-regulated growth to reproduce realistic stellar masses.

In the high mass regime, galaxy growth is regulated by AGN feedback. Though the physical mechanisms by which AGN feedback couples to the ISM is poorly understood, with several channels considered such as radiation pressure on free electrons and dust grains, or high velocity jets, the large scale effects from AGN are observed and show large scale high velocity outflows of ionised and molecular gas (Maiolino et al., 2012). Implementations of AGN feedback via subgrid prescriptions in simulations generally assume a fraction $\epsilon_f$ of the radiated luminosity of an SMBH accretion disc couples to the surrounding ISM, and the AGN feedback energy is quantified by $E_{\mathrm{AGN}} = \epsilon_f \epsilon_r \dot{M}_{\mathrm{BH}} c^2$ for an accretion rate $\dot{M}_{\mathrm{BH}}$. The energy can then be injected into the surrounding fluid elements via thermal and/or kinetic models, the simplest of which assume an injection rate proportional to the accretion rate.

## 2.6 Structure and Halo Identification

A critical part of post-processing the simulation snapshots involves identifying haloes and galaxies from the particle data. The most common structure finding methods are based on friends-of-friends (FOF) algorithms and spherical overdensity (SO) algorithms (Knebe et al., 2011). FOF algorithms use a linking length and form groups of particles by linking a particle to other particles found within the linking length distance; particles are directly linked to all other particles within their linking length, i.e. the friends, and indirectly linked to all particles linked to their friends, i.e. the friends-of-friends, forming a network of particles that constitute the group. SO algorithms identify peaks in the density field and group all the particles found around this centre within a sphere of a given enclosed density.

Modern structure finders have expanded upon these methods by identifying subhaloes that reside in and are gravitationally bound to a host halo. Subhalo finders also most commonly fall into two categories between 3D configuration-space finders and 6D phase-space finders; the former includes subhalo finders like SUBFIND (Springel et al., 2001) which identify locally overdense, self-bound particle groups within the larger parent group and the latter includes finders like VELOCIRAPTOR (Elahi et al., 2019) which use extra velocity information to identify overdensities in 6D phase space.

# 3

# Columba: The Technical Bits

## 3.1 Region Selection

A key aim of this thesis is to examine isolated dwarf galaxy populations formed in regions spanning a diversity of cosmic environments. With a sidelength of $L = 100\,\mathrm{cMpc}$ and $1504^3$ DM particles of mass of $9.70 \times 10^6 \mathrm{M}_\odot$ (Schaye et al., 2015), the EAGLE simulation lacks not only the resolution but also diversity in environment to authoritatively address many of the questions surrounding the formation and evolution of these low mass galaxies. The zoom-in technique enables a smaller region of interest to be resimulated at a higher resolution with full hydrodynamics whilst the surrounding volume is simulated at a coarser resolution; this enables detailed analysis of specific regions of interest whilst retaining the long-range gravitational forces and hence large-scale cosmological context (Katz & White, 1993). In this Chapter I describe the development of the Columba zoom-in simulation suite. I aim to bridge the resolution limitations of simulations of representative volumes and the lack of environmental diversity in high resolution zoom-ins of individual galaxies by adopting a similar methodology to the GIMIC simulations (Crain et al., 2009) by using overdensity as an environmental diagnostic to sample the full range of cosmic environments isolated dwarf galaxies form in.

### 3.1.1 Parent Volume

For the selection of zoom-in regions, I ran a parent simulation of sidelength $L = 400\,\mathrm{cMpc}$ with periodic boundary conditions and mass distribution realised with $N = 3008^3$

composite particles representing baryonic matter and dark matter. The parent simulation adopts the maximum posterior likelihood cosmological parameter values reported by the Dark Energy Survey team from year three data (Abbott et al., 2022, their '3 × 2pt + All Ext.' CDM parameters), for which $\Omega_\mathrm{m} = 0.306$ yielding a mean particle mass of $m_\mathrm{particle} = 9.22 \times 10^7\,\mathrm{M}_\odot$. The simulation's ICs were generated[1] following the same procedure adopted by Schaye et al. (2023, see their Section 2.4) for the FLAMINGO simulations, using the third-order Lagrangian perturbation software MONOFONICIC (Hahn et al., 2020), which suppress discreteness noise by perturbing particle masses rather than coordinates positions, coupled to the PANPHASIA (Jenkins, 2013) Gaussian random noise field. The phase descriptor of this volume is [Panph6,L18,(200557,163876,161484),S1,KK1025,CH-999,COLIBRE400]. As per FLAMINGO, I adopt linear power and transfer functions computed with CLASS (Lesgourgues, 2011; Lesgourgues & Tram, 2011), with separate transfer functions computed and applied for baryons, DM, and neutrinos. The ICs were created at $z = 127$ and evolved to $z = 0$ assuming collisionless dynamics using the open-source cosmology, gravity, hydrodynamics, and galaxy formation software SWIFT (Schaller et al., 2023, see Section 3.3). Haloes were identified in simulation outputs using the VELOCIraptor structure finder (Elahi et al., 2019). I show the projected surface mass density of the parent simulation in Figure 3.4.

Figure 3.1 shows the halo mass function (HMF) of the parent simulation. Halo mass is defined by the mass of a sphere enclosing 200 times the critical density of the Universe, $M_{200,\mathrm{crit}}$. For comparison, I show HMFs presented by Tinker et al. (2008) and Bocquet et al. (2016) in the solid and dashed grey lines respectively. The solid grey curve is a universal mean density HMF fit for the adopted cosmogony of our simulations, created with the COLOSSUS package (Diemer, 2018) based on the Tinker et al. (2008) HMFs assuming SO haloes with mean internal density contrast $\Delta_\mathrm{crit} = 200$. This HMF is based on fits to ensembles of collisionless simulations, and does not account for the influence of galaxy formation, hence showing great agreement with the results from our collisionless simulation. The dashed grey curve shows the HMF presented by Bocquet et al. (2016) taken from an ensemble of hydrodynamic simulations from the Magneticum suite (Hirschmann et al., 2014), these include the effects of baryonic processes which have been shown to influence the mass and internal structure of haloes. Such effects are mass

---

[1] The ICs for the parent simulation were generated by Matthieu Schaller.

dependent and subject to ongoing study (see e.g. Stanek et al., 2009; Velliscig et al., 2014; Schaller et al., 2015a; Bocquet et al., 2016; Lovell et al., 2018), and contribute to the offset from the HMF presented here which does not include the influence of galaxy formation. The shaded region indicates 100 particles. Objects of interest in the selection procedure of resimulation regions, specifically objects within the lower mass range $5\times10^{10}\mathrm{M}_\odot < \mathrm{M}_{200,\mathrm{crit}} < 7\times10^{11}\mathrm{M}_\odot$ that are comparable to the Large Magellanic Cloud (LMC), are sufficiently resolved.

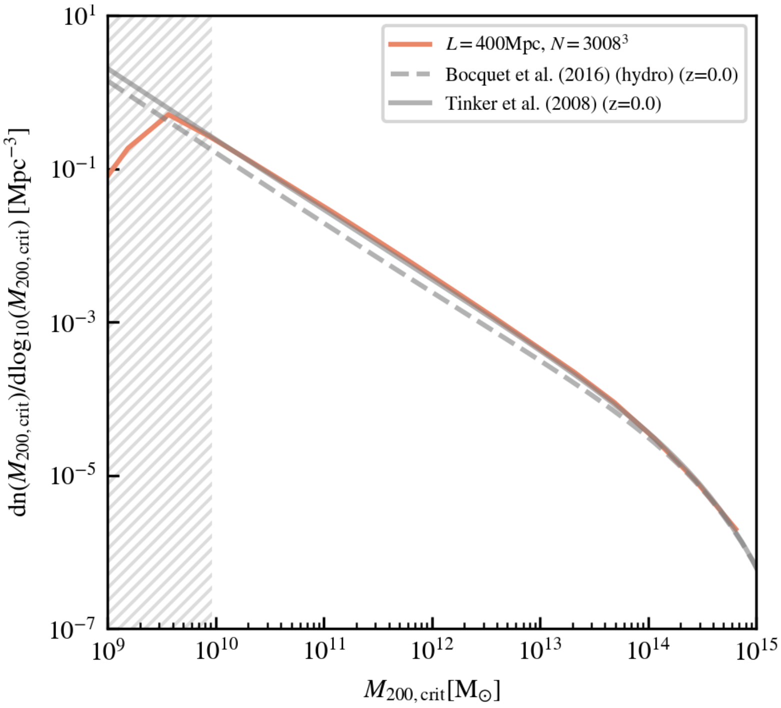


FIGURE 3.1: The halo mass function of the parent simulation from which zoom-in regions are selected. The parent simulation is a periodic volume with a sidelength of $L = 400\,\mathrm{cMpc}$ and mass distribution realised with $N = 3008^3$ composite particles and mass resolution of $m_\mathrm{particle} = 9.22 \times 10^7\,\mathrm{M}_\odot$. For comparison, the HMFs from Tinker et al. (2008) and Bocquet et al. (2016) are shown by the solid and dashed grey lines.

The region selection methodology is built on a sample of $2 \times 10^5$ randomly drawn coordinates, which form $2 \times 10^5$ randomly-centred spheres within the parent volume. This ensures that on average the entire parent volume is sampled at least once by spheres of $r = 5$cMpc. I define various mass ranges that are important in constraining the region selection in Table 3.1 and use the coordinate of the most bound particle of haloes computed by VELOCIraptor to record the minimum distance from each random coordinate to each object type.

| Object | Definition |
|---|---|
| Large Magellanic Cloud (LMC) mass object | $5 \times 10^{10} < M_{200,\mathrm{crit}} < 7 \times 10^{11}\mathrm{M}_{\odot}$ |
| Milky-Way mass object | $7 \times 10^{11} < M_{200,\mathrm{crit}} < 3 \times 10^{12}\mathrm{M}_{\odot}$ |
| ≥Milky-Way mass object | $M_{200,\mathrm{crit}} > 7 \times 10^{11}\mathrm{M}_{\odot}$ |
| ≥Group mass object | $M_{200,\mathrm{crit}} > 10^{13}\mathrm{M}_{\odot}$ |
| ≥Cluster mass object | $M_{200,\mathrm{crit}} > 10^{14}\mathrm{M}_{\odot}$ |

Table 3.1: The mass ranges, where halo mass is given by $M_{200,\mathrm{crit}}$, of object classes that have been used to record the diagnostics and build the selection criteria of the zoom-in regions for the Columba simulations.

### 3.1.2 Choice of Region Size

One of the initial choices in the development of this simulation suite was selecting the optimal size of the spherical zoom-in regions. Larger high resolution regions yield a smaller ratio between surface area and volume and hence a smaller fraction of the high resolution volume is affected by contamination from low resolution boundary particles. Larger volumes also naturally contain more objects and hence the sampling of objects per zoom-in simulation is superior. These advantages are offset by the higher computational expense that's incurred simulating larger and more populous regions, requiring more computational time per zoom-in simulation compared to a smaller region of interest. Choosing to simulate smaller regions may result in a lower sampling of objects per simulation, however, this can be offset by running a higher number of simulations and sampling a greater diversity of environments in the final simulation suite.

This simulation suite is designed to focus on dwarf galaxy populations that are isolated from the evironmental effects of nearby massive objects, I therefore elect to select regions devoid of massive haloes and investigate the average distance to such objects within a cosmologically representative volume. Figure 3.2 shows the distance from randomly-positioned coordinates in the parent simulation to objects of different masses: ≥MW mass objects, ≥group mass objects, and ≥cluster mass objects. The specific mass ranges are defined in table 3.1. Naturally arising from the HMF, more massive objects are rarer and lie, on average, at larger distances. I find that the average distance from a given position to a >MW-mass object is $\sim$ 5cMpc. As a result, regions characterised by a radius of $r = 10$cMpc containing no MW mass objects are rare and imposing a criterion of no >MW-mass galaxies within the region will lead to a significantly smaller and more highly biased sample compared to regions with a radius of $r = 5$cMpc.

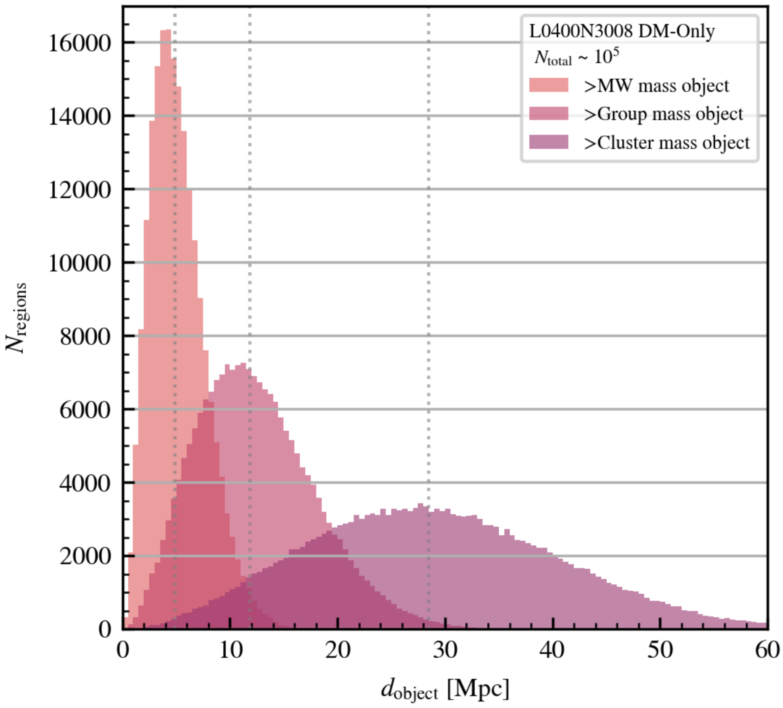


FIGURE 3.2: The distributions of distances to different mass objects from $N_{\rm total} = 2 \times 10^5$ random coordinates sampled from the $L$ =400cMpc parent volume. The distribution of distances to ≥MW, ≥group, and ≥cluster mass objects are shown in orange, pink, and purple respectively and the median distance to each object is shown by the vertical dotted lines. The exact mass ranges for these objects are defined in table 3.1 and distances were calculated using the coordinates of the most bound particle of haloes.

This result is further illustrated in Figure 3.3 which shows the distribution of overdensities in spherical regions with radii of $r = 5\,$cMpc and $r = 10\,$cMpc. I calculate the overdensity, $\delta_{\rm region}$, of a spherical region as,

$$\delta_{\rm region} = \frac{\rho_{\rm region}}{\bar{\rho}} \tag{3.1}$$

where $\bar{\rho}$ is the cosmic mass density, defined relative to the critical density $\rho_{\rm crit}$ ($\bar{\rho} = \Omega_{\rm m}\rho_{\rm crit}$), and $\rho_{\rm region}$ is the mean mass density enclosed by a sphere of radius $r$,

$$\rho_{\rm region} = \frac{m_{\rm region}}{\frac{4}{3}\pi R^3_{\rm region}} \tag{3.2}$$

Note that $\delta$ is often used to represent the *overdensity*, i.e. $\delta = (\rho - \bar{\rho})/\bar{\rho}$, but here I use $\delta = \rho/\bar{\rho}$ because the large dynamic range of densities we explore is conveniently compressed with $\log_{10}$, requiring that $\delta \geq 0$.

For spheres of a larger radius of $r$=10cMpc, regions containing precisely one or no ≥MW

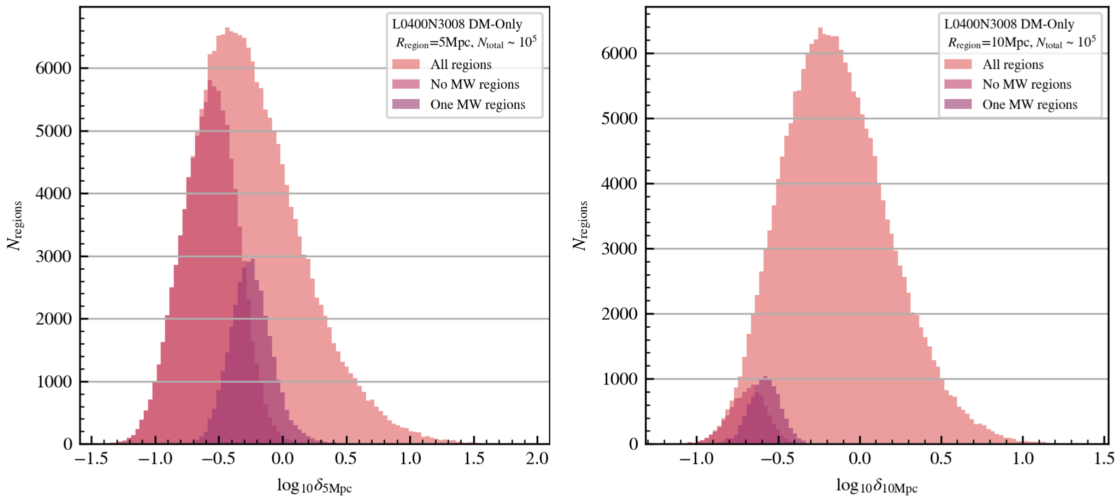


FIGURE 3.3: *Left:* The distribution of mean enclosed overdensity $\delta_{\rm 5Mpc}$ within $N_{\rm total} = 2 \times 10^5$ randomly placed spheres of radius $r$ = 5cMpc from the $L$ = 400Mpc periodic volume. The orange distribution represents all of the regions sampled, whereas the pink and purple distributions represent subsamples of regions that enclose 0 or 1 MW-mass object. Regions that are devoid of MW-mass objects are naturally biased to low overdensities whilst maintaining a large range. *Right:* The same as the left but the overdensity is calculated for a sphere of radius $r$ = 10cMpc. The number of regions that enclose 0 or 1 MW-mass object, and the subsequent diversity in overdensity of these subsamples, is greatly reduced compared to spheres of $r$ = 5cMpc.

objects represent a very small subsample of the total population which is highly biased towards very low overdensity voids with little diversity. These regions are therefore rarer and less representative of the diversity of isolated environments that dwarf galaxy populations form in. Using a radius of $r$=5cMpc whilst imposing a criterion of one or no $\geq$MW objects in the region retains a large sample of regions that sample a range of overdensities well.

Though the precise size of the zoom-in region is not critical, I find that the choice of spheres of $r$ = 5cMpc yields a good balance between probing the 'extremes' of the density distribution populated by isolated dwarf galaxies, whilst ensuring that even the most underdense regions yield a meaningful population of well-sampled galaxies. A related benefit of simulating relatively small volumes is that they can usually be accommodated within the memory footprint (512 gigabytes to 1 terabyte) of the individual nodes that comprise many modern high-performance computing facilities, simplifying parallelism.

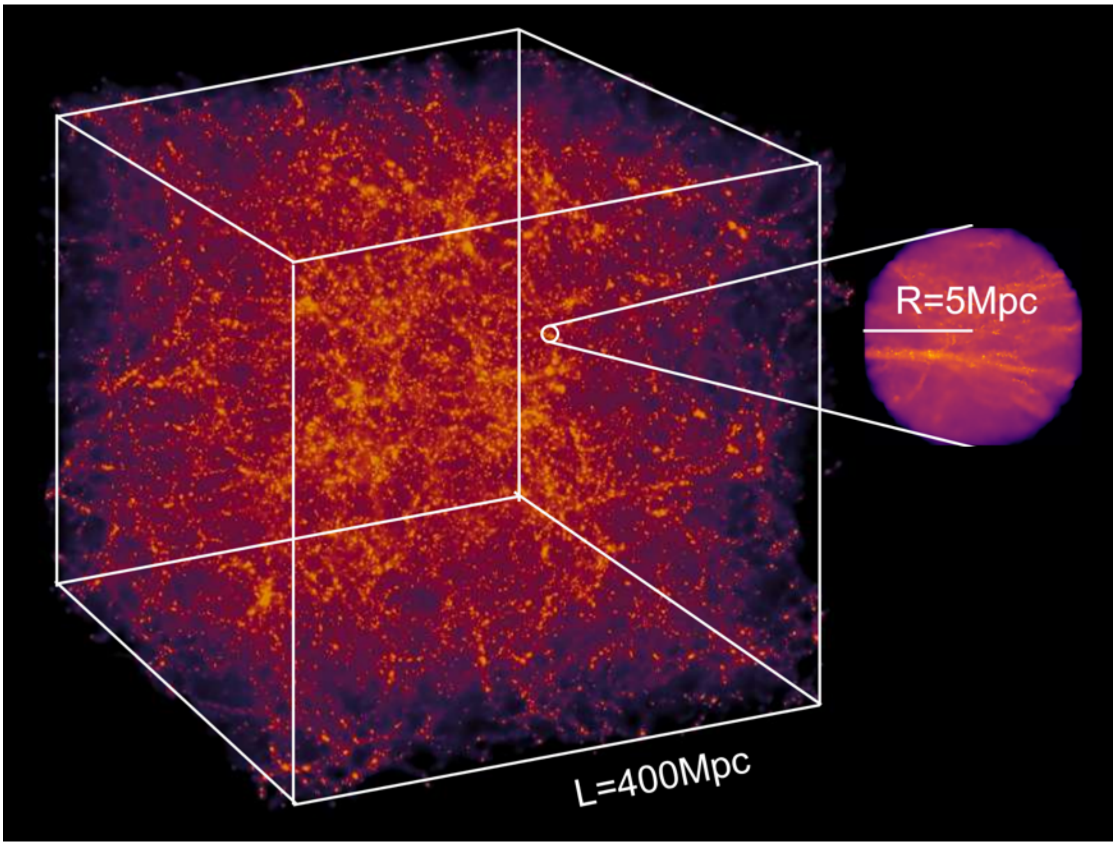


Figure 3.4: A projection of the surface mass density of the $L = 400$Mpc parent volume and an example $r = 5$Mpc spherical zoom-in region.

### 3.1.3 Distance Isolation Criteria

As well as excluding MW-mass objects from the high resolution region of interest, it is also desirable to exclude them, and more massive objects, from the immediate boundary and local surroundings of the zoom-in region. I investigate the environments based on their properties at z=0, having further isolation criteria ensures that the objects in the region are unlikely to have interacted with more massive objects prior to z=0. This reduces the possibility of having backsplash galaxies that have once resided inside a cluster, have been subject to ram pressure stripping, dynamical friction, and tidal forces, and have migrated to the clusters outskirts. Additionally, the distances used to ensure regions are devoid of massive objects are computed to halo centres meaning I have not excluded massive haloes whose virial radius overlaps with the zoom-in region. It is computationally inefficient to simulate massive objects that will not be included in the results if they are formed from lower resolution boundary particles. I therefore limit the distances to MW-mass objects from region centres to be $d_{\mathrm{centre,>MW}}$ >6Mpc and objects more massive than groups $d_{\mathrm{centre},>\mathrm{group}}$ >10Mpc which correspond to the approximate average distances to such objects computed in Figure 3.2.

These distance constraints are shown in Figure 3.5 which illustrate the overdensities of regions excluded by this criterion. In general, as the overdensity of a region decreases, the further a region centre is from a >group or >MW mass object and hence the regions excluded by this distance criteria are predominantly higher overdensity regions which are less suitable regions for the aims of this thesis. Moreover, this allows us to examine environmental effects on isolated field dwarf galaxies without the influence of a massive nearby object and have some resemblance to isolated dwarf galaxies around the Local Group which lies at a mean distance of 16.5 ± 0.1Mpc (Mei et al., 2007) from the Virgo cluster.

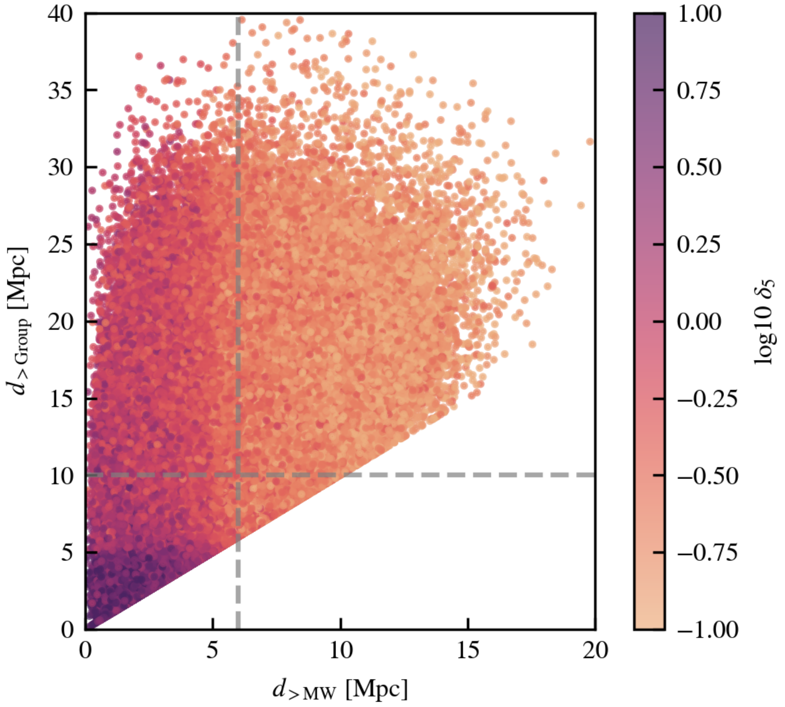


FIGURE 3.5: The distance from a region centre to a >MW mass object against the distance to a >group mass object coloured by the overdensity enclosed in a $r = 5$Mpc sphere from that centre. The distance selection criteria ($d_{>\mathrm{MW}} > 6$Mpc and $d_{>\mathrm{group}} > 10$Mpc) for isolated zoom-in regions naturally excludes high overdensity regions. The exact mass ranges for these objects are defined in table 3.1 and distances were calculated using the coordinates of the most bound particle of haloes in the $L = 400$Mpc parent simulation.

### 3.1.4 Overdensity Variation

Following the distance isolation criteria, I investigate the range of overdensities that form in the remaining regions. The overall density distribution of the $2 \times 10^5$ randomly centred spheres of $r = 5$cMpc is shown as the lightly-shaded histogram in Figure 3.6. Note that this unbiased sample yields, for a sufficiently large sample of regions, a mean density of $\bar{\rho}$, but as a volume-weighted measure the distribution is not symmetric, and the median and peak values are less than unity ($\delta_5^{\rm median} = -0.304$, $\delta_5^{\rm peak} \simeq -0.434$).

The darker, foreground histogram shows the density distribution of the subset of regions that are devoid of haloes with mass comparable to or more massive than that of the Milky Way, and which are not in close proximity to a massive halo that would host a galaxy group or cluster. As determined in Section 3.1.3, I require specifically that there are no haloes of dynamical mass $M_{200} > 7.0 \times 10^{11}\,{\rm M}_{\odot}$ within 6 Mpc of the region centre, and no halo of mass $M_{200} > 10^{13}\,{\rm M}_{\odot}$ within 10 Mpc of the centre. These criteria are satisfied by $5.8 \times 10^4$ of the initial $2 \times 10^5$ regions, and are necessarily biased towards lower densities, yielding mean, median and peak densities of -0.653, -0.645 and -0.568, respectively. Around the median, there is a significant spread in the density of environments that form isolated populations of low mass haloes, including densities exceeding the mean density of the full sample. To fully explore the diversity of these environments I select regions from the $\pm 2\sigma$ of this distribution. The vertical bands on Figure 3.6 denote the densities corresponding to the (-2, -1, 0, +1, +2)$\sigma$ percentile rank ordering of a Gaussian distribution, with values $\delta = (-1.064, -0.859, -0.645, -0.449, -0.282)$, respectively. The width of the bands corresponds to 1 percentile.

### 3.1.5 Probing the High Mass End of Dwarf Galaxies (N(LMC))

To add another dimension of diversity in environment to the region selection, I examine the number of low mass haloes within a mass range of $7 \times 10^{10} \leq M_{\rm 200,crit} \leq 5 \times 10^{11} {\rm M}_{\odot}$, approximately the mass of the Large Magellanic Cloud (LMC), that reside within each region. I refer to this parameter as $N_{\rm LMC}$ and it probes the high mass end of the halo mass function for the subsample of underdense regions.

Despite its inherent bias towards low density regions, the sample of regions with no MW-mass haloes nevertheless exhibits significant diversity in the number of haloes enclosed

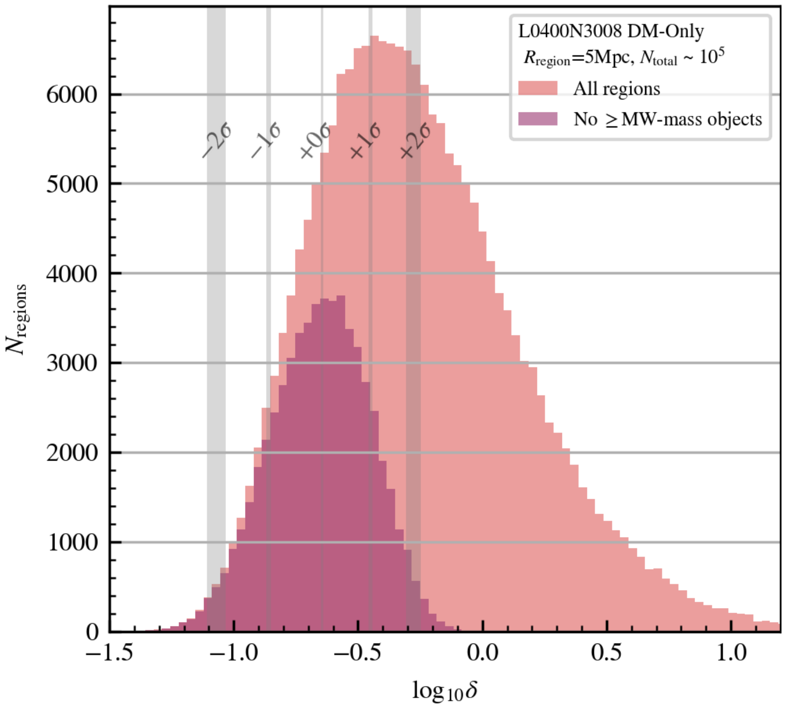


FIGURE 3.6: The mean enclosed overdensity $\delta_5$ distribution of spheres with $r = 5$Mpc, $\delta_5 = \rho(< 5\text{Mpc})/\rho_{\rm mean}$, selected from the $L = 400$Mpc parent simulation. The lighter distribution in the background corresponds to the densities in $2 \times 10^5$ random spheres, and the darker histogram shows the subsample of them that contain no objects with $M_{200,\rm crit} > 7 \times 10^{11}\rm M_\odot$ and follow the distance isolation criteria. The vertical bands mark the mark the median, $+/-1\sigma$ and $+/-2\sigma$ percentiles of the latter distribution and their width corresponds to 1 percentile around these values.

that have total mass comparable to that of the Large Magellanic Cloud (LMC). Such haloes are amongst the most massive permitted by our selection criteria, so the diversity is necessarily greatest for those regions at the upper end of the density distribution. Figure 3.7 shows a 2-dimensional histogram of the number of regions as a function of density, $\delta$, and the number of LMC-mass haloes, $N_{\rm LMC}$. $N_{\rm LMC}$ correlates positively with density, varying between 0 and 22 in the $5.8 \times 10^4$ regions defining our MW-free sample, but with large scatter at fixed density. I characterise the diversity of $N_{\rm LMC}$ as a function of density for the MW-free sample of regions by computing, for haloes within the 1 percentile-wide density bins denoted by the vertical bands (repeated from Figure 3.6), the values of $N_{\rm LMC}$ corresponding to the Gaussian (-2, -1, 0, +1, +2)$\sigma$ percentile rank ordering. These values are denoted by white dots overlaid on the vertical bands (where fewer than five dots are visible, two or more of the lowest-$\sigma$ percentiles correspond to $N_{\rm LMC} = 0$). For example, in the regions comprising the +2$\sigma$ density sample, the number of LMC-like haloes corresponding to (-2, -1, 0, +1, +2)$\sigma$ are $N_{\rm LMC} = (4, 6, 8, 11, 13)$, respectively.

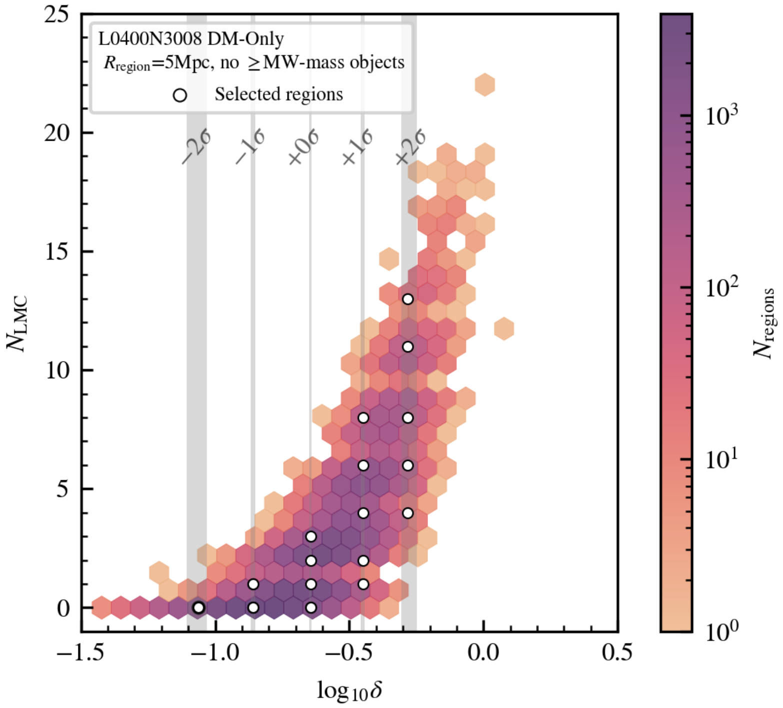


FIGURE 3.7: The number of LMC-mass objects ($M_{\rm halo} \sim 10^{11}{\rm M}_{\odot}$) in spherical regions of $r = 5$Mpc, as a function of the regions' mean overdensity, $\delta_5$. The regions are devoid of massive haloes ($M_{\rm 200,crit} > 5 \times 10^{11}{\rm M}_{\odot}$) and follow the distance isolation criteria. The pixels are colour coded by the number of regions. The vertical bands are the same as in Fig.3.6, and the white circles indicate (-2, -1, 0, +1, +2)$\sigma$ of the $N_{\rm LMC}$ distribution at the fixed overdensities marked by vertical lines. Additionally, the white circles mark the final selection of regions which overlap where there are more than one region.

### 3.1.6 Final selection

The final selection of regions is completely devoid of objects more massive than the MW and follows isolation criteria for the distance to a ≥MW-mass object ($d_{\rm MW} > 6$Mpc) and ≥group mass object ($d_{\rm group} > 10$Mpc). From the remaining regions, I sample the full range of the overdensity distribution by choosing regions from the (-2, -1, 0, +1, +2)$\sigma$ percentiles (vertical lines in Figure 3.6 and 3.7). At each of the 5 overdensities, I select 5 regions that vary in $N_{\rm LMC}$ and span the (-2, -1, 0, +1, +2)$\sigma$ percentiles of that distribution. These are represented by the white circles in Figure 3.7, and overlap where the $\sigma$ percentiles are the same. These form the final suite of 25 spherical volumes, with $r$=5cMpc, that comprise the Columba zoom-in simulations. I summarise the selection procedure in Figure 3.8. The centre coordinates of the regions in the parent volume at z=0 that were selected for re-simulation are given in Table 3.2.

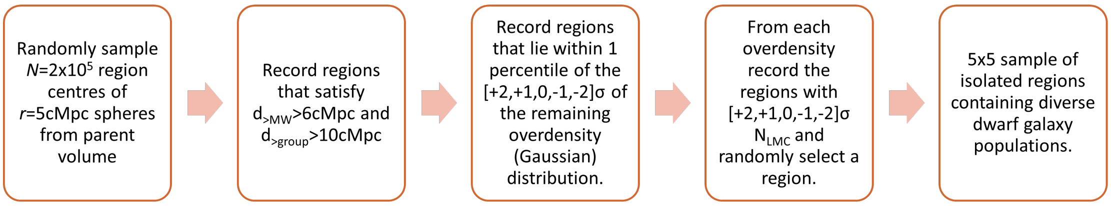


FIGURE 3.8: The final methodology for selecting the zoom-in regions for the Columba simulations; from the distance criteria to massive objects, sampling diverse overdensities, and sampling the range of LMC-like objects enclosed in $r = 5$Mpc spheres.

The projected surface mass density in these 25 regions[2] are shown in Figure 3.9. Each row represents a fixed overdensity, and the overdensity increases from the top to bottom. The rows, from top to bottom, correspond to (-2, -1, 0, +1, +2)$\sigma$ in density, and the columns, from left to right, correspond to (-2, -1, 0, +1, +2)$\sigma$ in $N_{\rm LMC}$. The increasing average density from top to bottom is clearly evident with regions in the top row resembling voids with very little structure, and richer and more complex cosmic structures appearing as I sample higher overdensity regions.

## 3.2 Initial Conditions

Generating initial conditions for zoom-in cosmological simulations involves constructing multi-resolution initial conditions that captures the region of interest in high resolution whilst simulating the large scale structure in coarser resolution. Here I describe the full process for generating the zoom-in initial conditions for the Columba simulation suite from masking the selected region of interest, generating a multi-resolution particle load, generating the gas particles for hydrodynamical simulations, and imposing the Gaussian random fields with a specified power spectrum. I generate ICs for all 25 selected regions at two resolutions, corresponding to high-resolution particle masses of $\sim 10^5 \mathrm{M}_\odot$ (‘m5’) and $\sim 10^6 \mathrm{M}_\odot$ (‘m6’), to enable convergence testing, the full details of each resolution is given in Table 3.3. The generation of these zoom-in initial conditions was based on the pipeline developed by Yannick Bahé[3].

---

[2] For superior clarity we show the surface mass density from the resulting zoom simulations, rather than the low-resolution parent volume.

[3] The repository of the pipeline developed by Yannick Bahé can be accessed via https://github.com/ymbahe/zoom-setup

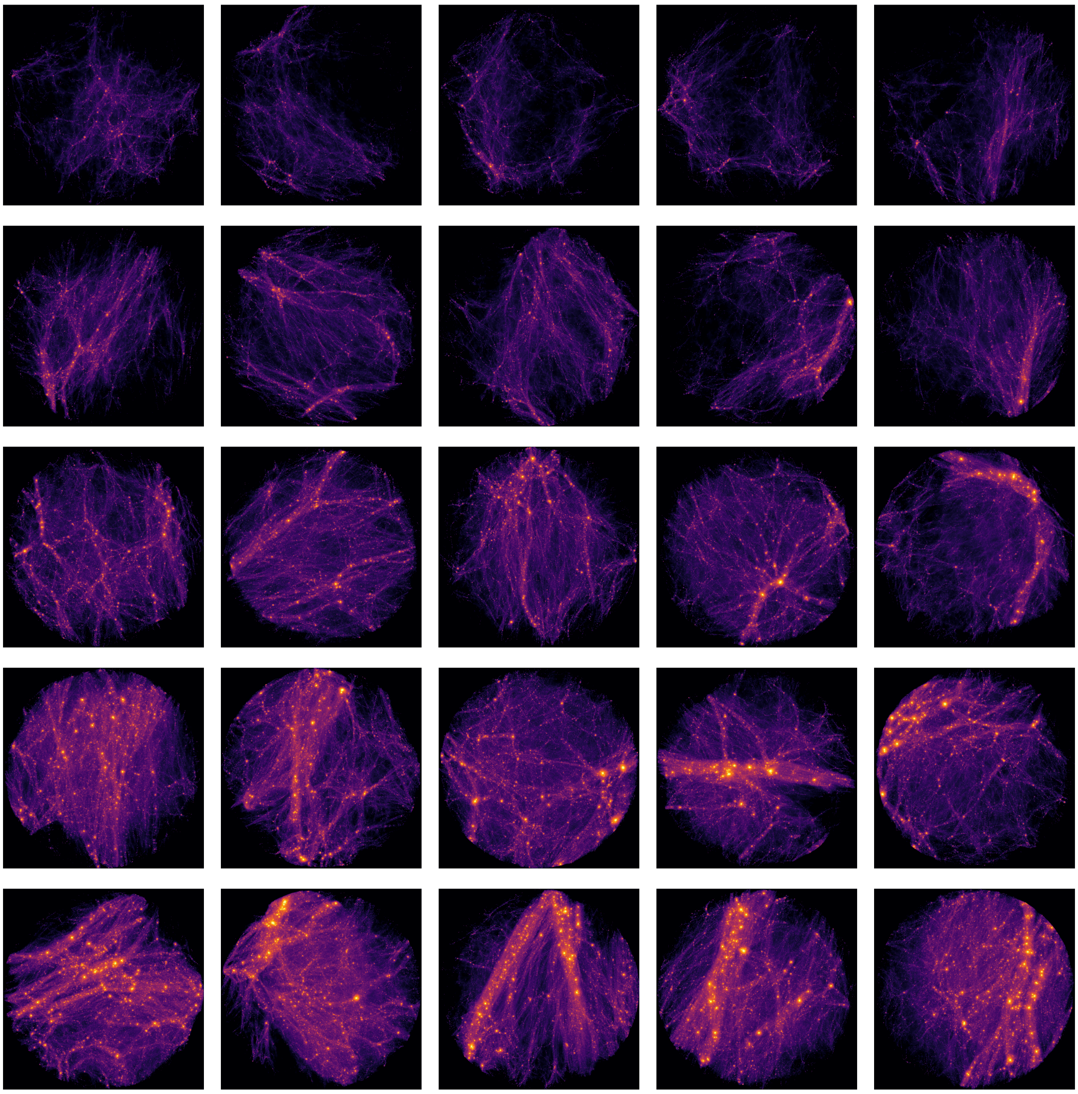


Figure 3.9: Surface mass density projections of the 25 selected spherical regions of $r = 5\,\mathrm{cMpc}$. From top to bottom each row corresponds to the $(-2, -1, 0, +1, +2)\sigma$ density band, respectively. From left to right, each column corresponds to the $(-2, -1, 0, +1, +2)\sigma$ value of $N_{\mathrm{LMC}}$. For example the bottom row shows the highest density regions that we simulate (the $+2\sigma$ sample, with median density $\log_{10}\delta_5 = -0.282$ and these regions exhibit, from left to right, $N_{\mathrm{LMC}} = (4, 6, 8, 11, 13)$, respectively.

### 3.2.1 Mask Generation

Once the region of interest has been selected, the first step in the initial condition generation is to create a mask that defines the region of the parent volume in the initial unperturbed volume that is to be realised at a higher resolution than the rest of the volume. The mask is constructed to contain all particles that form the high resolution region at low red-shift and to ensure that the final z=0 region contains no low resolution particles. Furthermore, it is useful to shield the high-resolution region

| Volume | $X_1$ [Mpc] | $Y_1$ [Mpc] | $Z_1$ [Mpc] |
|---|---|---|---|
| $(-2\sigma\delta)_1$ | 22.554 | 61.397 | 182.388 |
| $(-2\sigma\delta)_2$ | 285.056 | 380.053 | 5.246 |
| $(-2\sigma\delta)_3$ | 170.254 | 49.515 | 378.008 |
| $(-2\sigma\delta)_4$ | 252.884 | 61.142 | 226.887 |
| $(-2\sigma\delta)_5$ | 303.710 | 387.559 | 259.295 |
| $(-1\sigma\delta)_6$ | 247.260 | 223.085 | 118.065 |
| $(-1\sigma\delta)_7$ | 360.984 | 370.986 | 104.061 |
| $(-1\sigma\delta)_8$ | 113.541 | 105.269 | 295.703 |
| $(-1\sigma\delta)_9$ | 202.167 | 22.568 | 326.103 |
| $(-1\sigma\delta)_{10}$ | 327.347 | 244.685 | 199.035 |
| $(+0\sigma\delta)_{11}$ | 206.470 | 289.345 | 337.098 |
| $(+0\sigma\delta)_{12}$ | 164.727 | 370.529 | 389.946 |
| $(+0\sigma\delta)_{13}$ | 60.793 | 270.998 | 111.084 |
| $(+0\sigma\delta)_{14}$ | 191.408 | 185.739 | 41.606 |
| $(+0\sigma\delta)_{15}$ | 382.208 | 57.656 | 160.808 |
| $(+1\sigma\delta)_{16}$ | 211.498 | 78.734 | 381.412 |
| $(+1\sigma\delta)_{17}$ | 77.794 | 100.334 | 185.028 |
| $(+1\sigma\delta)_{18}$ | 236.746 | 148.260 | 82.034 |
| $(+1\sigma\delta)_{19}$ | 124.015 | 262.135 | 382.086 |
| $(+1\sigma\delta)_{20}$ | 356.896 | 181.048 | 185.781 |
| $(+2\sigma\delta)_{21}$ | 134.931 | 23.480 | 373.677 |
| $(+2\sigma\delta)_{22}$ | 96.567 | 46.163 | 314.870 |
| $(+2\sigma\delta)_{23}$ | 101.094 | 182.843 | 370.797 |
| $(+2\sigma\delta)_{24}$ | 211.361 | 265.825 | 313.522 |
| $(+2\sigma\delta)_{25}$ | 51.271 | 95.723 | 98.979 |

TABLE 3.2: The positions of the centres of the spherical zoom-in regions ($r = 5$Mpc) at z=0 in the $L = 400$Mpc parent simulation. The naming convention follows the overdensity of the region and region ID, for example, $(+0\sigma\delta)_{13}$ is volume 13 characterised by the median $(+0\sigma\delta)$ overdensity from Figure 3.6 and median $(+0\sigma)$ $N_{\rm LMC}$ from Figure 3.7 (projected in the centre row and centre column of Figure 3.9).

| Resolution | $m_{\rm b}$ [M$_\odot$] | $\epsilon_{\rm com,b}$ [ckpc] | $\epsilon_{\rm max,b}$ [pkpc] | $m_{\rm DM}$ [M$_\odot$] | $\epsilon_{\rm com,DM}$ [ckpc] | $\epsilon_{\rm max,DM}$ [pkpc] |
|---|---|---|---|---|---|---|
| m6 | $1.84 \times 10^6$ | 1.790 | 0.700 | $2.42 \times 10^6$ | 2.091 | 0.819 |
| m5 | $2.30 \times 10^5$ | 0.895 | 0.350 | $3.02 \times 10^5$ | 1.0455 | 0.4095 |

TABLE 3.3: The target masses of gas and DM particles at each of the three resolution levels we adopt, with corresponding Plummer-equivalent gravitational softening ($\epsilon_{\rm com}$) length and maximum proper value ($\epsilon_{\rm max}$).

from lower resolution boundary particles by constructing a larger mask with a padding region which I discuss in Section 3.2.1.1; this buffer layer of high resolution padding particles means that undesirable artefacts, such as an artificial pressure gradient, at the boundary of the high resolution region do not affect the region of scientific interest. Note that during the initial condition generation the boundary particles are treated the same as high resolution particles in the region of interest.

The mask is generated via the following steps:

1. The spherical region ($r$ =5cMpc) of interest of the parent volume is defined and all of the enclosed particles are identified as tagged particles. A projection of the particles selected for the mask of volume $(+0\sigma\delta)_{13}$ is illustrated in Figure 3.10.
2. The tagged particles are traced back to the initial conditions using their particle IDs and their Lagrangian coordinates are computed.
3. A bounding box is constructed that encompasses all the tagged particles, including extra padding, and is split into cubic cells. The size of the cubic cell can be controlled by setting the maximum acceptable size of the mask cells, I use $L_{\text{cell,max}} = 2$ mean inter-particle spacings.
4. All cells that are near a particle are tagged, these cells form the mask. This can be controlled by setting the minimum number of particles that must lie within a cell in the Lagrangian coordinates for it to count as active, I use $N_{\text{min}} = 1$. A visualisation of the mask is shown in Figure 3.11.
5. Shape regularisation is applied to the full mask to fill holes in the mask and to round the edges of the mask.
6. The coordinates of active mask cells are written to the mask file.

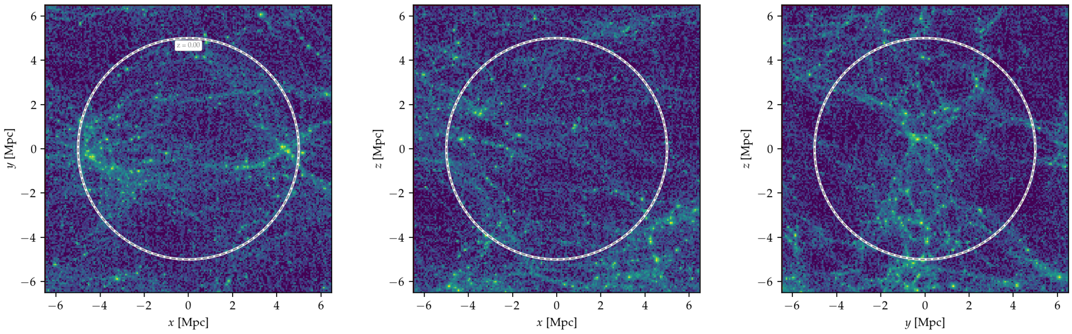


FIGURE 3.10: Projections of the z=0 zoom-in region $(+0\sigma\delta)_{13}$. The dashed circle indicates the sphere in the parent volume in which enclosed particles are tagged and used for making the mask in initial condition generation.

#### 3.2.1.1 Boundary Particle Contamination

The red markers in Figure 3.11 show the mask region occupied by padding particles. Padding particles are particles that surround the region of interest that are simulated at high resolution and the size of this padding area can be tuned to minimise the amount of contamination from lower resolution particles in the region of interest.

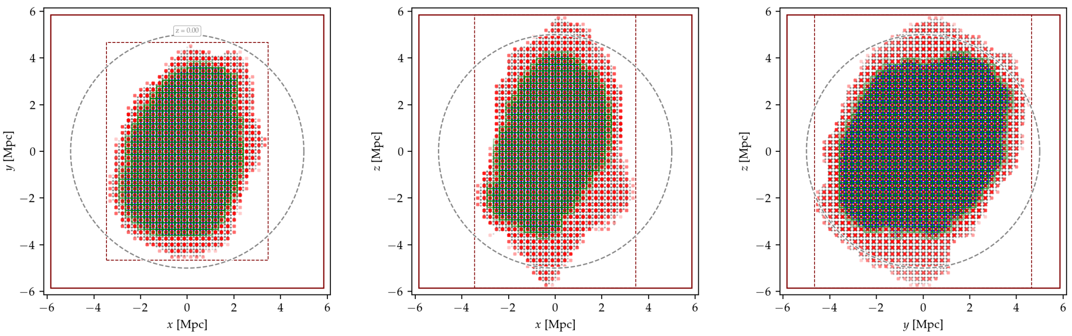


FIGURE 3.11: Coordinate projections of the mask for the zoom-in region $(+0\sigma)_{13}$. The dashed circle indicates the high-resolution sphere at z=0. The solid maroon square show the cubic and cuboidal bounding box used for computing the mask which contain the high resolution region cells and the entire mask volume, including the padding region. The mask region is shown in red and the target, high resolution region is overlaid in green forming the dark central region of the mask. The opacity effectively indicates how extended the mask is along the line of sight (generally faintest along the edge of the padding region, where it's often only one cell along the entire line of sight).

I investigated the optimal amount of padding for this simulation suite to minimise the contamination of the high resolution region whilst balancing the computational expense of simulating a larger high resolution region. For simplicity, I implement the padding as a uniform shell that extends the desired high resolution sphere to a larger radius from the region centre at z=0. I ran tests that systematically varied the padding radius from 0.5 to 2.5Mpc and examined the number of low resolution particles in the high resolution region of interest at z=0. In Figure 3.12 I show the number of boundary particles as a function of distance from the region centre at z=0 for simulations of all $(+0\sigma\delta)$ regions at z=0 with boundary particles in shells of $r = 1.0, 1.5,$ and 2.0Mpc from left to right. I find a shell of $r = 1.5$Mpc is sufficient to prevent significant contamination.

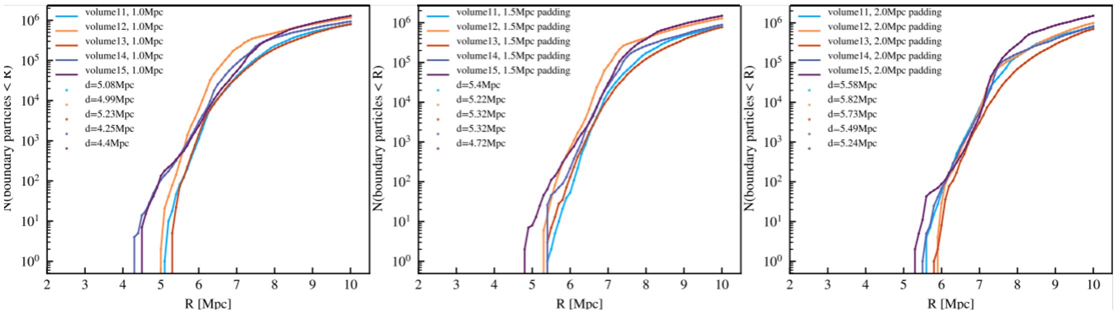


FIGURE 3.12: The number of low resolution boundary particles enclosed as function of radius for a padding shells of $r = 1.0, 1.5,$ and 2.0Mpc from left to right for simulations of $(+0\sigma\delta)$ regions at z=0.

### 3.2.2 Particle Distribution

The particle load represents the unperturbed Universe at z→ ∞, before it is modified to represent the perturbed Universe at finite redshift in the initial conditions. For periodic boxes, creating the particle load can be as simple as generating a uniform, periodic lattice of particles, however, for zoom-in simulations the structure of the particle load takes a more complex form with high resolution particles confined to the region defined by the mask and gradually reducing the resolution, and hence increasing the particle masses, in the surrounding volume.

After defining the mask region, a new unperturbed particle distribution must be generated at the desired, higher resolution and ideally free of any net gravitational forces, the following method for generating the particle load was used. The simulation box is centred to the mask region and is shifted back to its original coordinates at the end of the particle load generation. A cubic region, or gcube, that encompasses the mask region is overlaid and is divided into uniform cubic cells (gcells) depending on the desired resolution and the number of particles to be assigned to each cell. Each cell of the gcube that intersects with a mask cell is assigned to Zone I which is populated by the highest resolution particles, the remaining cells of the gcube form Zone II and contain progressively fewer particles per cell the further the cell is from Zone I. The remaining volume, Zone III is populated with concentric cubic cells that increase in size outwards, the overall setup of this configuration is illustrated in Figure 3.13.

The particle load of the gcube is constrained by the mask, the desired resolution of the zoom-in region, and the particle load per gcell. The resolution of the high resolution region is quantified by the number of particles per dimension that the simulation would have if the parent volume was realised at this high resolution, $n_{\mathrm{equiv}}$. The particle load per gcell is the number of particles contained within one cell denoted $n_{\mathrm{cell}}$. Combined, the size of one gcell, $l_{\mathrm{cell}}$, is computed by $l_{\mathrm{cell}} = L_{\mathrm{box}} n_{\mathrm{cell}} / n_{\mathrm{equiv}}$ and fit to encompass the full mask including the padding region.

#### 3.2.2.1 Zone I: Generating Gas

For DM only zoom simulations, the Lagrangian region can be modelled using a uniformly spaced grid of particles and for hydrodynamical simulations using SPH, the addition of

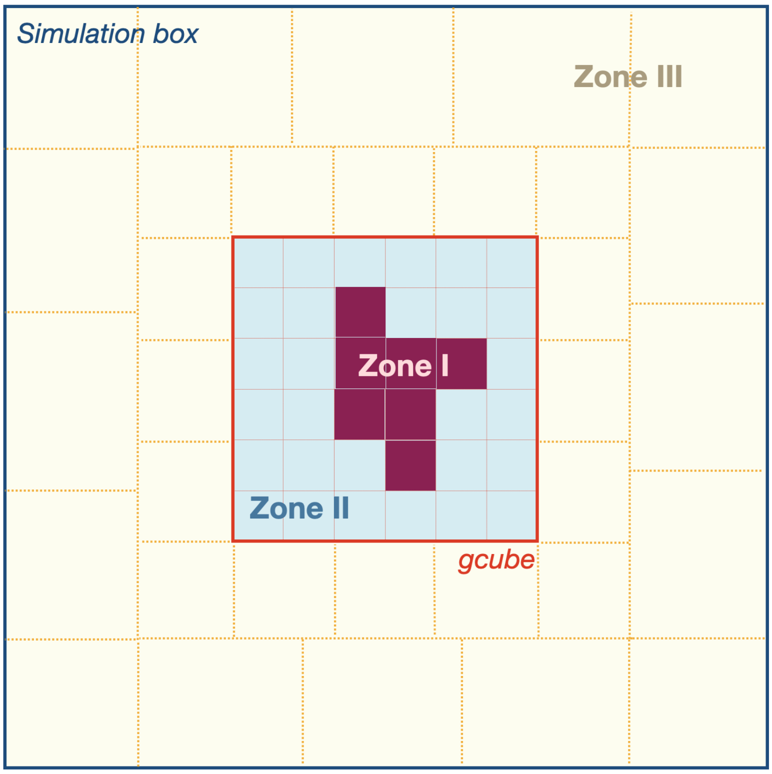


FIGURE 3.13: The cell structure of the particle load by Yannick Bahé. Zone I is formed of gcube cells that intersect with mask cells and is populated by high resolution particles. The remaining gcube, Zone II, is populated by nearby boundary particles of gradually decreasing resolution. The remaining volume, Zone III, is populated by distant boundary particles in a nested concentric shell structure.

gas is often implemented by splitting each particle into a gas particle and DM particle. The split pair of particles have a total mass equal to the DMO particle and are configured to have a centre of mass at the position of the DMO particle. This setup yields equal numbers of gas and DM particles whose masses are set by the cosmological parameters, specifically the particle mass ratio is given by $m_{\rm DM}/m_{\rm gas} = \Omega_{\rm DM}/\Omega_{\rm b}$. This is implemented for many cosmological simulations which have therefore had particles of different species with different masses, specifically, DM particles that are more massive than their gas particle counterparts.

Interactions between particles are subject to 2-body scattering and Ludlow et al. (2019) show that by using unequal mass particles, the spurious transfer of kinetic energy to the less massive species has a significant effect on the interpretation of galaxy sizes. Stars are treated as collisionless particles and are subject to 2-body scattering when collisions cannot be ignored, leading to energy equipartion in systems of unequal mass, as is the convention in many cosmological simulations. Using simulations which differ only in the

DM particle mass, Ludlow et al. (2019) show that galaxy sizes at z=0 (quantified by their projected half-mass radii $R_{50}$) are systematically increased for low mass systems ($M_{\rm star} < 2000 M_{\rm gas}$) due to the diffusion of stellar particles out of the dense central regions of galaxies. This effect can be minimised by increasing resolution, or using similar mass particles to represent different species of particles which can be achieved by varying the ratio of the number of DM to gas particles.

SWIFT gives the option to directly generate gas particles, however, in order to assign different numbers of gas and DM particles this must be done during the particle load stage in initial condition generation. Therefore, these initial conditions are set up to contain 4 times the number of DM particles than gas particles, this yields a particle mass ratio of $m_{\rm DM}/m_{\rm gas} = \Omega_{\rm DM}/4\Omega_{\rm b}$ and hence reduces the effects of spurious heating from the transfer of energy between unequal mass particles.

The extra DM particles are configured in the gcell using a replicative assignment, a similar technique is discussed for a DM to gas particle number ratio of 7 in Richings et al. (2021). For a number ratio of 4, a face replicative assignment is used as follows: 1 DM particle is offset by [+0.5, +0.5, +0.5] mean inter-particle spacings for a body-centred cubic lattice (this method is implemented by Schaller et al. (2023) in SWIFT), the remaining 3 DM particles are offset by +0.5 mean inter-particle spacings along the x, y, and z axes respectively. Starting with a grid distribution, this assignment is designed to maintain a high degree of symmetry so the forces on each particle cancel out. The gcells are then tiles within Zone I which form the particle load of the high resolution zoom-in region.

#### 3.2.2.2 Zone II and III

Beyond Zone I, the resolution is gradually decreased and populated by only DM particles. Zone II is populated by reducing the reducing the number of particles within each gcell as a function of their distance to Zone I. I set the desired mass increase in Zone II gcells per Mpc distance from Zone I to 1.0 with a minimum particle mass ratio between Zone II to Zone I of 1.5.

The volume beyond the gcube, that encompasses the majority of the parent volume, is configured in a nested-shell-structure and denoted Zone III. Zone III is populated by the more distant boundary particles; these particles are arranged in a series of self-similar cubic shells with a gradual increase in mass. The nested-shell-structure is setup such that each shell encloses the lower level shell within it and each shell has the same number of cells, resulting in a ratio of cell sizes between shells given by $n_{\rm cells}/(n_{\rm cells}-2)$. The nested-cell-structure most often will not naturally align with the edge of the parent simulation box. Therefore, the outermost shell is allowed to increase or decrease the number of cells by 10 for better alignment.

### 3.2.3 Fluctuations

I generate multi-resolution 'zoomed' initial conditions of our regions embedded within the parent volume using the second-order Lagrangian perturbation theory software IC_2LPT_GEN (Jenkins, 2010), coupled to PANPHASIA. The IC_2LPT_GEN software assumes the cosmic matter density is comprised only of baryons and DM, therefore the zoom-in ICs are generated with cosmological parameters that differ very slightly from those of the parent box, subsuming the matter density of massive neutrinos into that of the dark matter. The cosmological parameters of the zoomed ICs are therefore: $\Omega_{\rm CDM} = 0.256, \Omega_{\rm b} = 0.0486, \Omega_\Lambda = 0.695, h = 0.681$, and$\sigma_8 = 0.807$. I also make the simplifying assumption that baryons and DM follow the same total matter power spectrum, which is justified as we focus on spatial scales shorter than that on which the baryon power spectrum is suppressed (e.g. Semboloni et al., 2011). I generate the linear power spectrum with these parameters using CLASS. The $z = 127$ density perturbations are applied by displacing the particles from their unperturbed coordinates using the method described by Power et al. (2003).

## 3.3 SWIFT

SWIFT (SPH With Inter-dependent Fine-grained Tasking) (Schaller et al., 2023) is a fully open-source coupled hydrodynamics, gravity, cosmology, and galaxy formation code that was designed with a hydrodynamics-first approach to maximise the performance of galaxy formation simulations and scalability of the code for use on supercomputer architecture. This is achieved by balancing the workload using task-based parallelism within compute nodes.

Astrophysics simulation codes in the past have typically used a ‘branch-and-bound’ approach to parallelisation whereby operations are carried out in a pre-specified order and each compute unit performs the same operation simultaneously. SWIFT utilises task-based parallelisation which has 2 major performance advantages; firstly, pre-defining the distribution of workload can lead to a work imbalance whereby small parts of the volume dominate the total work. Task-based parallelisation allows the task scheduler to dynamically assign fewer cells (groups of particles) to an individual computing unit if they turn out to have a high computational cost, and vice versa, meaning the workload is dynamically balanced over all the available compute cores. Secondly, task-based parallelisation allows gravity and hydrodynamics calculations to be performed at the same time, this means SPH calculations can begin before the gravity operations have completed, and vice versa, reducing the total idle time (Schaller et al., 2023).

### 3.3.1 Gravity and Hydrodynamics

The zoom-in ICs for the columba suite of simulations were evolved to $z = 0$ using SWIFT. As per the collisionless simulation of the parent volume, gravitational forces were solved using a split fast multipole plus particle-mesh approach. The hydrodynamical equations of motion of the gas were solved using the SPHENIX (Borrow et al., 2021) density-energy formulation of smoothed particle hydrodynamics (SPH) adopting the quartic spline kernel with 65 effective neighbours within the kernel support radius. The SPH solver includes artificial viscosity and conduction terms, with in-built limiters to mitigate numerical energy losses. The time-step limiter of Durier & Dalla Vecchia (2012) is used to ensure accurate evolution of the fluid even during extreme shock events. Time integration uses a leapfrog scheme in which a particle’s timestep is the minimum of its gravitational timestep or, if applicable, that dictated by the hydrodynamical Courant-Friedrichs-Lewy condition with parameter $C_{\rm CFL} = 0.2$. The gravitational timestep of a particle with acceleration $a$ is $\Delta t = (0.025\epsilon/a)^{1/2}$, where $\epsilon$ is the gravitational softening length. The timesteps of gas particles cannot exceed that of any of their neighbours by more than a factor of 4. All high-resolution particles adopt the same softening length (see Table 3.3), and the smoothing length is limited to a minimum $h_{\rm min}/\epsilon_{\rm gas} = 10^{-5}$.

Stellar and BH particles that form during the simulations do not experience hydrodynamical forces but use an SPH-like kernel to find their neighbours during the operation of subgrid models. The mass of gas particles can increase due to the donation of mass by neighbouring stellar populations or decrease due to the accretion of mass by neighbouring BHs. Gas particles that exceed their initial mass by a factor of 4 are split into two particles of equal mass, whilst any reaching half of their initial mass become ineligible for mass accretion onto BHs.

## 3.4 Galaxy Formation Model

The simulations appeal to subgrid models to approximate the macroscopic effects of unresolved physical processes. The models used for the Columba simulations are either ported directly from the EAGLE simulations, or represent an evolution of the equivalents used by EAGLE. In this sub-section I describe the models used to treat radiative cooling and heating, star formation and the ISM, the evolution of stellar populations and Type Ia SNe, feedback from young stars and core collapse SNe, the seeding and growth of BHs and feedback from AGN. These models are implemented into the SWIFT engine.

### 3.4.1 Radiative Cooling and Heating

Radiative cooling and heating rates for hydrogen, helium, free electrons from these species and $H_2$ molecules are computed with the non-equilibrium thermochemistry solver CHIMES (Richings et al., 2014a,b). There is no artificial pressure floor, and gas is allowed to cool to a minimum temperature of 10 K. The model explicitly tracks the radiative cooling and heating of 9 metal species (C, N, O, Ne, Mg, Si, S, Ca and Fe) with the Wiersma et al. (2009b) implementation adopted by EAGLE, which assumes that these species are in equilibrium with the cosmic microwave background (CMB) and the Haardt & Madau (2001) model of a spatially-uniform, time-evolving UV/X-ray background from galaxies and quasars that is switched on at $z = 7.5$. At earlier times the CMB is supplemented only by an $H_2$ photodissociating background. As with EAGLE, this model accounts for optically-thick photoheating during reionisation by instantly adding 2 eV per proton mass as thermal energy. For HI this is done instantly at $z = 7.5$, for HeII the heating is distributed in redshift with a Gaussian function centred on $z = 3.5$ of width $\sigma(z) = 0.5$.

### 3.4.2 Star Formation

These simulations adopt the star formation model detailed by Nobels et al. (2023), for which gas acquires a non-zero star formation rate when it becomes locally unstable against gravitational collapse, specifically that the Jeans mass is smaller than the mass within the SPH kernel. The model has been shown to reproduce the spatially-resolved observed Kennicutt-Schmidt relations for neutral, atomic, and molecular gas, and exhibits good convergence over a wide dynamic range in adopted gas mass resolution.

Gravitational instability is assumed when the gravitational binding energy of a fluid element exceeds its kinetic energy due to thermal and turbulent motion:

$$\alpha \equiv \frac{\sigma_{\rm th}^2 + \sigma_{\rm turb}^2}{G \langle N_{\rm ngb} \rangle^{2/3} m_{\rm gas}^{2/3} \rho^{1/3}} < 1, \tag{3.3}$$

where $\sigma_{\rm th}$ and $\sigma_{\rm turb}$ are the thermal and turbulent velocity dispersions respectively, $\langle N_{\rm ngb} \rangle \simeq 65$ is the weighted number of neighbours within the SPH kernel. The thermal velocity dispersion of each gas element follows from its thermal pressure and density,

$$\sigma_{\rm th} = \sqrt{\frac{3P}{\rho}}, \tag{3.4}$$

whilst the turbulent velocity dispersion of gas element $i$ is:

$$\sigma_{\rm turb,i} = \frac{1}{\rho_i} \sum_j m_j |\mathbf{v}_i - \mathbf{v}_j|^2 W(|\mathbf{r}_i - \mathbf{r}_j|, h_i), \tag{3.5}$$

where the sum runs over the neighbours $j$. The vectors $\mathbf{r}_{\rm i}$ and $\mathbf{r}_{\rm j}$ are the position vectors of $i$ and $j$, $\mathbf{v}_i$ and $\mathbf{v}_j$ are their velocity vectors, $W_{i,j}$ is the SPH kernel function centred on gas element $i$, and $h_i$ is its smoothing length. Unstable gas particles are assigned a star formation rate (SFR), $\dot{m}_\star$, that follows from the Schmidt (1959) law,

$$\dot{\rho}_\star = \epsilon_{\rm SF} \frac{\rho}{t_{\rm ff}} \longrightarrow \dot{m}_\star = \epsilon_{\rm SF} \frac{m_{\rm gas}}{t_{\rm ff}}, \tag{3.6}$$

where $t_{\rm ff} = \sqrt{3\pi/32G\rho}$ is the density-dependent free-fall timescale. The efficiency $\epsilon_{\rm SF} = 0.01$ follows from observations of star formation within Giant Molecular Clouds (e.g. Krumholz & Tan, 2007). The SFR is used as a probability for the stochastic conversion of the gas particle into stellar particle during each timestep of duration $\Delta t$:

$$p_{\rm SF} = \frac{\dot{m}_\star}{m_{\rm gas}} \Delta t. \tag{3.7}$$

### 3.4.3 Stellar Evolution and Mass Loss

These simulations adopt the same (Wiersma et al., 2009b) stellar evolution and mass loss model used by EAGLE, in which stellar particles are treated as simple stellar populations with a Chabrier (2003) stellar initial mass function (IMF) defined between the limits 0.1 and $100\,{\rm M}_\odot$. Stellar particles inherit the element abundances of their parent gas particle. It tracks the time-dependent release of 11 species (H, He, C, N, O, Ne, Mg, Si, S, Ca and Fe) from core collapse and thermonuclear (Type Ia) SNe, and the asymptotic giant branch (AGB) phase of evolved stars. The assumed rate of Type Ia SNe (SNIa) is also very similar to the EAGLE implementation, differing only in the inclusion of an initial delay, $t_{\rm delay}$, between the formation of a stellar particle and the detonation of the first SNIa. The rate of SNIa per unit initial stellar mass formed is therefore:

$$\dot{N}_{\rm SNIa}(t) = \mathcal{H}(t - t_{\rm delay})\,\frac{\nu}{\tau}\,\exp\left(-\frac{t - t_{\rm delay}}{\tau}\right), \tag{3.8}$$

where $\mathcal{H}$ is the Heaviside step function, $\nu$ is the total number of SNIa per unit initial stellar mass and $\exp(-{\rm t}/\tau)/\tau$ is a normalised, empirical delay time distribution function. We use $t_{\rm delay} = 40\,{\rm Myr}$, $\tau = 2\,{\rm Gyr}$ and $\nu = 2{\times}10^{-3}\,{\rm M}_\odot\,{\rm yr}^{-1}$. Schaye et al. (2015) showed that these parameters yield a good match to the observed SNIa rate.

### 3.4.4 Early Stellar Feedback

The simulations include three pre-supernova feedback processes from massive stars: stellar winds, direct radiation pressure, and HII regions, whose numerical implementation will be described in detail by a forthcoming article (Ploeckinger et al. in prep). The energy, momentum, and ionising flux transferred to neighbouring gas is calculated using the Binary Population and Spectral Synthesis (BPASS) stellar evolution and spectral synthesis models (Eldridge et al., 2017; Stanway & Eldridge, 2018), again adopting a Chabrier (2003) IMF defined between 0.1 and $100\,{\rm M}_\odot$. The cumulative momentum of wind ejecta determined by the BPASS tables, as a function of a stellar population's age and birth metallicity, is used to determine the number of stochastic kicks with velocity $v_{\rm kick} = 50\,{\rm km\,s^{-1}}$ a stellar particle will impart to its gas neighbours to model stellar winds. Radiation pressure feedback is similarly determined by the photon energy spectrum determined by BPASS, and the photon momentum exerted onto the gas is calculated using the wavelength-dependent optical depth extracted from the Ploeckinger & Schaye (2020) modified tables based on the local Jeans column density. In HII regions, young star particles stochastically ionise and heat surrounding gas particles to $T = 10^4\,{\rm K}$ following a Strömgren (1939) sphere approximation. Young stellar particles select a new set of gas neighbours as HII regions every 2 Myr, based on the density of the ambient medium and the stellar particle's ionising photon flux as determined by BPASS. Gas particles comprising HII regions are ineligible for star formation, even if they satisfy the star formation criterion.

### 3.4.5 Core Collapse Supernovae

The core collapse supernovae (CC SNe) feedback model is a mildly-modified version of the thermal-kinetic implementation of Chaikin et al. (2023), which is designed for cosmological simulations that marginally resolve the cold ISM by injecting feedback energy via both thermal and kinetic channels. The combination of these modes generates strong galactic winds and the hot ISM through periodic powerful injections of thermal energy, and turbulence in the neutral ISM through small, frequent kinetic kicks (Chaikin et al., 2023).

The total amount of energy released by a stellar particle in a timestep is given by,

$$\Delta E_{\rm CCSN} = 10^{51}{\rm erg} f_{\rm E} m_\star \int_{m_{\rm d}(t+\Delta t)}^{m_{\rm d}(t)} \Phi(m) dm \tag{3.9}$$

where $\Phi(m)$ is the Chabrier (2003) IMF and $m_{\rm d,t}$ is the mass of the star(s) that explode as CCSN at age $t$, computed using the metallicity dependent stellar lifetime tables from Portinari et al. (1998). The function $m_{\rm d}(t)$ is non-zero for zero-age main sequence masses between $m_{\rm min,CCSN} = 8$ and $m_{\rm max,CCSN} = 10^2{\rm M}_\odot$. The parameter $f_{\rm kin}$ determines the amount of CC SNe energy injected kinetically, $f_{\rm kin}\Delta E_{\rm CCSN}$, and thermally, $(1 - f_{\rm kin})\Delta E_{\rm CCSN}$. For simplicity I assume a constant energy fraction of $f_{\rm E} = 1$, which differs from the density and metallicity-dependent energy fraction adopted by EAGLE (Crain et al., 2015).

#### 3.4.5.1 Thermal Channel of CCSN Energy Injection

The thermal channel for CCSN is implemented using the stochastic model of Dalla Vecchia & Schaye (2012), which was utilised by the EAGLE simulations (Schaye et al., 2015). Stellar particles inject energy to their neighbouring gas particles with a probability of $P_{\mathrm{SN,heat}}$ and the amount of energy transferred, $\Delta E_{\mathrm{SN,heat}}$, is governed by the desired temperature increase of the gas particle $\Delta T_{\mathrm{SN}}$. The relation between this energy and the heating temperature is given by,

$$\Delta E_{\mathrm{heat}}(m_{\mathrm{gas}}, \Delta T_{\mathrm{SN}}) = \frac{k_{\mathrm{B}} \Delta T_{\mathrm{SN}}}{(\gamma - 1)} \frac{m_{\mathrm{gas}}}{\mu m_{\mathrm{p}}} \tag{3.10}$$

where $m_{\mathrm{p}}$ is the proton mass, $k_{\mathrm{B}}$ is the Boltzmann constant, $m_{\mathrm{gas}}$ is the mass of the target gas particle being heated, $\gamma = 5/3$ is the ratio of specific heats for an ideal monatomic gas, and $\mu = 0.59$ is the mean molecular weight of a fully ionised gas. Within EAGLE, this was a fixed, constant temperature increase of $\Delta T_{\mathrm{SN}} = 10^{7.5}$K.

The probability of a stellar particle heating one of its gas particles in a given timestep between $t$ and $t + \Delta t$ is computed by the ratio of the available thermal SN energy $(1 - f_{\mathrm{kin}}) \Delta E_{\mathrm{CCSN}}(t, \Delta t, m_\star, f_{\mathrm{E}})$ to the energy required to increase the temperature of the gas mass contained within the stellar particle kernel $\Delta E_{\mathrm{heat}}(m_{\mathrm{ngb}}, \Delta T_{\mathrm{SN}})$, where $m_{\mathrm{ngb}}$ is the sum of the masses of gas particles found in the stellar particle kernel. This is calculated as shown in Equation 3.11.

$$P_{\mathrm{SN,heat}} = (1 - f_{\mathrm{kin}}) \frac{\Delta E_{\mathrm{CCSN}}(t, \Delta t, m_\star, f_{\mathrm{E}})}{\Delta E_{\mathrm{heat}}(m_{\mathrm{ngb}}, \Delta T_{\mathrm{SN}})} \tag{3.11}$$

After computing this probability, a random number, $0 < r < 1$, is drawn $N_{\mathrm{ngb}}$ times where $N_{\mathrm{ngb}}$ corresponds to the number of gas particle neighbours contained within the stellar particle kernel. The number of energy injections per timestep is then determined by how many times the random number is smaller than the probability, $r < P_{\mathrm{SN,heat}}$. For each of these occasions, the stellar particle injects energy into its gas particle neighbours which are chosen by the isotropic neighbour selection algorithm described by Chaikin et al. (2022) with a maximum number of rays set to 8.

In EAGLE, thermal energy injections were distributed among gas neighbours with an equal probability, however, this method, hereafter referred to as the mass-weighted algorithm, preferentially selects particles that are in denser regions of the kernel that are made up of a higher number of particles due to the Lagrangian nature of SPH (Chaikin et al., 2022). Chaikin et al. (2022) demonstrated that the bias towards high-density gas and resulting radiative losses from the mass-weighted scheme can be minimised by adopting an isotropic neighbour selection algorithm and an illustrative comparison of the schemes is shown in Figure 3.14. The isotropic neighbour algorithm ensures that the SN energy injections are statistically distributed isotropically as seen by the star and is fully detailed in Section 2.2 by Chaikin et al. (2022).

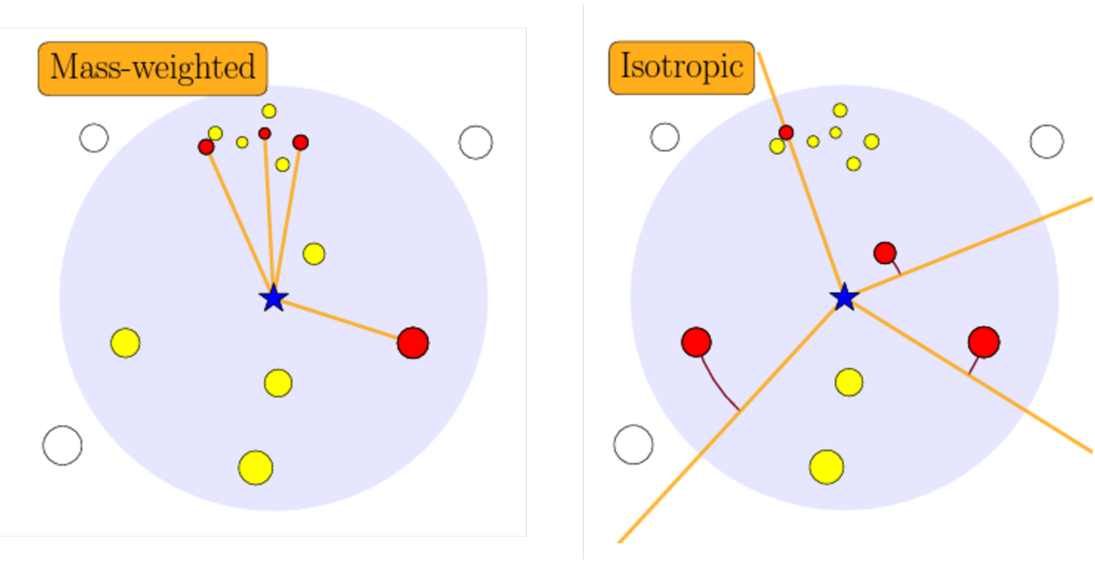


FIGURE 3.14: A comparison of the mass-weighted and isotropic gas-neighbour selection algorithms for energy injection in SN feedback by Chaikin et al. (2022). The shaded circular regions represent the SPH kernel of the star particle depicted by the blue star in the centre. Gas neighbours are represented by circles, coloured yellow if they are within the SPH kernel and coloured red if selected for energy injection. The size of the circles increases with decreasing density.

#### 3.4.5.2 Density-Dependent Heating Temperature

In contrast to the constant heating temperature used in EAGLE, I use a variable density-dependent heating temperature. In these simulations the heating temperature is allowed to vary within the range $10^{6.5} < \Delta T_{\rm SN} < 10^{8}$K and monotonically increases with the gas density. The choice of heating temperature is motivated by a balance between good sampling of the feedback history of a stellar population and ensuring each feedback event is numerically efficient. Using a minimum heating temperature that is lower than $10^{7.5}$K gives increased sampling of SN feedback events in low-mass galaxies formed with fewer star particles and hence SN energy injections. A lower heating temperature also makes SN feedback less destructive in lower density gas environments which may alleviate the problem identified by Bahé et al. (2016) where galaxies in EAGLE exhibit large unphysical SN-driven holes in their gas distributions.

Specifically, the heating temperature $\Delta T_{\mathrm{SN}}$ depends on the average physical gas density at the location of the star particle, $\rho_{\mathrm{SN}}$, which is computed by,

$$\rho_{\mathrm{SN}} = \sum_{i=1}^{N_{ngb}} m_i W(|r_\star - r_i|, h_\star) \tag{3.12}$$

here $m_i$ is the mass of gas particle $i$, $r_i$ and $r_\star$ are the coordinates of the given gas and star particle respectively, and $W$ is the SPH kernel over the star particle's smoothing length, $h_\star$.

At a given timestep, the average gas density enclosed by the star particle's SPH kernel is computed and converted into a hydrogen number density, $n_{\mathrm{H,SN}}$, assuming primordial abundances and a hydrogen mass fraction of $X_{\mathrm{H}} = 0.756$. This density is then used to evaluate the heating temperature through,

$$\Delta T_{\mathrm{SN}}(n_{\mathrm{H,SN}}) = \Delta T_{\mathrm{SN,0}} \left( \frac{n_{\mathrm{H,SN}}}{n_{\mathrm{SN,0}}} \right)^{2/3} \tag{3.13}$$

where $\Delta T_{\mathrm{SN,0}}$ and $n_{\mathrm{SN,0}}$ are free parameters and I fix $\Delta T_{\mathrm{SN,0}} = 10^{6.5}$K. The slope of 2/3 is implemented to prevent overcooling in dense regions; by considering the ratio of the cooling time to the sound-crossing time across a resolution element Dalla Vecchia & Schaye (2012) derive a maximum density ($n_{H,t_c}$) at which thermal feedback can be efficient in the high temperature regime where dominated by Bremsstrahlung cooling: $n_{H,t_c} \propto T^{3/2}$.

#### 3.4.5.3 Kinetic Channel of CCSN Energy Injection

The remaining CCSN energy is released via a kinetic mechanism based on the stochastic kinetic model of Dalla Vecchia & Schaye (2008), with the full algorithm detailed in Chaikin et al. (2023). Stellar particles inject kinetic energy with probability $P_{\mathrm{kick,pair}}$

given by,

$$P_{\text{kick,pair}} = f_{\text{kin}} \frac{E_{\text{CCSN}}(t, \Delta t, m_\star, f_{\text{E}})}{2\Delta E_{\text{kick}}(m_{\text{ngb}}, \Delta v_{\text{kick}})} \tag{3.14}$$

where $\Delta E_{\text{kick}}(m_{\text{ngb}}, \Delta v_{\text{kick}}) = m_{\text{ngb}} \Delta v_{\text{kick}}^2/2$ and $\Delta v_{\text{kick}}$ is the kick velocity. The number of kicks delivered by a given stellar particle is determined by drawing random numbers between $0 < r < 1$ for $N_{\text{ngb}}$, and demanding $r < P_{\text{kick,pair}}$ is satisfied. Each kick event is then implemented by kicking two neighbouring gas particles in opposite directions to ensure conservation of linear momentum. The model additionally ensures the conservation of energy and angular momentum and injects the energy in a stastically isotropic distribution. The choice of $\Delta v_{\text{kick}} = 50\,\text{km}\,\text{s}^{-1}$, injecting much less energy per event than the thermal channel, drives turbulence in the neutral ISM and helps to yield a scaling relation between HI velocity dispersion and SFR surface density similar to that observed in nearby galaxies.

### 3.4.6 Black Holes

In this galaxy formation model supermassive black holes (SMBHs) are treated as collisionless BH particles that are seeded within haloes which are then able to grow in mass by accreting surrounding gas or merging with other BH particles.

#### 3.4.6.1 Seeding

SMBHs are seeded during the simulation using an on-the-fly FoF group finder using a linking length of 0.2 times the main inter-particle separation, the scale-factor ratio between consecutive FoF black hole seeding calls is $\Delta\log_{10}\text{a} = 1.00751$ from a starting cosmic scale-factor of $a = 0.05$. A BH particle is seeded in haloes which meet a set of criteria: (i) the halo FoF mass must exceed the mass threshold of $10^{10}\text{M}_\odot$, (ii) the halo must contain at least 256 particles and (iii) the halo must not already hold a BH particle.

The seeding process involves identifying the densest gas particle in the FoF halo and converting it into a BH particle, inheriting the parent gas particle's dynamical mass, velocity, and position. The BH particle is assigned a subgrid mass that allows the BH mass to be below the simulation particle mass. The subgrid mass is initially set to the seed mass and the difference between the subgrid and dynamical mass can be considered a subgrid gas reservoir around the SMBH. The subgrid mass is then used in subsequent subgrid calculations instead of the dynamical mass of the particle.

#### 3.4.6.2 Gas Accretion

SMBHs grow continuously by accreting gas from their surroundings, for the resolution limit of our simulations the mass accretion rate onto a BH particle is estimated using the modified Bondi-Hoyle-Lyttleton formula from Krumholz et al. (2006),

$$\dot{m}_{\rm BH} = 4\pi G^2 \frac{m_{\rm BH}^2 \rho_{\rm gas}}{c_{\rm sound}^2} \left[ \frac{(1+\mathscr{M}^2)^4}{1.1^2+\mathscr{M}^2} + \frac{1}{(0.34 f_\star)^2} \right]^{-1/2} \tag{3.15}$$

Here $\mathscr{M}^2$ is the square of the Mach number defined by $\mathscr{M}^2 = (\sigma_{\rm turb}/\sigma_{\rm sound})^2 + (v_{\rm gas}/c_{\rm sound})^2$, $f_\star$ is the correction for vorticity, $\omega$, in the gas flow defined by $f_\star = 1/\left[1+(\omega r_{\rm Bondi}/c_{\rm sound})^{0.9}\right]$ and $r_{\rm Bondi} = G m_{\rm BH}/c_{\rm sound}^2$ is the Bondi radius for a BH particle subgrid mass $m_{\rm BH}$. The gas turbulent velocity dispersion $\sigma_{\rm turb}$, the gas bulk velocity $v_{\rm gas}$, and the vorticity $\omega \equiv |\nabla \times v_{gas}|$ are calculated as mass weighted averages over all gas neighbours within the kernel of the BH particle. The gas mass density is calculated using the standard SPH formulism for the gas particles within the BH kernel.

The total gas mass accreted onto the BH in a given timestep $\Delta t$ is therefore $\dot{m}_{\rm BH}\Delta t$, and the fraction received by the BH is given by

$$\Delta m_{\rm BH} = (1-\epsilon_{\rm r})\dot{m}_{\rm BH}\Delta t \tag{3.16}$$

based on the radiative efficiency $\epsilon_{\rm f}$. The mass fraction not received by the BH, $\epsilon_{\rm r}\dot{m}_{\rm BH}\Delta t$, is lost from the dynamical mass, converted to energy and assumed to have escaped the SMBH via radiation. The subgrid mass of the BH is updated to give $m_{\rm BH}^{\rm new} = m_{\rm BH} + \Delta m_{\rm BH}$. If its dynamical mass is greater than the new subgrid mass $m_{\rm BH}^{\rm new}$, the accreted mass is assumed to come from the sub-grid gas reservoir from when the SMBH was seeded.

In the case that the updated subgrid mass exceeds the dynamical BH mass, the dynamical mass is increased to be equal to the subgrid mass. The difference in mass, $m_{\rm BH} + \Delta m_{\rm BH} - m_{\rm BH}^{\rm dynamical}$, is 'nibbled' from neighbouring gas particles within the BH kernel following the method of Bahé et al. (2022), to ensure mass conservation. The nibbling technique involves transferring a small fraction of mass from each gas particle, $\delta m_i$, computed by weighting the gas particles by their contribution to the gas density at the location of the SMBH. This is described by,

$$\delta m_i = \frac{m_{\rm BH}^{\rm new} - m_{\rm BH}^{\rm dynamical}}{1-\epsilon_{\rm r}} \frac{W(|r_{\rm BH} - r_i|, h_{\rm BH}) m_i}{\sum_j W(|r_{\rm BH} - r_j|, h_{\rm BH}) m_j} \tag{3.17}$$

where $W(|r_{\rm BH} - r_i|, h_{\rm BH})$ is the kernel weight of particle $i$ at position $r_i$, with a mass of $m_i$, within the BH smoothing length $h_{\rm BH}$ from the position of the BH particle $r_{\rm BH}$.

In these simulations the mass accretion rate is capped such that the maximal allowed accretion rate $\dot{m}_{\rm BH,max}$ is 10 times the mass accretion rate at the Eddington luminosity, $\dot{m}_{\rm Edd}$, expressed as,

$$\dot{m}_{\rm BH,max} = 10\dot{m}_{\rm Edd} = 10\frac{4\pi G m_{\rm BH} m_{\rm p}}{\epsilon_{\rm r}\sigma_{\rm T} c} \tag{3.18}$$

where $\sigma_{\rm T}$ is the Thompson cross section for electron scatter and $c$ is the speed of light.

#### 3.4.6.3 BH Merging

Through a galaxy merger a halo may contain more than one BH and under certain circumstances the BHs can become gravitationally bound and merge, in the simulations this is implemented in a simplified manner by merging the BH particles and increasing the remaining BH particle's mass. For BH particles to merge, the particles must satisfy distance and velocity criteria: (i) the separation of the BH particles $\Delta r_{\rm BH}$ must be less than 3 times the gravitational softening length, $\Delta r_{\rm BH} < 3\epsilon_{\rm soft,gas}$, following Bahé et al. (2022), and (ii) their relative velocity must be below $\Delta v_{\rm BH} < 2G(M+m)/\Delta r$, where $M$ and $m$ are the dynamical masses of the larger and smaller BH particles respectively. When both criteria are satisfied, the BH particles are merged instantaneously. The BH particle with the smaller subgrid mass transfers its properties to the BH particle with the larger subgrid mass and is then removed from the simulation.

#### 3.4.6.4 BH Repositioning

SMBHs are subject to dynamical friction which causes the BH to slow down and sink toward the centre of their host halo, however, dynamical friction cannot be resolved within the simulation and its effects cannot be predicted reliably enough to implement via a subgrid technique, therefore, the net effect of the BH moving towards the local minimum gravitational potential is modelled. This is applied in these simulations by BH particle repositioning as described by Bahé et al. (2022); for each timestep $\Delta t$, BH particles search their neighbouring gas particles that lie within their smoothing kernel and 3 gravitational softening lengths, and then re-position themselves to the coordinates of the gas particle with the lowest gravitational potential energy. During the computation of the gravitational potentials, the contribution from the BH particle is excluded in order to avoid the BH becoming trapped within its own potential.

#### 3.4.6.5 AGN Feedback

Active Galactic Nuclei (AGN) feedback from SMBHs is implemented via a thermal and stochastic mechanism. Each BH carries a reservoir of feedback energy that increases in a given timestep,

$$\Delta E_{\rm AGN} = \epsilon_{\rm f}\epsilon_{\rm r}\dot{m}_{\rm BH}c^2\Delta t \tag{3.19}$$

here $\dot{m}_{\rm BH}$ is the SMBH's gas accretion rate given by Equation 3.15 and $\epsilon_{\rm f}$ is the coupling efficiency set to $\epsilon_{\rm f} = 0.05$.

Injecting small amounts of energy can lead to numerical overcooling, therefore, following Booth & Schaye (2009) the energy injection is delayed until a sufficient amount of energy is accumulated by the BH. Once the energy reservoir reaches a threshold energy $\Delta E_{\rm AGN,thr}$, this is injected into the surrounding gas within the BH's kernel and subsequently subtracted from the BH energy reservoir. The energy threshold is defined based on the desired temperature increase of the gas $\Delta T_{\rm AGN}$ such that $\Delta E_{\rm AGN,thr} \equiv \Delta E_{\rm heat}(< m_{\rm gas} >, \Delta T_{\rm AGN})$ where $< m_{\rm gas} >$ is the average gas particle mass within the kernel and $\Delta E_{\rm heat}(< m_{\rm gas} >, \Delta T_{\rm AGN})$ is computed using Equation 3.10. For these simulations the heating temperature is set to $\Delta T = 10^9$K.

In cases where the BH particle accretes rapidly, and/or the timestep is long, the energy reservoir can exceed $N_{\rm AGN}\Delta E_{\rm AGN,thr}$ where $N_{\rm AGN} > 1$. Here, the energy is injected into $N_{\rm AGN}$ gas neighbours and the energy reservoir is reduced by $N_{\rm AGN}\Delta E_{\rm AGN,thr}$. The gas neighbour that receives the energy is selected based on the minimum distance algorithm by Chaikin et al. (2022), and if more than one injection is required then the $N_{\rm AGN}$ closest neighbours each recieve one injection. Following Bahé et al. (2022), the maximum number of injections or ‘rays’ carried by the BH is set to $N_{\mathrm{BH,rays}} = 50$ which limits the number of energy injections that can be delivered simultaneously by the BH particle. In rare events where the required $N_{\mathrm{AGN}}$ exceeds $N_{\mathrm{BH,rays}}$, $\Delta T_{\mathrm{AGN}}$ is increased by $N_{\mathrm{AGN}}/N_{\mathrm{BH,rays}}$ and the $N_{\mathrm{BH,rays}}$ closest neighbours are heated by the updated $\Delta T_{\mathrm{AGN}}$. Additionally, in cases where the number of gas particles within the BH’s kernel, $N_{\mathrm{ngb}}$is lower than $\min(N_{\mathrm{AGN}}, N_{\mathrm{BH,rays}})$, $\Delta T_{\mathrm{AGN}}$ is increased by $\min(N_{\mathrm{AGN}}, N_{\mathrm{BH,rays}})/N_{\mathrm{ngb}}$ and all $N_{\mathrm{ngb}}$ particles receive energy according to the updated $\Delta T_{\mathrm{AGN}}$.

## 3.5 Halo and Galaxy Identification

Haloes are identified in the simulation outputs using the VELOCIRAPTOR structure finder (Elahi et al., 2019). VELOCIRAPTOR uses a 3-dimensional friends-of-friends (3D FOF) algorithm to identify field haloes followed by a 6D FOF algorithm to identify subhaloes within each halo, in particular, subhaloes deep within host haloes where the difference in density relative to the background is negligible. The object with the largest distinction from the background density field is identified by the 3D FOF algorithm and labelled the field galaxy.

Galaxy measurements are computed using the Spherical Overdensity and Aperture Processor (SOAP). SOAP is a tool, introduced by Schaye et al. (2023), that calculates numerous (sub)halo properties for a range of apertures using the centres of potential and particle group memberships computed by VELOCIRAPTOR.

# 4

# Columba I: Environmental Results

## 4.1 Introduction

As some of the oldest, most dark matter dominated systems, dwarf galaxies have shallow gravitational potentials and are particularly sensitive to both internal and external astrophysical processes, making them an interesting test for both galaxy formation and cosmological models. An important influence on the properties of a dwarf galaxy is their local environment which can range from overdense regions in close proximity to more massive objects, to isolated field environments, that can shape their evolution. Dwarf galaxies that are found in isolation are free from the gravitational influences of large neighboring galaxies that affect satellites. This isolation makes them ideal laboratories for studying the intrinsic properties of galaxies without the confounding effects of interactions. However, even isolated dwarf galaxies are not entirely free from environmental influences. Various external factors, such as cosmic reionisation (e.g. Gnedin, 2000), and local cosmic web structures (e.g. Hellwing et al., 2021), can impact their formation and evolution. Understanding these environmental effects is essential for a comprehensive picture of galaxy formation, particularly in the low-mass regime.

How environment can shape a dwarf galaxy can be seen in the dichotomy between satellite and isolated dwarf galaxies in the Local Group. Satellites are typically dwarf spheroidals with little to no recent star formation whereas observed isolated dwarf galaxies are typically gas rich with ongoing star formation and irregular morphology. These environmental differences have been shown to affect other global statistics such as the SMHM relation which has significant scatter at the low mass end. Christensen et al. (2024) investigate the SMHM relation for galaxies in the Marvel-ous Dwarfs and D.C. Justice League simulations as a function of the distance to a massive ($M_{\mathrm{vir}} > 10^{11.5}\mathrm{M}_{\odot}$) galaxy and find that the scatter in the SMHM relation is correlated with this distance; more isolated dwarf galaxies have systematically lower stellar mass for a given halo mass. For less isolated systems, the scatter in this relation is increased due to the tidal stripping from more massive nearby objects. The gravitational influence from the nearby object strips the smaller object of its matter, starting with the dark matter halo that extends beyond the luminous galaxy (Smith et al., 2016). At a fixed stellar mass this reduces the halo mass and pushes the system to the left of the SMHM relation increasing the scatter (Fattahi et al., 2018). Additionally, such environments can play a role in shaping the stellar assembly history of dwarfs. Gallart et al. (2015) derive the star formation histories (SFHs) of nearby dwarf galaxies whose colour-magnitude diagrams can be resolved, finding that dwarf galaxies can be broadly categorised as having fast or slow formation histories: the former characterised by a short period of intense star formation at early epochs, and the latter by more continuous star formation, often persisting to the present epoch. They report that dwarfs with fast formation histories are preferentially located close to the MW or M31, and conclude that environment influences the present-day properties of dwarf galaxies primarily via assembly bias, i.e. that haloes of fixed mass in high-density environments tend to form earlier than those in low-density environments (e.g. Sheth & Tormen, 2004; Gao et al., 2005), rather than via direct environmentally-driven interactions at late times.

The large scale environment of a halo and galaxy has been shown to also affect properties such as their halo concentration, assembly history, and their morphology. Analyses of galaxies in different cosmic web environments, classified into nodes, walls, filaments, and voids, show that below a mass threshold of $M_{200,\mathrm{crit}} < 6.12 \times 10^{10} h^{-1}\mathrm{M}_{\odot}$, halo concentration and halo formation time show an increasing trend with web environment (Hellwing et al., 2021). Hellwing et al. (2021) find that galaxies residing in voids exhibit lower halo concentrations and later halo formation times than in all other environments. They also probe the halo mass distribution of galaxies using the shape of their velocity curve, enabling analysis of even smaller galaxies with an observable measure. For a denser inner region, and hence steeper inner mass profile of a halo, a smaller $R_{\mathrm{max}}$ is expected at fixed $V_{\mathrm{max}}$. A signal from the cosmic web environment is found in this measure of the internal dynamics of the halo where void galaxies exhibit the largest $R_{\mathrm{max}}$ at a given $V_{\mathrm{max}}$. Liao & Gao (2019) find that compared to their isolated field counterparts, dwarf galaxies that have formed in filaments have higher baryon and stellar fractions at $z \sim 2.5$ and Zheng et al. (2022) find that filament dwarf galaxies in the Auriga simulations tend to have higher star formation rates than field dwarfs of a fixed halo mass, though do not exhibit different galaxy $g$-$r$ colours. These findings highlight the role of filaments in assisting gas accretion at high redshift and enhancing star formation in dwarf galaxies.

In this Chapter I investigate the effect of large scale environment on key properties of low mass field haloes and galaxies. Specifically, I examine the isolated dwarf galaxies found in the Columba simulations and investigate how changes in the overdensity of their large-scale environment influences their population diagnostics. As shown in the different rows of Figure 3.9, the change in overdensity between regions in the Columba simulations corresponds to changes in the cosmic web structure, spanning from void-like environments in the lowest density regions to filamentary structures as the density increases.

## 4.2 Methodology

In this Chapter I utilise the Columba simulations, a suite of 25 zoom-in regions embedded in an $L = 400$Mpc periodic parent simulation. Full details of the initial conditions, gravity and hydrodynamics implementation, and galaxy formation model are described in Chapter 3. The analyses in this chapter focus on results from the high 'm5' resolution simulations which are detailed in Table 3.3. Haloes are identified in the simulation outputs using the VELOCIRAPTOR structure finder (Elahi et al., 2019) and the properties of haloes and the galaxies they host are computed using the Spherical Overdensity and Aperture Processor (SOAP).

### 4.2.1 Halo Properties and Matching

Throughout the following chapters, halo mass is given by $M_{200,\mathrm{crit}}$ which is defined as the total mass enclosed in $R_{200,\mathrm{crit}}$, the radius enclosing 200 times the critical density of the Universe at z=0. The velocity, $V_{200,\mathrm{crit}}$, is hence given by $V^2_{200,\mathrm{crit}} = GM_{200,\mathrm{crit}}/R_{200,\mathrm{crit}}$. The circular velocity profile of a halo is given by $V^2_{\mathrm{circ}} = GM(r)/r$ and the characteristic velocity peak is denoted by $V_{\mathrm{max}}$.

Haloes within the hydrodynamical simulations are matched with their counterparts in the DM-only simulations using the bijective matching technique between particle IDs described by Schaller et al. (2015b). I use matched DM-only properties to separate the entanglement of cause and effect that can arise from the inclusion of baryons; by using the matched DM-only properties I can reliably state that results relating to properties like halo concentration are a causation and not an effect from the inclusion of baryons.

In Figure 4.1 I show halo properties of the same haloes matched between hydrodynamical and DM only simulations of the m5 Columba simulation suite. In the left panel I compare the matched halo masses: in the well resolved regime, halo masses in the hydro simulations are smaller than those in the DM only simulations. This suppression is also seen in the ratio of $V_{\rm max}$ between hydro and DM only runs in the right hand panel. These effects occur due to the loss of baryons due to reionisation and supernova feedback which propagates a reduced halo growth rate and reduction in $V_{\rm max}$ (Sawala et al., 2013; Schaller et al., 2015a). This has been shown before in simulations, for example, in the APOSTLE simulations Sawala et al. (2016b) show a similar level of suppression ($\sim 15\%$) in $V_{\rm max}$ below $\sim$ 100km/s measured in the hydrodynamical compared to DM only simulations.

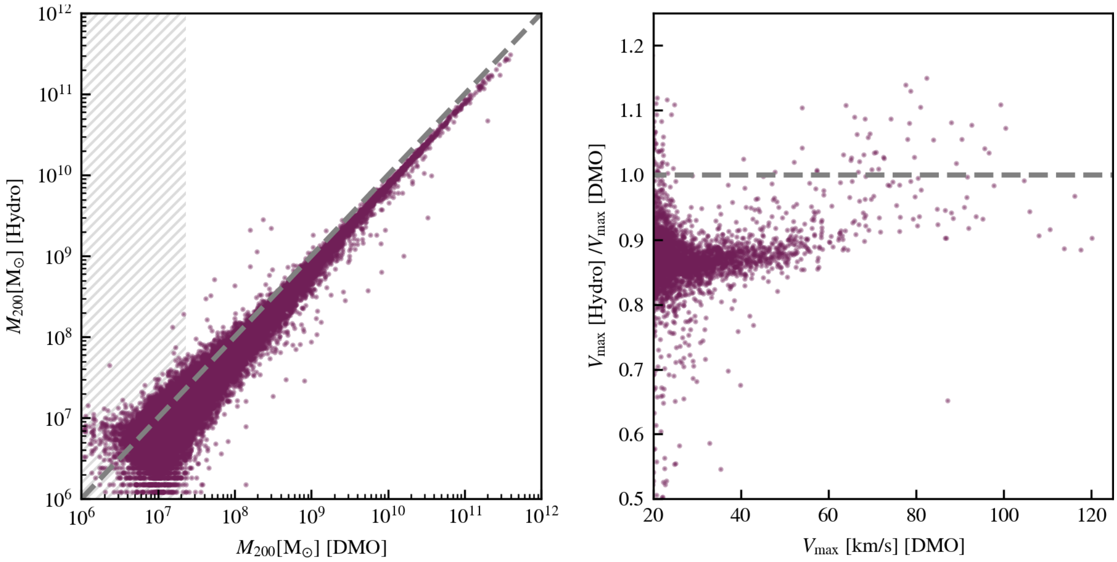


FIGURE 4.1: Properties of matched haloes between the DM only and hydrodynamical simulations in the m5 resolution Columba suite. *Left:* A comparison of between the halo masses $M_{200,\rm crit}$ in the matched DM only and hydrodynamical simulations. *Right:* The ratio between $V_{\rm max}$ measured in the hydrodynamical simulations to that measured in the DM only simulations. The halo masses and $V_{\rm max}$ of low mass haloes are suppressed by the inclusion of baryons in the hydrodynamical simulations.

These effects highlight the need to use matched DM only properties when examining properties like halo concentration, $c_{\rm NFW}$. Halo concentration is defined as the dimensionless shape parameter of the Navaro-Frenk-White (NFW) density profile (Navarro et al., 1996b), where $c_{\rm NFW} = R_{200}/R_{\rm s}$ and $R_{\rm s}$ is the NFW scale radius. Rather than performing parametric density profile fits to recover $c_{\rm NFW}$, I estimate the (intrinsic) halo concentration using the ratio of $V_{\rm max}/V_{200}$ from the matched DM-only simulations, via:

$$\frac{V_{\rm max}^2}{V_{200}^2} \approx 0.216 \frac{c_{200}}{A(c_{200})}, \tag{4.1}$$

where,

$$A(c_{200}) \equiv \ln(1 + c_{200}) - \frac{c_{200}}{1 + c_{200}}. \tag{4.2}$$

Bullock et al. (2001) advocate use of this relationship as it enables recovery of a concentration estimate for low-mass haloes that may lack the particle sampling required for a robust density profile fit.

My analyses consider only those haloes/galaxies residing within the central $r = 5\,$cMpc of each region at $z = 0$, guaranteeing that the structures I examine are uncontaminated by boundary particles. In this work I do not construct galaxy samples at earlier epochs, at which time the Lagrangian region of the central $z = 0$ sphere does not have a simple morphology, complicating the selection procedure.

## 4.3 Model Validation and Convergence

The macroscopic efficiencies of feedback mechanisms are governed by microphysics acting on spatial scales several orders of magnitude below the resolution scale of my simulations (e.g. Orlando et al., 2005). These efficiencies must therefore be implemented on resolved scales by the adopted subgrid models. The appropriate parametrisation of the subgrid models is difficult to predict a priori, or to infer from observations, and is generally resolution dependent. It has therefore become common practice to calibrate the parameters of subgrid feedback models, so that simulations broadly reproduce judiciously-chosen properties of the galaxy population (e.g. Schaye et al., 2015; Pillepich et al., 2018; Davé et al., 2019, see also the review by Crain & van de Voort 2023). Such diagnostics include for example the galaxy stellar mass function.

In the absence of strong observational constraints in the low-mass regime that is my main focus, I have elected against calibrating the fraction of energy from CC SNe that couples to the ISM, and instead use $f_{\rm E} = 1$ at all times. It is therefore imperative to first validate the simulations via comparison with observational measurements of fundamental galaxy properties, and to assess how robust these quantities are to changes to numerical resolution. I first combine the 25 simulations at each resolution to yield a composite galaxy population (albeit one from a collection of regions that are lower density than a random sample, as per the region selection), in order to examine the galaxy stellar mass function, the stellar mass - halo mass relation, and galaxy sizes. I test the convergence of the simulations by comparing the properties of the galaxy population at m5 and m6 resolution. This comparison is for a fixed galaxy formation model, or a 'strong convergence test' in the terminology of Schaye et al. (2015). Many of the results I examine in Chapters 4 and 5 follow from the intrinsic properties of DM haloes, which do not depend on the implementation of the galaxy formation model, therefore I examine the numerical convergence of the DM halo mass function and the inner structure of DM haloes separately, in Section 4.3.4.

I note that each simulation has only been run once and is therefore subject to uncertainties from the stochastic processes in the adopted subgrid prescriptions and that re-simulations of the same initial conditions with the same model can yield different properties of individual galaxies (e.g. Genel et al., 2019; Keller et al., 2019; Davies et al., 2021; Grand et al., 2021; Borrow et al., 2023), however, this thesis focuses on properties of large populations over various simulation volumes which is more robust against random variability and this uncertainty is manifest primarily as a source of scatter in scaling relations.

### 4.3.1 Galaxy Stellar Mass Function

Figure 4.2 shows the present-day GSMF of the composite galaxy population formed from all 25 spherical regions of $r = 5\,$cMpc. The purple and orange curves correspond to the m5 and m6 simulations, respectively. The grey hatched region indicates the regime corresponding to 10 or fewer stellar particles at m5 resolution, the grey arrow extends to the mass scale corresponding to 10 stellar particles at m6 resolution. Since I simulate regions that are biased to low-density, one cannot make a direct comparison

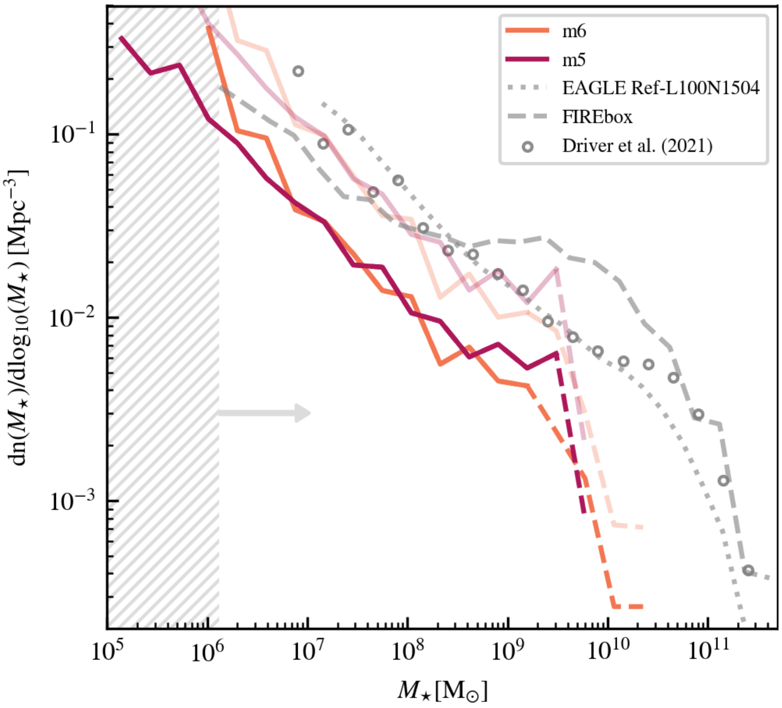


FIGURE 4.2: A strong convergence test of the GSMF with numerical resolution between m5, in purple, and m6, in orange. At the high-mass end, the lines become dashed when the bin contains fewer than 10 galaxies. The circles show observations from Driver et al. (2022). The dotted and dashed lines show the results from the EAGLE (Schaye et al., 2015) and FIREbox (Feldmann et al., 2023) simulations. The low opacity lines show the GSMFs scaled by their overdensity. The shaded region indicates 10 m5 resolution star particles and extends to 10 m6 resolution star particles indicated by the arrow.

of the resulting GSMF with observational measurements, nor with GSMFs recovered from simulations of average density volumes. I therefore apply a simple correction to the simulated GSMF, rescaling the number density at fixed mass by the reciprocal of the (normalised) mean density of the 25 regions. These rescaled GSMFs are drawn with low-opacity curves, enabling a more meaningful comparison with the GSMF inferred by Driver et al. (2022) from the fourth data release of the GAMA survey, shown as circles. For comparison with the GSMF recovered from a simulation of a large periodic volume, I show the EAGLE $z = 0$ GSMF (Schaye et al., 2015) as a dotted line, and for comparison with a cosmological simulation including an explicit cold ISM phase, I show the FIREbox $z = 0$ GSMF (Feldmann et al., 2023) as a dashed line.

In the regime where galaxies comprise at least 10 or more particles at m6 resolution and the stellar mass bins are sampled by at least 10 galaxies (broadly, $2 \times 10^7\,\mathrm{M}_\odot \lesssim \mathrm{M}_\star \lesssim 2 \times 10^9\,\mathrm{M}_\odot$), the simulated GSMF exhibits very good convergence between m5 and m6 resolutions. At the high-mass end of the well-sampled regime of the GSMF (corresponding to a number density of $\simeq 5 \times 10^{-3}$ cMpc) galaxies reach a slightly greater stellar mass in the m5 simulations, likely due to their higher resolution enabling star formation to begin at an earlier epoch in the progenitors of these galaxies. In the poorly-sampled more massive bins, the handful of galaxies tend to reach greater masses in the lower-resolution m6 simulations. The same BH seed mass is used by the m5 and m6 simulations, and it has been shown that BHs enter the 'rapid growth phase' (e.g. Bower et al., 2017) at earlier epochs in higher resolution simulations (Booth & Schaye, 2009). The BHs of these massive galaxies become able to deliver enough feedback energy to influence the evolution of their host galaxies at m5 resolution, but not at m6. It is likely that better convergence could be achieved by calibrating the BH seed mass as a function of resolution.

The preferential selection of low density regions leads to the unscaled GSMFs residing a lower number density (by approximately a factor of 3) at fixed stellar mass, relative to the GAMA GSMF. In the regime where simulated galaxies are resolved by at least 10 stellar particles and the stellar mass bins are well sampled, the rescaled GSMFs of the simulations reproduce the GAMA GSMF, with the number density at fixed stellar mass differing by $\lesssim 0.27\,\mathrm{dex}$. This is comparable to the agreement achieved by the EAGLE flagship Ref-L100N1504 simulation, but here this level of agreement extends down to stellar masses a factor of 8 lower for the m5 simulations, which is remarkable considering that I did not calibrate the CC SNe feedback efficiency to achieve this correspondence. The agreement with the observational data is also significantly better than that achieved by the FIREbox simulation in the mass interval I consider here. Even after rescaling, the number density of galaxies of mass $M_\star \gtrsim 2 \times 10^9\,\mathrm{M}_\odot$ is significantly lower than inferred by GAMA: although the simulations yield too few galaxies to enable an authoritative comparison, it is likely that this low number density of massive galaxies is genuine, stemming from a paucity of massive DM haloes in the selected regions, as can be seen by comparison of the halo mass functions of the regions with that of mean density volumes (see Figure 4.5, also Crain et al., 2009).

### 4.3.2 Stellar to Halo Mass Relation

The GSMF is effectively a convolution of the DM HMF with the scaling relation specifying the stellar mass of DM haloes, which can be thought of as a halo mass-dependent galaxy formation 'efficiency'. Figure 4.3 shows the present-day stellar mass - halo mass

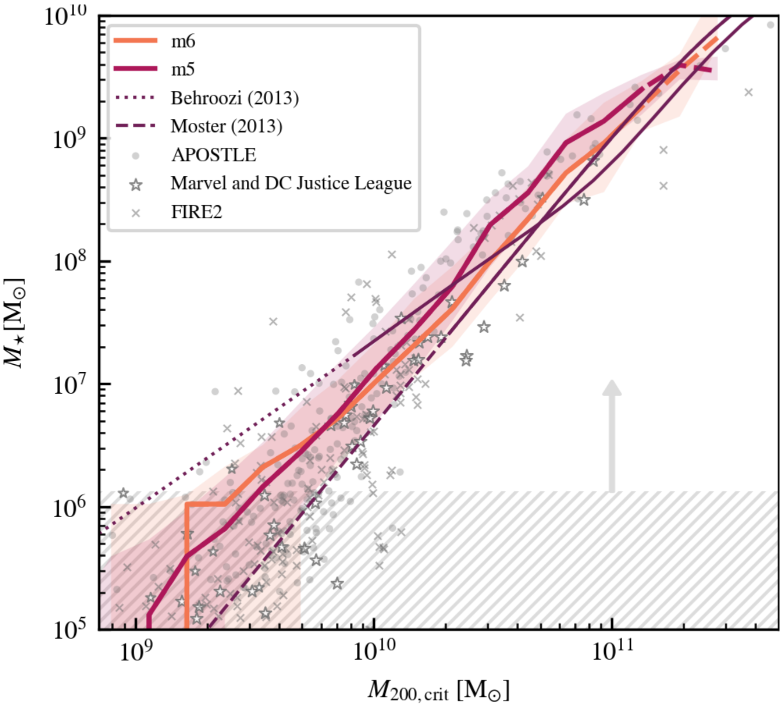


FIGURE 4.3: A strong convergence test of the stellar mass - halo mass relation with numerical resolution between m5, in purple, and m6, in orange. At the high-mass end, the lines become dashed when the bin contains fewer than 10 haloes. The 10th-90th percentile scatter is shown by the coloured shaded region for each resolution. I show the scatter of field galaxies from various other simulations: APOSTLE (Sawala et al., 2016a) are shown by the grey points, the Marvel (Munshi et al., 2021) and DC Justice league suites (Applebaum et al., 2021) are shown by the grey stars, and the Latte and ELVIS zoom-ins (Wetzel et al., 2016; Garrison-Kimmel et al., 2019) and various zoom-ins of individual dwarfs using FIRE-2 (Hopkins et al., 2018; Wheeler et al., 2015, 2019; Fitts et al., 2017) are shown by the grey crosses. The dashed and dotted dark purple lines denote the extrapolated present-day SMHM relations from Moster et al. (2013) and Behroozi et al. (2013), respectively. The shaded region indicates 10 m5 resolution star particles and extends to 10 m6 resolution star particles indicated by the arrow.

(SMHM) scaling relation of the composite galaxy population formed from all 25 spherical regions of $r = 5\,\mathrm{cMpc}$. I show the median stellar mass as a function of halo mass in bins of width 0.2 dex, and the corresponding $10^{\mathrm{th}} - 90^{\mathrm{th}}$ percentile scatter. Per Figure 4.2, the purple and orange curves correspond to the m5 and m6 simulations, respectively, and the grey hatched region indicates the regime corresponding to 10 or fewer stellar particles at m5 resolution, with the grey arrow extending to the mass scale corresponding to 10 stellar particles at m6 resolution. The solid and dotted purple lines denote the extrapolated present-day SMHM relations of the semi-empirical models of Moster et al. (2013) and Behroozi et al. (2013), respectively. The median of the SMHM relations more closely resembles the slope of the abundance matching relation of Moster et al. (2013) although the galaxies generally tend to have higher stellar masses. I show the scatter of field galaxies from various other simulations: APOSTLE (Sawala et al., 2016a) are shown by the grey points, the Marvel (Munshi et al., 2021) and DC Justice league suites (Applebaum et al., 2021) are shown by the grey stars, and the Latte and ELVIS zoom-ins (Wetzel et al., 2016; Garrison-Kimmel et al., 2019) and various zoom-ins of individual dwarfs using FIRE-2 (Hopkins et al., 2018; Wheeler et al., 2015, 2019; Fitts et al., 2017) are shown by the grey crosses. The scatter of the results encompass and show good agreement with the results from other models. I note that observational constraints on the SMHM relation at the faint end are very uncertain. Overall our results lie within a similar level of agreement than is found between other simulations and abundance matching studies.

The good convergence of the GSMF between the m5 and m6 resolution is evident in Figure 4.2 translates into similarly good convergence of the SMHM relation. In the regime for which galaxies are resolved by at least 10 stellar particles at m6 resolution ($M_\star \gtrsim 2 \times 10^7\,\mathrm{M}_\odot$) and halo mass bins are sampled by at least 10 galaxies, the scaling relation is well converged. As one might infer from the analysis of the GSMF, the poorest correspondence is for the relatively massive galaxies, with the m5 simulations exhibiting greater stellar masses at fixed halo mass, until one reaches a mass scale of $M_{200} \sim 10^{11}\,\mathrm{M}_\odot$, at which point the greater characteristic mass of central BHs in these galaxies at m5 resolution results in a suppression of their stellar mass.

### 4.3.3 Galaxy sizes

Crain et al. (2015) showed that it was possible to reproduce the present-day GSMF in the EAGLE simulations with a simple $f_\mathrm{E} = 1$ model (as I adopt here), but at the cost of a poor reproduction of galaxy sizes, stemming primarily from spurious radiative losses in high-density gas. I therefore consider galaxy sizes as an important validation diagnostic. Figure 4.4 shows the sizes of all galaxies in the composite sample as a function of stellar mass at the present day, where I equate 'size' to the 3-dimensional stellar half-mass radius, $R_\star$. I show the median size as a function of stellar mass in bins of width 0.2 dex. Line colours, line styles and shading relating to these simulations are consistent with their definitions in preceding plots. The solid grey curve shows the corresponding scaling relation from the EAGLE Ref-L100N1504 simulation.

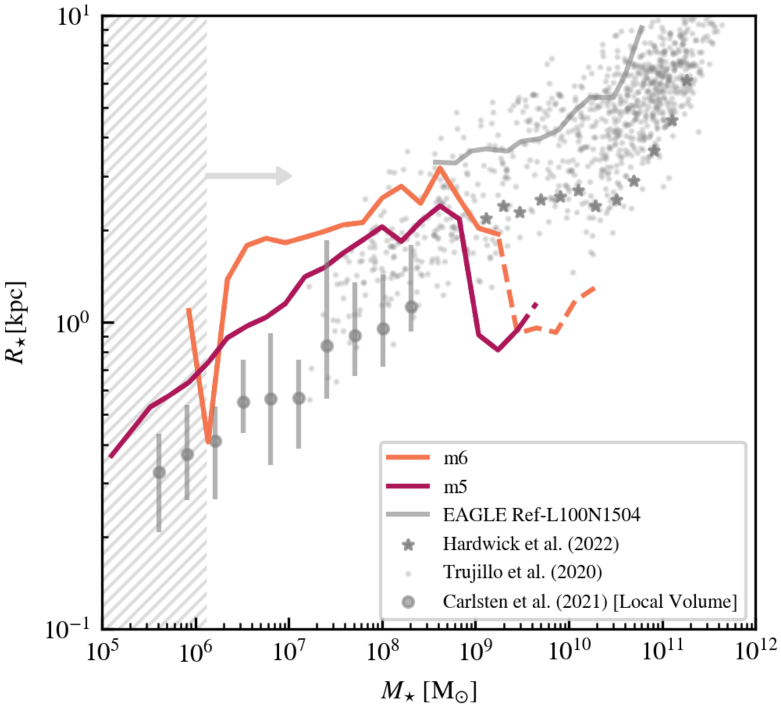


FIGURE 4.4: A strong convergence test of the galaxy stellar sizes as a function of stellar mass, with numerical resolution between m5, in purple, and m6, in orange. For comparison, I show the observed mass-size relations from the xGASS survey (Hardwick et al., 2022), Trujillo et al. (2020), and ELVES survey (Carlsten et al., 2021). I also show the calibrated mass-size relation of EAGLE galaxies (Schaye et al., 2015). The shaded region corresponds to 10 stellar particles at m5, extended to 10 stellar particles at m6 by the arrow.

In the regime for which each simulation's galaxies are well resolved, the simulations yield a broadly monotonic size-mass relation up to a mass scale of $M_\star \simeq 5 \times 10^8\,\mathrm{M}_\odot$, beyond which galaxies quickly become compact, qualitatively reminiscent of the decline in sizes reported by Crain et al. (2015) for variants of the EAGLE model that do not scale $f_\mathrm{E}$ with the density of natal gas particles. They find that this model yields a cosmic star formation history that is weighted too strongly towards early cosmic epochs and results in galaxies that too compact and too quiescent at present-day due to numerical overcooling. The m6 simulations yield larger galaxies at fixed mass, a result qualitatively consistent with the strong convergence test of galaxies sizes in EAGLE reported by Furlong et al. (2017, see their Appendix B). At the mass scale corresponding to 10 particles at m6 resolution, $M_\star \simeq 2 \times 10^7\,\mathrm{M}_\odot$, the median galaxy size at m6 resolution (1.98 kpc) is approximately 1.3 times that at m5 (1.51 kpc). The offset declines towards greater masses: at $M_\star = 7 \times 10^8\,\mathrm{M}_\odot$ the corresponding sizes are 2.52 kpc at m6 and 2.17 kpc at m5. This highlights that galaxy sizes generally require significantly better sampling (at least $\sim 10^2$ particles) to achieve convergence, and that they are more sensitive to the details of the galaxy formation model than galaxy (stellar) masses.

The simulated galaxies are systematically larger than inferred by Carlsten et al. (2021) from observations of dwarf galaxies in the Local Group. Although it is not trivial to compare 3-dimensional half-mass radii with isophotal sizes (see e.g. Genel et al., 2018), a naïve comparison indicates that galaxies in the m5 simulations are broadly twice the size of observed counterparts in the mass range $2 \times 10^6\,\mathrm{M}_\odot < \mathrm{M}_\star < 5 \times 10^8\,\mathrm{M}_\odot$. I show the median half-mass galaxy sizes from the xGASS survey (Hardwick et al., 2022) and the half-mass galaxy sizes from Trujillo et al. (2020). Though there is agreement with the scatter from Trujillo et al. (2020) between $10^8 < M_\star < 10^9\mathrm{M}_\odot$, in the high mass regime sampled by the Columba simulations galaxies with $M_\star \gtrsim 10^9\mathrm{M}_\odot$ fall beneath the observed relations and form too compact galaxies. This inability to reproduce the observed size-mass relation highlights the importance of calibration; as I show here, a successful reproduction of the GSMF does not guarantee realistic galaxy sizes and emphasises the importance of calibrating models with observational diagnostics that supplement the GSMF.

### 4.3.4 Convergence of Halo Properties

#### 4.3.4.1 Halo Mass Function

I look at the convergence of the halo mass function (HMF) from the hydrodynamical simulations between m5 and m6 resolution, shown in Figure 4.5 by the purple and orange lines, respectively. The shaded region indicates the mass of 100 DM particles at m5 resolution, extended to 100 m6 DM particles by the arrow. The turn over at the low mass end of the HMF is a resolution artefact where haloes are not well resolved by number of particles. Above this scale, the convergence of the HMF between numerical resolutions is extremely good. For comparison, I also show the HMF from Tinker et al. (2008) in grey. By design, the regions sampled in this simulation suite are low overdensity and do not contain any objects above the mass of $M_{200,\mathrm{crit}} > 5 \times 10^{11}\mathrm{M}_\odot$, this is reflected in the HMFs which do not extend to higher halo masses and lie below the global HMF from Tinker et al. (2008). I also scale the mass functions by their enclosed density before summing over all the regions which results in an upwards shift in the HMF and is shown in Figure 4.5 for a better comparison. The scaled HMFs show good agreement with the global HMFs from Tinker et al. (2008) and the vertical offset is fully accounted for when scaling the enclosed mass to the cosmic mean. There is still a downturn in HMFs from the reference data at the high mass end resulting from the region selection criteria and the fact that more massive objects are unable to form in the low overdensity environments.

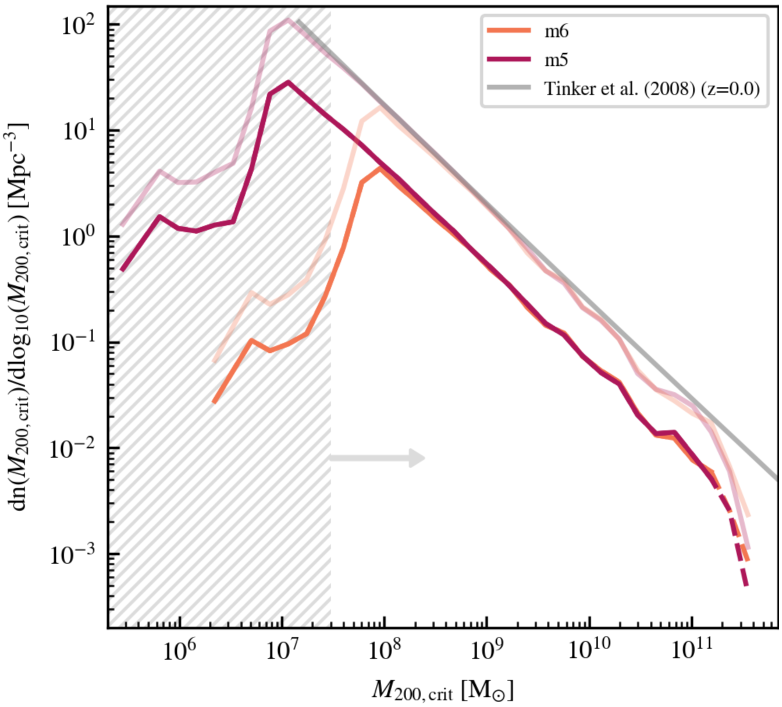


FIGURE 4.5: A convergence test of the hydrodynamical HMF with numerical resolution between m5, in purple, and m6, in orange. For comparison, the grey solid line shows the halo mass function from Tinker et al. (2008). The dashed line indicates where halo mass bins contain fewer than 10 haloes. The low opacity lines show the HMFs scaled by their overdensity. The shaded region indicates 100 m5 resolution DM particles and extends to 100 m6 resolution DM particles indicated by the arrow.

#### 4.3.4.2 Mass - Concentration Relation

In Figure 4.6 I show a convergence test between m5 and m6 resolutions of the halo concentration - halo mass relation where halo concentration is computed from matched halo counterparts in the DMO simulations using the method described in 4.2.1. The shaded region indicates the mass of 100 DM particles at m5 resolution, extended to 100 DM particles at m6 resolution by the arrow. Halo concentration is generally well converged above a halo mass of $\sim 10^8\mathrm{M}_\odot$ which coincides with the regime where haloes are well sampled in particle numbers at both resolutions. The halo concentrations are systematically lower than by Ludlow et al. (2016), however as will be discussed later in this chapter, this is likely due to the low density environments sampled by the Columba suite.

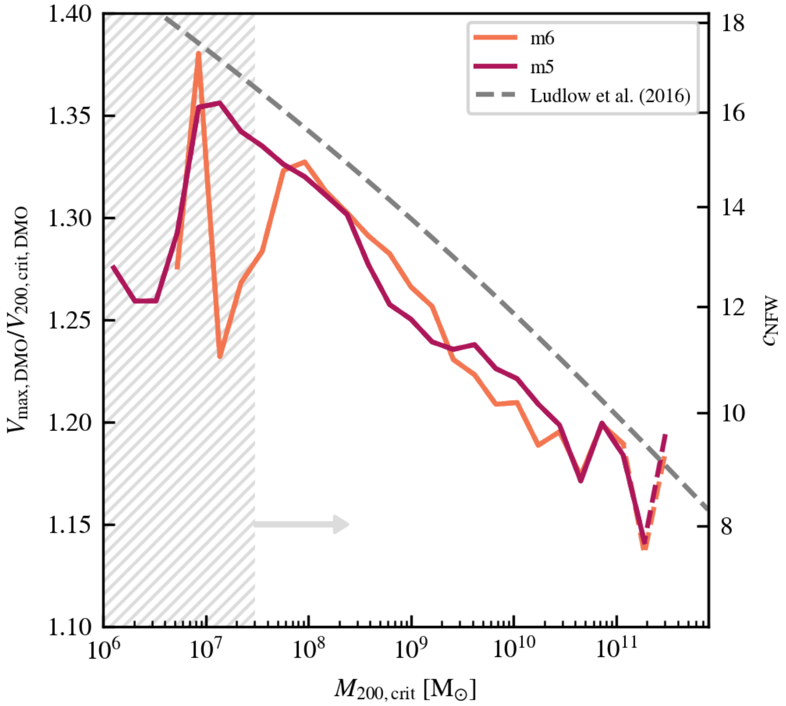


FIGURE 4.6: A convergence test of the mass-concentration relation with numerical resolution between m5, in purple, and m6, in orange. For comparison, the grey dashed line shows the relation from Ludlow et al. (2016). The lines become dashed when the halo mass bin contains fewer than 10 haloes. The shaded region indicates 100 m5 resolution DM particles and extends to 100 m6 resolution DM particles indicated by the arrow.

## 4.4 Halo Mass Function

In Figure 4.7 I show the halo mass functions from the hydrodynamical Columba simulations split by the large-scale spherical overdensity ($r = 5$Mpc, $\delta_5$) of the regions sampled in the Columba suite. Each overdensity, spanning from the $\pm 2\sigma\delta$ in Figure 3.6, are represented by different colours. Each line represents a composite galaxy population formed from the 5 regions at each overdensity and the mass function is constructed from the composite population of field haloes. All of the regions' HMFs fall below Tinker et al. (2008) and Bocquet et al. (2016) (grey lines), this is expected given that the highest density regions ($+2\sigma$) are still underdense with $\log(\delta_5) \sim -0.282$. Between overdensities, there is a systematic shift in the HMFs at all masses with higher density regions having more haloes at a given mass. At the low mass end, i.e. the power law regime, the shift can be fully explained by the ratio of overdensities. This is shown in Figure 4.8 where

HMFs have been normalised to $\delta_5$ then multiplied by $M_{200,\rm crit}$ to reduce the dynamic range and highlight differences in the slope. Below a halo mass of $\sim 10^9 \rm M_\odot$, all the lines overlap, meaning the slope of the HMFs at the low mass end does not depend on the environment of our regions, and the normalisation in the HMFs reflect the overdensity (i.e. enclosed mass) of each region. Fattahi et al. (2020) report similar results for Local Group-like environments: there is a tight correlation between the number of low mass haloes and the total mass inside a 3Mpc sphere centred on the Local Group-like environments.

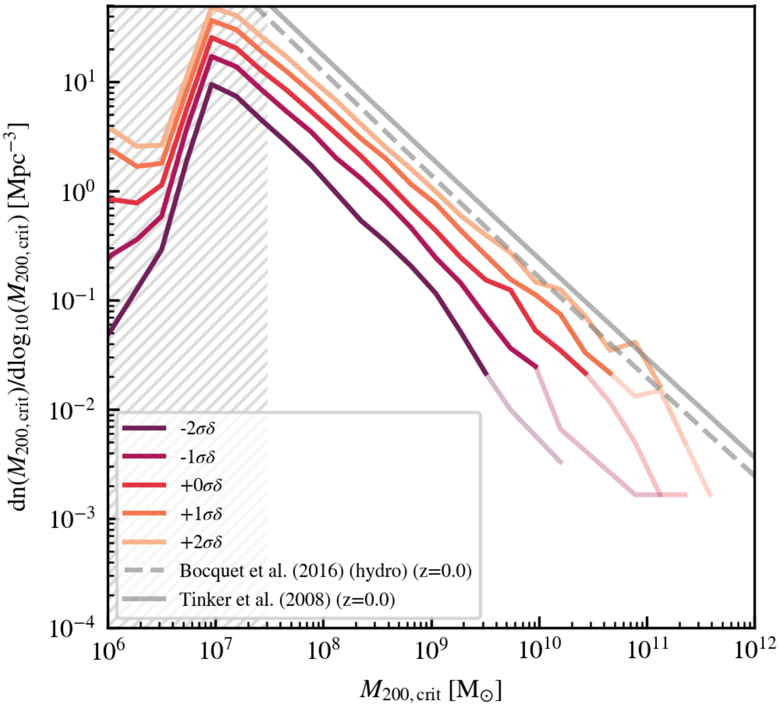


Figure 4.7: Halo mass function in the m5 hydrodynamical simulations at z=0, for all field haloes. Coloured lines represent regions with different overdensities ($\delta_5$) and each line includes 5 of the simulated volumes at a fixed overdensity. The lines show a systematic trend, with higher density regions having larger number density of haloes. The shaded region at the low mass end mark <100 DM particles for m5 resolution, hence the downward trend at the low mass end is due to the resolution. At the high-mass end, the lines have lower opacity when the bin contains fewer than 10 haloes. For reference, the grey solid and dashed lines show halo mass function from Tinker et al. (2008) and Bocquet et al. (2016), respectively.

At the high mass end, Figure 4.7 shows that the highest mass haloes are formed in regions with higher densities. Moreover, there are hints that the characteristic mass of the HMF (the ‘knee’ in the HMF) is shifting systematically with $\delta_5$. This is shown more clearly in the Figure 4.8 where the lines deviate from the power law at the high mass end, at different masses which correlate with the regions’ overdensity. I note however, the lines are noisy due to low number statistics at higher masses. These results are

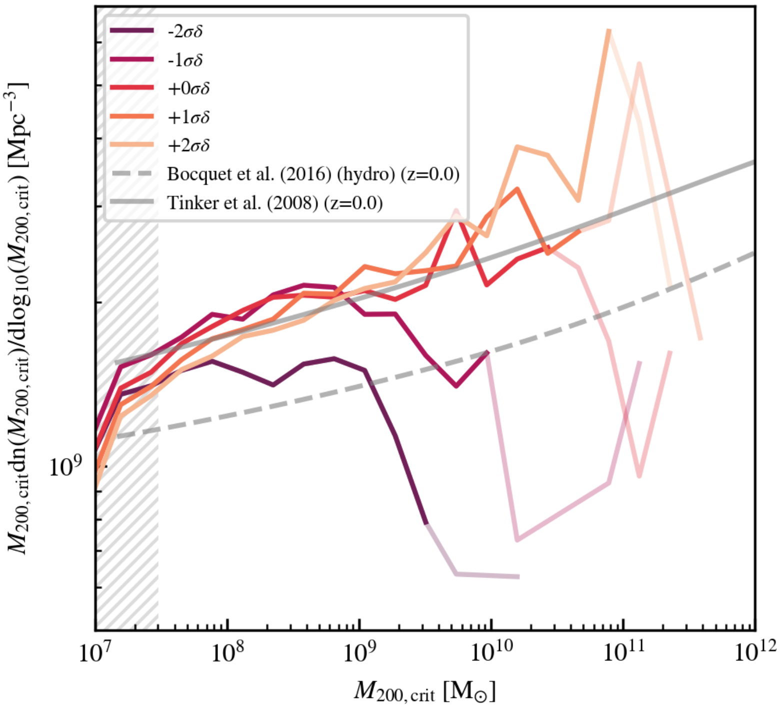


FIGURE 4.8: The HMFs from Figure 4.7 multiplied by $M_{200,\mathrm{crit}}$ $(Mdn/d(logM))$ to reduce the dynamic range, and divided by the overdensity of the regions to put all HMFs on the same scale. The shaded region indicates 100 DM particles and the lines are drawn with lower opacity when there are fewer than 10 haloes in a bin.

consistent with those from the GIMIC simulations (Crain et al., 2015) though they were probing overdensities and volumes larger than our regions. The behaviour at the high mass end is expected from the hierarchical growth of structure in ΛCDM as higher mass haloes form later in cosmic time and lower density regions are less dynamically advanced in their evolution (Bond et al., 1996).

The previous results were based on the HMFs in the hydrodynamical simulations. In Figure 4.9, I compare the HMFs in the hydrodynamical simulations with their DM-only (DMO) counterparts. The top panel is shows the hydrodynamical HMFs from Figure 4.7 with the addition of the HMFs from DMO simulations, shown as dotted lines. The bottom panel shows the ratio of hydro to DMO halo mass functions. Firstly, there is a systematic change in the HMFs between hydro and DMO with the abundance of haloes at a fixed halo mass decreased in the hydro runs. This effect has been shown and discussed in various simulations (Sawala et al., 2013; Schaller et al., 2015b) and arises due to the fact that low mass haloes lose the majority of their baryons (roughly $\Omega_{\mathrm{bar}}/\Omega_{\mathrm{m}} \sim 0.25$ of their mass budget) due to cosmic reionisation. The effect is enhanced by these haloes accreting less mass over time due to their lower total mass. The bottom panel shows that the suppression of the halo mass function is not sensitive to the large scale environment. However, I caution that within our model cosmic reionisation is homogeneous and happens at the same time in all the simulations and environments. If reionisation happens at different redshift in different overdensities, these results may change.

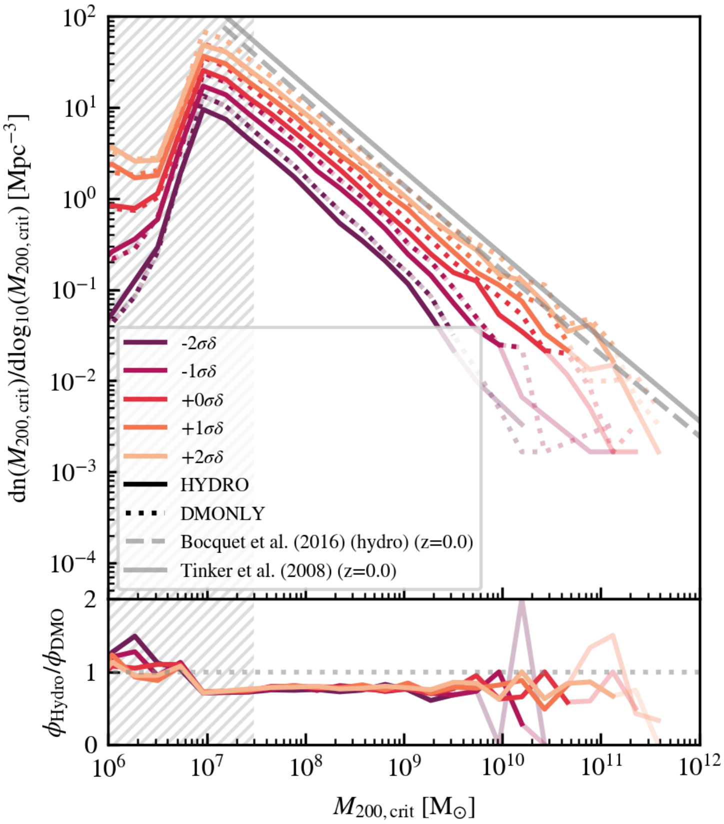


Figure 4.9: *Top:* The halo mass function of all centrals in the DM-only simulations shown by the dotted lines, and in the hydrodynamics simulations shown by the solid lines. *Bottom:* The ratio of the DM-only and hydrodynamical halo mass functions ($\phi$) showing the suppression of the HMF due to the presence of baryons; haloes in the hydrodynamical runs have lower masses causing the a systematic shift in the HMF to the left. At the high-mass end, the lines have lower opacity when the bin contains fewer than 10 haloes and the shaded region at the low mass end represents 100 DM particles at m5 resolution.

## 4.5 Stellar Mass Function

Turning to the properties of field dwarf galaxies, Figure 4.10 shows the galaxy stellar mass functions (GSMF) in the Columba simulations, grouped by their large scale overdensity, $\delta_5$. Similar to the halo mass functions, the GSMFs systematically shift according to the overdensity of the regions. Similarly, the GSMFs are generally below the observations Driver et al. (GAMA, 2022) or simulations (EAGLE and FIRE; Schaller et al., 2015b; Feldmann et al., 2023) due to the underdense nature of the regions sampled in the Columba suite. Following from the HMFs, lower overdensity environments are also unable to form more massive galaxies reflected by the decline in maximum stellar mass found in these regions as $\delta_5$ decreases.

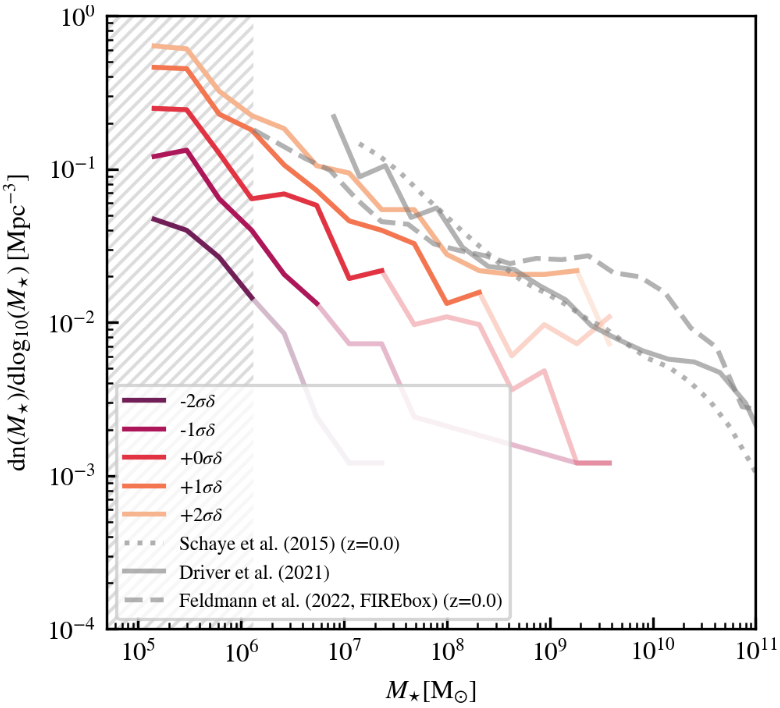


Figure 4.10: The galaxy stellar mass function of central galaxies at z=0 in our simulations, where each line corresponds to 5 simulated volumes at a fixed overdensity, and different colours represent the change in overdensity. The coloured solid lines have lower opacity at the higher-mass end when there are fewer than 10 galaxies in the bin. The shaded region indicates where galaxies have fewer than 10 star particles. The solid grey line show observations from Driver et al. (2022). The dotted and dashed lines show the results from EAGLE simulations (Schaye et al., 2015) and FIREbox (Feldmann et al., 2023).

Similar to the previous section, in order to examine the effect of environment on the GSMF more closely, I show the GSMFs multiplied by $M_\star$ and then scaled by the regions' overdensities in Figure 4.11. The slope of the GSMF does not change significantly with the environment. However, there remains a small systematic offset between SMFs even after normalising them by the density (mass) of the regions. In other words, this suggests that the galaxy formation efficiency per unit mass increases with increasing the density of regions in the stellar mass range explored here. This correlation becomes less significant at the lowest stellar mass bins. Given the uncertainties due to resolution a robust conclusion cannot be drawn regarding the SMF below stellar mass of $\sim 10^6 \mathrm{M}_\odot$, corresponding to 10 star particles.

I attribute the correlation between regions' overdensity to the differences in halo abundances, as opposed to a systematic bias in forming brighter galaxies per halo. As will be discussed in the following Section, the average stellar mass - halo mass relation is similar between regions with different overdensities. Moreover, the stellar mass explored here ($M_\star > 10^6 \mathrm{M}_\odot$) corresponds to a halo mass $M_{200,\mathrm{crit}} \gtrsim 10^9 \mathrm{M}_\odot$. This is precisely the scale where differences occur in the normalised HMF shown in Figure 4.8.

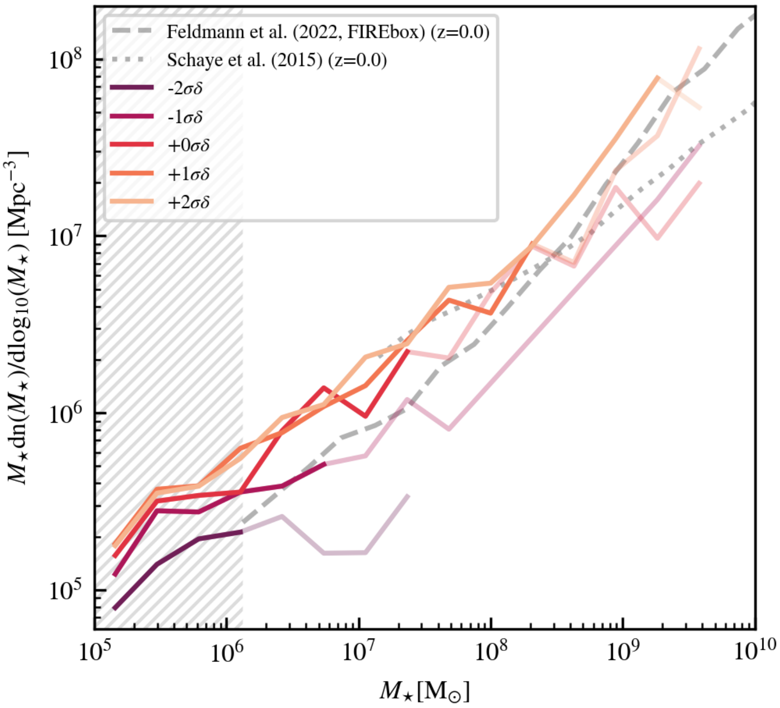


FIGURE 4.11: The SMFs from Figure 4.10 multiplied by $M_\star$ to reduce the dynamic range ($Mdn/d(logM)$), and divided by the overdensity of the regions to put all SMFs on the same scale. The shaded region indicates 10 star particles and the lines are drawn with lower opacity when there are fewer than 10 galaxies in a bin.

## 4.6 Stellar Mass Halo Mass Relation

The shape of the stellar mass-halo mass (SMHM) relation at the faint end is very uncertain and a hindering factor in robust predictions for the properties of dwarf galaxies. Sales et al. (2022) collate predictions for the SMHM relation at the faint end from various cosmological hydrodynamical simulations which show a general agreement between the models in the range $M_\star > 10^5 \mathrm{M}_\odot$ and with extrapolated abundance matching relations, however, the slope and scatter show significant variance between studies. I examine here whether the large scale environment influences the SMHM relation for field dwarf galaxies in the Columba simulations.

In Figure 4.12 I show the SMHM relation of field dwarf galaxies in the Columba simulations at z=0, separated by the overdensity of the regions, $\delta_5$. Individual galaxies are colour coded according to the $\delta_5$ of their parent region. I exclude satellites (at z=0) whose evolution may be affected by ram pressure stripping of gas and mass loss due to tidal stripping which increase the scatter in the SMHM relation. Potential backsplash galaxies, which have been satellites in the past and are currently outside of the virial radius of the halo, have not been removed from the field sample. I expect there would be a very small number of such objects since there is no massive host in our simulated regions. The horizontal lines in the scatter at the low mass are due to discrete particle numbers. The shaded regions indicate where galaxies contain less than 10 stellar particles, or less than 100 DM particles. There is no obvious dependence on environment amongst the scatter of individual galaxies.

The solid lines with varying colours show the median SMHM relation for the 5 different region overdensities. These lines show the median stellar mass at a fixed halo mass and include haloes with no stars. The lines are truncated at varying stellar and halo masses for different overdensities. As demonstrated by the HMFs and SMFs in Figures 4.7 and 4.10, lower density regions are devoid of higher mass haloes, hence their average lines stop at smaller masses. I do not find significant variation in the slope and normalisation of the median SMHM relation between overdensities, indicating the larger scale environment is not influencing this relation in the regime explored here. This lack of dependence of SMHM on regions' overdensities indicates that the differences found in normalised stellar mass functions between various regions in Figure 4.11 is due to the differences in halo mass function, rather than in the efficiency of galaxy formation in the host dark matter halos.

The absence of any significant variation in the SMHM relation with regions' overdensities suggests that for dwarfs forming in isolation and with $M_\star > 10^5\mathrm{M}_\odot$, the environment on scales of $\sim$ 5Mpc does not significantly change their evolution. However, this does not rule out the influence of environment on even larger or smaller scales on galaxy formation. Wu et al. (2024) utilise a machine learning model with galaxies ($M_\star > 10^9 M_\odot$) from

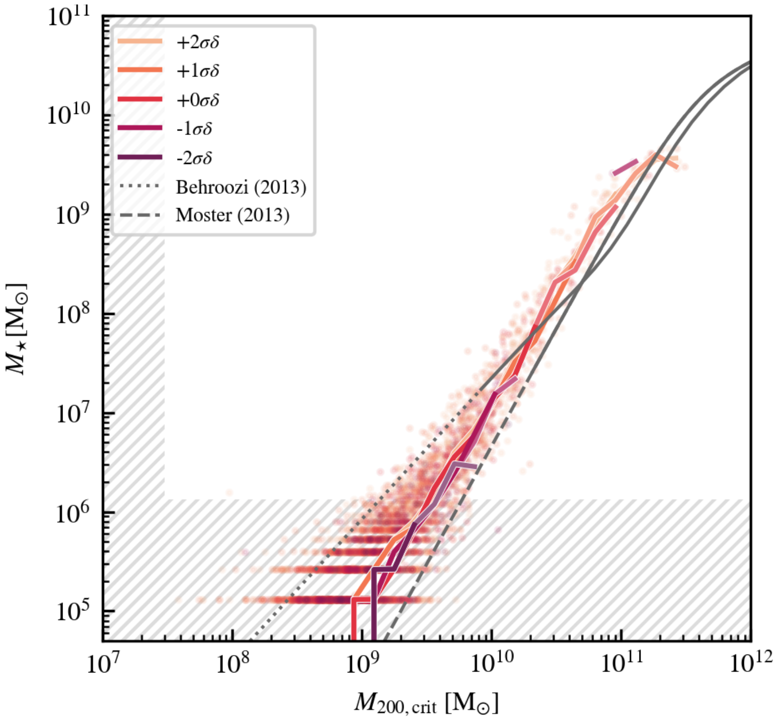


Figure 4.12: The stellar mass - halo mass (SMHM) for galaxies in the simulation suite. Only central galaxies are included, to minimise the effect of potential tidal stripping. The lines represent the median of galaxies stacked by the large scale overdensity which shows no clear influence on the slope or amplitude of the SMHM relation. The shaded region indicates where galaxies have fewer than 10 star particles and where haloes have fewer than 100 DM particles. The dashed and dotted lines show the extrapolated abundance matching relations of Moster et al. (2013) and Behroozi et al. (2013).

the Illustris TNG-300 simulation (Nelson et al., 2019; Pillepich et al., 2017) to examine the influence of spherical overdensity on the galaxy-halo connection. They investigate spherical overdensities over various scales and find that the optimal length influencing the galaxy-halo connection is $L = 3$Mpc. Above this length the information is smoothed out over too large a scale and may be lost. This effect may contribute in explaining the apparent difference between my results and those of Meshveliani et al. (2024) who find a (weak) correlation between SMHM relation and spherical overdensity over $R = 3$Mpc spheres. I note however that the range of overdensities sampled here (less than 1-dex) is smaller than the range investigated by Wu et al. (2024) and Meshveliani et al. (2024), and weak correlations might not show up.

I conclude that environments on the scales of $r = 5$Mpc or larger have no significant influence on the SMHM relation of isolated dwarfs. This raises a question about what contributes to the scatter in the SMHM relation of field dwarf galaxies which I investigate further in Chapter 5.

## 4.7 Mass-Concentration Relation

The concentration of DM haloes is anti-correlated with their mass (e.g. Ludlow et al., 2014), this mass-concentration relation arises due to the hierarchical structure formation in ΛCDM (Navarro et al., 1996b) where the characteristic density of a halo is proportional to the density of the universe at its initial collapse time. Low mass haloes have higher concentrations reflecting their collapse at higher densities and higher redshift, making concentration a reasonable proxy for the formation time of a halo (Neto et al., 2007). This also implies that regions with different overdensities might introduce biases in the halo mass-concentration relation, an effect I investigate in this Section.

In Figure 4.13 I show the median halo mass - concentration relation in our hydrodynamical simulations, grouped by $\delta_5$ of their parent regions. The concentrations are estimated using the method described in Section 4.2.1 and correspond to the concentration of the matched counterparts in the DMO simulations. I adopt the concentration from the DMO counterparts in order to remove the influence of baryons on the inner structure of haloes, as I am investigating the influence of cosmological environment on the concentration of halos, rather than the effect of baryons. The solid lines in Figure 4.13 show the average relation for all field haloes. I show the CDM halo mass-concentration relation from the COCO zoom-in simulations at z=0 (Ludlow et al., 2016), this simulation follows a high resolution sphere $R \sim 18h^{-1}$Mpc embedded within a periodic box of $L = 70.4h^{-1}$Mpc. The zoom region contains a cosmologically representative sample of MW-mass halos and hence follows an average relation close to the cosmic mean density of the Universe. The Columba simulations follow lower density environments and exhibit a similar slope with systematically lower concentrations. Below a halo mass of $\sim 10^9\mathrm{M}_\odot$, the mass-concentration relation shows a systematic shift between overdensities with more dense regions forming higher concentration haloes for a given halo mass. This is consistent with the statements above on the relation of concentration and the density at the time of collapse.

Hellwing et al. (2021) investigate the mass-concentration relation across different cosmic web environments and find systematic differences for haloes in node, filament, wall, and void environments at low halo masses. They find that, on average, haloes in voids have lower concentrations than those in denser environments. A similar trend is seen in these simulations where the lower density, and hence void-like regions, have lower halo concentrations than higher density regions which are more likely to be in or close to a filamentary structure. Furthermore, the low density regions sampled here are likely the cause of the low halo concentrations which are shown to diverge from the relation of Ludlow et al. (2016) towards lower halo masses in void and filment environments.

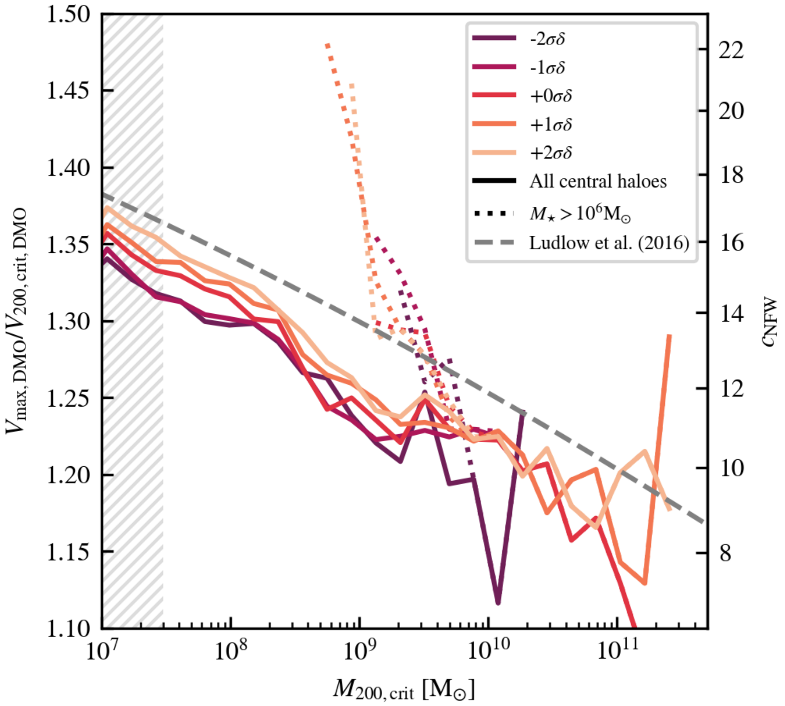


FIGURE 4.13: The median mass - concentration relation for all field haloes shown by the solid lines. The lines are coloured by the large scale overdensity $\delta_5$. The dotted lines show the relation of 'luminous' haloes that host a stellar mass $M_\star > 10^6\mathrm{M}_\odot$ and exhibit elevated concentrations. The shaded region indicates 100 DM particles.

Figure 4.13 additionally shows the mass - concentration relation of haloes that host a galaxy with a stellar mass, $M_\star > 10^6\mathrm{M}_\odot$, using dotted lines. The stellar mass threshold corresponds to 10 star particles, above which the predicted stellar mass functions are reliable. Interestingly, I find that haloes that have been able to form 'luminous' galaxies exhibit a strong bias in their concentrations. They have a much steeper mass-concentration relation, and have higher concentrations, at a fixed halo mass, than average haloes. This deviation in concentration occurs at halo masses below $\sim 10^{9.5}\mathrm{M}_\odot$ and steeply diverges towards lower masses. This suggests that higher concentration, and hence earlier forming, haloes are more likely to host a 'luminous' dwarf galaxy with $M_\star > 10^6\mathrm{M}_\odot$. Despite the effect of environment on the mass-concentration for *all* haloes, the relation for luminous ones do not show any significant change as a function of $\delta_5$ of their parent region. This implies that the galaxy formation efficiency of a given halo is mostly influenced by its assembly history (traced by the concentration here) rather than the environment.

The strong bias seen in the mass-concentration relation of 'luminous' haloes ($M_\star > 10^6\mathrm{M}_\odot$), is mainly the result of the shape of the SMHM relation in our simulations. The SMHM relation, see for example, Figure 4.3, indicates that $M_\star > 10^6\mathrm{M}_\odot$ galaxies live on average in $\sim 3 \times 10^9\mathrm{M}_\odot$ haloes, the same mass scale where mass-concentration relation of luminous haloes start to deviate from the average. The steep shape of the mass-concentration relation for luminous halos, compared to the average, can be attributed to the steep shape of the SMHM relation at the low masses. Small changes in the halo mass, results in a big change in stellar mass. Low mass haloes below $M_{200,\mathrm{crit}} \sim 3\times10^9\mathrm{M}_\odot$ must have biases in their assembly history to be able to host galaxies more luminous than the threshold. In particular, haloes in this relevant mass scale are affected by reionisation and those that have higher mass pre-reionisation are more likely to have higher stellar masses by $z = 0$. Rey et al. (2019) shows this effect using EDGE hydroydynamical simulations.

The stellar mass threshold chosen to define the 'luminous' haloes above is motivated by the resolution of the simulations; $10^6\mathrm{M}_\odot$ corresponds to 10 star particles and the simulated SMFs showed good convergence above this minimum number of star particles. The effect of changing the stellar mass threshold on the mass-concentration is shown in Figure 4.14. I find that by changing the threshold to $M_\star > 10^5\mathrm{M}_\odot$ or $M_\star > 10^7\mathrm{M}_\odot$, the divergence from the average mass-concentration relation starts at higher halo mass ($M_{200,\mathrm{crit}} \sim 10^7\mathrm{M}_\odot$) or lower halo mass ($M_{200,\mathrm{crit}} \sim 10^9\mathrm{M}_\odot$), respectively, in accordance with the average SMHM relation. Furthermore, I do not find a strong dependence of the above results on the resolution of the simulations.

## 4.8 Luminous Fraction

Another angle of investigating the mass-concentration results is by looking at the fraction of 'luminous' haloes ($M_\star > 10^6\mathrm{M}_\odot$) as a function of halo mass. The top panel of Figure 4.15 shows the luminous fraction as a function of halo masses, separated by the overdensity of the regions, $\delta_5$. Overall, all regions follow a similar trend: all haloes with $M_{200,\mathrm{crit}} > 10^{10}\mathrm{M}_\odot$ host a 'luminous' galaxy (an exception is the $-2\sigma$ regions

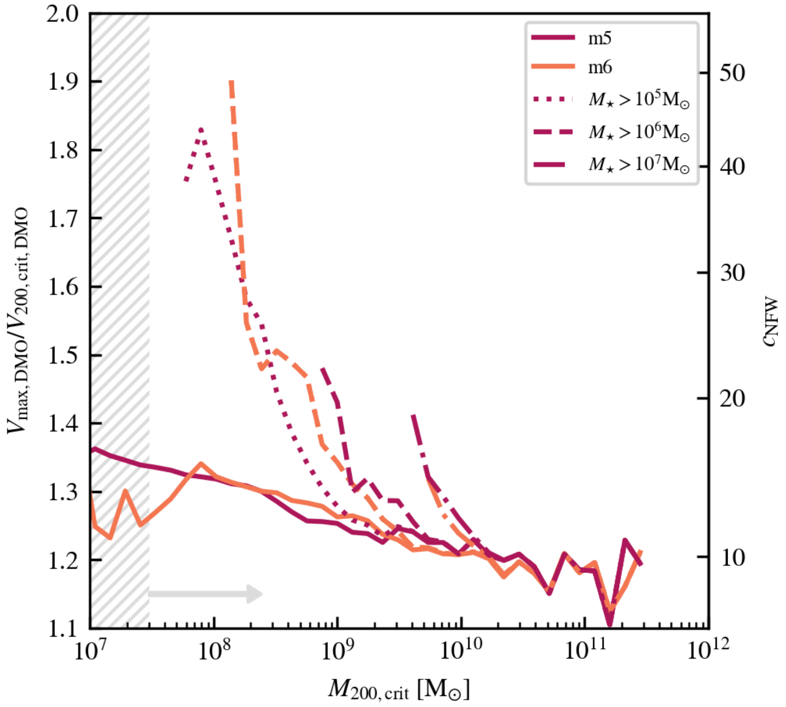


Figure 4.14: The effect of changing the definition of 'luminous' haloes on the mass - concentration relation of field haloes in the Columba simulations. The solid lines include all field haloes and the dotted, dashed, and dash-dotted lines correspond to stellar mass thresholds of $M_\star > 10^5, 10^6$, and $10^7$M$_\odot$ respectively.

which have very few galaxies at higher masses and are dominated by noise); almost no halo below $10^9$M$_\odot$ has been able to form a luminous galaxy. The mass scale where this luminous fraction is at 50 per-cent is around $3 \times 10^9$M$_\odot$ which is the same mass scale where the mass-concentration of luminous haloes start to diverge from the average mass-concentration relation.

I notice a weak trend in luminous fraction with environment, $\delta_5$, which is more noticeable at the lower mass end. The higher density regions have slightly higher luminous fractions at halo masses close to $\sim 10^9$M$_\odot$. I attribute this systematic difference to the earlier findings, shown in Figure 4.13, that the average mass-concentration for *all* haloes shift systematically with $\delta_5$ of their parent region. In other words, higher density regions host higher concentration haloes (earlier forming halos) on average, which in turn have higher chance of hosting 'luminous' galaxies.

The important role of halo concentration on the luminous fraction is highlighted in the bottom panel of Fig. 4.15, where I show the luminous ($M_\star > 10^6$M$_\odot$) fraction of haloes, separated into bins of high, intermediate, and low halo concentrations measured from their DMO counterparts. The three concentration bins correspond to the terciles of the halo concentration distribution; high concentration haloes have $c > 18.1$, intermediate concentration haloes have $12.7 < c < 18.1$, and low concentration haloes have $c < 12.7$. The luminous fraction changes significantly with halo concentration. For example, at $M_{200,\mathrm{crit}} = 2 \times 10^9 \mathrm{M}_\odot$, the occupation fraction changes from $\sim 10\%$ to $\sim 80\%$ when going from the high concentration bin to the low concentration. These results, again, show that the formation and evolution of low mass dwarf galaxies are tied to their halo assembly history which I explore using halo concentration.

I emphasize that by using a stellar mass threshold of $M_\star > 10^6 \mathrm{M}_\odot$ in the previous discussion to define 'luminous' galaxies, I am not probing the absolute halo occupation fraction. In Figure 4.15 I show how the luminous fraction changes when changing the stellar mass threshold to $10^7 \mathrm{M}_\odot$ and $10^5 \mathrm{M}_\odot$, represented by the dotted-dashed as dashed lines, respectively. As expected the luminous fraction curve shifts to higher and lower halo masses. Given the resolution limit of the simulations, $M_{\mathrm{star}} > 10^5 \mathrm{M}_\odot$ is lowest mass galaxies that are able to form in the simulation, and the corresponding luminous fraction curve should be viewed as an upper limit on halo mass for the true occupation fraction.

There are large uncertainties in the literature about the true occupation fraction at these low masses. Two mass scales are often discussed in the literature affecting the occupation fraction. Reionisation affecting the evolution of haloes with present day halo mass of $M_{200,\mathrm{crit}} \lesssim 10^{9.5} \mathrm{M}_\odot$ and causing the occupation fraction to drop below 1 on those scales. Another critical mass threshold is the hydrogen atomic cooling limit, $M_{200,\mathrm{crit}} \sim 10^9$ at z=0, setting the mass scale above which haloes can cool gas and form stars (most relevant before reionisation). Most haloes with present-day halo mass of $M_{200,\mathrm{crit}}(z = 0) < 10^{8.5} \mathrm{M}_\odot$ have never been above atomic hydrogen cooling limit in their history and are predicted to remain completely devoid of stars to z=0, independent of the resolution of simulations. Using these important mass scales, Benitez-Llambay & Frenk (2020) derive the true occupation fraction, which is dependent on the redshift of reionisation.

Comparing to their occupation fraction for $z_{\mathrm{reion}}$ =6 and 8, where $z_{\mathrm{reion}} = 7.5$ in the Columba simulations, the Columba $M_\star > 10^5 \mathrm{M}_\odot$ luminous fraction is shifted towards lower halo masses from their prediction. Considering our results should be seen as an upper limit to the true occupation fraction, they are in tension with the model of

Benitez-Llambay & Frenk (2020). In particular, haloes in our model appear to form stars even below the hydrogen atomic cooling limit. Investigating the source of this is beyond the scope of the current Chapter and can be investigated in future work.

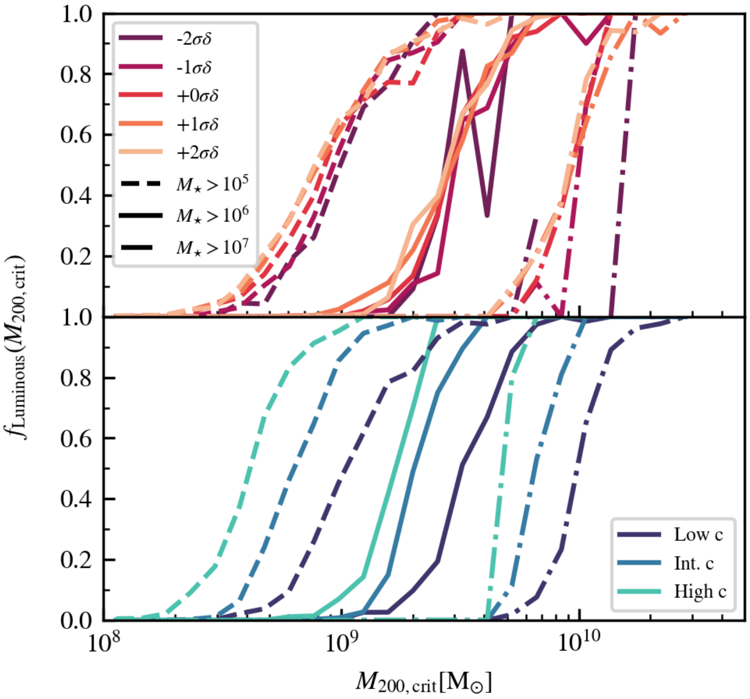


FIGURE 4.15: *Top:* The luminous fraction of all central haloes as a function of halo mass, split between overdensities. A luminous halo is defined as a halo hosting a stellar mass $M_\star > 10^6 M_\odot$ in solid lines and the luminous fraction is the number ratio of luminous to total haloes ($N_{\mathrm{Luminous}}/N_{\mathrm{total}}$) in a mass bin. The dashed and dash-dotted lines show different definitions of 'luminous', $M_\star > 10^5\mathrm{M}_\odot$ and $M_\star > 10^7\mathrm{M}_\odot$ respectively. *Bottom:* The luminous fraction of all central haloes divided into low, intermediate, and high concentration bins. Halo concentration is calculated as described in Section 4.2.1 and the bin limits are determined by the terciles of the total concentration distribution.

## 4.9 Conclusion

In this Chapter I compare key diagnostics from the Columba simulations, introduced in Chapter 3, to a combination of observations and other simulations and models to test the validity of our model. The simulations adopt a galaxy formation model whose components are either ported directly from the EAGLE model, or which represent a substantial evolution of the equivalent component in EAGLE. The key developments are the inclusion of an explicit model of cold, dense neutral gas, a star formation model that accounts for gas turbulence, a treatment of pre-supernova feedback from massive stars, more sophisticated methods of distributing the energy liberated by SNe and active galactic nuclei (AGN), and updated treatments of BH growth and dynamics. In contrast to EAGLE, the subgrid efficiency of feedback from SNe is not calibrated to reproduce one or more observable properties of galaxies. Nevertheless the simulations, when renormalised to account for their preferential sampling of underdense regions, reproduce the present-day galaxy stellar mass function (GSMF) in the stellar mass regime they resolve and sample adequately to within 0.27dex. The SMHM relation is within a similar level of agreement as is achieved between extrapolations from abundance matching relations and predictions from other simulations in the dwarf galaxy regime. However, Columba produces galaxies that are too compact above $M_\star > 5 \times 10^8 \mathrm{M}_\odot$ and too large below, and are poorly converged between resolutions m5 and m6, highlighting the importance of calibration which is necessary to match additional observables.

The regions probed by the simulations span 5 different densities and are characterised by various cosmic web environments from void to filaments. I quantify environment using the region's spherical density, which correspond to $(-2\sigma, -1\sigma, +0\sigma, +1\sigma, +2\sigma)\delta$ from Figure 3.6, collating 5 simulations per environment, and investigate their influence on properties of isolated dwarf galaxy populations. The HMFs from each environment shift systematically based on the density of the region. I find that the shift is mostly accounted for when scaling by their density and show that the slope of the HMFs are similar between different environments. There are suggestions that the characteristic mass of the HMF (the 'knee') shifts systematically with the density of the region, however, the poor sampling in the lowest density regions makes this difficult to assess. The GSMFs also shift according to the density of the region. Scaling the GSMFs by density shows that they follow similar slopes but there is still an offset in the scaled GSMFs between the high and low density regions. I attribute the correlation between regions' overdensity to the differences in halo abundances, as opposed to a systematic bias in forming brighter galaxies per halo as further examination shows that the slope of the SMHM relation does not show a dependence on the large scale environment.

These findings appear to be in tension with those of Zehavi et al. (2018), who examine the influence of halo assembly bias on the galaxy population predicted by a semi-analytic model when applied to the Millennium Simulation (Springel et al., 2005), and of Artale et al. (2018) who performed a similar study using the EAGLE Ref-L100N1504 simulation. In both cases, a significant correlation between the median stellar mass at fixed halo mass with the total mass density measured within a sphere of $r = 5\mathrm{Mpc}$ is found for haloes of mass $M_{\text{halo}} \gtrsim 10^{11}\text{M}_{\odot}$. This halo mass range is not well sampled by the Columba simulations as can be seen in Figure 4.7 and hence a direct comparison cannot be made, however there are several significant factors that likely contribute to the difference in conclusions. Both studies examine a significantly larger dynamic range in density (approximately 3 dex in the Millennium Simulation and 2 dex in EAGLE) than is sampled by the Columba simulations (approximately 0.8 dex). In both cases the positive correlation between $M_{\star}(M_{200,\text{crit}})$ and the environmental density is reflected strongly by an increased formation efficiency for galaxies located in high-density regions, which are absent from my simulation suite by construction. Additionally, these studies measure the environmental density in a sphere centred on each galaxy, and thus perform a different test to that presented in Figure 4.12. A more direct comparison to the findings of Zehavi et al. (2018) and Artale et al. (2018) can be achieved with galaxy-specific measures of (or proxies for) environment. Furthermore, I cannot rule out the influence of environment on different scales on dwarf galaxy evolution, and note that Wu et al. (2024) recently concluded from an analysis of the IllustrisTNG-300 simulation that the galaxy-halo connection is most significantly influenced by the environment measured on a scale of $\simeq 3\,\text{Mpc}$, smaller than examined in this Chapter.

Although this measure of environment does not show a signal in the SMHM relation, halo concentrations at fixed halo mass show a gradient with region density on $r = 5\text{Mpc}$ scales, reflecting the environmental density in which they formed. The halo concentrations of isolated low mass haloes in the Columba simulations are lower than predicted by Ludlow et al. (2016), but are consistent with findings by Hellwing et al. (2021) who find that haloes in voids and filaments have on average lower concentrations. I show that the mass-concentration relation of luminous haloes ($M_{\star} > 10^{6}\text{M}_{\odot}$) is steeper below $M_{200} < 3 \times 10^{9}\text{M}_{\odot}$ suggesting that low mass haloes must have biases in their assembly history to be able to host galaxies more luminous than the threshold. This effect is corroborated by the luminous fraction of haloes which when examined as a function of halo concentration shows a significant difference; at a fixed mass of $M_{200} \sim 3 \times 10^{9}\text{M}_{\odot}$, the luminous fraction differs by $\sim 60\%$ between low and high halo concentration bins. Therefore, high concentration, and hence earlier forming, haloes are more likely to host a luminous component and this effect dominates over any variance between environments. This implies that halo occupancy is mostly influenced by its assembly history for which

I use concentration here as a proxy, rather than its large-scale environment. This provides insight into the effects of reionisation, suggesting that present-day low-mass haloes whose main progenitors formed earlier and had greater masses before reionisation are better able to retain their gas and consequently form a significant stellar mass by the present day. A similar conclusion was reached by Rey et al. (2019) from an analysis of the EDGE simulations.

The findings presented in this Chapter highlight that the influence of large-scale environment (defined on $\simeq 5\,\mathrm{Mpc}$ scales) on the properties of low-mass field galaxies is driven primarily by the dissimilar distribution of halo properties (such as halo concentration or halo formation time) that emerge in differing environments. In Chapter 5 I therefore focus on the correlation between scatter in the low-mass regime of the SMHM relation at fixed halo mass with, firstly, halo concentration, and secondly, the corresponding formation time of the galaxy.

# 5

# Columba II: Formation Time and Stellar Assembly

## 5.1 Introduction

Dwarf galaxies, the smallest and most numerous galaxies in our Universe, play a critical role in our understanding of galaxy formation and evolution. In particular, isolated dwarf galaxies are insulated from the environmental effects from high density regions which can impact the gas, stellar, and dark matter distributions, and can therefore offer unique insights into the interplay between dark and baryonic matter. A fundamental parameter in this context is halo concentration which characterises the DM distribution of the halo and is strongly linked to its formation time (Neto et al., 2007). This parameter has been shown to significantly influence properties of galaxies from their assembly history (Wechsler et al., 2002) to their ability to host galaxies as demonstrated in Chapter 4. This chapter aims to directly explore how variations in halo concentration affect key properties of dwarf galaxies.

Understanding galaxy formation involves understanding the luminous component of the Universe, the collapse of dark matter haloes under ΛCDM cosmology, and the link between these processes. This galaxy-halo connection can be quantified by the stellar mass halo mass (SMHM) relation which specifies the mass of galaxies that occupy dark matter haloes of a given mass. Dark matter halo masses are challenging to measure observationally, however, various modelling approaches including halo occupation models, abundance matching, and hydrodynamical simulations have helped to quantify this relation with good constraints above a halo mass of $M_{200,\mathrm{crit}} \gtrsim 10^{11}\mathrm{M}_{\odot}$ (e.g. Wechsler & Tinker, 2018). With decreasing halo mass, star formation efficiency decreases and below a certain mass threshold, galaxies are increasingly unlikely to form. In the low mass range, numerical models show a high degree of scatter in the SMHM relation; different simulations loosely follow extrapolations of abundance matching models, however, in detail the slope and scatter varies (Sales et al., 2022), and how low mass galaxies occupy haloes is generally uncertain. Several studies have found that halo concentration influences the scatter in the SMHM relation. Matthee et al. (2016) examine simulated galaxies from the EAGLE simulation with halo mass $M_{200,\mathrm{DMO}} > 10^{11.1}\mathrm{M}_{\odot}$ and show that halo concentration is a significant driver of scatter. Similarly, Fitts et al. (2017) utilise zoom-in simulations of isolated dark matter haloes $M_{\mathrm{halo}} \sim 10^{10}\mathrm{M}_{\odot}$ with the FIRE-2 feedback model, and find a spread in stellar mass at z=0 which is strongly correlated with the central halo density, emphasising the role of halo concentration in the absence of confounding environmental effects.

The importance of formation time on dwarf galaxies has also been linked to their ability to retain their gas content and resist quenching during reionisation. In this low mass regime, observations of Local Group dwarf galaxies of mass scale $M_{\star} \sim 10^{9}\mathrm{M}_{\odot}$ show an absence of quenched field galaxies and has led to theories surrounding a stellar mass threshold for internal quenching at $M_{\star} < 10^{9}\mathrm{M}_{\odot}$. There are however observed field dwarf galaxies in the Local Group, such as Tucana and Cetus, which exhibit little to no recent star formation. Tensions have also arisen from simulations, Fitts et al. (2017) find two isolated dwarf galaxies below this mass threshold in their zoom-in simulations that are self-quenched with no SF or cold gas at z=0. Results from the MARVELous Dwarfs and D.C.Justice League simulations are consistent with observations between $10^{8}\mathrm{M}_{\odot} < \mathrm{M}_{\star} < 10^{9}\mathrm{M}_{\odot}$, whereby no galaxies were quenched beyond 350kpc from a massive galaxy (Christensen et al., 2024). At lower masses of $10^{4}\mathrm{M}_{\odot} < \mathrm{M}_{\star} < 10^{6}\mathrm{M}_{\odot}$, their quenched fraction increases to 100% in the regime $M_{\star} < 10^{6}\mathrm{M}_{\odot}$ where reionisation is thought to be the likely cause of quenching (Bovill & Ricotti, 2011).

Results from Chapter 4 highlight the importance of the halo concentration, hence halo assembly, on the stellar mass (or luminous fraction) of objects. I investigate this further in this chapter by looking into the influence of halo concentration on the scatter of the stellar mass - halo mass (SMHM) relation, the assembly histories of isolated dwarf galaxies, and their morphologies.

## 5.2 Methodology

In this chapter I utilise the Columba simulations which comprise 25 simulations of isolated environments containing dwarf galaxy populations, the development and details of which are described in Chapter 3. Following Chapter 4, I calculate halo concentration from the matched DMO counterparts of haloes using their $V_{\rm max}/V_{200}$ ratio and approximate their NFW concentration via equation 4.2.1. In this chapter I also utilise a kinematic morphology measure following the method of Correa et al. (2017) who calculate the corotation parameter, $\kappa_{\rm co}$, as the fraction of corotational to total kinetic energy,

$$\kappa_{\rm corot} = \frac{K_{\rm corot}}{K} \tag{5.1}$$

where the total kinetic energy $K$ is defined as,

$$K = \frac{1}{2}\sum_{i}^{r<30{\rm kpc}} m_i v_i^2 \tag{5.2}$$

and the corotational kinetic energy $K_{\rm corot}$ is,

$$K_{\rm rot} = \frac{1}{2}\sum_{i, L_{z,i}>0}^{r<30{\rm kpc}} \frac{L_{z,i}^2}{m_i R_i^2} \tag{5.3}$$

These quantities are summed over all stellar particles within $r < 30$kpc of the VELOCIraptor minimum centre of potential. $m_i$ and $v_i$ are the mass and velocity of each stellar particle, $L_{z,i}$ is the particle angular momentum along the direction of the total angular momentum of the stellar component of the galaxy, and $R_i$ is the projected distance to the axis of rotation. Following Correa et al. (2017), $K_{\rm corot}$ is computed by considering only star particles that follow the direction of rotation of the galaxy ($L_{z,i} > 0$). $\kappa_{\rm co}$ has been shown to be a reasonable measure of visual morphology (Correa et al., 2017; Thob et al., 2019).

## 5.3 Correlation with Halo Concentration

In Figure 5.1 I show the SMHM relation of field galaxies at z=0 in the Columba simulation suite, coloured by their DM halo concentration which are measured from the matched counterparts in DMO runs as described in Section 4.2.1. Since the SMHM relation is steep at the low mass end, I have normalised the stellar mass axis by the halo mass to reduce the dynamic range and show the scatter more clearly. The solid grey line shows the running median of $M_\star/M_{200,\mathrm{crit}}$ as a function of $M_{200,\mathrm{crit}}$, computed using the locally weighted scatter plot smoothing method (LOWESS, Cleveland, 1979). The scatter shows a clear correlation with halo concentration; at a fixed halo mass, higher concentration haloes host a galaxy with greater stellar mass.

I quantify the statistical significance of this in the bottom panel of Figure 5.1. I compute residuals about a running median for $c_{\mathrm{DMO}}$ as a function of $M_{200,\mathrm{crit}}$ and denote this $\Delta c_{\mathrm{DMO}}$. The residuals of $M_\star/M_{200,\mathrm{crit}}$ from the running median are denoted $\Delta M_\star/M_{200,\mathrm{crit}}$. The bottom panel shows a running value of the Spearman rank correlation coefficient ($\rho$) between $\Delta c_{\mathrm{DMO}}$ and $\Delta M_\star/M_{200,\mathrm{crit}}$, which exhibits a significant positive correlation ($p < 0.01$) between halo concentration and stellar mass at a fixed halo mass, across the resolved mass range. Simply, this means that deviations away from the median concentration are strongly correlated with the deviation from the median stellar mass (normalised by halo mass) at a given halo mass. The strongest correlation is characterised by a maximum coefficient of $\rho = 0.850$ at $M_{200,\mathrm{crit}} = 1.2 \times 10^{11}\mathrm{M}_\odot$.

These findings extend the results of Matthee et al. (2016), who find that halo concentration is partly responsible for the scatter in the SMHM relation of EAGLE galaxies with masses between $10^{11} < M_{200,\mathrm{crit,DMO}} < 10^{12}\mathrm{M}_\odot$, with the strongest effect in their low mass range, and that the scatter can be reduced by up to 0.04 dex by correcting for concentration. Matthee et al. (2016) speculate that higher concentration galaxies in this mass range are likely to be in a more advanced stage of their evolution and/or because it is harder for feedback to drive winds out of haloes that are more concentrated. For dwarf galaxies in the MARVEL-ous Dwarfs and D.C. JUSTICE LEAGUE simulations, Christensen et al. (2024) find a similar trend with concentration (although calculated for galaxies at high-z ($t \sim 2.6$Gyr)) and the residuals from the SMHM relation. This trend is stronger when only considering isolated galaxies, with galaxies lying above the average SMHM relation having higher concentrations at earlier times.

In the lower mass regime, this effect may also be attributed to reionisation which is expected to quench star formation in dwarf galaxies via photoevaporation, prolonged cooling times, and preventing gas inflows (Katz et al., 2020). Furthermore, dwarf galaxies that have higher concentrations and hence earlier formation times have accumulated

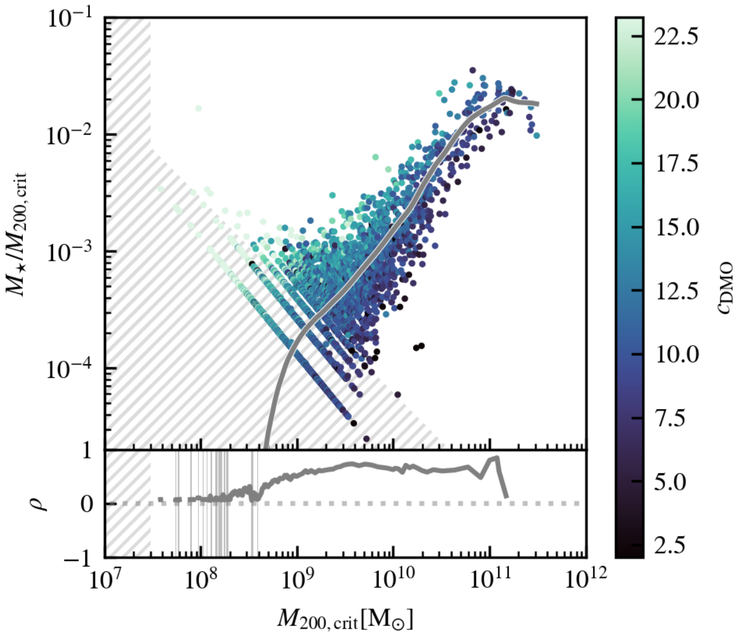


Figure 5.1: *Top*: The stellar mass-halo mass relation of the central simulated galaxies, coloured by the NFW concentration of their dark matter haloes. The median stellar mass at a fixed halo mass is calculated by a running LOWESS method and shown as a grey line. The scatter at a fixed halo mass shows a clear trend with concentration, with higher stellar masses being associated with higher concentration haloes. The shaded region indicates where galaxies have fewer than 10 star particles or 100 DM particles. *Bottom*: The running Spearman rank correlation coefficient, $\rho$, between the residuals $\Delta c_{\mathrm{DMO}}$ and $\Delta M_\star/M_{200,\mathrm{crit}}$. Positive values denote a positive correlation. The shaded regions indicate when the correlation has low significance ($p > 0.01$).

more mass at the time of reionisation and are more likely to retain or accrete gas following reionisation and grow to larger stellar masses by present-day.

### 5.3.1 Stellar Mass - $V_{\mathrm{max}}$ Relation

I also investigate the relation between stellar mass and $V_{\mathrm{max}}$, and the influence of halo concentration. I show stellar mass against $V_{\mathrm{max}}$ in Figure 5.2 with the grey line representing the running LOWESS median and the scatter coloured by halo concentration. The scatter is significantly reduced using $V_{\mathrm{max}}$, which accounts for concentration at a fixed halo mass. Similar to Figure 5.1, I calculate the residuals about the running median for $c_{\mathrm{DMO}}$ as a function of $V_{\mathrm{max}}$, denoted $\Delta c_{\mathrm{DMO}}$ and the bottom panel shows the running Spearman rank correlation coefficient ($\rho$) between $\Delta c_{\mathrm{DMO}}$ and $\Delta V_{\mathrm{max}}$. The stellar mass - $V_{\mathrm{max}}$ scatter does not show a strong correlation with halo concentration, implying that the stellar mass of a galaxy can be more reliably inferred from its $V_{\rm max}$ than halo mass without knowledge of its host halo concentration or formation time.

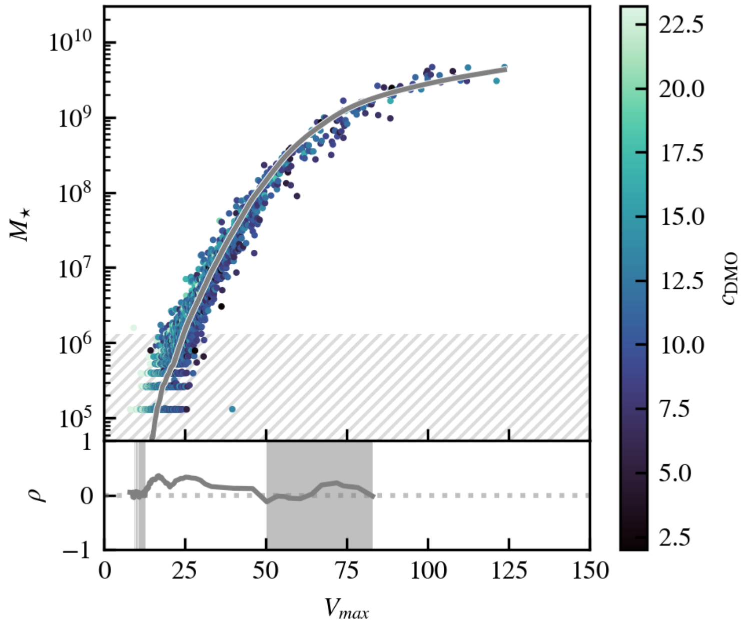


Figure 5.2: *Top:* The relation between stellar mass and $V_{max}$ of field galaxies in the Columba suite, points are coloured by NFW concentration. *Bottom:* The running Spearman rank coefficient between $\Delta c_{\rm DMO}$ and $\Delta V_{\rm max}$. Unlike the $M_\star - M_{200}$ relation, the scatter in the $M_\star - V_{\rm max}$ relation shows no significant correlation with concentration in the well resolved regime.

## 5.4 Correlation with Stellar Assembly

Given the strong connection between concentration and formation time (e.g. Wechsler et al., 2002; Zhao et al., 2009; Ludlow et al., 2014), this clear trend with stellar mass and concentration at a fixed halo mass indicates that formation time influences the assembly of stellar mass and how a galaxy forms stars over its lifetime. Here I investigate whether the concentration of the dwarf galaxy halo shows any signal in the star formation history (SFH).

Figure 5.3 shows cumulative SFHs, normalised by z=0 stellar mass, averaged over galaxies in bins of stellar mass and halo concentration. Galaxies are initially binned by their stellar mass into 3 mass ranges corresponding to low mass $10^7 < M_\star < 10^8 {\rm M}_\odot$, intermediate mass $10^8 < M_\star < 10^9 {\rm M}_\odot$ and high mass $10^9 < M_\star < 10^{10} {\rm M}_\odot$ dwarf galaxies. Within each mass range, I examine the distribution of their halo concentrations and divide the samples into 3 groups based on the terciles of the concentration distributions, hereafter referred to as high, intermediate, and low concentration bins. Figure 5.3 shows the median cumulative SFH of each concentration bin for each stellar mass range, with shaded regions showing the interquartile range.

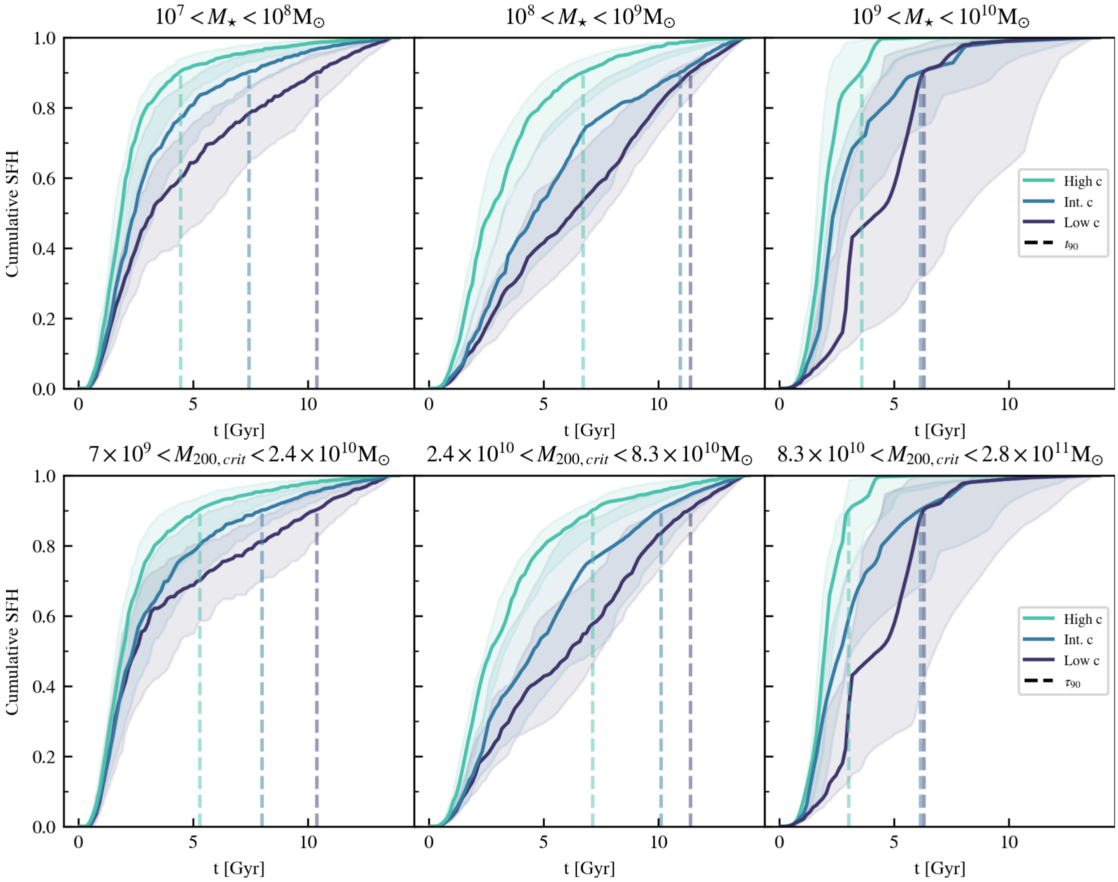


FIGURE 5.3: *Top:* The cumulative star formation histories (SFH) of central galaxies in the Columba simulation, binned by their stellar mass at z=0 and their dark matter halo concentration. Each panel corresponds to a given stellar mass bin: $10^7 < M_\star < 10^8 M_\odot$, $10^8 < M_\star < 10^9 M_\odot$ and $10^9 < M_\star < 10^{10} M_\odot$ from left to right, respectively. In each stellar mass bin, the sample is split into three bins based on the terciles of the halo concentration distribution, corresponding to different colours. The cumulative SFHs are normalised to the z=0 stellar mass first, and then median and scatter (25th-75th percentiles) are computed and shown here by the shaded regions. The median times at which 90% of the final stellar mass has formed, $t_{90}$, are shown by the vertical dashed lines. Consistent over the full mass range, higher concentration galaxies have earlier SF, reflected by earlier $t_{90}$. *Bottom:* The same as the panel above except galaxies are binned by their halo mass ($M_{200,\mathrm{crit}}$) before binning by halo concentration. The trend is consistent with the top panel.

Consistent across the stellar mass bins, the average SFH for higher concentration haloes is characterised by a steeper cumulative SFH at earlier times, compared to low concentration haloes which have a slower accumulation of stellar mass over cosmic time. As a quantitative measure of the large scale SFHs I investigate $t_{90}$, $t_{50}$, and $t_{10}$, the time at which 90%, 50% and 10% of star formation has occurred, respectively. $t_{90}$ probes the late stages of galaxy evolution when most of the star formation has completed which observationally can be linked to the colour of the galaxy. The $t_{90}$ values from the average SFHs are shown in Figure 5.3 by the vertical dashed lines and illustrate the slower evolution that occurs in galaxies of a lower halo concentration compared to galaxies in the higher concentration bins. In general, higher concentration galaxies exhibit systematically earlier $t_{10}$, $t_{50}$, and $t_{90}$ values over each stellar mass range, increasingly diverging at the end of their SFHs with the greatest distinction exhibited by $t_{90}$.

I repeat this examination in the bottom panels of Figure 5.3 and divide the sample into halo mass bins instead of stellar mass bins to investigate whether the signal comes from a change in halo mass rather than concentration. Following the same method, galaxies are binned by halo mass, $7 \times 10^9 < M_{200} < 2.4 \times 10^{10} \mathrm{M}_\odot$, $2.4 \times 10^{10} < M_{200} < 8.3 \times 10^1 0 \mathrm{M}_\odot$, and $8.3 \times 10^{10} < M_{200} < 2.8 \times 10^{11} \mathrm{M}_\odot$, and then concentration. The trend is consistent, in bins of halo mass, higher concentration haloes assemble their stellar mass earlier than low concentration haloes.

I have shown halo concentration shows a positive correlation with the scatter of the SMHM relation and propagates a signal through the large scale SFHs of galaxies. Halo concentration shows the largest influence on $t_{90}$ which probes late time stellar assembly, therefore, I investigate whether the SMHM relation shows a similar trend with $t_{90}$. In the top panel of Figure 5.4 I show the SMHM relation coloured by $t_{90}$ which exhibits a clear trend in the scatter; at a given halo mass, higher stellar masses exhibit earlier $t_{90}$. The bottom panel shows the running Spearman rank correlation coefficient between the residuals of $M_\star / M_{200,\mathrm{crit}}$ and $t_{90}$ and demonstrates a significant anti-correlation above $M_{200,\mathrm{crit}} \gtrsim 5 \times 10^9 \mathrm{M}_\odot$. This result follows that of halo concentration, given that high concentration haloes are more likely to have an earlier $t_{90}$, as shown in Figure 5.3, and high concentration haloes are more likely to host more massive galaxies. The anti-correlation peaks with a minimum correlation coefficient of $\rho = -0.49$ at $M_{200,\mathrm{crit}} = 1.0 \times 10^{10} \mathrm{M}_\odot$, this is not as strong as the correlation with halo concentration which peaked with $\rho = 0.85$.

I demonstrate this trend more explicitly in Figure 5.5 which shows the time it takes for 90%, 50%, and 10% of the cumulative star formation to occur as a function of stellar mass, each point is an individual field galaxy and is coloured by their halo concentration.

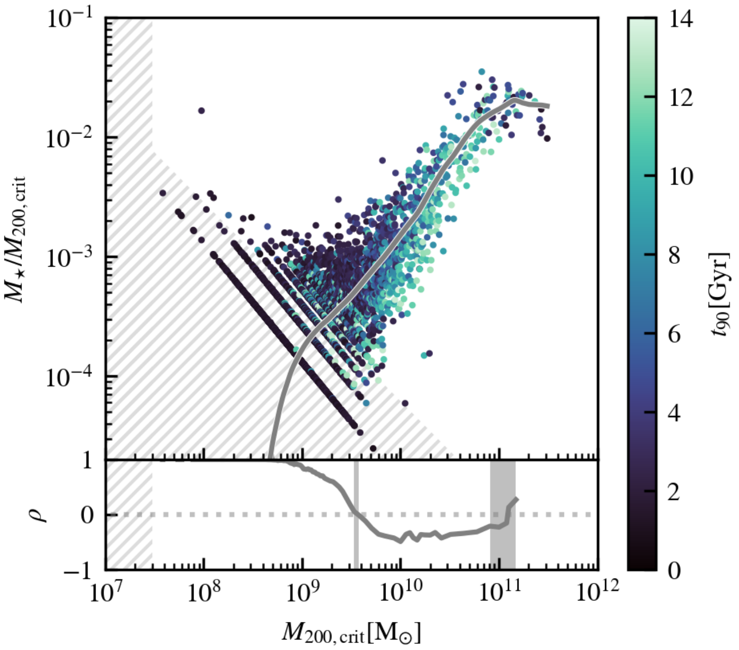


FIGURE 5.4: *Top*: The stellar mass halo mass relation at z=0 coloured by $t_{90}$, defined as the time at which 90% of the total stellar mass of a galaxy has formed. At a given halo mass, galaxies that form earlier have higher stellar mass. The grey solid line indicates the LOWESS median. *Bottom*: The running Spearman rank between the residuals $\Delta t_{90}$ and $\Delta M_\star/M_{200,\rm crit}$; above a halo mass of $M_{200,\rm crit} \gtrsim 10^{10} M_\odot$ the SMHM scatter shows a negative correlation with $t_{90}$.

Though most obvious when examining $t_{90}$, the trend with halo concentration is also visible when examining $t_{50}$ and $t_{10}$, and further demonstrates that earlier forming haloes assemble their stellar mass earlier.

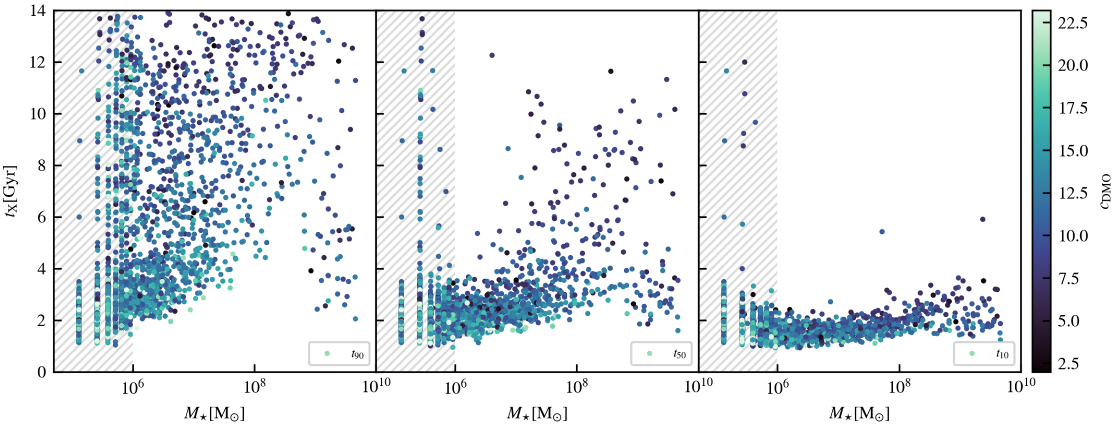


FIGURE 5.5: The time it takes for 90% ($t_{90}$), 50% ($t_{50}$), and 10% ($t_{10}$) of the cumulative star formation to occur as a function of stellar mass shown in the left, centre, and right panels respectively. Each point is coloured by the galaxy's halo concentration. The times are computed from the cumulative SFH of the galaxy normalised by its z=0 stellar mass. The shaded regions indicate 10 star particles at m5 resolution.

Turning to the overall distribution of $t_{90}, t_{50}$, and $t_{10}$ as a function of stellar mass, the star formation occurs at much earlier times than expected for isolated, field dwarf galaxies. Most galaxies below $M_\star < 10^8 \mathrm{M}_\odot$ have formed half ($t_{50}$) of their stellar mass by $t$=4Gyr. Digby et al. (2019) examine dwarf galaxies in the APOSTLE and AURIGA simulations which are zoom-ins of the Local Group and MW analogue galaxies, and adopt the definition of isolated field galaxies as a distance to their nearest primary of 300kpc $< r_{\mathrm{primary}} <$ 2Mpc. They show $t_{90}$ as a function of $M_\star$ for isolated field dwarfs and generally find two distinct groups of late and early $t_{90}$ with little population of galaxies between ($5 < t_{90} <$ 10Gyr). They find that all galaxies of stellar mass $M_\star > 10^7 \mathrm{M}_\odot$ form 90% of their stellar mass later than $t_{90} \sim$ 7.5Gyr, in this regime in the Columba simulations I see a significantly larger spread in $t_{90}$ toward earlier SF with no late time SF. The presence of galaxies with such early $t_{90}$ in the Columba simulations suggests that many of these galaxies have little to no late time star formation and are likely quenched.

## 5.5 Quenched Fraction

Figure 5.5 demonstrates that many of these isolated field dwarf galaxies assemble their stellar components very early and their galaxy growth is slower in later periods which raises questions about the number of systems that are actively star forming at present-day. In this Section I investigate the fraction of galaxies that are considered quenched at z=0. I use the following definitions for 'quenched': (i) defined by a specific star formation rate (sSFR) below sSFR $< 10^{-11}\mathrm{yr}^{-1}$, and (ii) the absence of HII. Figure 5.6 shows the quenched fraction, defined as the fraction of quenched galaxies to total galaxies in a stellar mass bin, as a function of stellar mass, adopting the former (i) and latter (ii) definitions of quenched in the top and bottom panels respectively. I also show the fraction computed from both m5 and m6 resolutions shown by the solid and low opacity lines. The different definitions of quenched show a general agreement. I find that the quenched fraction decreases towards higher stellar masses, reaches a minimum of $f_{\mathrm{quenched}} \sim 20\%$ around $M_\star \sim 5 \times 10^8 \mathrm{M}_\odot$, and increases again to $f_{\mathrm{quenched}} \sim 80\%$. This is consistent with the distribution of $t_{90}$ in Figure 5.5.

From SDSS DR8 data Geha et al. (2012) observe that isolated field galaxies with no active star formation in a stellar mass range between $10^7 < M_\star < 10^9 \mathrm{M}_\odot$ are extremely rare. They define a quenched galaxy by the absence of H$\alpha$ emission and show that the fraction of quenched galaxies decreases as the distance to their nearest luminous galaxy increases, plateauing at a distance of $d_{\mathrm{host}} > 1.5$Mpc. Examining the quenched fraction as a function of stellar mass, they conclude that there is a complete absence of quenched isolated field galaxies below a stellar mass of $M_\star < 10^9 M_\odot$ and that such galaxies are unable to quench SF internally. The quenched fractions in Figure 5.6 are inconsistent with these observations. Between $10^7 < M_\star < 10^9\mathrm{M}_\odot$ the Columba simulations produce a large population of field galaxies that are quenched at z=0 where observations find none: approximately half of the galaxies at $M_\star \sim 10^8\mathrm{M}_\odot$ have no current SF or HII present. This fraction increases towards lower stellar mass, and below $M_\star \sim 10^7\mathrm{M}_\odot$ nearly all galaxies are quenched. This may signal the influence of numerical effects in low mass galaxies, such as poor sampling of SF histories and/or inaccuracies in the galaxy formation model. These results are also in tension with findings from other simulations, which show better agreement with local observations. For example, using AURIGA zoom-in simulations of MW analogues, Simpson et al. (2018) investigate the quenched fraction of satellites as a function of the distance to their host halo. They investigate two definitions for quenched galaxies: (i) galaxies with a HI mass $< 10^5\mathrm{M}_\odot$ and (ii) galaxies whose youngest star particle's age is >100Myr and whose final SFR is zero. They consistently find little to no quenched satellites of $M_\star > 10^7\mathrm{M}_\odot$ at a distance between 0.6-1.0Mpc from the MW host.

Crain et al. (2015) show that low mass galaxies ($M_\star \sim 10^9\mathrm{M}_\odot$) in simulations with an uncalibrated model that injects a constant amount of energy per unit stellar mass ($f_E = 1$) for CCSNe feedback have sSFRs at z=0.1 that are both lower than observed and lower than the same results from the calibrated EAGLE model. They found that the SFHs were weighted towards earlier times due to excess numerical losses and result in galaxies that are too quiescent at the present epoch. The Columba simulations use the same uncalibrated subgrid implementation for CCSNe and I speculate that this may influence the high quenched fraction. I also caution that the convergence of the quenched fraction between resolutions is poor for both definitions of 'quenched'. The low resolution m6 simulations produce fewer quenched galaxies between $10^9 < M_\star < 10^{10}\mathrm{M}_\odot$ where there are differences of up to $\sim$ 80%, though this regime is affected by lower number statistics. It is possible that this effect results from the fixed BH seed mass at m5 and m6 resolution discussed in Section 4.3, the potential effect of this on the GSMF function that demonstrates that galaxies are more massive in the m6 resolution where the AGN feedback is less effective.

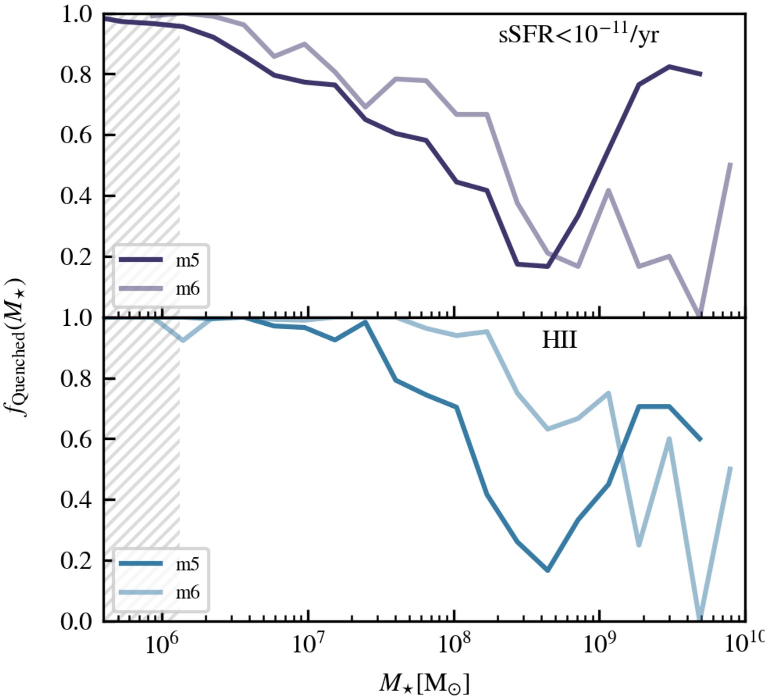


FIGURE 5.6: *Top:* The quenched fraction of field dwarf galaxies as a function of stellar mass. Quenched is defined using a specific star formation rate threshold of sSFR$<10^{-11}$yr$^{-1}$. m5 resolution simulations are denoted by the solid line and m6 by the lower opacity line. The shaded region indicates 10 stellar particles at m5 resolution. *Bottom:* The same as the top panel except quenched galaxies are defined by the absence of HII.

I investigate other potential sources of the high quenched fraction by examining the gas fraction of galaxies at present-day. In Figure 5.7 I show the sSFR of galaxies as a function of $M_{200}$ coloured by their circumgalactic mass fraction $f_{\rm CGM}$ normalised to the cosmic average baryon fraction, $\Omega_{\rm b}/\Omega_0$. The CGM gas mass of a galaxy is computed by the sum of gas that is not star forming within $r_{200}$, $f_{\rm CGM}$ is then given the ratio of CGM gas mass to halo mass ($M_{\rm CGM}/M_{200}$). Davies et al. (2020) show that galaxies in the EAGLE and Illustris-TNG simulations with higher CGM mass fractions are more likely to be star forming (sSFR$>10^{-11}$yr$^{-1}$) whilst galaxies with lower CGM mass fractions have a higher probability of being quenched. A similar trend is seen in the Columba simulations from Figure 5.7, at a fixed halo mass, galaxies with higher CGM fractions have higher sSFR and are less likely to be quenched.

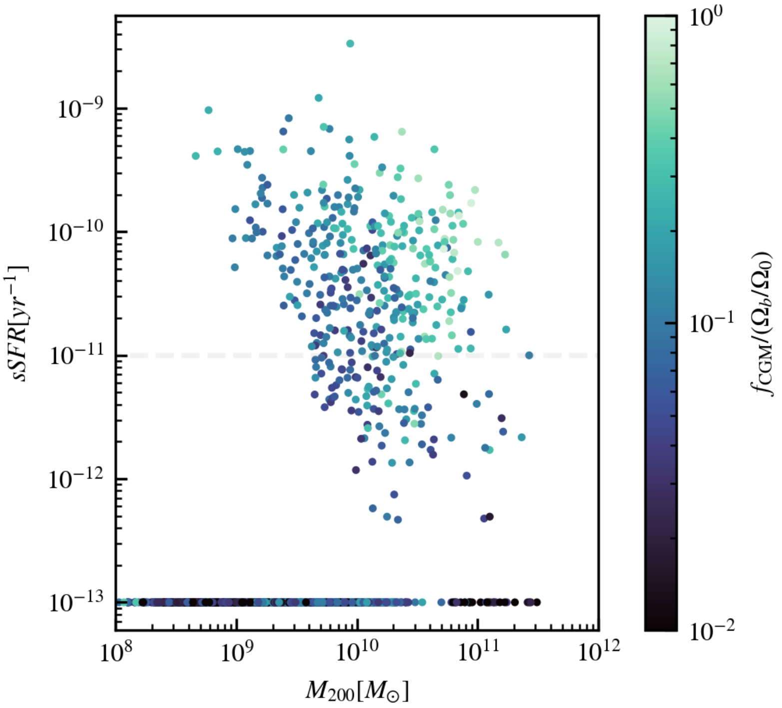


FIGURE 5.7: The specific star formation rate (sSFR) of central dwarf galaxies as a function of halo mass, each galaxy is coloured by their circumgalactic (CGM) gas fraction normalised to the cosmic average baryon fraction, $\Omega_{\mathrm{b}}/\Omega_0$. $f_{\mathrm{CGM}}$ is computed as the ratio of non-SF gas in $r_{200}$ to halo mass ($M_{\mathrm{CGM}}/M_{200}$). Galaxies with no SF are all set to sSFR$= 10^{-13}\mathrm{yr}^{-1}$.

## 5.6 Morphology

In this Section I investigate the influence of halo concentration, and hence formation time, on the morphology of central galaxies at z=0 in the simulations. I adopt a kinematic morphology measure following the method of Correa et al. (2017) who calculate the corotation parameter, $\kappa_{\mathrm{co}}$, as the fraction of corotational to total kinetic energy which gives a measure of how much of the stellar kinetic energy is invested in ordered corotation, therefore, galaxies with higher $\kappa_{\mathrm{co}}$ have more ordered corotation and are morphologically more disky and extended. $\kappa_{\mathrm{co}}$ has also been shown to correlate with the colour of both satellite and central galaxies in EAGLE above a mass $M_\star > 10^{10}\mathrm{M}_\odot$, with higher $\kappa_{\mathrm{co}}$ being bluer in color (Correa et al., 2017). Correa et al. (2017) use a threshold of $\kappa_{\mathrm{co}} = 0.4$ to delineate between disk ($\kappa_{\mathrm{co}} > 0.4$) and elliptical ($\kappa_{\mathrm{co}} < 0.4$) galaxies above $M_\star > 10^{10}\mathrm{M}_\odot$. I show example images of galaxies with stellar mass $10^8 < M_\star < 10^9\mathrm{M}_\odot$ from the Columba suite that span a range of $\kappa_{\mathrm{co}}$ in Figure 5.8, the top and bottom rows show the galaxies face and edge on respectively. From left to right, the value of $\kappa_{\mathrm{co}}$ decreases as $\kappa_{\mathrm{co}}$ =0.538, 0.364, and 0.147 and the visual morphology changes from more disk-like to elliptical systems.

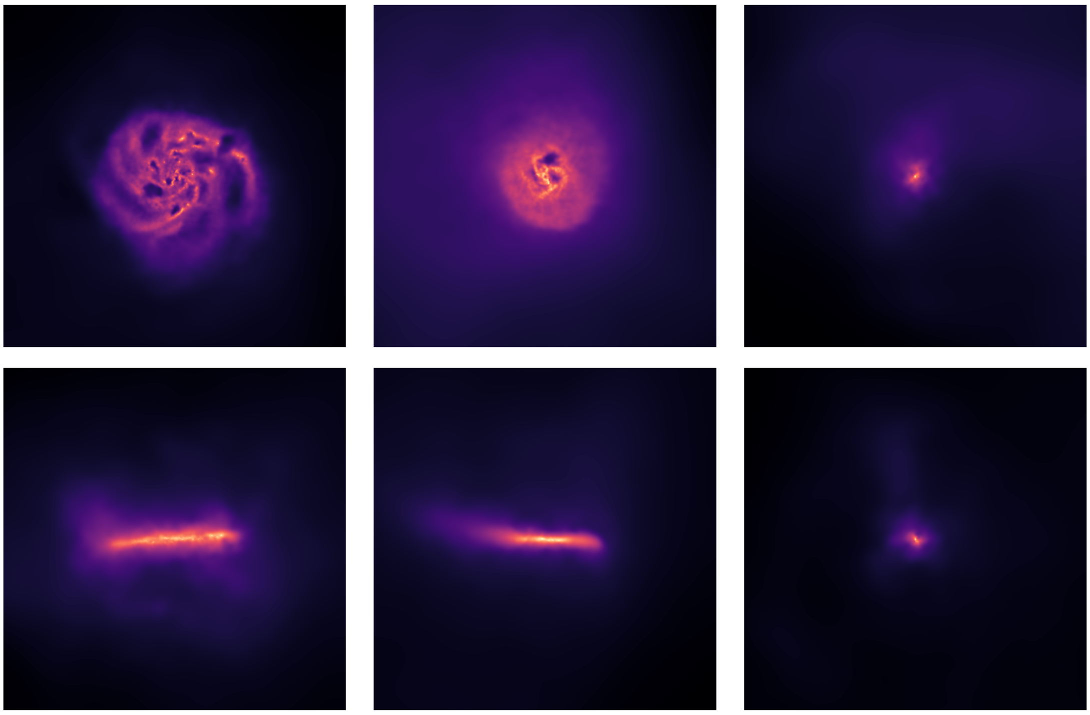


Figure 5.8: The surface density of gas ($r$=50kpc) in galaxies of stellar mass $10^8 < M_\star < 10^9 \mathrm{M}_\odot$ of different $\kappa_{\mathrm{co}}$ parameters. The top and bottom rows show the galaxies face and edge on respectively. The corotation parameter decreases from left to right as $\kappa_{\mathrm{co,star}}$ =0.538, 0.364, and 0.147, visible inspection shows that the galaxies go from more disk to elliptical morphology as the fraction of ordered corotation decreases.

In Figure 5.9 I show the distribution of $\kappa_{\mathrm{co}}$ corotation parameters in bins of stellar mass and halo concentration. The overall distribution in each stellar mass bin is shown in grey and I find the population of disk systems increases with stellar mass. The lowest mass bin is dominated by low $\kappa_{\mathrm{co}}$ systems which peak around $\kappa_{\mathrm{co}} \sim 0.2$ and there are hints of a bimodal distribution with a smaller peak around $\kappa_{\mathrm{co}} \sim 0.5$. This result is clearer for galaxies in the central stellar mass bin which exhibits two distinct distributions of galaxies peaking at $\kappa_{\mathrm{co}} \sim 0.2$ and $\kappa_{\mathrm{co}} \sim 0.5$. This signal is lost for the galaxies in the highest mass bin which do not display a clear bimodal distribution; the sample is spread over $0.1 < \kappa_{\mathrm{co}} < 0.8$ with a peak around $\kappa_{\mathrm{co}} \sim 0.5$, though the smaller number statistics in this bin introduces noise into this distribution.

I also subdivide each stellar mass range by halo concentration into high, intermediate, and low concentration displayed in the cyan, blue, and purple distributions with the median of each distribution indicated by the dashed vertical lines. The concentration bins are calculated from the terciles of the halo concentration distribution in each stellar mass bin following Section 5.4. Where there is a clear bimodal distribution, the high $\kappa_{\rm co}$ population is dominated by low concentration haloes. In the highest mass bin, concentration does not show a correlation with kinematic morphology of the galaxy, the medians of the low and intermediate-high concentration distributions are distinct but all three distributions occupy the full range of $\kappa_{\rm co}$.

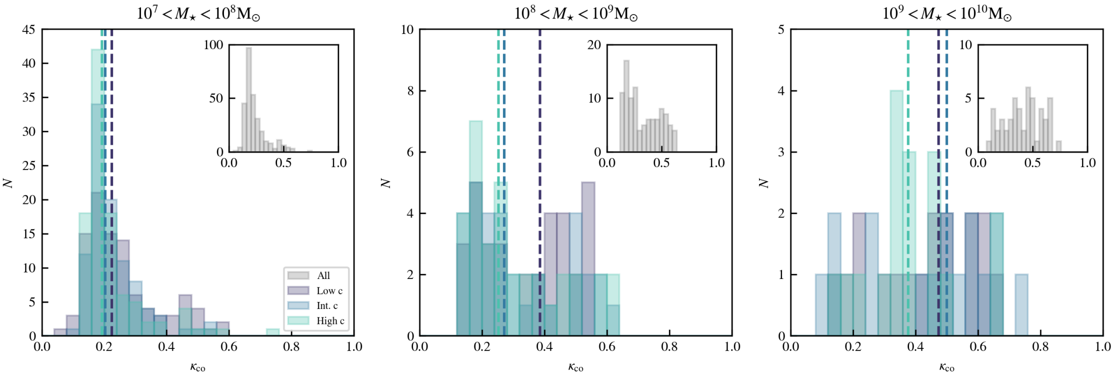


FIGURE 5.9: The $\kappa_{\rm co}$ distribution of field dwarf galaxies in the Columba simulations binned by stellar mass and halo concentration. From left to right the stellar mass ranges follow $10^7 < M_\star < 10^8{\rm M}_\odot$, $10^8 < M_\star < 10^9{\rm M}_\odot$, and $10^9 < M_\star < 10^{10}{\rm M}_\odot$. The inset axes show the distribution of all galaxies in these mass bins in grey. In each stellar mass bin, the sample is divided into high, intermediate, and low concentration bins based on the terciles of the concentration distribution, shown in cyan, blue, and purple respectively. The medians of these distributions are indicated by the vertical dashed lines.

I investigate the potential correlation more quantitatively in Figure 5.10 which shows $\kappa_{\rm co}$ against stellar mass coloured by halo concentration. The bottom panel shows the running Spearman rank coefficient between the residuals $\Delta c_{\rm DMO}$ and $\Delta\kappa_{\rm co}$ which shows no correlation across the full stellar mass range. Furthermore, though low concentration haloes are more likely to host more disk-like galaxies, the simulations do not show a simple relation between the halo formation time and their kinematic morphology at present-day.

## 5.7 Conclusion

Previous results have emphasised the important role of halo assembly, manifested through halo concentration, on the final stellar masses of dwarf galaxies. Motivated by these, I have examined the effect of halo concentration on the present-day properties and stellar assembly of isolated field dwarf galaxies in the Columba suite of simulations. I

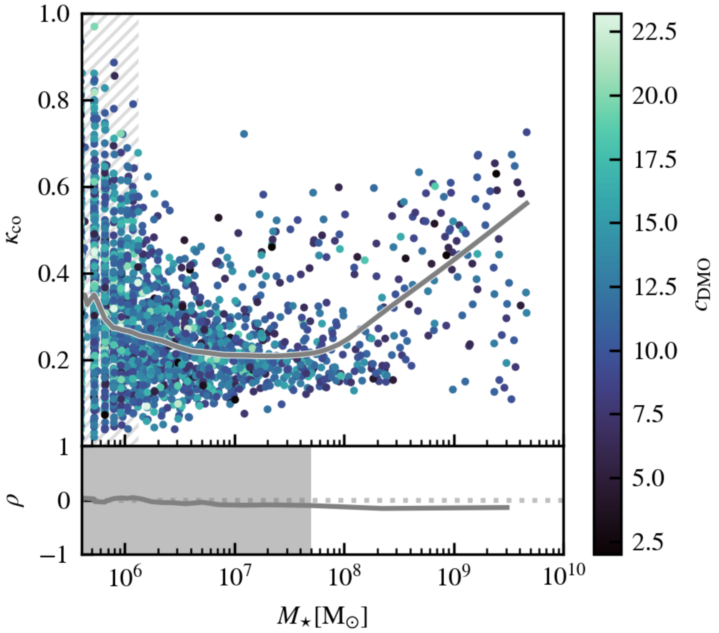


FIGURE 5.10: *Top:* The corotation parameter, $\kappa_{\mathrm{co}}$, of field galaxies in the Columba simulations as a function of stellar mass coloured by their halo concentration. The grey line shows the running LOWESS median. The shaded regions indicate 10 star particles at m5 resolution. *Bottom:* The running Spearmann rank coefficient between the residuals, there is no correlation between halo concentration and $\kappa_{\mathrm{co}}$ at a given stellar mass.

have investigated the scatter of the SMHM relation and demonstrated that it exhibits a strong, significant correlation with the DM halo concentrations of galaxies. This positive correlation peaks at a maximum Spearman rank correlation coefficient of $\rho = 0.85$ at $M_{200} = 1.2 \times 10^{11}\mathrm{M}_\odot$. A direct implication of this correlation is that the scatter in the $M_\star - V_{\mathrm{max}}$ relation is smaller at a fixed halo mass compared to the SMHM relation and exhibits a low correlation between halo concentration and the scatter. These findings are in overall agreement with predictions from theoretical models (Benitez-Llambay & Frenk, 2020), and other high resolution hydrodynamical simulations of dwarf galaxies (Fitts et al., 2017; Rey et al., 2019). In the latter cases, the inference was based on a limited number of simulations at fixed halo mass whereas here I show that similar conclusions can be drawn using a statically large sample of dwarf galaxies, and that use consistent galaxy formation models compared to inference based on compilations of different simulations (Christensen et al., 2024). Moreover, these results follow the same trends found at higher masses ($M_{\mathrm{halo}} > 10^{11}\mathrm{M}_\odot$), where large scale galaxy surveys have shown that stellar mass of galaxies is more tightly correlated with $V_{\mathrm{max}}$ than halo mass (see Wechsler & Tinker, 2018, for a review) and studies using large hydrodynamical simulations and semi-analytical models also point towards the strong correlation of the scatter in the SMHM relation with halo formation time or halo concentrations (Matthee et al., 2016; Zehavi et al., 2018; Artale et al., 2018).

I show that the stellar assembly of simulated dwarf galaxies in the Columba sample correlates strongly with halo concentration at both fixed halo mass and fixed stellar mass. High concentration, and hence earlier forming, haloes assemble their stellar mass earlier than low concentration haloes and this can be quantified by the time it takes for 90% of the z=0 stellar mass to assemble, $t_{90}$. The link between stellar assembly and halo concentration or halo assembly leaves its imprint in various relations, for example, this manifests as an anti-correlation between $t_{90}$ with the scatter of the SMHM relation. For galaxies of similar stellar mass, those who live in higher mass haloes have later stellar assembly times (higher $t_{90}$). At the same time, at fixed halo masses, higher stellar mass galaxies form the majority of their stars earlier (smaller $t_{90}$). This anti-correlation peaks at $\rho = -0.49$ for $M_{200} = 1.0 \times 10^{10}\mathrm{M}_{\odot}$.

Additionally, I show a general positive trend between $t_{90}$ and $M_{\star}$ in agreement with other works (Digby et al., 2019). The scatter in the relation is, however, linked to the halo concentrations; such that galaxies that form their stars later live in lower concentration and later forming haloes. This is in qualitative agreement with the trend found in higher stellar masses in the EAGLE simulations where Matthee & Schaye (2019) show that the scatter in $sSFR - M_{\star}$ relation of the EAGLE galaxies is linked to the formation time of their haloes.

The correlations between $M_{\star}, M_{200}, c_{\mathrm{NFW}}$, and $t_{90}$ give insight into how these properties are all interlinked. The significance of these correlations lies in their ability to connect observable properties of dwarf galaxies, i.e. their stellar mass and star formation histories, to their dark matter halo properties which are less easily accessible. Moreover, these are clear predictions which can be tested with upcoming measurements of properties of large sample of isolated dwarf galaxies in the local universe, using Euclid, DESI, and future Rubin and Roman Observatories.

Whilst the Columba simulations give insight into the low mass galaxy-halo connection, there are tensions with observations which highlight limitations of the galaxy formation model. Field galaxies in the Columba simulations are overquenched compared to observed isolated dwarf galaxies. At $M_{\star} \sim 10^{9}\mathrm{M}_{\odot}$, there are no quenched dwarf galaxies in the Local Group outside of the MW and M31 haloes (Geha et al., 2012), whereas $\sim 40\%$ of galaxies in the Columba simulations are quenched at this mass scale. Additionally, below $M_\star \sim 4 \times 10^8 \mathrm{M}_\odot$, the quenched fraction increases from 20% to 80% at $M_\star \sim 10^7$. The excess quenched fraction is likely influenced by the use of a constant $f_\mathrm{E}$ implemented in the subgrid CCSNe routine which has been shown to skew star formation to earlier epochs as a result of numerical overcooling.

Finally, the Columba simulations produce a range of disk and elliptical dwarf galaxies in isolated environments. Quantified by the corotation parameter $\kappa_\mathrm{co}$, galaxies in the stellar mass range $10^7 < M_\star < 10^8 \mathrm{M}_\odot$ are predominantly elliptical whereas between $10^8 < M < 10^9 \mathrm{M}_\odot$ the $\kappa_\mathrm{co}$ distribution is bimodal with peaks at $\kappa_\mathrm{co} \sim 0.2$ and $\kappa_\mathrm{co} \sim 0.5$. Though the halo concentration shows a signal in the SFH of these galaxies, $\kappa_\mathrm{co}$ does not show a correlation with the halo concentration of the galaxy. The mechanisms driving the population of elliptical low-mass systems in underdense, and in some cases void-like, environments within these simulations remain unclear. This may indicate the presence of dispersion-dominated dwarf galaxies forming in situ, as recent observations of voids suggest (de los Reyes et al., 2023), challenging the tidal stirring hypothesis which requires nearby tidal forces from neighbouring objects to evolve from rotation to dispersion dominated systems.

# 6

# Conclusions

## 6.1 Summary

This thesis is centred around the outstanding questions surrounding the formation and evolution of dwarf galaxies. Isolated dwarf galaxies represent a ‘clean’ test for cosmology and galaxy formation as they are insulated from the effects of nearby massive neighbours which have inextricable effects on their evolution. Large representative cosmological simulations lack the resolution to address the small-scale questions and current state of the art zoom-in simulations often focus on Local Group or MW systems, or individual dwarf galaxies. Furthermore, the goal of this thesis was to complement existing simulations with a suite that contains a large sample of isolated dwarf galaxies whilst spanning the full range of environments they form and evolve in. To this aim, I designed and ran the Columba suite of zoom-in simulations. To investigate the influences driving the evolution of isolated dwarf galaxies I investigated the diversity of regions these objects form in using a large cosmological volume ($L = 400$Mpc). After imposing distance criteria on the distance to the nearest MW mass and group mass objects ($d_{>\mathrm{MW}} > 6$Mpc and $d_{>\mathrm{group}} > 10$Mpc) I found a wide spread in the overdensity and enclosed dwarf galaxy populations. I systematically selected regions from the $(-2\sigma, -1\sigma, +0\sigma, +1\sigma, +2\sigma)$ of the overdensity distribution and then explored the number of LMC-like objects enclosed in these regions, probing the high mass end of dwarf galaxies, and found a significant scatter at fixed overdensity. The final simulation suite is comprised of 25 regions that span the diversity of cosmic environments that isolated dwarf galaxy populations are able to form in. I generated the initial conditions with similar mass DM and gas particles at m5 and m6 resolution, corresponding to gas particle masses of $m_{\mathrm{gas}} \sim 10^5\mathrm{M}_\odot$

and $m_{\rm gas} \sim 10^6 {\rm M}_\odot$, respectively. These simulations are run with a galaxy formation model that represents an evolution of that used for EAGLE which explicitly follows the cold, dense, neutral gas of the ISM, adopts a star formation model that accounts for turbulence, a treatment of pre-supernova feedback from massive stars, more sophisticated methods of distributing the energy liberated by SNe and active galactic nuclei (AGN), and updated treatments of BH growth and dynamics. I examined key diagnostics from the model against observational data and found that, despite the lack of calibration, the simulations reproduce the observed GSMF from Driver et al. (2022), however were unsuccessful in matching the galaxy size - mass relation.

In Chapter 4 I examined how the environment influences key properties of galaxies in the Columba simulations. The Columba suite samples a range of overdensities corresponding to changes in the cosmic web, from low-density voids to filamentary structures. While the amplitudes of the HMFs and SMFs depend on the region density, their slopes, when normalized to the cosmic mean, do not show an environmental dependence. Similarly, the SMHM relation does not depend on the region's environment, maintaining consistent slopes despite density variations. However, low-mass halo concentrations exhibit a slight dependence on the environment, with higher concentration halos forming in denser regions. Although the mass-concentration relation itself shows only a minor environmental dependence, it has a much steeper slope for luminous halos. This indicates a strong correlation between halo concentration and the luminous fraction, highlighting that earlier-forming halos are more likely to host a galaxy.

The role of halo concentration on the properties of dwarf galaxies in the Columba suite was explored more thoroughly in Chapter 5. The scatter in the SMHM relation at a fixed halo mass exhibits a strong correlation with halo concentration, extending the results of Matthee et al. (2016) to lower mass. Haloes that form earlier are also likely to have a faster stellar assembly, reflected by steeper SFHs and earlier $t_{90}$ (the time at which 90% of their z=0 stellar mass has assembled). The early SF and hence low $t_{90}$ translates into a high quenched fraction of these field galaxies and is inconsistent with observed galaxies in the Local Group (Geha et al., 2012). I attribute this to the implementation of CCSNe feedback using a constant $f_E = 1$, instead of scaling with the density of natal gas particles, which Crain et al. (2015) have demonstrated can result in earlier SFHs and too quiescent galaxies at present-day.

## 6.2 Future Work

In Chapter 4 I examined key properties of the Columba simulation suite with reference to observed properties. I highlight that the subgrid parameters used in the galaxy formation model of the Columba simulation are not calibrated to match these observables, including the GSMF as is common approach. The Columba simulations successfully reproduce the GSMF at present-day, however, the simulations form galaxy sizes that are too compact in the mass range $M_\star > 5 \times 10^8 \mathrm{M}_\odot$ and too large at lower masses. This stresses the importance of calibration, and in particular, using observables like galaxy sizes in addition to the GSMF to produce realistic galaxies. For simulations like EAGLE (Schaye et al., 2015), calibration was done by hand by varying the subgrid parameters within reasonable bounds until the simulation showed good agreement with the calibration targets. More recently, this process has been made more efficient by the use of machine learning, specifically in the form of emulators using Gaussian processes. Emulator-based methods have been used in conjunction with semi-analytic models of galaxy formation (e.g. Bower et al., 2010), and Kugel et al. (2023) have utilised this method to calibrate the FLAMINGO simulations. In summary, the emulator is trained on a sample of input simulations that strategically cover the range of input parameters of interest, the emulator is then able to predict the outcome, or observable, for a new set of input parameters. Furthermore, a focus of future work would be calibrating the subgrid models used in the Columba simulations using emulators, which would enable the simulations to produce more realistic galaxy populations.

In addition to calibration, increasing the resolution of the Columba simulations is desirable. Following the methodology of Chapter 3, I have generated DM only and hydrodynamical initial conditions of the full Columba suite at an ‘m4’ resolution, corresponding to a gas particle mass of $m_{\mathrm{gas}} \sim 10^4 \mathrm{M}_\odot$ and running these is an immediate focus of future work. Running the simulations at increased resolution would enable further assessment of the numerical convergence of the simulation outputs, and would allow me to probe increasingly smaller scales and lower mass galaxies. Higher resolution would improve the sampling of objects in the lowest density regions of the Columba simulations by resolving lower masses. This increased sampling would allow me to examine the SMHM relation and luminous fraction of lower mass galaxies and verify there are no large scale environmental trends towards lower halo masses.

The Columba simulations adopt the standard ΛCDM cosmological framework, however, there remains small scale tensions between theory and observations (see e.g. Sales et al., 2022) and no particle candidate for CDM has ever been directly detected. A possible avenue for extensions to the Columba simulations would be to adopt different flavours of dark matter, for example a self interacting dark matter. Correa et al. (2022) adopt a ΛSIDM cosmology for the TangoSIDM project and demonstrate how SIDM models may offer a promising explanation for the diversity in the density and velocity profiles of observed dwarf galaxies. Additionally, exploring warm dark matter (WDM) could address some of the small-scale structure problems inherent in ΛCDM. WDM particles suppress the formation of low-mass halos, potentially providing a better match to the observed abundance of satellite galaxies and the cores of dwarf galaxies; for example, Meshveliani et al. (2024) use the EAGLE simulations with a WDM model and suggest a sterile neutrino DM candidate has the potential to explain the suggested deficit of faint Local Group dwarfs.

The isolated dwarf galaxies in the Columba suite show a diversity in morphology, SFR, and stellar mass at a fixed halo mass. Halo concentration, and hence formation time, has been shown to drive some of the variations for example in the SMHM relation, however, does not strongly affect their morphology. An interesting avenue for future work is to examine different definitions of environment on these properties. This thesis has focused on large scale ($r$=5Mpc) overdensity, however, Wu et al. (2024) have shown that smaller scales $L \sim 3$Mpc may be more informative. Varying the aperture of the density measurement may emphasise potential correlations that are too weak to detect on the large scale investigated in this thesis. Beyond overdensity, local environment can be quantified by many other measures such as the number of nearest neighbours (Muldrew et al., 2012) and the cosmic web environments of a given galaxy can be more authoritatively categorised. For a comprehensive understanding of how environment links to the diversity in isolated field galaxies it is not sufficient to investigate only one measure and future work would explore different definitions of ‘environment’.

# A

# Appendix

## A.1 Simulation Outputs

Figure A.1 shows the spacing between snapshot outputs as a function of expansion factor ($a$), red-shift ($z$), and cosmic time ($t$) from left to right. I investigated using simulation outputs at fixed time intervals (the dotted line) and fixed intervals of $\ln a$ (dashed line). Fixed time intervals are preferable at late cosmic time to capture the rapidly evolving physics and fixed $\ln a$ intervals follows structure collapse at early time, therefore, I implemented a spline curve, in purple, with 5 interpolation points (including a point at z=7.5 so the time of reionisation is captured) that smoothly transitions from the curve of fixed ln($a$) to the curve of fixed time intervals. The final curve chosen to assign simulation outputs is shown in orange which follows the spline and transitions to regular intervals in cosmic time ($\delta t \sim$ 150Myr) with a total number of snapshots $n_{\rm snap}$ =128 and $n_{\rm snap}$ =64 at m6 and m5 resolution respectively.

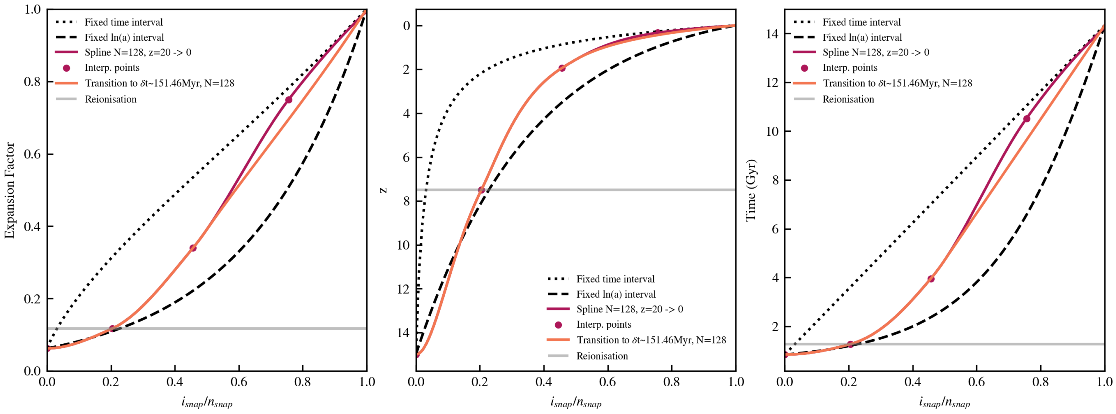


FIGURE A.1: The spacing between simulation outputs as a function of expansion factor ($a$), redshift ($z$) and cosmic time ($t$) from left to right. The dashed and dotted lines show fixed $\ln a$ and time intervals respectively. The purple line shows a spline (and interpolation points) between the dotted and dashed lines. The orange shows the final choice of simulation outputs with follows the spline and transitions to fixed time intervals of $\delta t \sim 150$Mpc.

## A.2 Runtime Analyses

I show the final wallclock times of the DMO (left) and hydrodynamical (right) simulations as a function of scale factor for m6 and m5 resolution levels in Figure A.2. Each simulation was run using SWIFT on 1 node (32 cores) on the PROSPERO supercomputer[1]. In Figure A.3, I show a breakdown of the time spent on various SWIFT operations (left) and SWIFT task categories (right) in the m5 hydrodynamical run of volume$(+0\sigma)_{13}$. Operations were dominated by gravity and 'dead' or idle time. Currently SWIFT is not optimised for zoom simulations which involve simulating large volumes but require the computation to focus on a small volume of interest. For the Columba simulations this is a spherical volume of $r = 5$Mpc embedded in a $L = 400$Mpc periodic volume. These inefficiencies can be reduced by introducing a second high-resolution grid in the zoom region and is the focus of SWIFT-zoom. Preliminary tests have shown improvements in the computational performance of the Columba simulations which will make resimulations of the Columba suite and other zoom-in simulations more computationally efficient with SWIFT.

---

[1] See full specifications for PROSPERO at: https://prospero-docs.readthedocs.io/en/latest/specifications.html

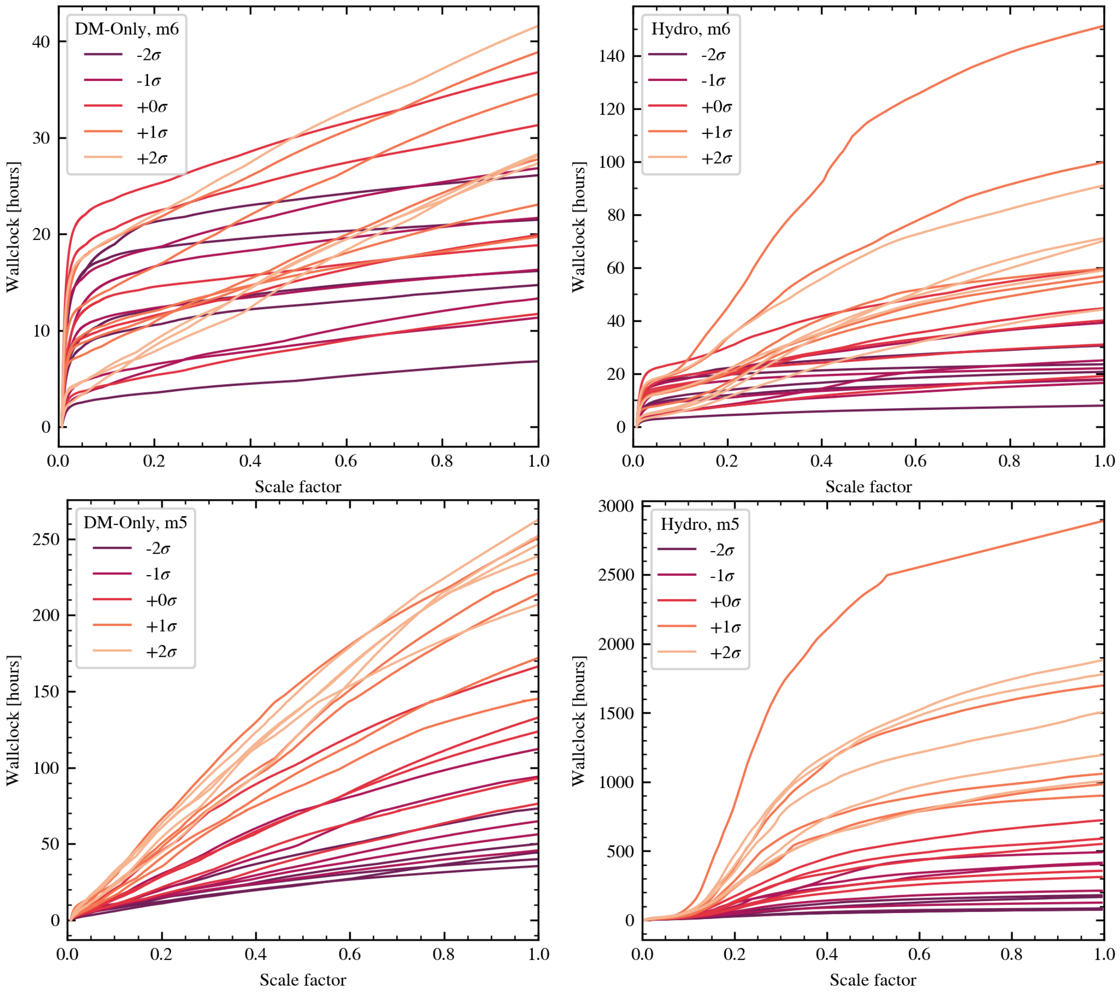


FIGURE A.2: *Top:* The final wallclock times of the Columba simulations ran with DMO (left) and full hydrodynamics (right) at m5 resolution. *Bottom:* The same as above for m6 resolution.

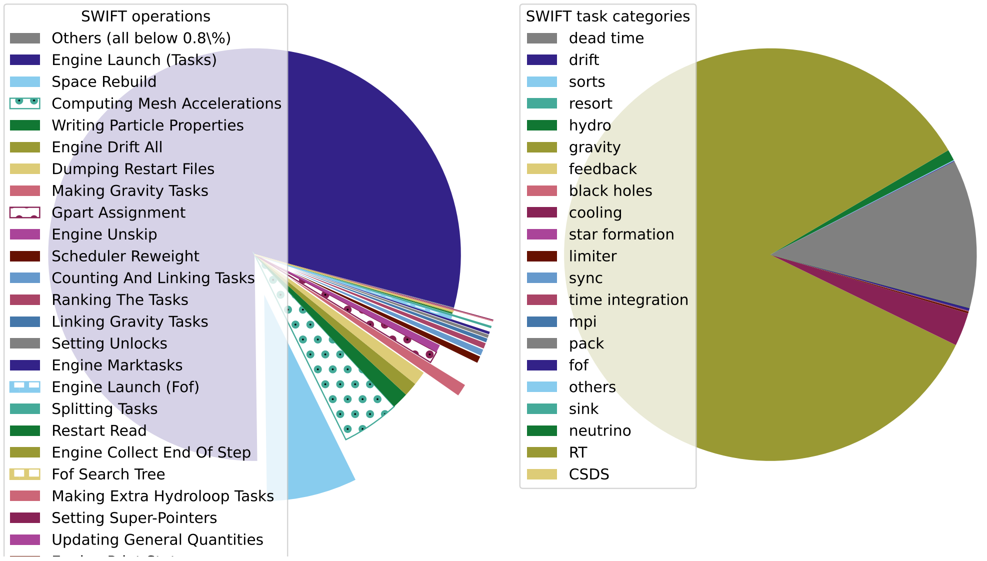


FIGURE A.3: A decomposition of time spent on SWIFT operations (*left*) and tasks *right* from the m5 hydrodynamical simulation of Columba region $(+0\sigma\delta)_{13}$.

*“For my part, I know nothing with any certainty, but the sight of the stars makes me dream.”*

Vincent Van Gogh, 1888

# Bibliography

Abbott T. M. C., et al., 2022, , 105, 023520

Abel T., Bryan G. L., Norman M. L., 2002, Science, 295, 93

Applebaum E., Brooks A. M., Christensen C. R., Munshi F., Quinn T. R., Shen S., Tremmel M., 2021, The Astrophysical Journal, 906, 96

Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, Monthly Notices of the Royal Astronomical Society, 480, 3978

Bahé Y. M., et al., 2016, Monthly Notices of the Royal Astronomical Society, 456, 1115

Bahé Y. M., et al., 2022, Monthly Notices of the Royal Astronomical Society, 516, 167

Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, The Astrophysical Journal, 600, 681

Barnes J., Hut P., 1986, Nature, 324, 446

Bastian N., Covey K. R., Meyer M. R., 2010, Annual Reviews, 48, 339

Behroozi P. S., Wechsler R. H., Conroy C., 2013, The Astrophysical Journal, 770, 57

Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, Monthly Notices of the Royal Astronomical Society, 488, 3143

Bellovary J. M., Cleary C. E., Munshi F., Tremmel M., Christensen C. R., Brooks A., Quinn T. R., 2019, Monthly Notices of the Royal Astronomical Society, 482, 2913

Benitez-Llambay A., Frenk C., 2020, Monthly Notices of the Royal Astronomical Society, 498, 4887

Bertschinger E., 1998, Annual Reviews, 36, 599

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, Monthly Notices of the Royal Astronomical Society, 456, 2361

Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603

Booth C. M., Schaye J., 2009, Monthly Notices of the Royal Astronomical Society, 398, 53

Borrow J., Schaller M., Bower R. G., Schaye J., 2021, Monthly Notices of the Royal Astronomical Society, 511, 2367

Borrow J., Schaller M., Bahé Y. M., Schaye J., Ludlow A. D., Ploeckinger S., Nobels F. S. J., Altamura E., 2023, Monthly Notices of the Royal Astronomical Society, 526, 2441

Bovill M. S., Ricotti M., 2009, The Astrophysical Journal, 693, 1859

Bovill M. S., Ricotti M., 2011, The Astrophysical Journal, 741, 17

Bower R. G., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, Monthly Notices of the Royal Astronomical Society, 407, 2017

Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, Monthly Notices of the Royal Astronomical Society, 465, 32

Boylan-Kolchin M., 2021, Nature Astronomy, 5, 1188

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, Monthly Notices of the Royal Astronomical Society: Letters, 415, L40–L44

Bryan G. L., et al., 2014, The Astrophysical Journal Supplement Series, 211, 19

Buck T., Obreja A., Macciò A. V., Minchev I., Dutton A. A., Ostriker J. P., 2020, Monthly Notices of the Royal Astronomical Society, 491, 3461

Bullock J. S., Boylan-Kolchin M., 2017a, Annual Reviews, 55, 343

Bullock J. S., Boylan-Kolchin M., 2017b, Annual Reviews, 55, 343

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, Monthly Notices of the Royal Astronomical Society, 321, 559

Burkert A., 1995, Astrophyics Journal, 447, L25

Carlsten S. G., Greene J. E., Greco J. P., Beaton R. L., Kado-Fong E., 2021, The Astrophysical Journal, 922, 267

Cassinelli J. P., 1979, Annual Reviews, 17, 275

Chabrier G., 2003, PASP, 115, 763

Chaikin E., Schaye J., Schaller M., Bahé Y. M., Nobels F. S. J., Ploeckinger S., 2022, Monthly Notices of the Royal Astronomical Society, 514, 249

Chaikin E., Schaye J., Schaller M., Bení tez-Llambay A., Nobels F. S. J., Ploeckinger S., 2023, Monthly Notices of the Royal Astronomical Society, 523, 3709

Christensen C. R., Brooks A. M., Munshi F., Riggs C., Van Nest J., Akins H., Quinn T. R., Chamberland L., 2024, The Astrophysical Journal, 961, 236

Cleveland W. S., 1979, Journal of the American Statistical Association, 74, 829

Clowe D., Gonzalez A., Markevitch M., 2004, The Astrophysical Journal, 604, 596

Correa C. A., Schaye J., Clauwens B., Bower R. G., Crain R. A., Schaller M., Theuns T., Thob A. C. R., 2017, Monthly Notices of the Royal Astronomical Society, 472, L45

Correa C. A., Schaller M., Ploeckinger S., Anau Montel N., Weniger C., Ando S., 2022, Monthly Notices of the Royal Astronomical Society, 517, 3045

Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, , 112, 839

Crain R. A., van de Voort F., 2023, Annual Reviews, 61, 473

Crain R. A., et al., 2009, Monthly Notices of the Royal Astronomical Society, 399, 1773

Crain R. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1937

Dalgarno A., McCray R. A., 1972, Annual Reviews, 10, 375

Dalla Vecchia C., Schaye J., 2008, Monthly Notices of the Royal Astronomical Society, 387, 1431

Dalla Vecchia C., Schaye J., 2012, Monthly Notices of the Royal Astronomical Society, 426, 140

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, Monthly Notices of the Royal Astronomical Society, 486, 2827

Davies J. J., Crain R. A., Oppenheimer B. D., Schaye J., 2020, Monthly Notices of the Royal Astronomical Society, 491, 4462

Davies J. J., Crain R. A., Pontzen A., 2021, Monthly Notices of the Royal Astronomical Society, 501, 236

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, The Astrophysical Journal, 292, 371

Dehnen W., 2000, Astrophyics Journal, 536, L39

Diemer B., 2018, , 239, 35

Digby R., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 5423

Driver S. P., et al., 2022, Monthly Notices of the Royal Astronomical Society, 513, 439

Dubois Y., Teyssier R., 2008, Astronomy & Astrophysics, 477, 79

Durier F., Dalla Vecchia C., 2012, Monthly Notices of the Royal Astronomical Society, 419, 465

Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87

Elahi P. J., Cañas R., Poulton R. J. J., Tobar R. J., Willis J. S., Lagos C. d. P., Power C., Robotham A. S. G., 2019, Publications of the Astronomical Society of Australia, 36

Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, PASA, 34, e058

Fabian A. C., 2012, Annual Reviews, 50, 455

Fattahi A., Navarro J. F., Frenk C. S., Oman K. A., Sawala T., Schaller M., 2018, Monthly Notices of the Royal Astronomical Society, 476, 3816–3836

Fattahi A., Navarro J. F., Frenk C. S., 2020, Monthly Notices of the Royal Astronomical Society, 493, 2596

Feldmann R., et al., 2023, Monthly Notices of the Royal Astronomical Society, 522, 3831

Ferland G. J., et al., 2017, , 53, 385

Fitts A., et al., 2017, Monthly Notices of the Royal Astronomical Society, 471, 3547

Flores R. A., Primack J. R., 1994, Astrophyics Journal, 427, L1

Friedmann A., 1922, Zeitschrift fur Physik, 10, 377

Furlong M., et al., 2017, Monthly Notices of the Royal Astronomical Society, 465, 722

Gallart C., et al., 2015, Astrophyics Journal, 811, L18

Gao L., Springel V., White S. D. M., 2005, Monthly Notices of the Royal Astronomical Society, 363, L66

Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014, Monthly Notices of the Royal Astronomical Society, 438, 2578

Garrison-Kimmel S., et al., 2019, Monthly Notices of the Royal Astronomical Society, 487, 1380

Geha M., Blanton M. R., Yan R., Tinker J. L., 2012, The Astrophysical Journal, 757, 85

Genel S., et al., 2018, Monthly Notices of the Royal Astronomical Society, 474, 3976

Genel S., et al., 2019, The Astrophysical Journal, 871, 21

Gingold R. A., Monaghan J. J., 1977, Monthly Notices of the Royal Astronomical Society, 181, 375

Gnedin N. Y., 2000, The Astrophysical Journal, 542, 535

Grand R. J. J., et al., 2017, Monthly Notices of the Royal Astronomical Society, 467, 179

Grand R. J. J., et al., 2021, Monthly Notices of the Royal Astronomical Society, 507, 4953

Guedes J., Callegari S., Madau P., Mayer L., 2011, The Astrophysical Journal, 742, 76

Gunn J. E., Gott J. Richard I., 1972, The Astrophysical Journal, 176, 1

Gutcke T. A., Pakmor R., Naab T., Springel V., 2021, Monthly Notices of the Royal Astronomical Society, 501, 5597

Guth A. H., 1981, Phys. Rev. D, 23, 347

Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. p. 64 (`arXiv:astro-ph/0106018`), doi:10.48550/arXiv.astro-ph/0106018

Hahn O., Michaux M., Rampf C., Uhlemann C., Angulo R. E., 2020, MUSIC2-monofonIC: 3LPT initial condition generator, Astrophysics Source Code Library, record ascl:2008.024

Hardwick J. A., Cortese L., Obreschkow D., Catinella B., Cook R. H. W., 2022, Monthly Notices of the Royal Astronomical Society, 509, 3751

Hellwing W. A., Cautun M., van de Weygaert R., Jones B. T., 2021, , 103, 063517

Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, Monthly Notices of the Royal Astronomical Society, 442, 2304

Hopkins P. F., et al., 2018, Monthly Notices of the Royal Astronomical Society, 480, 800

Hoyle F., 1948, Monthly Notices of the Royal Astronomical Society, 108, 372

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

Jenkins A., 2010, Monthly Notices of the Royal Astronomical Society, 403, 1859

Jenkins A., 2013, Monthly Notices of the Royal Astronomical Society, 434, 2094

Karachentsev I. D., et al., 2001, Astronomy & Astrophysics, 379, 407

Karachentsev I. D., Makarova L. N., Makarov D. I., Tully R. B., Rizzi L., 2015, Monthly Notices of the Royal Astronomical Society, 447, L85

Katz N., White S. D. M., 1993, The Astrophysical Journal, 412, 455

Katz N., Weinberg D. H., Hernquist L., 1996, , 105, 19

Katz N., Keres D., Dave R., Weinberg D. H., 2003, in Rosenberg J. L., Putman M. E., eds, Astrophysics and Space Science Library Vol. 281, The IGM/Galaxy Connection. The Distribution of Baryons at z=0. p. 185 (`arXiv:astro-ph/0209279`), doi:10.1007/978-94-010-0115-1˙34

Katz H., et al., 2020, Monthly Notices of the Royal Astronomical Society, 494, 2200

Keller B. W., Wadsley J. W., Wang L., Kruijssen J. M. D., 2019, Monthly Notices of the Royal Astronomical Society, 482, 2244

Kennicutt Robert C. J., 1998, The Astrophysical Journal, 498, 541

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, Monthly Notices of the Royal Astronomical Society, 363, 2

King L., et al., 1998, Monthly Notices of the Royal Astronomical Society, 295, L41

Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, The Astrophysical Journal, 522, 82

Knebe A., et al., 2011, Monthly Notices of the Royal Astronomical Society, 415, 2293

Kormendy J., Ho L. C., 2013, Annual Reviews, 51, 511

Kroupa P., 2001, Monthly Notices of the Royal Astronomical Society, 322, 231

Krumholz M. R., Tan J. C., 2007, The Astrophysical Journal, 654, 304

Krumholz M. R., McKee C. F., Klein R. I., 2006, The Astrophysical Journal, 638, 369

Kugel R., et al., 2023, Monthly Notices of the Royal Astronomical Society, 526, 6103

Lemaître G., 1927, Annales de la Soci&eacute;t&eacute; Scientifique de Bruxelles, 47, 49

Lesgourgues J., 2011, arXiv e-prints, p. arXiv:1104.2932

Lesgourgues J., Tram T., 2011, Journal of Cosmology and Astroparticle Physics, 2011, 032–032

Liao S., Gao L., 2019, Monthly Notices of the Royal Astronomical Society, 485, 464

Lovell M. R., et al., 2018, Monthly Notices of the Royal Astronomical Society, 481, 1950

Lovell C. C., Vijayan A. P., Thomas P. A., Wilkins S. M., Barnes D. J., Irodotou D., Roper W., 2021, Monthly Notices of the Royal Astronomical Society, 500, 2127

Lucy L. B., 1977, , 82, 1013

Ludlow A. D., Navarro J. F., Angulo R. E., Boylan-Kolchin M., Springel V., Frenk C., White S. D. M., 2014, Monthly Notices of the Royal Astronomical Society, 441, 378

Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, Monthly Notices of the Royal Astronomical Society, 460, 1214

Ludlow A. D., Schaye J., Schaller M., Richings J., 2019, Monthly Notices of the Royal Astronomical Society: Letters, 488, L123–L128

Madau P., Dickinson M., 2014, Annual Reviews, 52, 415

Maiolino R., et al., 2012, Monthly Notices of the Royal Astronomical Society, 425, L66

Makarov D., Makarova L., Sharina M., Uklein R., Tikhonov A., Guhathakurta P., Kirby E., Terekhova N., 2012, Monthly Notices of the Royal Astronomical Society, 425, 709

Matthee J., Schaye J., 2019, Monthly Notices of the Royal Astronomical Society, 484, 915–932

Matthee J., Schaye J., Crain R. A., Schaller M., Bower R., Theuns T., 2016, Monthly Notices of the Royal Astronomical Society, 465, 2381–2396

McConnachie A. W., 2012, , 144, 4

Mei S., et al., 2007, The Astrophysical Journal, 655, 144

Meshveliani T., Lovell M. R., Crain R. A., Pfeffer J., 2024, Monthly Notices of the Royal Astronomical Society, 532, 1296

Milgrom M., 1983, The Astrophysical Journal, 270, 365

Moore B., 1994, Nature, 370, 629

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, The Astrophysical Journal, 524, L19–L22

Moster B. P., Naab T., White S. D. M., 2013, Monthly Notices of the Royal Astronomical Society, 428, 3121

Muldrew S. I., et al., 2012, Monthly Notices of the Royal Astronomical Society, 419, 2670

Munshi F., Brooks A. M., Applebaum E., Christensen C. R., Quinn T., Sligh S., 2021, The Astrophysical Journal, 923, 35

Navarro J. F., Eke V. R., Frenk C. S., 1996a, Monthly Notices of the Royal Astronomical Society, 283, L72–L78

Navarro J. F., Frenk C. S., White S. D. M., 1996b, The Astrophysical Journal, 462, 563

Navarro J. F., Frenk C. S., White S. D. M., 1996c, The Astrophysical Journal, 462, 563

Nelson D., et al., 2019, Computational Astrophysics and Cosmology, 6, 2

Neto A. F., et al., 2007, Monthly Notices of the Royal Astronomical Society, 381, 1450

Nobels F. S. J., Schaye J., Schaller M., Ploeckinger S., Chaikin E., Richings A. J., 2023, Tests of subgrid models for star formation using simulations of isolated disk galaxies (`arXiv:2309.13750`)

Nomoto K., Kobayashi C., Tominaga N., 2013, Annual Reviews, 51, 457

Olive K. A., Steigman G., Walker T. P., 2000, , 333, 389

Oman K. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 452, 3650

Orlando S., Peres G., Reale F., Bocchino F., Rosner R., Plewa T., Siegel A., 2005, Astronomy & Astrophysics, 444, 505

O'Shea B. W., Nagamine K., Springel V., Hernquist L., Norman M. L., 2005, The Astrophysical Journal Supplement Series, 160, 1–27

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

Peebles P. J. E., 1982, Astrophyics Journal, 263, L1

Penzias A. A., Wilson R. W., 1965, The Astrophysical Journal, 142, 419

Pillepich A., et al., 2017, Monthly Notices of the Royal Astronomical Society, 475, 648

Pillepich A., et al., 2018, Monthly Notices of the Royal Astronomical Society, 473, 4077

Ploeckinger S., Schaye J., 2020, Monthly Notices of the Royal Astronomical Society, 497, 4857

Pontzen A., Governato F., 2012, Monthly Notices of the Royal Astronomical Society, 421, 3464–3471

Portinari L., Chiosi C., Bressan A., 1998, Astronomy & Astrophysics, 334, 505

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, Monthly Notices of the Royal Astronomical Society, 338, 14

Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleyna J. T., 2006, Monthly Notices of the Royal Astronomical Society, 366, 429

Rey M. P., Pontzen A., Agertz O., Orkney M. D. A., Read J. I., Saintonge A., Pedersen C., 2019, Astrophyics Journal, 886, L3

Rey M. P., Pontzen A., Agertz O., Orkney M. D. A., Read J. I., Rosdahl J., 2020, Monthly Notices of the Royal Astronomical Society, 497, 1508

Richings A. J., Schaye J., Oppenheimer B. D., 2014a, Monthly Notices of the Royal Astronomical Society, 440, 3349

Richings A. J., Schaye J., Oppenheimer B. D., 2014b, Monthly Notices of the Royal Astronomical Society, 442, 2780

Richings J., Frenk C., Jenkins A., Robertson A., Schaller M., 2021, Monthly Notices of the Royal Astronomical Society, 501, 4657–4668

Robertson B. E., Ellis R. S., Dunlop J. S., McLure R. J., Stark D. P., 2010, Nature, 468, 49

Rubin V. C., Ford W. Kent J., 1970, The Astrophysical Journal, 159, 379

Rubin V. C., Ford W. K. J., Thonnard N., 1980, The Astrophysical Journal, 238, 471

Sales L. V., Wetzel A., Fattahi A., 2022, Baryonic solutions and challenges for cosmological models of dwarf galaxies (`arXiv:2206.05295`)

Salpeter E. E., 1955, The Astrophysical Journal, 121, 161

Sawala T., Frenk C. S., Crain R. A., Jenkins A., Schaye J., Theuns T., Zavala J., 2013, Monthly Notices of the Royal Astronomical Society, 431, 1366

Sawala T., et al., 2016a, Monthly Notices of the Royal Astronomical Society, 457, 1931

Sawala T., et al., 2016b, Monthly Notices of the Royal Astronomical Society, 457, 1931–1943

Schaller M., et al., 2015a, Monthly Notices of the Royal Astronomical Society, 451, 1247

Schaller M., et al., 2015b, Monthly Notices of the Royal Astronomical Society, 451, 1247

Schaller M., et al., 2023, arXiv e-prints, p. arXiv:2305.13380

Schaye J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 446, 521–554

Schaye J., et al., 2023, The FLAMINGO project: cosmological hydrodynamical simulations for large-scale structure and galaxy cluster surveys (`arXiv:2306.04024`)

Schmidt M., 1959, The Astrophysical Journal, 129, 243

Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, Monthly Notices of the Royal Astronomical Society, 417, 2020

Sheth R. K., Tormen G., 2004, Monthly Notices of the Royal Astronomical Society, 350, 1385

Simon J. D., 2019, Annual Reviews, 57, 375

Simon J. D., Geha M., 2007, The Astrophysical Journal, 670, 313–331

Simpson C. M., Grand R. J. J., Gómez F. A., Marinacci F., Pakmor R., Springel V., Campbell D. J. R., Frenk C. S., 2018, Monthly Notices of the Royal Astronomical Society, 478, 548

Smith R., Choi H., Lee J., Rhee J., Sanchez-Janssen R., Yi S. K., 2016, The Astrophysical Journal, 833, 109

Spekkens K., Urbancic N., Mason B. S., Willman B., Aguirre J. E., 2014, Astrophyics Journal, 795, L5

Springel V., 2005, Monthly Notices of the Royal Astronomical Society, 364, 1105

Springel V., 2010, Annual Review of Astronomy and Astrophysics, 48, 391–430

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, Monthly Notices of the Royal Astronomical Society, 328, 726

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

Stanek R., Rudd D., Evrard A. E., 2009, Monthly Notices of the Royal Astronomical Society, 394, L11

Stanway E. R., Eldridge J. J., 2018, Monthly Notices of the Royal Astronomical Society, 479, 75

Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, Monthly Notices of the Royal Astronomical Society, 373, 1074

Strömgren B., 1939, The Astrophysical Journal, 89, 526

Sutherland R. S., Dopita M. A., 1993, , 88, 253

Teyssier R., 2002, Astronomy &amp; Astrophysics, 385, 337–364

Teyssier M., Johnston K. V., Kuhlen M., 2012, Monthly Notices of the Royal Astronomical Society, 426, 1808

Thob A. C. R., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 972

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, The Astrophysical Journal, 688, 709

Trujillo I., Chamba N., Knapen J. H., 2020, Monthly Notices of the Royal Astronomical Society, 493, 87

Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Dalla Vecchia C., 2014, Monthly Notices of the Royal Astronomical Society, 442, 2641

Wang J., Frenk C. S., Navarro J. F., Gao L., Sawala T., 2012, Monthly Notices of the Royal Astronomical Society, 424, 2715

Wechsler R. H., Tinker J. L., 2018, Annual Reviews, 56, 435

Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, The Astrophysical Journal, 568, 52

Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, Astrophyics Journal, 827, L23

Wheeler C., Oñorbe J., Bullock J. S., Boylan-Kolchin M., Elbert O. D., Garrison-Kimmel S., Hopkins P. F., Kereš D., 2015, Monthly Notices of the Royal Astronomical Society, 453, 1305

Wheeler C., et al., 2019, Monthly Notices of the Royal Astronomical Society, 490, 4447

White S. D. M., 1994, Formation and Evolution of Galaxies: Les Houches Lectures (`arXiv:astro-ph/9410043`)

Wiersma R. P. C., Schaye J., Smith B. D., 2009a, Monthly Notices of the Royal Astronomical Society, 393, 99

Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, Monthly Notices of the Royal Astronomical Society, 399, 574

Wu J. F., Kragh Jespersen C., Wechsler R. H., 2024, arXiv e-prints, p. arXiv:2402.07995

Zehavi I., Contreras S., Padilla N., Smith N. J., Baugh C. M., Norberg P., 2018, The Astrophysical Journal, 853, 84

Zel'dovich Y. B., 1970, Astronomy & Astrophysics, 5, 84

Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, The Astrophysical Journal, 707, 354

Zheng H., Liao S., Hu J., Gao L., Grand R. J. J., Gu Q., Guo Q., 2022, Monthly Notices of the Royal Astronomical Society, 514, 2488

de los Reyes M. A. C., Kirby E. N., Zhuang Z., Steidel C. C., Chen Y., Wheeler C., 2023, The Astrophysical Journal, 951, 52

van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011, Monthly Notices of the Royal Astronomical Society, 414, 2458