Stability and Convergence of N-body Simulations for Galaxy Formation.

Julian Onions

The University of Nottingham

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy, September 2015

"To answer why did it happen, I suggest that our Universe is simply one of those things which happen from time to time."

–Edward P. Tryon

"In physics, you don't have to go around making trouble for yourself – nature does it for you."

–Frank Wilczek

"In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move."

–Douglas Adams

"Those who can't do, simulate.

Supervisor: Prof. Frazer Pearce

Examiners: Prof. Shaun Cole
University of Durham
Prof. Christopher Conselic
University of Nottingham

Submitted: 7th Sept 2015
Examined: 19th Nov 2015
Final version: 27 Jan 2016

Abstract

Galaxy formation is still a current topic in astronomy. An important tool to understanding it is through simulation, which allows galaxies to be studied from all angles and across time. It allows us to explore the gap between observation and theory, but only if the results are sufficiently accurate. In this thesis I look at the majority of the simulation pipeline from running through the various stages of analysis, and some of the limits of their accuracy, and the fidelity of the subsequent analysis tools.

It starts by looking at running simulations from initial conditions, and what influence changing parameters and simulation engines has on the outcome. Then I look in detail at how successful subhalo detection is by comparing a number of substructure finders, and examining their strengths and weaknesses. Following this I focus on a single parameter recovered for such haloes, the spin, and how well it was recovered, and what it tells us about the spin of substructures. Following this I investigated the building of merger trees, by writing my own merger tree program, and comparing it with some of the established ones. Then I look at using these processes as input to semi-analytic models, and how mass changes could affect the outcome. Finally I used a number of these tools to investigate the fate of some of the larger haloes formed at early times in an attempt to show where ultra-compact dwarf galaxies are formed and their fate.

Acknowledgements

The work included in this thesis has been stimulating and interesting and I have very much enjoyed the time. This is of course dependant on a lot of help, encouragement and support throughout this period.

I would like to start by thanking my supervisor, Professor Frazer Pearce. Rather than the conventional way of taking on a PhD straight after a relevant BSc, there was a gap of nearly 30 years since my original undergraduate days, which were also at Nottingham. After working in the computer industry for many years, I decided to do an Open University degree in biology in my spare time. As part of the optional modules I took some 2nd year courses on astronomy, and fell in love with the detail of a subject that I have always liked. I wandered in off the street, so to speak, to see if there was a possibility of taking my studies further in a part-time way, and Frazer seemed happy to take on a part-time, mature, recently qualified biology graduate, for which I am very grateful.

A PhD is necessarily very focused on a single topic, and the initial few weeks are quite difficult, keeping up with the new flow of ideas, terminology and concepts. I found all the staff in the department very friendly, supportive and happy to help, and the other PhD and postdocs also approachable, despite being in awe of these “real students” who obviously knew much more than I. So I would like to thank the department which continues to be a great and stimulating place to work in, and in particular Stuart Muldrew, Hanni Lux, Lyndsay Old and Thomas Yue for being a supportive group of simulators.

I have had occasion to work with some fantastic people outside the department too, of which the list is long. I would like to single out Alexander Knebe who has always been most helpful in response to questions, has seemingly endless enthusiasm and energy,

and gets stuff done, in turn inspiring me to push onwards with difficult tasks.

I also have to thank my family and in particular my wife for putting up with late night working and my struggles over equations, papers, self-doubt and results. Also to my friend Peter Wigmore who I've shared a number of pub lunches with, and has been encouraging and a wonderful time to let off steam when things have been less than perfect.

My employers have also been supportive of my part-time studies and the occasional requirements for conferences and workshops that have arisen.

The universities High Performance Computer cluster has been instrumental in a number of the simulations undertaken herein, and I acknowledge the use of such a facility. Likewise, the ICC at Durham's provision of the Dirac computer node COSMA has also been used to run analysis, and the Piz Daint super computer which ran the big simulations.

I would like to thank Juli Furniss, Elizabeth Atia and Alice-Amanda Kay for proof-reading some of this manuscript and providing some very helpful corrections.

Finally, I have to thank the online community, many of whom have helped me through some of the tough times. Some who similarly went from part-time Open University degrees to taking and achieving PhD's, have served as an inspiration. Others I journeyed with on the degree course, have also supported me in my subsequent studies, and to all these, I am extremely grateful.

Published Work

This Thesis is a piece of original reseach performed by myself, unless otherwise stated, with sections having previously been published in the journal *Monthly Notices of the Royal Astronomical Society* and *The Astrophysical Journal*. During this Thesis I have led or contributed to the following work:

1. *SubHaloes going Notts: The SubHalo-Finder Comparison Project*
   Julian Onions, Alexander Knebe, Frazer R. Pearce, Stuart I. Muldrew, Hanni Lux, Steffen R. Knollmann, Yago Ascasibar, Peter Behroozi, Pascal Elahi, Jiaxin Han, Michal Maciejewski, Manuel E. Merchn, Mark Neyrinck, Andrs N. Ruiz, Mario A. Sgr, Volker Springel, Dylan Tweed. 2012, MNRAS, 423, 1200

2. *Subhaloes gone Notts: Spin across subhaloes and finders*
   Julian Onions, Yago Ascasibar, Peter Behroozi, Javier Casado, Pascal Elahi, Jiaxin Han, Alexander Knebe, Hanni Lux, Manuel E. Merchn, Stuart I. Muldrew, Mark Neyrinck, Lyndsay Old, Frazer R. Pearce, Doug Potter, Andrs N. Ruiz, Mario A. Sgr, Dylan Tweed, Thomas Yue. 2013, MNRAS, 428, 2039

3. *Streams Going Notts: The tidal debris finder comparison project*
   Pascal J. Elahi, Jiaxin Han, Hanni Lux, Yago Ascasibar, Peter Behroozi, Alexander Knebe, Stuart I. Muldrew, **Julian Onions**, Frazer Pearce. 2013, MNRAS, 433, 1537

4. *Structure Finding in Cosmological Simulations: The State of Affairs*
   Alexander Knebe, Frazer R. Pearce, Hanni Lux, Yago Ascasibar, Peter Behroozi, Javier Casado, Christine Corbett Moran, Juerg Diemand, Klaus Dolag, Rosa Dominguez-Tenreiro, Pascal Elahi, Bridget Falck, Stefan Gottloeber, Jiaxin Han,

Anatoly Klypin, Zarija Lukic, Michal Maciejewski, Cameron K. McBride, Manuel E. Merchan, Stuart I. Muldrew, Mark Neyrinck, **Julian Onions**, Susana Planelles, Doug Potter, Vicent Quilis, Yann Rasera, Paul M. Ricker, Fabrice Roy, Andres N. Ruiz, Mario A. Sgro, Volker Springel, Joachim Stadel, P. M. Sutter, Dylan Tweed, Marcel Zemp. 2013, MNRAS, 435, 1618

5. *Galaxies going MAD: The Galaxy-Finder Comparison Project*
   Alexander Knebe, Noam I. Libeskind, Frazer Pearce, Peter Behroozi, Javier Casado, Klaus Dolag, Rosa Dominguez-Tenreiro, Pascal Elahi, Hanni Lux, Stuart I. Muldrew, **Julian Onions**. 2012, MNRAS, 428, 2039

6. *Sussing Merger Trees: The Merger Trees Comparison Project*
   Chaichalit Srisawat, Alexander Knebe, Frazer R. Pearce, Aurel Schneider, Peter A. Thomas, Peter Behroozi, Klaus Dolag, Pascal J. Elahi, Jiaxin Han, John Helly, Yipeng Jing, Intae Jung, Jaehyun Lee, Yao Yuan Mao, **Julian Onions**, Vicente Rodriguez-Gomez, Dylan Tweed, Sukyoung K. Yi. 2013, MNRAS, 436, 150

7. *Subhaloes gone Notts: the clustering properties of subhaloes*
   Arnau Pujol, Enrique Gaztanaga, Carlo Giocoli, Alexander Knebe, Frazer R. Pearce, Ramin A. Skibba, Yago Ascasibar, Peter Behroozi, Pascal Elahi, Jiaxin Han, Hanni Lux, Stuart I. Muldrew, Mark Neyrinck, **Julian Onions**, Doug Potter, Dylan Tweed. 2014, MNRAS, 438, 3205

8. *Sussing Merger Trees: the influence of the halo finder*
   Santiago Avila, Alexander Knebe, Frazer R. Pearce, Aurel Schneider, Chaichalit Srisawat, Peter A. Thomas, Peter Behroozi, Pascal J. Elahi, Jiaxin Han, Yao-Yuan Mao, **Julian Onions**, Vicente Rodriguez-Gomez. 2014, MNRAS, 441, 3488

9. *Subhaloes gone Notts: Subhaloes as tracers of the dark matter halo shape*
   Kai Hoffmann, Susana Planelles, Enrique Gaztanaga, Alexander Knebe, Frazer R. Pearce, Hanni Lux, **Julian Onions**, Stuart I. Muldrew, Pascal Elahi, Peter Behroozi, Yago Ascasibar, Jiaxin Han, Michal Maciejewski, Manuel E. Merchan, Mark Neyrinck, Andrs N. Ruiz, Mario A. Sgro. 2014, MNRAS, 442, 1197

10. *The life and death of cosmic voids* P.M. Sutter, Pascal Elahi, Bridget Falck, **Julian Onions**, Nico Hamaus, Alexander Knebe, Chaichalit Srisawat, Aurel Schneider. 2014, MNRAS, 445, 1235

11. *Sussing Merger Trees : The Impact of Halo Merger Trees on Galaxy Properties in a Semi-Analytic Model*
    Jaehyun Lee, Sukyoung K. Yi, Pascal J. Elahi, Peter A. Thomas, Frazer R. Pearce, Peter Behroozi, Jiaxin Han, John Helly, Intae Jung, Alexander Knebe, Yao-Yuan Mao, **Julian Onions**, Vicente Rodriguez-Gomez, Aurel Schneider, Chaichalit Srisawat, Dylan Tweed. 2014, MNRAS, 445, 4197

12. *Solving the puzzle of subhalo spins* Yang Wang, Weipeng Lin, Frazer R. Pearce, Hanni Lux, Stuart I. Muldrew, **Julian Onions**. 2015, ApJ, 801, 93

13. *nIFTy Cosmology: Comparison of Galaxy Formation Models*
    Alexander Knebe, Frazer R. Pearce, Peter A. Thomas, Andrew Benson, Jeremy Blaizot, Richard Bower, Jorge Carretero, Francisco J. Castander, Andrea Cattaneo, Sofia A. Cora, Darren J. Croton, Weiguang Cui, Daniel Cunnama, Gabriella De Lucia, Julien E. Devriendt, Pascal J. Elahi, Andreea Font, Fabio Fontanot, Juan Garcia-Bellido, Ignacio D. Gargiulo, Violeta Gonzalez-Perez, John Helly, Bruno Henriques, Michaela Hirschmann, Jaehyun Lee, Gary A. Mamon, Pierluigi Monaco, **Julian Onions**, Nelson D. Padilla, Chris Power, Arnau Pujol, Ramin A. Skibba, Rachel S. Somerville, Chaichalit Srisawat, Cristian A. Vega-Martinez, Sukyoung K. Yi. 2015, MNRAS, 451, 4029.

14. *Matter power spectrum and the challenge of percent accuracy*
    Aurel Schneider, Romain Teyssier, Doug Potter, Joachim Stadel, **Julian Onions**, Darren S. Reed, Robert E. Smith, Volker Springel, Frazer R. Pearce.
    In Preparation

15. *Sussing Merger Trees: Stability and Convergence*
    Yang Wang, Frazer R. Pearce, Alexander Knebe, Aurel Schneider, Chaichalit Srisawat, Dylan Tweed, Intae Jung, Jiaxin Han, John Helly, **Julian Onions**, Pascal J. Elahi, Peter A. Thomas, Peter Behroozi, Sukyoung K. Yi, Vicente

Rodriguez-Gomez, Yao-Yuan Mao, Yipeng Jing, and Weipeng Lin.
In Preparation

Specifically Chapter 3 was published as *(1.)*. Chapter 4 was published as *(2.)*, and *(12.)* is a follow up paper to this. Chapter 5 makes reference to much of the work published in *(6.)*, *(8.)* and *(11.)*. Chapter 2 contains a lot of the preliminary work leading to *(14.)*. The other papers contain work to which I have contributed data or analysis to, and some material presented in this thesis is taken from those papers on occasion for illustration, but does not form a full chapter.

Contents

List of Figures

List of Tables

Stability and Convergence of N-body Simulations for Galaxy Formation.

Chapter 1

Introduction

Our place in the universe is unremarkable, a fact that has been continually reinforced over the millenia. We used to think of Earth as the centre of the universe, which is a reasonable assumption given a limited viewpoint. However increasingly accurate observations by Tycho Brahe of the planets, combined with mathematical models by Kepler led to a refined heliocentric view that fitted the facts much better. With better telescopes and mathematical models we found we were orbiting a fairly average star, circling around in an average galaxy some distance from its centre.

As we look further out, we find our galactic environment is nothing special either, being somewhat of a backwater compared to the large super-clusters of galaxies. The large surveys that have been, or are being undertaken, such as Two-degree-Field Galaxy Redshift Survey (2dFGRS) (Colless *et al.*, 2001),Two Micron All-Sky Survey (2MASS) (Skrutskie *et al.*, 2006), and the Sloan Digital Sky Survey (Alam *et al.*, 2015); plus the upcoming ones such as Mapping Nearby Galaxies at APO (MANGA) (Bundy *et al.*, 2015), the Dark Energy Survey (Frieman, 2011), EUCLID (Laureijs *et al.*, 2014) and the Large Synoptic Survey Telescope (Ivezic *et al.*, 2008) all show or reinforce a large scale structure that is isotropic and homogeneous. They reveal the structure on the largest of scales, giving us the founding stones of the science of cosmology, where we study the origin, evolution and fate of the universe as a whole. So we see further out, and further back in time, letting us see how galaxies evolve.

Meanwhile, the Cosmic Microwave Background (CMB), through such missions as

COBE (Smoot *et al.*, 1992), WMAP (Jarosik *et al.*, 2011) and PLANCK (Planck Collaboration *et al.*, 2015) let us see the universe close to its very beginnings. This lets us view the conditions very soon after the big bang, and leads to a picture of the universe early on when it was a lot simpler.

Astronomy is not an experimental science, and therefore is not open to taking physical systems and setting up carefully monitored interactions to find out what happens. I have often described it in talks for the public as akin to *"Trying to learn Chinese by listening to a Chinese radio station"*. All one can do is listen to what is received and try and interpret it using what we already know, looking for patterns and hints. In a sense it is worse than that, as about 95% of the mass-energy of the universe is made of substances unknown, so it is rather as though we have been listening to a single radio station that is devoted to flower arranging, unaware that there are many radio stations available.

However, it is possible to bridge the gap between initial conditions found in the CMB data (and other sources), and the large scale structure found in these increasingly detailed surveys. With models derived from the initial conditions evolved from CMB findings, and then set to evolve forwards using the laws of physics, we have a way of experimenting with our understanding. The models, when complete, can be compared using statistical and other techniques, to see if they match with observational data. If they give results that match with what is observed, our confidence in the models grow. If they fail to match, then the assumptions need to be reconsidered, and perhaps extra parameters or physics included. We can also play "what if" games with theories and ideas, and while they are not physical experiments, they can guide us to new understandings.

However, the models and analysis are only useful if they have a requisite level of fidelity. The models have to approximate the physics in the observed universe closely enough to be useful. The analysis tools have to pick up the items of interest from the output data with increasing accuracy as both observations and model precision improve. Thus, in this thesis, I study the accuracy achievable by a number of common tools to determine the reliability that can be placed on their output.

FIG. 4*a*

FIG. 4*b*

FIG. 4*a*.—Tidal deformations corresponding to parabolic motions, clockwise rotations, and a distance of closest approach equal to the diameters of the nebulae. The spiral arms point in the direction of the rotation.

FIG. 4*b*.—Same as above, with the exception of counterclockwise rotations. The spiral arms point in the direction opposite to the rotation.

Figure 1.1: An illustration from the original paper of Holmberg (1941) including original caption showing the interaction of two spiral galaxies.

1.1 The history of astronomical N-body simulations

The history of running large scale simulations of astronomical structures dates back to Holmberg (1941) with the attempt to run models of 37 particle systems using light-bulbs to simulate gravity and measurement with galvanometer and photocell. The lamps are then moved stepwise and the "gravity" measured based on the intensity of the light, which reduces with the inverse square in the same way as gravity does.

Although this gives a sense of what might happen, and was relatively groundbreaking for the time as shown in Figure 1.1, the resolution is necessarily poor, the details were complex to keep the system balanced, bulbs had to be carefully calibrated, and then moved by hand, and the results were of limited use, although a worthy attempt to start the field of galaxy simulation.

With the advent of computers, it was possible to run simulations using test particles to approximate cosmological objects, although initially with very poor resolution both in space and mass (Aarseth & Hoyle, 1963). These simulations worked with point masses, solving the $N \times N$ force equations across all particles in a stepwise scheme

to approximate the full interactions. Resolution and particle number slowly improved, but the physics of radiative transfer, cooling and many other factors makes such simulations, even today, non-trivial.

A new finding in the structure of the Universe, dark matter, changed simulations in their nature.

1.1.1 The evidence for dark matter

Dark matter was first hinted at in findings by Jan Oort in 1932 (Oort, 1932) attempting to find the location of the Sun within the Milky-Way. Whilst examining the rotation velocity of stars, in an attempt to locate the centre of the galaxy, he discovered that only about one third of the gravitational mass needed to keep the stars in their orbits was present. He therefore suggested there was some "missing mass" that was not so far visible to the telescopes of the day, which must be present to counteract this velocity and keep the galaxy stable.

The following year, in 1933 Fritz Zwicky (Zwicky, 1933, 1937) made a calculation based on the properties of the Coma Cluster, a cluster of galaxies assumed to be in virial equilibrium. By comparing the mass to light ratio of the cluster, with that of the mass required to keep it in virial equilibrium, he found the two numbers obtained for the mass differed by a factor of about four hundred. Therefore substantial extra mass, that was so far undetectable was required. This invisible mass he referred to as *dunkle materie* or "dark matter", and so coined the phrase.

Although Zwicky's calculated figure is much higher than modern estimates of around to 5 to 6 times the missing mass (e.g., Lokas & Mamon (2003)), the two figures are still clearly out of balance.

Horace Babcock found similar results by looking at the rotation curves of the Andromeda galaxy in 1939 and published them in his thesis (Babcock, 1939), but the results were largely ignored as the spectral resolution of the stars used was poor, and the extent to which the edges of the galaxy could be probed was limited.

Vera Rubin in the 1970s (Rubin & Ford, 1970) revisited the work of Babcock, with a much more precise spectrograph, and published the now commonly understood result

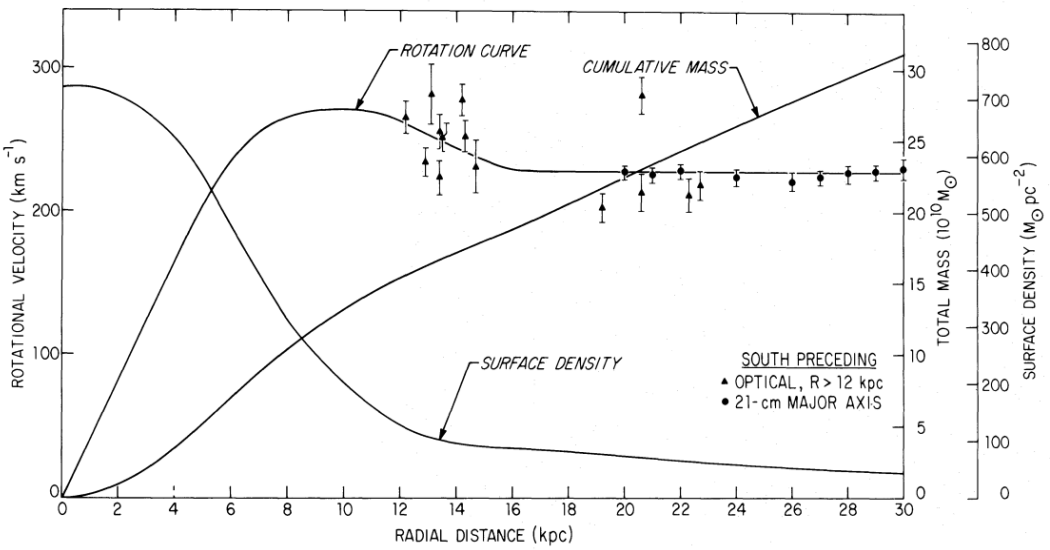

Figure 1.2: The rotation curve of M31 as measured by Roberts & Whitehurst (1975), which includes triangular points by Rubin in the optical, and 21-cm radio measurements. Extending out to 30kpc shows that the rotation curve is substantially flat, and not Keplerian as would be assumed by the mass to light ratio.

that most stars in spiral galaxies orbit at almost the same speed outside of the central part. This led to the unexpected flat rotation curves that were indicated by Babcock. Based on the visible mass, orbital velocities more like Keplarian should have been seen as the visible mass decreased. Rubin estimated that there was about six times as much dark matter present as regular matter.

With the advent of high precision radio telescopes that could trace the gas present in the outer reaches of galaxies, Roberts & Whitehurst (1975) showed that this flat curve extended well beyond the stellar matter. This indicates that whatever the dark matter is, it encloses a far bigger volume than the visible galaxy, as the outskirts of the galaxy are still rotating anomalously fast Figure 1.2. Thus was born the concept of the dark matter halo.

Big Bang nucleosynthesis, first proposed in the classic paper by Alpher, Bethe & Gamow (1948), and subsequently much refined to more modern values such as shown in Fields, Molaro & Sarkar (2014) makes specific predictions on the quantity of baryonic matter with respect to the critical density of the universe. The critical density of the universe is defined as

$$\rho_c = \frac{3H^2}{8\pi G} \tag{1.1}$$

assuming there is no curvature (for which there is good evidence the curvature is very

close to zero (Planck Collaboration *et al.*, 2015)). This is more often expressed as a density parameter, a ratio of the given density to the critical density,

$$\Omega = \frac{\rho}{\rho_c} = \frac{8\pi G\rho}{3H^2} \tag{1.2}$$

where ρ is the density, ρ_c the critical density, G the gravitational constant and H is the Hubble parameter. So the fraction of baryons is given by Ω_b for baryon density compared to the critical density,

$$\Omega_b = \frac{\rho_b}{\rho_c} \tag{1.3}$$

The critical density gives the escape velocity of the universe, and is the dividing line between a universe that will collapse back in on itself, and one that will expand forever. A universe below the critical density does not have enough matter within it to counteract expansion, and so expands eternally. A universe with more than the critical density finds gravity eventually dominates, and the universe will collapse in on itself in a "big crunch". This is without the dark energy factor, which changes the effect as the universe grows, exerting an outward pressure which seems to be a property of space itself.

Current values from BBN such as Fields, Molaro & Sarkar (2014) give the baryon fraction Ω_b as about 4.5 percent of the critical density, which corresponds extremely well with Cosmic Microwave Background (CMB) measurements. With the density of the universe shown to be almost exactly the critical density, this leaves 95.5% of the universe unaccounted for. Some 26% of this is in the form on non-baryonic matter, which is usually assumed to consist completely of dark matter Ω_{dm}.

The other constituents of the universe are the radiation pressure density – Ω_R, which was important early on in the universe, but is almost negligible now; dark energy – Ω_Λ, which is now the dominant player, but is not strong enough to overcome gravitational force on all but the largest of scales; and a curvature term, Ω_k, that seems to be almost exactly zero. Together these total up to the overall density parameter.

$$\Omega = \Omega_R + \Omega_b + \Omega_{dm} + \Omega_\Lambda + \Omega_k \tag{1.4}$$

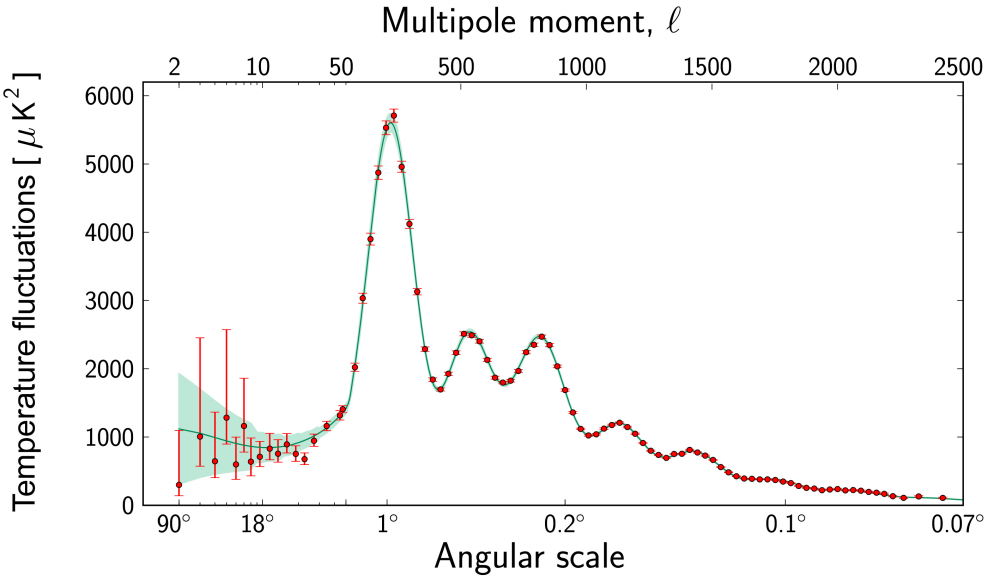

Figure 1.3: The cosmic microwave background power spectrum as measured by the ESA and the Planck Collaboration. (Planck Collaboration *et al.*, 2015)

Table 1.1: Cosmological large scale parameters from CMB temperature data, lensing, polarisation data, high-ℓ and BAO data (Planck Collaboration *et al.*, 2015).

Density	Value	Percentage
Cold dark matter	0.2622 ± 0.0049	26.2%
Baryon	0.04922 ± 0.00031	4.9%
Matter (Baryons + CDM)	0.3089 ± 0.0062	30.9%
Dark Energy	0.6911 ± 0.0062	69.1%
Curvature	$0.0008^{+0.0040}_{-0.0039}$	0.08%

Alternatively this can be viewed as

$$\frac{H(t)^2}{H_0^2} = \Omega_R a^{-4} + \Omega_b a^{-3} + \Omega_{dm} a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda \tag{1.5}$$

where a is the scale factor with value 1 today, H is the Hubble constant at given time t, and H_0 the Hubble constant today. This shows how radiation pressure falls away the fastest, followed by matter and then curvature, whilst dark energy remains constant.

The most precise measure of dark matter comes from the cosmic microwave background power spectrum as depicted in Figure 1.3. The first peak equates to a measurement of the curvature of space, and is consistent with zero curvature. The ratio of the second two peaks in this spectrum fit a model with a non-baryonic fraction of 0.262 of the total mass-energy of the universe, compared to 0.049 for baryonic matter. The errors on these figures are increasingly tiny and are summarised in Table 1.1.

All these results point in the same direction that dark matter is present, and dominates over regular matter.

However, so far all attempts to detect dark matter directly or indirectly, or to make it in particle accelerators, have been unfruitful. This does still leave the door open for other explanations, such as modified gravity, but using Occam's razor, dark matter particles seem to fit the facts in the simplest way so far.

With this knowledge, that it is extremely likely that dark matter not only exists, but is the dominant form of gravitational attraction in the universe, simulations of large scale structure were given a boost in performance and simplicity.

1.2 Dark matter simulations

Dark matter is thought to be a collision-less particle which is electrically neutral, stable and of indeterminate mass and scattering cross section. It is usually characterised as cold dark matter (CDM), meaning its speed of movement is well below relativistic velocities, although some work with warm dark matter (WDM) has also been trialled in simulations. If the particle were not collision-less, they would tend to pool together into much more compact haloes, or if they are their own anti-particle (for which several theories predict), they would annihilate and produce a strong energy signal.

Hot dark matter (HDM), where the particles move at relativistic speeds has been largely ruled out (Frenk & White, 2012), despite the neutrino being an otherwise ideal candidate in many respects. HDM predicts that structure would form in a top down way, with massive galaxies forming first, and fragmenting over time, which is not what we see observationally.

Dark matter particles, having only gravitational interaction, make for an almost perfect medium for simulations. The complications from electromagnetic interaction present in baryonic matter do not apply, and with CDM outweighing baryonic matter by a factor of about five to one, the approximation that large scale structure is dominated by dark matter is a fairly good one. Thus dark matter simulations have long been a favourite way to explore details on the super galactic scale.

Table 1.2: Dark matter simulations

Simulation	Box (h^{-1}Mpc)	N particles	Particle Mass $M_\odot h^{-1}$	Year	Code
Hubble Volume	3000	1000^3	2.25×10^{12}	2002	HYDRA
Millenium	500	2160^3	8.6×10^8	2005	GADGET-2
Millenium II	100	2160^3	6.885×10^6	2009	GADGET3
Horizon	2000	4096^3	7.7×10^{11}	2009	RAMSES
Horizon Run 3	10815	7210^3	2.44×10^{11}	2011	GOTPM
Bolshoi	250	2048^3	1.35×10^8	2011	ART
Millenium XXL	3000	6700^3	6.2×10^9	2012	GADGET3
MultiDark	1000	2048^3	8.63×10^9	2012	ART
DEUS FUR	21000	8192^3	1.2×10^{12}	2012	RAMSES-DEUS
Juropa Hubble Volume	6000	6000^3	7.49×10^{11}	2012	CUBEP3M
Illustris-Dark	75	1820^3	5.295×10^6	2014	AREPO
Galaxy simulations					
Via Lactea	90	616^3	2.09×10^4	2007	PKDGRAV
Via Lactea II	90	1032^3	4.1×10^3	2008	PKDGRAV2
GHALO	40	1458^3	1.0×10^3	2009	PKDGRAV2
Aquarius	100	1638^3	1.712×10^3	2009	GADGET3

Although a simulation with particles approximately equal to the expected mass would be ideal, this is far beyond the ability of current or planned computers, therefore simulations are usually run with particles in the range $10^7 - 10^{10}$ $M_\odot$. The progress of this is shown in Table 1.2 and graphically in Figure 1.4.

The earliest simulation runs invoked simple particle by particle (P-P) calculations. Whilst the interaction between two particles can be solved analytically, any number above this cannot, and has to be solved stepwise in an iterative fashion. This uses the basic gravitational laws given by Newton of

$$\vec{F} = \frac{GMm\vec{r}}{r^3} \tag{1.6}$$

However, this equation is not suitable for small scale interactions when approximating conglomerations of atomic sized particles by large point masses. Such masses will almost always pass through each other, being composed of collision-less particles of tiny cross section, whereas large point masses when they approach too closely tend to either orbit each other, or when colliding, the force tends to infinity. Therefore a modified version of this equation is used, with a softening constant chosen such that the force never rises too highly when particles come into close contact. The modified

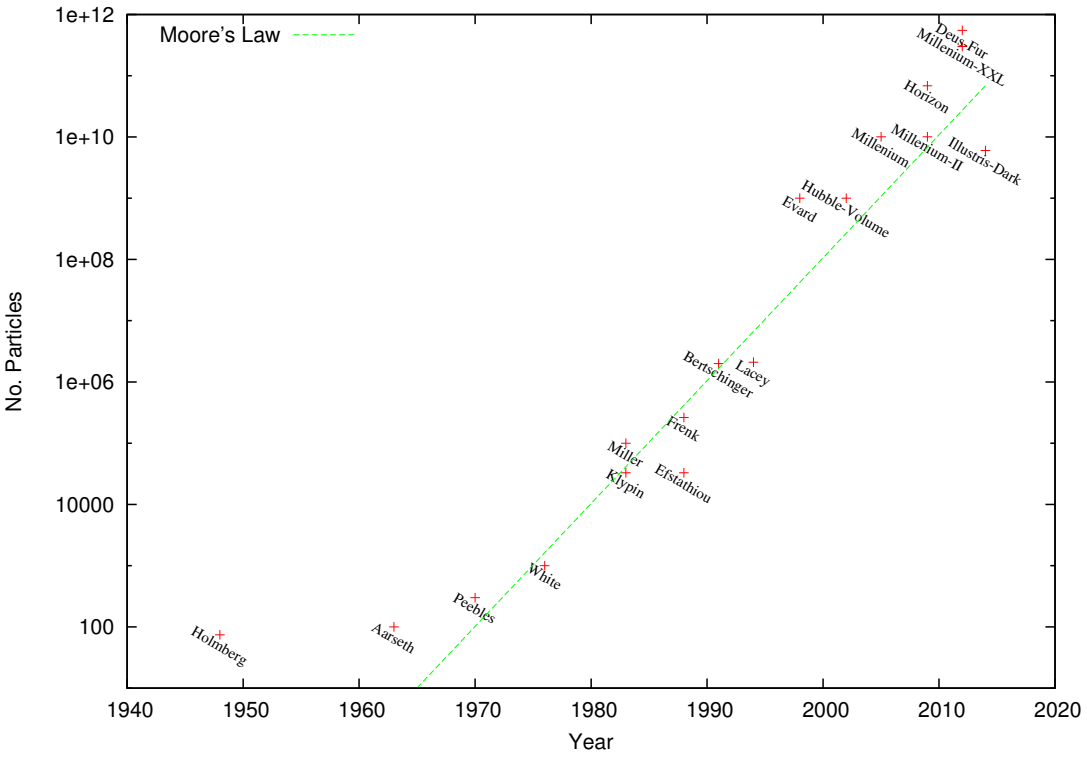

Figure 1.4: The increase in particle number since the first dark matter simulations. The rise approximately follows Moore's law (Moore, 1965) for computing increases, whereby computing power doubles approximately every 18 months.

variant is

$$\vec{F} = \frac{GMm\vec{r}}{(r^2 + \epsilon^2)^{3/2}} \tag{1.7}$$

where ϵ is the softening value. The value of ϵ is usually chosen to be some fraction of the mean inter-particle spacing, although as I show later in Chapter 2, changing this value can have quite dramatic effects on small scales.

More evidence for dark matter comes as a result of running such simulations. Without a large dark matter addition, normal baryonic matter would not have enough time to form the structures we see today (Davis *et al.*, 1985).

The particle-particle calculations quickly run out of resources, as it is an $N \times N$ calculation, so runs in $O(N^2)$ time. This scales extremely badly with increases in particle numbers.

So several other approaches based on approximations of this calculation are normally used, all of which have some limitations.

One such approach is the Particle-Mesh, or PM grid method (Efstathiou & Eastwood, 1981), it relies on smoothing the particles onto a grid, from which a gravitational po-

tential can be calculated using the Poisson equation:

$$\nabla^2\phi(x) = 4\pi G[\rho(x) - \bar{\rho}] \tag{1.8}$$

where $\phi(x)$ is the gravitational potential at x, G is the gravitational constant, and the part in square brackets is the difference in density at x, $\rho(x)$, compared to the average density $\bar{\rho}$.

The particles can be assigned to the grid structure in a number of ways, the simplest being assigning the particle to the grid point nearest to it. This can cause problems, with discontinuities between cells. So a smoothing method is usually invoked to spread the particle mass onto multiple grid cells. A common technique is the cloud-in-cell scheme, where the particle is proportioned over the nearest 8 cells in 3-dimensions giving fractional mass to each based on how close the particle is to each cell. A more refined scheme is the triangular-shaped-cloud method, which considers the 27 nearest cells for assignment, with a consequent penalty in speed, but improvement in the derivative being continuous.

The gravitational potential equation 1.8 is then solved by Fast Fourier Transform (FFT) techniques. This gives a scaling relation of $O(N_g \log(N_g))$, where N_g is the number of grid points. This improves significantly on the run time of P-P. The limitation of this method is the grid size, and if you wish to simulate particles undergoing infall into a high mass concentration the inaccuracies on the size of the grid become an issue.

An alternate technique is to arrange the particles in a hierarchical tree and walk the tree calculating forces using the Barnes-Hut methodology, as described in Barnes & Hut (1986). This presumes that forces from very distant particles do not need to be calculated very accurately, and attention can be focused on the dominant local environment. Thus, at small scales, forces between individual particles are calculated, whilst at larger scale forces between groups of particles in the tree structure are sufficient. The opening angle parameter θ is the parameter that determines at what point the code switches from approximation to full calculation using the inequality

$$d > \frac{l}{\theta} \tag{1.9}$$

where d is the distance of a particle from the cell, l is the size of the cell, and θ is the opening angle parameter. So if the inequality is true, the cell is used in place of the

particle. This runs in $O(N \log(N))$ time, so not as fast as the mesh code, but faster than the P-P code. It also has advantages in not running calculations in uninteresting regions such as voids.

Naturally there are combinations of these codes, with adaptive grid codes that attempt to refine the grid in dense regions. Combinations of particle-mesh and particle-mesh-tree have all been tried, and have advantages and disadvantages depending on the task at hand.

Much of the work shown in this thesis is derived from simulation runs using the GADGET3 code (Springel, 2005) which is a Tree-PM code. Some of the work has used the PKDGRAV code, which is a pure tree code, and other examples have been run with RAMSES a grid based code, and CUBEP3M, a P^3M code (Particle-Particle/Particle-Mesh) hybrid code.

As particle count increases and the resolution of the simulation improves, it becomes important to identify the structures generated in an automated way. Initially these were focused on looking for complete structures which represent dark matter gravitational wells that hold galaxies or galaxy clusters. As resolution improved there was a need to isolate and identify substructure amongst the larger structures. Such objects might hold dwarf and satellite galaxies of a main galaxy, including those undergoing merging.

Given a large dark matter simulation, it is then necessary to find the overdense regions that are the nuclei for the formation of galaxies. These can be picked out by eye, but as the simulations get larger, this is tedious and also error prone. A more rigorous definition of an overdense halo is required, and an automated method for finding such structures is essential with the larger simulations.

This leads on to how a halo is defined. Usually it is some anomalous overdensity with respect to the average density of the box. However the density may be taken from several definitions that can comprise the density of the box. These might be the critical density, the average density, a top hat collapse density or variations between these. But once this is defined, how many particles are required for a halo? Some programs go as low as 10 or 20 particles to identify a minimum halo size, although in practice around 300 or more particles are required to get convergent properties (Bett *et al.*, 2007).

Should all the particles be gravitationally bound to the halo? This depends somewhat on the application in mind. If only the mass is required, for cases such as gravitational lensing maps, then just the particles within the designated volume will suffice. However if properties of halos such as spin, size, density profile etc are required, then a tighter definition is needed. This involves unbinding the particles that are not gravitationally bound. This is an iterative process as each particle removed reduces the gravitational mass. Examples of issues with unbinding are discussed more in Chapter 4.

Also prevalent are issues to do with the temporal history of haloes. Two haloes passing through each other can appear very much like a single halo when viewed in one snapshot. Sometimes these can be untangled using other attributes of the particles, but not in all cases. This becomes particularly difficult as the density of the halo increases. Some of these issues are considered in Chapter 3.

Once dark matter simulations were completed, it soon became obvious that a means of populating the dark matter haloes with synthetic galaxies would be useful. To this end so called semi-analytic models (SAMs) were developed that could take the overall history of halo development, taking into account merger history and other parameters, and construct reasonable facsimiles of occupying galaxies.

This step necessarily requires the history of mergers over cosmic time to be represented as one of the inputs, thus requiring merger tree generation. Making a history of merger trees seems at face value quite a straightforward task, but there is hidden complexity in the process. Haloes can pass through one another, they can get stripped of particles, and they can merge or fragment. Some metric to work out what a temporal halo ancestor becomes is often required, and this is looked at in more detail in Chapter 5.

1.3 Layout of this thesis

This work starts by comparing simulation models for their possible accuracy. Then compares methods of subhalo detection techniques, followed by a look at the recovery of the spin of subhaloes (a dimensionless parameter related to angular momentum, further defined in Chapter 4) by the various finders. This was followed up by considering

the mechanics of building merger trees, and then some work on how mass definitions affect semi analytic models. Finally some applications of all parts of this work are brought together to study possible formation of ultra-compact dwarf galaxies.

1.3.1 Analysis pipeline

One of the common methods of analysing simulations is a multi step approach. This comprises of:

1. Generate a set of appropriate initial conditions from which the simulation can be run.

2. Run the simulation, saving snapshots as appropriate to the task. This results in a large amount of particle data, some of which can be analysed directly, such as producing the power spectrum or density maps.

3. The next procedure is usually to run a structure finder over the raw snapshots. These produce halo catalogs after identifying the overdense regions by one of a number of standard techniques, described in more detail in Chapter 3. The catalog typically contains a collection of attributes associated with each halo. This is usually the location, mass, size and velocity of the structure, but may also contain other derived information such as spin, orientation, shape and other properties.

4. With haloes detected and a catalog produced, this can be followed by the building of a merger tree. This process traces the history of each structure across the snapshots, so that for an individual halo it is possible to find all its progenitors, siblings and descendants. This process can be a goal in itself, but is often just the next step along the way to populating dark matter haloes with galaxies.

5. Finally the merger tree produced in the previous step, together with the halo catalog information, can be used as input for a semi-analytic model code. These codes use fixed formula and associated parameters to derive a model of how the resulting galaxies would evolve, based on their history of mergers and other interactions.

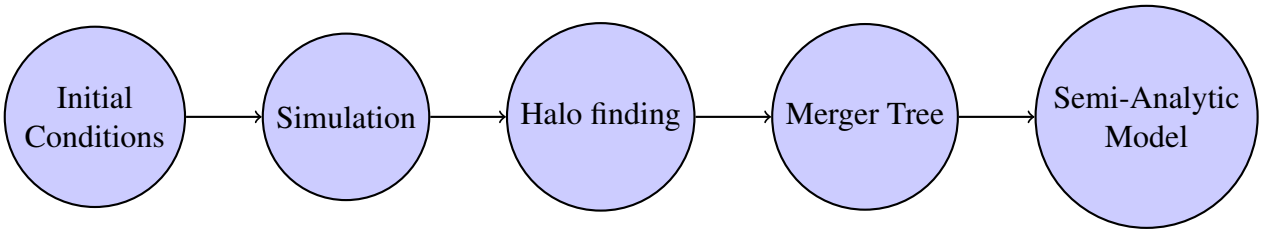

Figure 1.5: Typical flow of processes in a dark matter simulation pipeline

Figure 1.5 shows such a pipeline of processes that are followed in sequence to allow the analysis of a question under investigation to begin.

1.3.2 Initial Conditions

Initial conditions are generated normally by taking a random uncorrelated field of points representing particle mass, giving a white noise distribution. This distribution should have a constant power spectrum, so there is no bias in the points. Then a perturbation field is computed on these points using data from the CMB power spectrum, for which the temperature fluctuations are directly correlated with the density fluctuations. This leads to velocities that can be applied to the grid, using either just the first order terms (1LPT) – the Zel'Dovitch approximation, where particles simply evolve linearly to their new location based on initial velocity; or including the second order terms (2LPT) to give finer results. In most cases, the Zel'Dovitch approximation is found to be good enough, but with increased accuracy required, the 2LPT models are being increasingly examined (LHuillier, Park & Kim, 2014).

Effects of the initial conditions on the output of simulations have not been considered in this thesis,

1.3.3 Simulations and the goal of 1% accuracy

The ESA Euclid mission, (Laureijs *et al.*, 2014), is a satellite mission to map the geometry of the dark matter and dark energy of the universe. As part of this mission, which involves examining weak gravitation lensing, some investigation was carried out into the accuracy of the models that can currently be made. I looked at dark matter only simulations and compared three different simulation engines starting with the same

initial conditions.

Each simulation was run from the same initial conditions at a variety of resolutions, each run varying the major parameters that each engine allowed, and the results obtained compared is the subject of Chapter 2.

The Euclid consortium are looking for a 1% accuracy level for the matter-matter power spectrum, between $k = 0.1$ and $k = 50$. Based on the work done here, and also results derived from the halo finder projects, such accuracy seems ambitious at best for full baryonic physics, as dark matter only simulations struggle to get any where close to this value.

The baseline taken was the GADGET3 simulation engine, and compared to RAMSES, CUBEP3M, and PKDGRAV. In most cases the simulation engine was modified to take in GADGET3 format initial conditions and save in GADGET3 format for analysis.

1.3.4 Comparison of Subhalo finders

The research started with a comparison of the techniques for subhalo detection. This was carried out in a black-box type way, by providing participants with a number of dark matter only snapshots of a recognised simulation (Springel *et al.*, 2008a) and asking them to run their structure finders on the results, returning the data in a pre-agreed format. These were then analysed and compared to see what sort of agreement there was amongst the various different structure finders.

There are a number of different techniques popular in halo finding. It was thought that some of these would be better at detection of subhaloes, notably in high density environments, than others.

The types of finder can be broken up into the following different categories:

1.3.4.1 Spherical Overdensity

In this case, the finder is looking for an overdense region by considering the particles in a sphere. Usually the particles are collected using a friends-of-friends type of approach, where each particle and its immediate neighbours are considered up to some

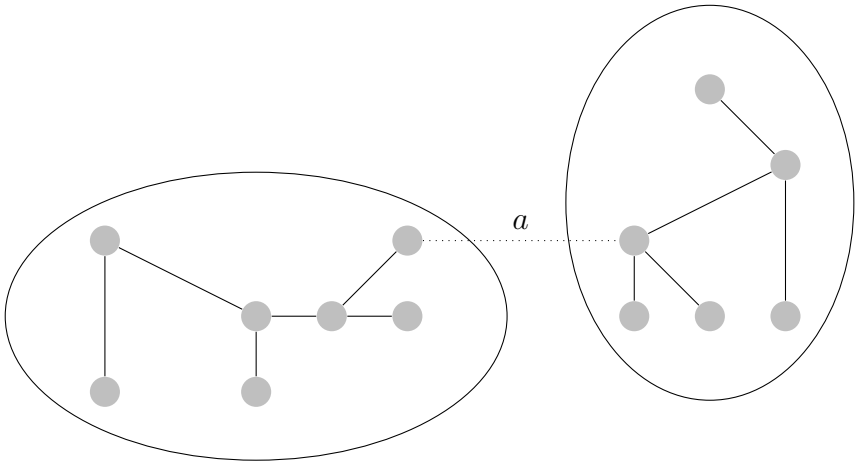

Figure 1.6: An example of the friends-of-friends algorithm with two haloes found, separated because of the distance between nodes lined by a exceeding the linking length.

pre-specified linking length (Huchra & Geller, 1982; Press & Davis, 1982; Eke *et al.*, 2004). A typical value is that of 0.2 of the mean interparticle separation and has been show to be independent of redshift and matter density (Jenkins *et al.*, 2001). This requires linked particles to be 5 times closer than they would be at random, and in practice leads to a halo with an approximate overdensity of about 180 times the box density, (M_{180}) (More *et al.*, 2011). This includes more particles than is required for an M_{200} halo – a halo with more than 200 times the critical density. Therefore this is a good starting point to work from to be able to remove non gravitationally bound particles leading to the m_{200} result. A halo of $M_{200_{\rho_{crit}}}$ is a size where a top-hat collapse will occur. A top hat collapse occurs when an overdensity concentrates matter compared to the background, to the point it exceeds the expansion rate. For an Einstein-de Sitter universe, this occurs at $18\pi^2 \approx 178$ times the critical density. Figure 1.6 shows an illustration of the friends-of-friends technique.

This gives a good starting point for most schemes of halo finding, not just spherical overdensity.

The spherical overdensity technique takes this list of candidate particles and assesses the centre, usually in an iterative way. Then working outwards adds particles to the list, calculating the density of the candidate halo at each step. At some point the density will fall below a candidate density, usually some factor of the critical density such as $\rho_{200_{crit}}$. The halo is then complete to a first approximation, but may be subject to further processing such as unbinding steps. This is described in Lacey & Cole (1994)

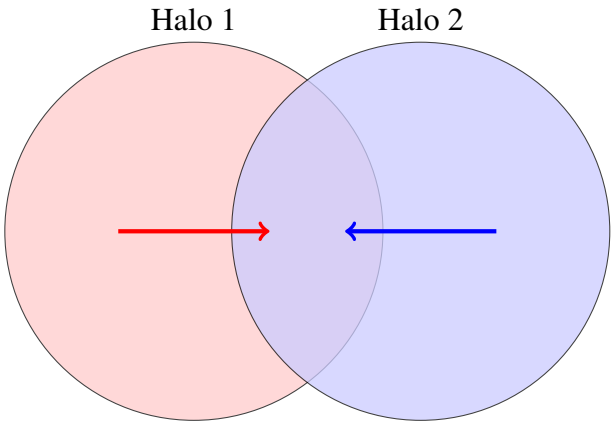

Figure 1.7: Two interpentrating haloes, which as they get closer, halo finders increasingly struggle to separate.

and improved in Tinker *et al.* (2008).

However other techniques can potentially give a better solution to the problem of detection, particularly when considering triaxial haloes, or streams. It somewhat depends on the what is being looked for.

1.3.4.2 Phase space detection

Typically, the data available for each particle is their position in 3-d space, their 3-d velocity, and usually their mass. Many times the simulations are run on a fixed particle mass setup, so individual masses are not required. Whilst much can be gained from just the positional data, using the friends-of-friends technique, there is clearly more information available in the velocities. Whilst this information is used in the unbinding stages, it can in principle be used earlier in the process.

Locating overdensities based on both position and velocity has the potential to be much more accurate in some circumstances. Consider for instance, two substructures that are passing through each other as shown in Figure 1.7.

In this case, most of the particles will be linked together by the friends-of-friends algorithm, an overdensity will be detected, and a structure found. In the subsequent unbinding steps, most of the particles from one of the halos will be rejected as not bound, if the velocity of the two structures is significantly different. This may result in the particles from one of the haloes being returned to the background set as unattached.

If instead the velocity information is used in the gathering set, by projecting the par-

ticles into phase-space, then this may allow the two haloes to be detected early on. Whilst the case of two independent haloes colliding in this manner may be rare, it is very common for substructures to be taken into larger haloes and travel through them for some period of time, before either being merged or stripped of particles. Therefore any scheme that can enhance the detection of substructures against a complex background will be ideal for examining the detail of substructure.

The hopes were high that so called phase space detectors may well do far better than simple overdensity detectors when it comes to looking for substructure. The results in Chapter 3 appear to show these hopes were not justified.

1.3.4.3 Other techniques

There are other techniques for locating haloes. One method is to consider Voronoi tessellation of the particles, surrounding each particle by a volume where each point in the volume is closer to a given particle than any other. This allows a local density to be calculated, and extended to form a structure.

One other technique started from the velocity data, using this as the first method of identifying co-moving groups of particles before using spatial data to consolidate them. This technique is particularly useful if the intent is to identify streams of particles that may be the result of an interaction of one structure with another.

1.3.5 Spin of Subhaloes

One of the follow up projects to the initial subhalo detection work investigated the spin properties of the subhaloes. Given the amount of data available on subhaloes from the initial project and the common processing pipeline it was possible to recover spin for all of the finders present in the first project. This allowed comparisons to be made between finders and to also present some new science with the results.

For this project all of the Aquarius haloes were analysed together with a different set of data provided from the GHALO project, a similar simulation of a milky way sized dark matter halo. The results highlight the difference in the finders recovery of the halo, and show that the spin parameter is a particularly good indicator of how well the finders

are recovering uncontaminated haloes. This work is presented in detail in Chapter 4.

1.3.6 Comparison of Merger Tree Builders

The construction of merger trees is an important step in the processing of simulation data. It can be used for a variety of purposes, but perhaps most commonly as input to semi-analytic models for galaxy population.

In preparation for a workshop on merger trees, I wrote a new merger tree construction program. This differed from the normal processes of making merger tree by not using particle data, instead using purely the halo catalog data. The workshop in July 2013 looked at comparing merger tree results(Srisawat *et al.*, 2013), so it was an ideal time to float a new idea and see how it performed.

The resulting merger tree builder (JMERGE- a very unoriginal name) resulted in an extremely fast method for making merger trees, but was, at least initially, a less precise technique. It is very dependant on temporal information to reconstruct the movements of the haloes between snapshots.

1.3.7 Comparing Semi-analytic model mass definition

At another workshop (nIFTy) on semi-analytic models (SAMs), most of the current SAM authors were present and a lot of data was available for analysis. The initial focus of the meeting was to compare the output of the models by looking at various data from the catalogues and produced the paper Knebe *et al.* (2015). As a spin off from this project, it seemed an interesting idea to compare the effect of the mass definition provided in the halo catalogue on the subsequent SAM catalog.

The workshop data provided five different mass definitions. These were

- Friends-of-friends mass. This is the mass from the assembly of a friends-of-friends halo containing all the particles gathered.
- Bound mass. This is the mass from the friends-of-friends mass, but with gravitationally unbound particles removed.

- $M200_c$. This is the mass found when extending the halo out until it is no longer 200 times the critical density of the universe.
- $M200_b$. This is the mass found when extending the halo out until it is no longer 200 times the background, or average density of the universe.
- BN. This uses the definition from (Bryan & Norman, 1998) and is largely equivalent to the spherical top-hat collapse mass.

Many of the SAM codes were run on the same halo catalog but just using the different mass definitions. This seemed a useful opportunity to see how the difference in mass definitions alters the outputs. This is looked at in Chapter 6.

1.3.8 Tracing the possible progenitors of ultra-compact dwarf galaxies

Finally using the tools and methods described, I investigated a possible history of Ultra-Compact Dwarfs (UCDs). UCDs are a fairly recent phenomena in astronomy, being discovered in the early 2000s (Phillipps *et al.*, 2001; Drinkwater *et al.*, 2003) as a category "structurally and dynamically distinct from globular star clusters and known types of dwarf galaxy". They are, as their name suggests, small and compact galaxies, they have luminosity comparable to faint dwarf galaxies, but very different morphologies. This is shown in the plot from the original paper included here as Figure 1.8.

In Chapter 7 I explore how these might have formed, investigating two possible scenarios. One in which tight compact galaxies are formed at early time, and starved of resources to grow. The other where the galaxies are formed early on, grow through mergers, and are then stripped of all the outer most stars to form the nucleated core often seen in clusters.

1.3.9 Conclusion

In the final chapter, I summarise the findings of each of these chapters and look at some possible future work that could follow up these findings.

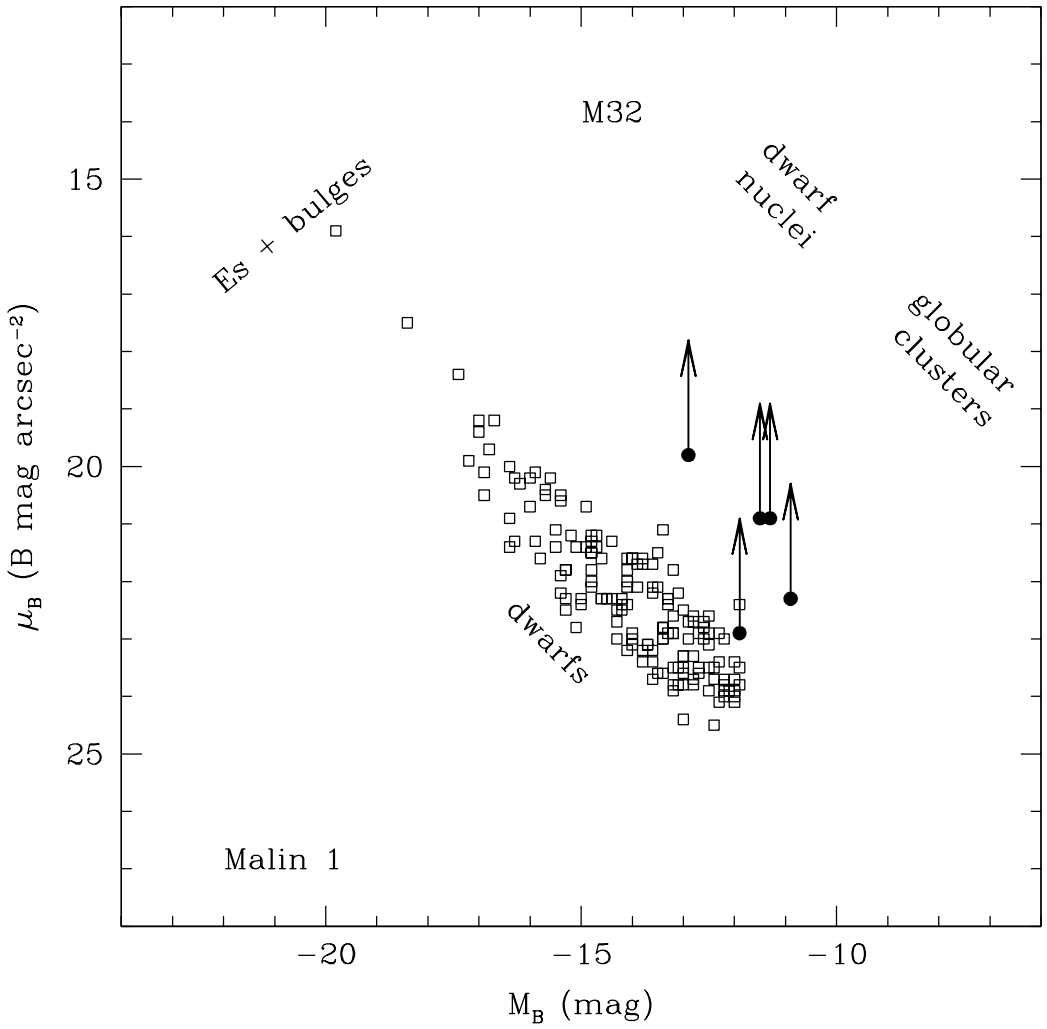

Figure 1.8: The original definition of UCD from the paper Phillipps *et al.* (2001) built on a plot from Ferguson & Binggeli (1994). The squares are their measurements of dwarf galaxies in the Fornax cluster, and the filled circles the UCDs found in Fornax. Surface brightness given for the UCDs is a lower limit, hence the upward pointing arrows.

Chapter 2

Can we get to 1% precision in simulations?

2.1 Introduction

Cosmological simulations have long been shown to reproduce the large scale structure of the universe to a good degree of statistical accuracy (Springel *et al.*, 2005). It is claimed we are entering a period of precision cosmology (Primack, 2005) for which theory and data are in good agreement. Whilst this is certainly true for much of the large scale structure of the universe, simulations still struggle to accurately predict the smaller scale structure.

Accuracy up to one percent is required for some techniques, such as weak lensing (Huterer *et al.*, 2006), and is viewed as a goal for the upcoming EUCLID mission (Laureijs *et al.*, 2014) for instance, but it is increasingly difficult to obtain with power spectrum wavenumbers beyond a k value of 1 $\mathrm{Mpc}^{-1}h$. Beyond this it is expected that baryonic and neutrino effects may be quite large (White, 2004; Guillet, Teyssier & Colombi, 2010), but to be able to start to consider these effects a solid understanding of the dark matter component is required.

In attempting to refine the models to this level of accuracy, there are a number of places that errors may arise. They may stem from the generation of initial conditions, they may arise from defects in the model code, or from issues with the parameters the

Table 2.1: Initial conditions of the different scaled simulations used.

Name	Boxsize Mpc/h	N	Mass (10^{10} $M_{\odot}$)
B1	62.5	256^3	0.109402
B2	125	512^3	0.109402
B3	250	1024^3	0.109402
B4	312	1280^3	0.109402
B5	375	1536^3	0.109402

models are run with, or indeed the analysis tools.

In Heitmann *et al.* (2010) they looked in particular at the power spectrum, which we also use here to evaluate the models, but in this case we are focused just on the simulation engine, and its parameters, and we consider four simulation engines.

In this project we ran a number of simulations using the different simulation engines, varying their parameters and comparing the results one with another such as density maps and power spectra, to see where the variations appear. All simulations were run from the same initial conditions and analysed with the same analysis tools. Thus the only variables were the simulation engine and their associated parameters.

2.2 The Data

The data was taken using a set of dark matter initial conditions prepared by Rob Smith with WMAP7(Jarosik *et al.*, 2011) cosmology ($\Omega_m = 0.270877$, $\Omega_\Lambda = 0.72913$, $\sigma_8 = 0.816$ and $h_0 = 0.704$). The initial conditions were available at varying resolutions shown in Table 2.1. All conditions were set at a redshift of 49 with fixed particle mass using second order Lagrangian perturbation theory (2LPT).

The data was run using different simulation engines all taking exactly the same initial conditions and running to redshift zero, making output as instructed. The programs were modified to directly read the gadget format initial conditions and to write out in gadget format, so a common analysis pipeline could be run regardless of the simulation. Snapshots were taken at regular intervals, although in most cases only the final snapshot at redshift of zero was used for the comparative analysis.

2.2.1 Simulation Engines

The first code to be examined was GADGET3 (Springel, 2005). GADGET3 takes a parameter file that can allow the run to vary a number of parameters. In particular we looked at two of the main parameters. The softening length, and the grid size. We also tried fixed time steps as well as letting the default variable timesteps take effect.

The softening length governs at what distance the calculations should switch from a $1/r^2$ force law, to one adapted to compensate for the fact that the particle resolution on small scales does not represent a fluid of particles, but instead large discreet particles. The forces therefore get very high when in close proximity and result in unrealistic interactions (see Binney, 1987; Dyer & Ip, 1993). Therefore the forces are reduced starting at the softening distance according to the formula to better simulate a collision-less medium. The common case is to replace the gravitational force from pure Newtonian (Equation 2.1) to one with a softening term identified by Plummer as shown in Equation 2.2 where ϵ is the softening length, which reduces to Newtonian at large distances.

$$\vec{F}(r) = -\frac{Gm_1m_2\vec{r}}{r^3} \tag{2.1}$$

$$\vec{F}(r) = -\frac{Gm_1m_2\vec{r}}{(r^2+\epsilon^2)^{3/2}} \tag{2.2}$$

The GADGET3 code uses a variant of this that has a spline interpretation, that reduces to the Newtonian formula at values above the softening length, where the Plummer code converges more slowly. See Springel, Yoshida & White (2001) for the exact details.

The grid size parameter is a compile time setting that specifies how big the FFT mesh size if for a tree-PM code as described in Section 1.2. This is used to trade off long distance calculations with intra-grid calculations. The grid size is usually tuned for the size of the simulation being performed. As more particles are used, a larger grid helps to improve the run time performance.

RAMSES is a simulation engine that works on an adaptive grid refinement technique (Teyssier, 2002). It has only a few parameters that can be set such as what grid size is a minimum and what maximum grid size to allow.

PKDGRAV is a tree based simulation code that constructs a k-d tree (Bentley, 1975), a

binary tree of particles in k dimensions (typically the 3 spatial dimensions), resulting in cells with a small number of particles in each cell (Stadel, 2001). This is then solved using the Barnes-Hut algorithm for tree walking applying a modified opening angle calculation given in 2.3 to decide when to descend

$$r_{open} = \frac{2B_{max}}{\sqrt{3}\theta} + B_{centre} \tag{2.3}$$

Here B_{max} and B_{centre} are the maximum distances to the centre of mass and centre of the cell, and θ is the traditional opening angle.

CUBEP3M is an particle-particle-particle-mesh N-body code that breaks the volume up into cubes, which are then assigned to processors. There are then a coarse and fine mesh into which the particles are divided. A restriction of this code which limited some of the experiments, is that it can only be run on an $\mathrm{integer}^3$ number of processors.

2.3 Results

After the runs were complete, the same analysis pipeline was used to evaluate each of the results. As all simulations were adapted to write out results in a common format, the analysis was the same for each. We used the power spectrum code provided by Robert Smith to produce independent analysis, using the same settings for each run. Density plots were produced by python scripts taking slices through the box of half depth and joining them up to get a double width in each case, with clearer depth resolution.

2.3.1 Gadget only analysis

To first get an overall feel of how the simulation evolved over time, we took a single run of GADGET3 with the default set of parameters – listed in Table 2.2, and examined how the matter-matter power spectrum evolved over time. A range of snapshots were analysed as shown in Figure 2.1, to show that the initial power distribution which had a local minimum to start with at around $k = 10.5$ evened out over time with transfer of power from larger scales to smaller scales as the simulation evolved. By a redshift of

Table 2.2: The default parameters used for the base runs, and subsequently altered across runs.

Parameter	Value
Softening	0.01
SofteningMaxPhys	0.01
ErrTolIntAccuracy	0.025
MaxSizeTimestep	0.025
MinSizeTimestep	1e-7
MaxRMSDisplacementFac	0.25
ErrTolTheta	0.5
ErrTolForceAcc	0.005

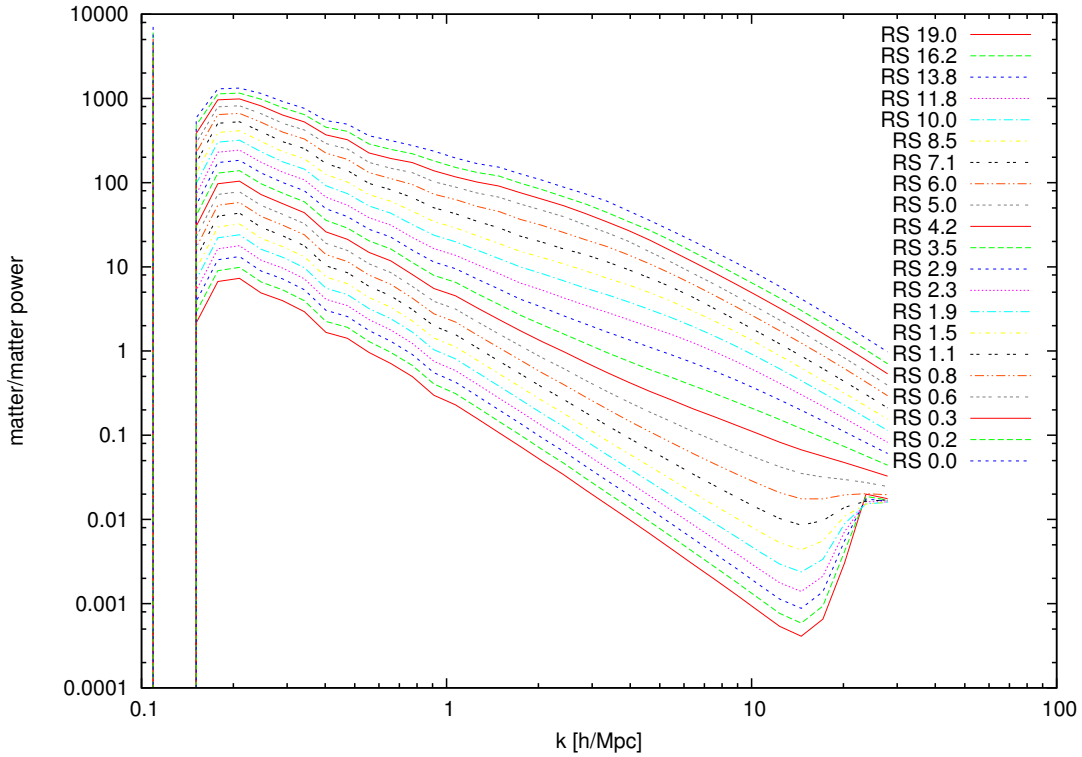

Figure 2.1: Power spectrum of GADGET3 snapshots evolving over redshift. This was run on the B1 initial conditions with 100 snapshots sampled at intervals here.

5-4, the curve becomes smooth and the transfer on the scales shown is complete.

The next procedure was to try running a single simulation engine on the same data but by varying some of the parameters. For this we started with GADGET3 and adjusted both the internal gridsize (PMGRID), and the softening length.

The results are shown in Figure 2.2, which shows numerous power spectra of GADGET3 with different grid sizes. The figure shows both the power spectrum, and the residual from the mean value of all spectra with a percentage scale. It is clear that on the biggest scales from $k = 0.1$ to $k = 1$ there is some small variation around the

mean. However as power moves to smaller scales even minor changes in the size of the grid makes for considerable change, especially if an accuracy of one percent is being aimed at. Furthermore, there is no immediately obvious correlation with the grid size. Whilst it is somewhat evident that smaller grids lead to slightly higher than average power, and bigger grids to smaller than average power, the largest sized grid size of 2048 hovers around the middle of the plot, diverging most at about a k value of 2.5 and then dropping at k values above 20. However the second largest grid size of 1024 deviates strongly at larger values of k, and exceeds the two percent boundary at a value of 10.

These results can be an issue, as in most cases the grid size of the GADGET3 simulation is chosen as the best trade off for performance for a given simulation size. Typically a larger grid is picked as the simulation size is increased, otherwise the performance advantages of the grid based approach are lost.

Next a GADGET3 run was done keeping the grid size fixed (PMGRID=512) and instead the softening length changing. The normal default softening length for such a simulation is 0.1, and here we tried a range of values from 0.001 to 0.04. The results are shown in Figure 2.3.

Again whilst on large scales there is a good agreement, there is a large divergence as values $k > 1$ are probed. The default value of 0.1 is somewhere close to the mean level, which is encouraging. There is also a small trend that smaller values of softening give higher values of power at high k and larger values a smaller value of power.

However, the two smallest values of softening at 0.00009 and 0.0001 seem to give a lower power at bigger k, but not one that is wildly different to the default value, being almost the mirror image of the default value. Even a small change from 0.0001 to 0.0002 up to 0.0005 seems to give extremely large values of power at higher values of k.

2.3.2 Gadget, Ramses, PkdGrav and cubep3m compared

Next some runs of the RAMSES, PKDGRAV, and CUBEP3M codes were attempted, and compared with the GADGET3 results. Figure 2.4 shows a comparative density plot of

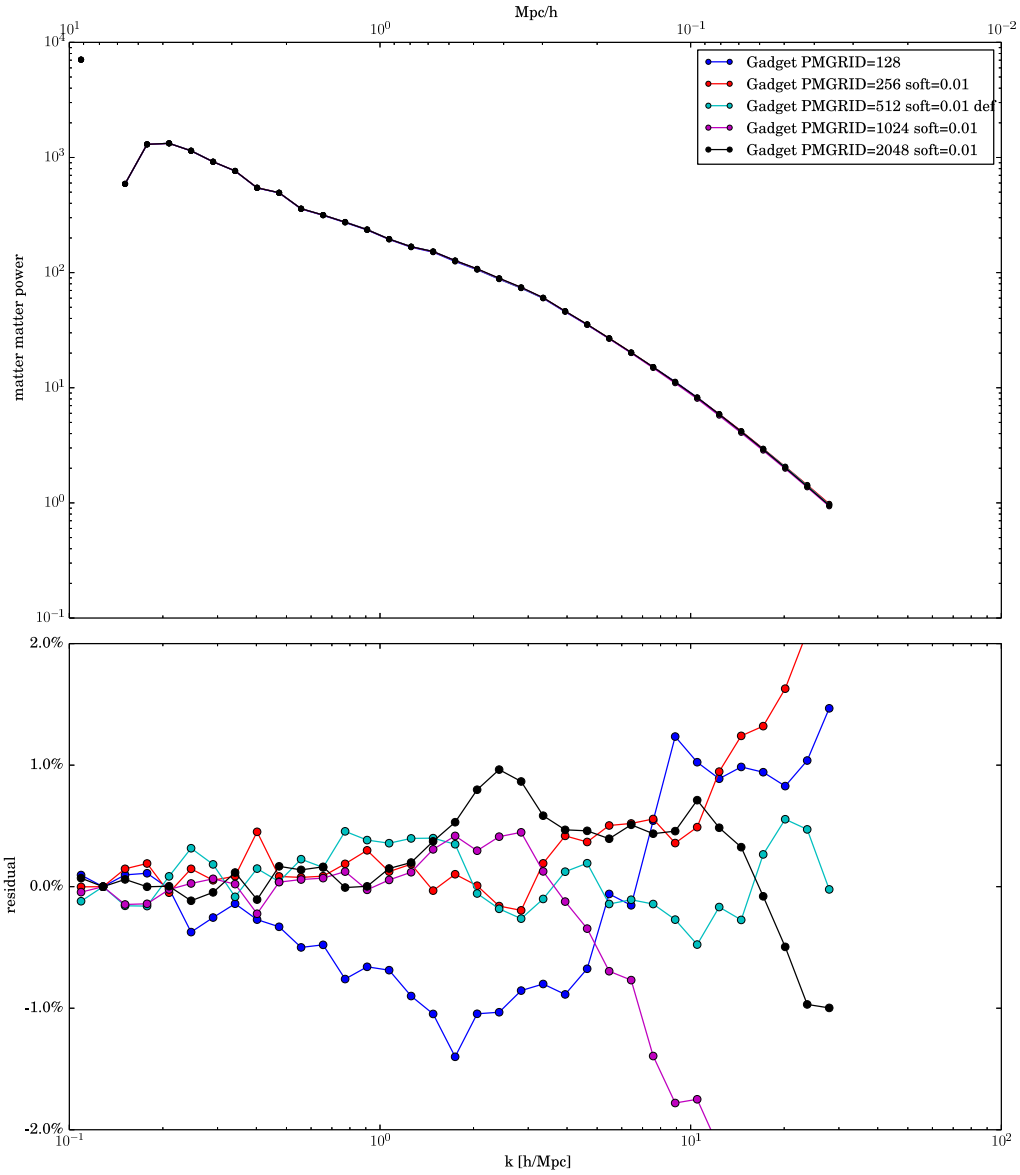

Figure 2.2: Power spectrum of GADGET3 variation with respect to changing the underlying grid size. The top panel shows the full spectrum, and the bottom panel the variation of the spectra as compared to an average of all spectra.

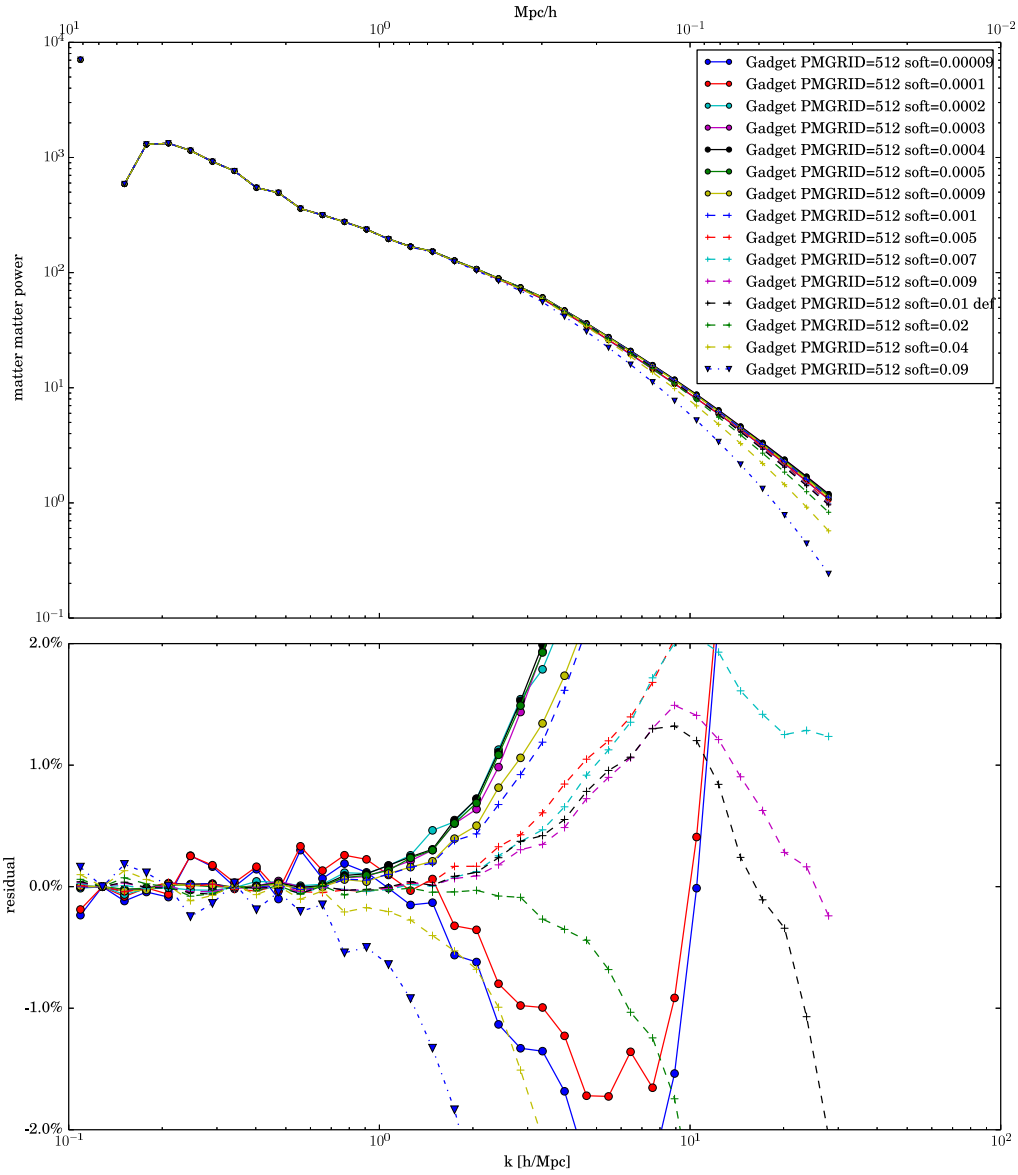

Figure 2.3: Power spectrum of GADGET3 variation with respect to changing the softening length but keeping all other parameters the same at red shift zero. The line marked "def" are the default parameters from 2.2.

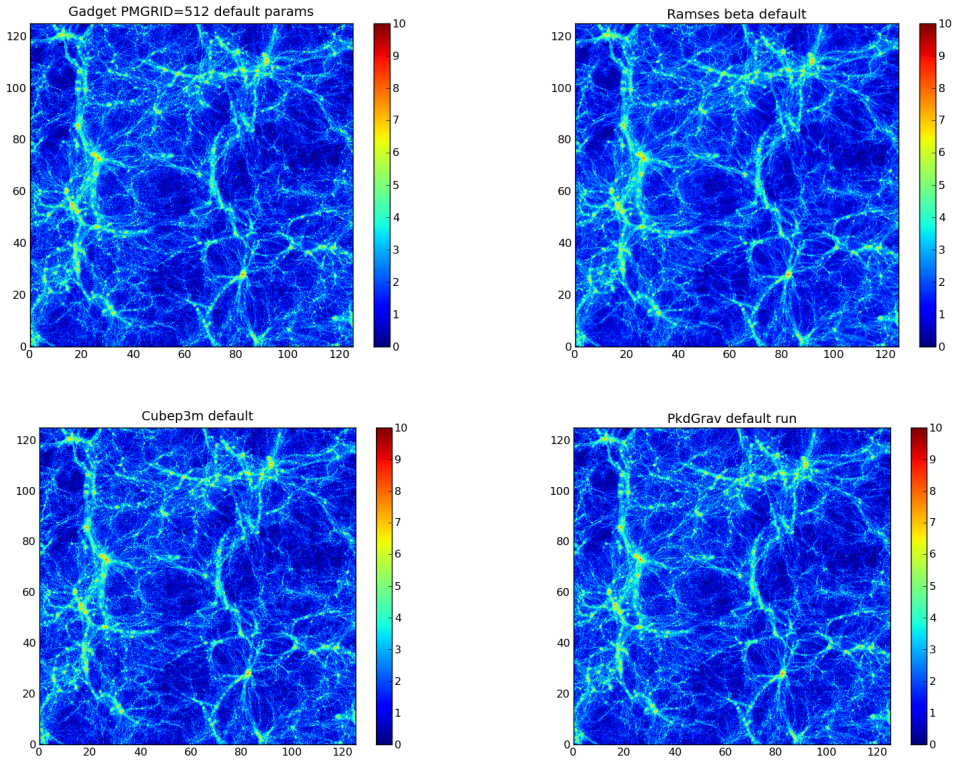

Figure 2.4: Density plots of GADGET3, RAMSES, CUBEP3M and PKDGRAV from the full box. These are half box slices wrapped to join up in two dimensions so as to spread out the structure. The scales are in h^{-1}Mpc and the density is log scaled.

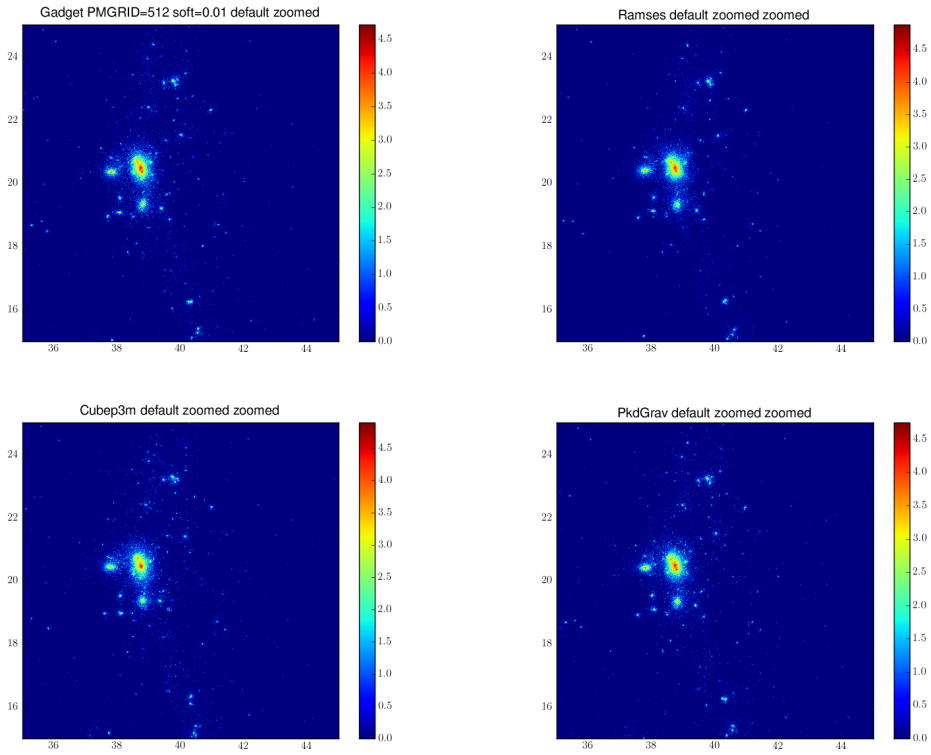

Figure 2.5: Density plots of GADGET3, RAMSES, CUBEP3M and PKDGRAV focusing on the same region.

the four simulations. Again the large scale structure is clearly very similar. The major clusters shown in red all occur in the same places and similarly the voids are shown to be coincident. Therefore the large scale structure is reflected very similarly amongst all the codes.

If instead we focus in on the small scale differences as can be seen in the zoomed plots shown in Figure 2.5, whilst again the general trends are in good agreement, the fine details differs slightly between the simulations. Some of the smaller clusters are less well defined, or have not fully formed in some cases. The density in some of the voids is less clear, and occasionally some of the really small clusters appear to be missing between plots.

2.3.3 Adaptive time steps compared to fixed

To examine in more detail a power spectrum was taken of each of the simulations to show in what can only faintly be seen in the density maps.

In particular it shows that the four agree reasonably well at largest scale and small k values. The PKDGRAV simulations shows the largest power on almost all scales, and is particularly obvious as it gets to larger values of k. RAMSES is mostly in the middle of the three, with GADGET3 producing the smallest power at large values of k.

To complete the comparison we finally compared fixed time steps with the more normal dynamic timesteps that the simulations use. In this case we ran the simulations with GADGET3, PKDGRAV and RAMSES, but unfortunately not for CUBEP3M. The runs were done with the default parameters, and left to complete. The CPU run times were somewhat increased with the timesteps fixed, but this allowed comparisons to be made.

All these codes use adaptive time steps, which attempt to speed up the simulation by taking as large a jump in time between steps as is reasonable. This may occasionally miss small changes, and lead to approximations, and this is shown in Figure 2.7 where three engines are compared when running with a fixed timestep contrasting to a variable timestep.

The effect of fixing the timestep for PKDGRAV is to reduce the power at larger k by a small amount. Otherwise the effect of a fixed time step show little noticeable

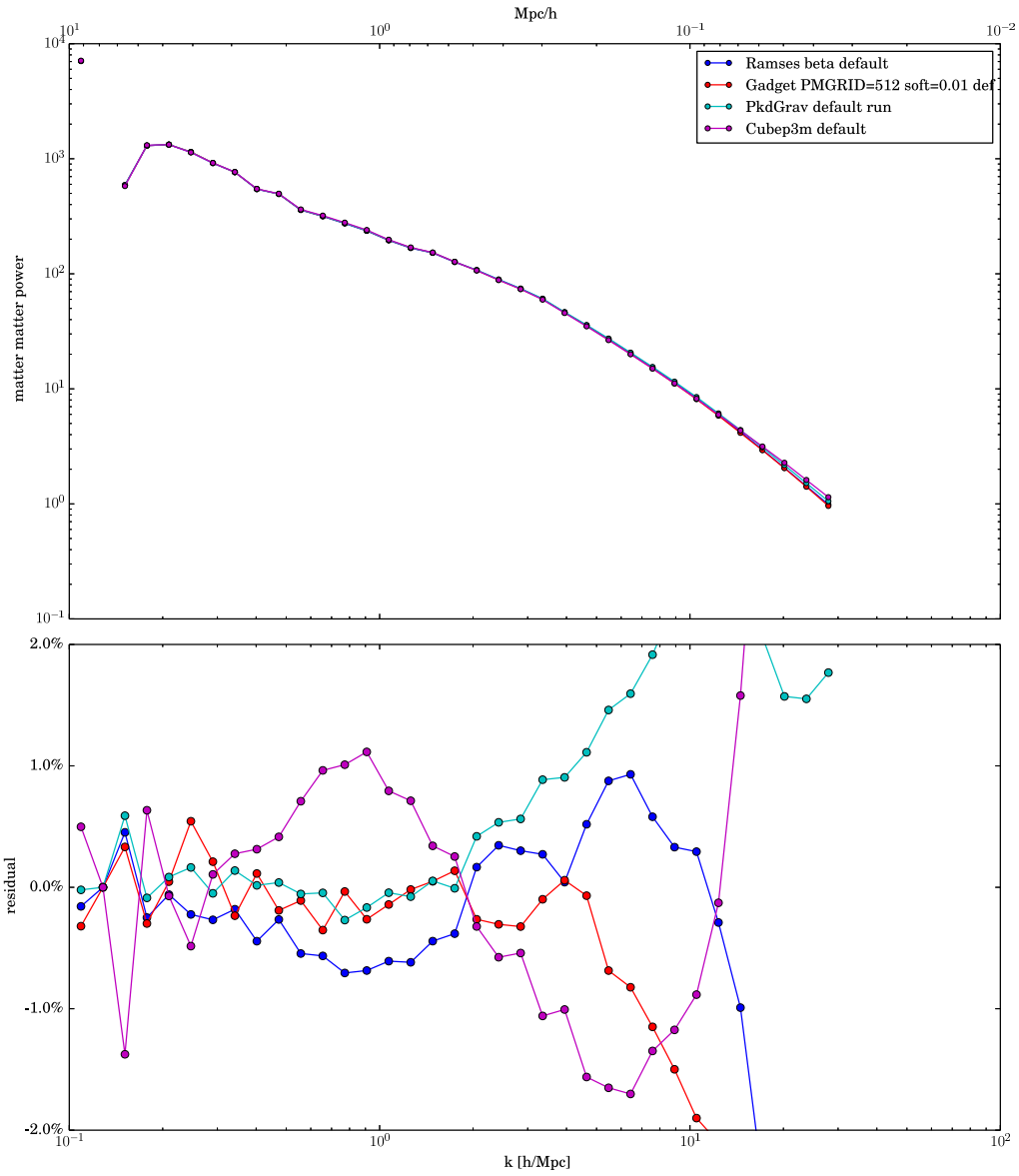

Figure 2.6: Power spectrum comparison of GADGET3, RAMSES PKDGRAV and CUBEP3M using adaptive time steps.

difference over what might be expected from earlier results. RAMSES shows a larger drop in power for the fixed time step case, and a wider difference between the two runs. When looking at the GADGET3 results, the two give very similar results up to a k value of about 5, and then start to diverge noticeably.

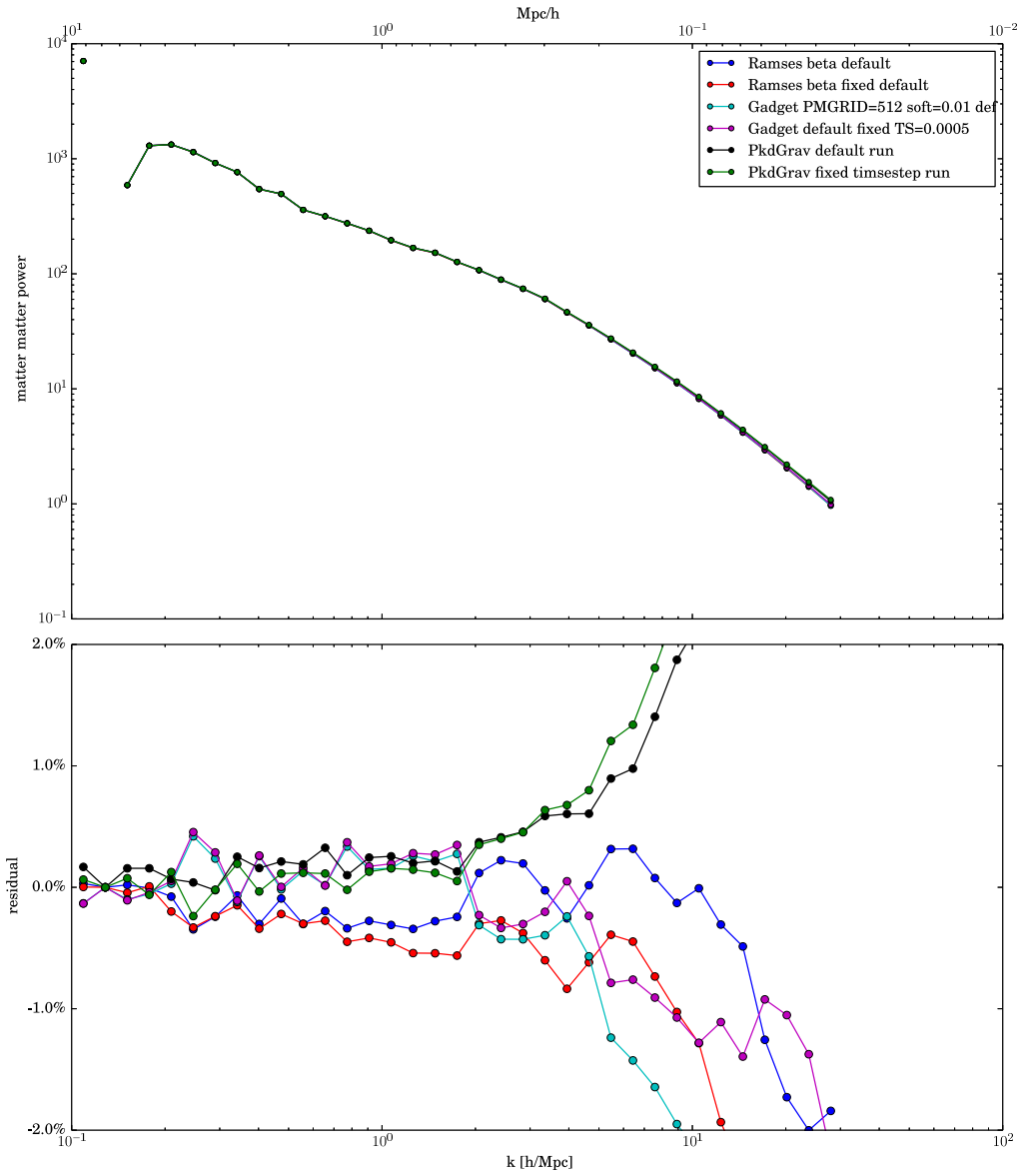

Figure 2.7: Power spectrum comparison of GADGET3, RAMSES and PKDGRAV using fixed time steps.

2.4 Summary and Conclusions

It seems clear from the results presented that attempts to reach one percent accuracy for simulations is still ambitious for small scale structures. Whilst agreement is good on large scale, and the images are almost indistinguishable on the grander scales, there is still much disagreement at smaller scales. In particular any changes to the parameters such as the softening level, the grid size or other parameters can have significant change when looking to a precision of a percentage point.

When comparing across simulation engines, the agreement is again good at large scales, but diverges at small scale. In some cases changing between fixed and dynamic timesteps also has an influence.

Of course it is not really clear which if any of these models are correct, when compared to the general small scale structure of the universe, but given they all use similar physics and dark matter it would seem important that they should all agree to a good level before going further.

These results are for dark matter only simulations, which are by far the easiest to achieve. They have none of the complex physics associated with baryons, such as radiative transfer, electromagnetic interactions and the issues of heating and cooling. Furthermore at smaller scales the impact of other features such as neutrinos may have effects, further distorting the picture.

So whilst there is still disagreement at the scales investigated here, it does seem that precision cosmology may be somewhat difficult for the current range of simulation engines when looking at small scales.

Chapter 3

SubHaloes going Notts: The SubHalo-Finder Comparison Project

3.1 Introduction

The growth of structure via a hierarchical series of mergers is now a well established paradigm (White & Rees, 1978). As larger structures grow they subsume small infalling objects. However the memory of the existence of these substructures is not immediately erased, either in the observable Universe (where thousands of individual galaxies within a galaxy cluster are obvious markers of this pre-existing structure) or within numerical models, first noted for the latter by Klypin *et al.* (1999a).

Knowing the properties of substructure created in cosmological N-body simulations allows the most direct comparison between these simulations and observations of the Universe. The fraction of material that remains undispersed and so survives as separate structures within larger haloes is an important quantity for both studies of dark-matter detection (Springel *et al.*, 2008b; Kuhlen, Diemand & Madau, 2008; Vogelsberger *et al.*, 2009; Zavala, Springel & Boylan-Kolchin, 2010) and the apparent over-abundance of substructure within numerical models when compared to observations (Klypin *et al.*, 1999b; Moore *et al.*, 1999). The mass and radial position of the most massive Milky Way satellites seem to raise new concerns for our standard ΛCDM cosmology (Boylan-Kolchin, Bullock & Kaplinghat, 2011b,a; di Cintio *et al.*, 2011;

Ferrero *et al.*, 2011), while differences between the simulated and observed internal density profiles of the satellites seems to have been reconciled by taking baryonic effects into account (e.g. Oh *et al.*, 2011; Pontzen & Governato, 2011). We are certain that between 5 per cent and 10 per cent of the material within simulated galactic sized haloes exists within bound substructures (e.g. Gao *et al.*, 2004; De Lucia *et al.*, 2004; Contini, De Lucia & Borgani, 2011) and a substantial part of the host halo has formed from disrupted subhalo material (e.g. Gill *et al.*, 2004; Knebe *et al.*, 2005; Warnick, Knebe & Power, 2008; Cooper *et al.*, 2010; Libeskind *et al.*, 2011).

Quantification of the amount of substructure (both observationally and in simulations of structure formation) is therefore an essential tool to what is nowadays referred to as "Near-Field Cosmology" (Freeman & Bland-Hawthorn, 2002) and attempts to do so in numerical models have followed two broad approaches: either a small number of individual haloes are simulated at exquisite resolution (e.g. Diemand *et al.*, 2008; Springel *et al.*, 2008a; Stadel *et al.*, 2009a, respectively the Via Lactea, Aquarius and GHalo projects) or a larger representative sample of the Universe is modelled in order to quantify halo-to-halo substructure variations (e.g. Angulo *et al.*, 2009; Boylan-Kolchin *et al.*, 2009; Klypin, Trujillo-Gomez & Primack, 2011, who used the Millennium simulation, Millennium II simulation and the Bolshoi simulation, respectively). As this chapter studies the convergence of halo-finders within a single halo we can add nothing to the topic of halo-to-halo substructure variations.

In a very comprehensive study that included 6 different haloes and 5 levels of resolution Springel *et al.* (2008a) utilised their substructure finder SUBFIND to detect around 300,000 substructures within the virial radius of their best resolved halo. They found that the number counts of substructures per logarithmic decade in mass falls with a power law index of at most 0.93, indicating that smaller substructures are progressively less dynamically important and that the central regions of the host dark matter halo are likely to be dominated by a diffuse dark matter component composed of hundreds of thousands of streams of tidally stripped material. Maciejewski *et al.* (2011) confirmed the existence and properties of this stripped material using a 6-dimensional phase-space finder HSF. A similar power law index was also found for the larger cosmological studies (Boylan-Kolchin *et al.*, 2009; Angulo *et al.*, 2009). For the Bolshoi

simulation Klypin, Trujillo-Gomez & Primack (2011) find results that are in agreement with their re-analysis of the Via Lactea II of Diemand *et al.* (2008) with an abundance of subhaloes falling as the cube of the subhalo rotation velocity. Rather than the present value of the maximum rotation velocity they prefer to use the value that the subhalo had when it first became a subhalo (i.e. on infall). This negates the effects of tidal stripping and harassment (the removing of particles in interactions) within the cluster environment but makes it difficult for us to directly compare as we have generally only used the final $z = 0$ snapshot for this comparison study.

In recent years there has not only been a number of different groups performing billion particle single-halo calculations, there has also been an explosion in the number of methods available for quantifying the size and location of the structures within such an N-body simulation (see, for instance, Fig.1 in Knebe *et al.*, 2011). In this chapter we extend the halo finder comparison study of Knebe *et al.* (2011) to examine how well these finders extract the properties of those haloes that survive the merging process and live within larger haloes. While this issue has already been addressed by Knebe *et al.* (2011) it was nevertheless only in an academic way where controlled set-ups of individual subhaloes placed into generic host haloes were studied; here we apply the comparison to a fully self-consistently formed dark matter halo extracted from a cosmological simulation. As the results of credible and reliable subhalo identification have such important implications across a wide range of astrophysics it is essential to ask how well the (sub-)halo finders perform at reliably extracting subhaloes. This still leaves open the question of how well different modern gravity solvers compare when performing the same simulation but at least we can hope to ascertain whether or not – given the same set of simulation data – the different finders will arrive at the same conclusions about the enclosed subhalo properties. We intend this chapter to form the first part of a series of comparisons. It primarily focuses on the most relevant subhalo properties, i.e. location, mass spectrum and the distribution of the value of the peak of the subhaloes' rotation curve.

In Section 3.2 we begin by summarising the eleven substructure finders that have participated in this study, focusing upon any elements that are of particular relevance for substructure finding. In Section 3.3 we introduce the Aquarius dataset that the de-

scribed finders analysed for this study. Both a qualitative and a quantitative comparison between the finders is contained in Section 3.4 which also contains a discussion of our results, before we summarise and conclude in Section 3.5.

3.2 The SubHalo Finders

In this section we present the (sub-)halo finders participating in the comparison project in alphabetical order. Please note that we primarily only provide references to the actual code description papers and not an exhaustive portrait of each finder as this would be far beyond the scope of this chapter. While the general mode of operation can be found elsewhere, we nevertheless focus here on the way each code collects and defines the set of particles belonging to a subhalo: as already mentioned before, those particle lists are subjected to a common post-processing pipeline and hence the retrieval of this list is the only relevant piece of information as far as the comparison in this particular chapter is concerned.

3.2.1 ADAPTAHOP (Tweed)

ADAPTAHOP is a full topological algorithm. The first stage consist in estimating a local density using a 20 particle SPH kernel. Particles are then sorted into groups around a local density maximum, and saddle points act as links between groups. All groups are first supposed to be one single entity, that we hierarchically divide into smaller groups, by using an increasing density threshold. Haloes are then defined as groups of groups linked by saddle points corresponding to densities higher than 80 times the mean DM density. By increasing the threshold we further detail the structure of the halo as a node structure tree. Where a node is either a local maxima, or a group of particles connecting higher level nodes. After using this bottom to top approach, the (sub)haloes are defined using a top to bottom approach, hierarchically regrouping nodes so that a sub(sub)halo has a smaller mass than its host (sub)halo. Each particle belongs to a single structure either a halo or a subhalo. The (sub)haloes' centres are defined as the position of its particles with the highest SPH density. We need to stress that no unbinding procedures are used in this algorithm, at the risk of overestimating the number/misidentification of

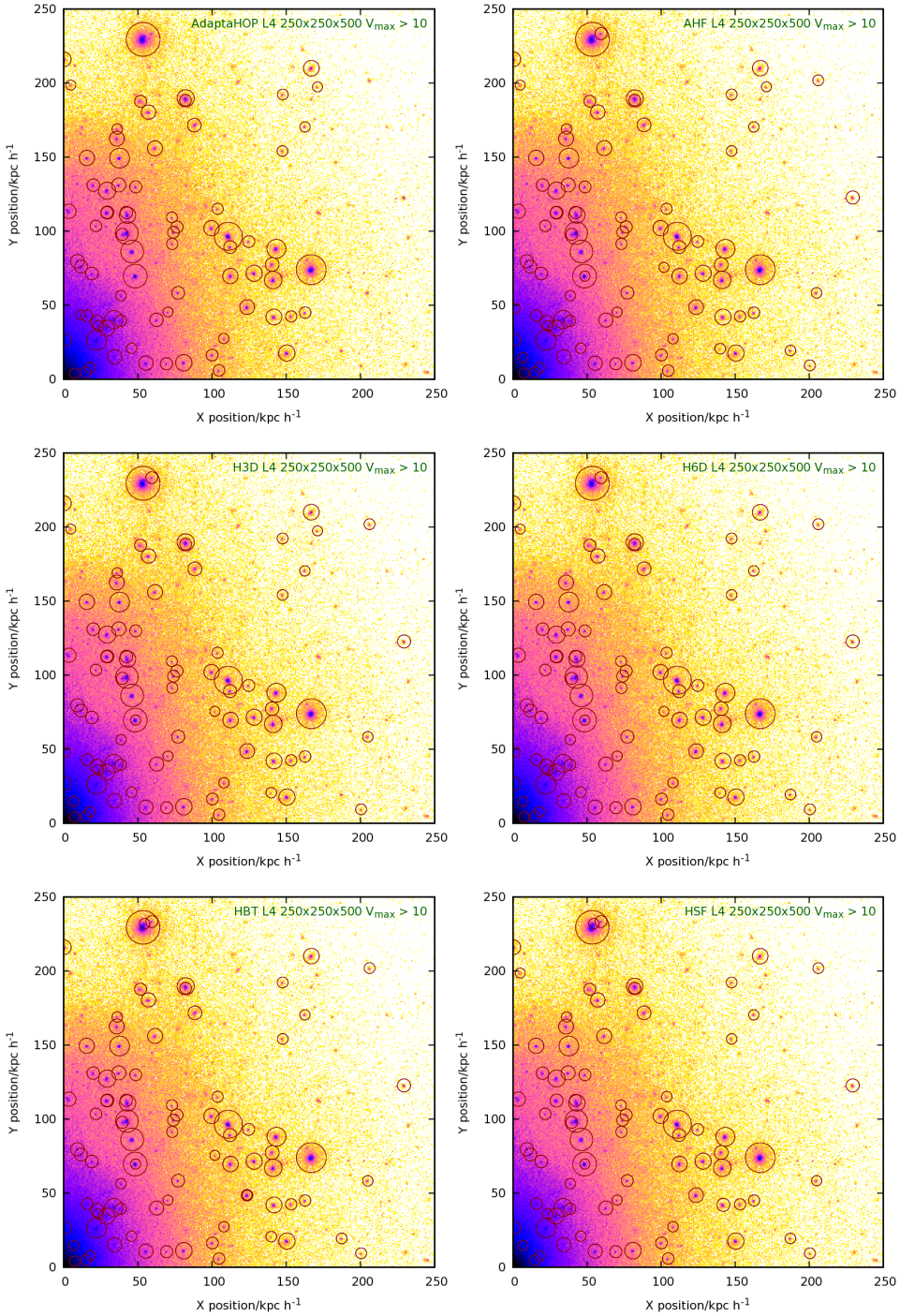

Figure 3.1: The images show the smoothed dark matter density within a quadrant at resolution level 4. In each panel the overplotted circles indicate the location of the recovered subhaloes for the finder labelled at the top of each panel. They are scaled proportionally using $v_{\max}$. Only subhaloes with a $v_{\max}$ greater than 10 km/s are shown.

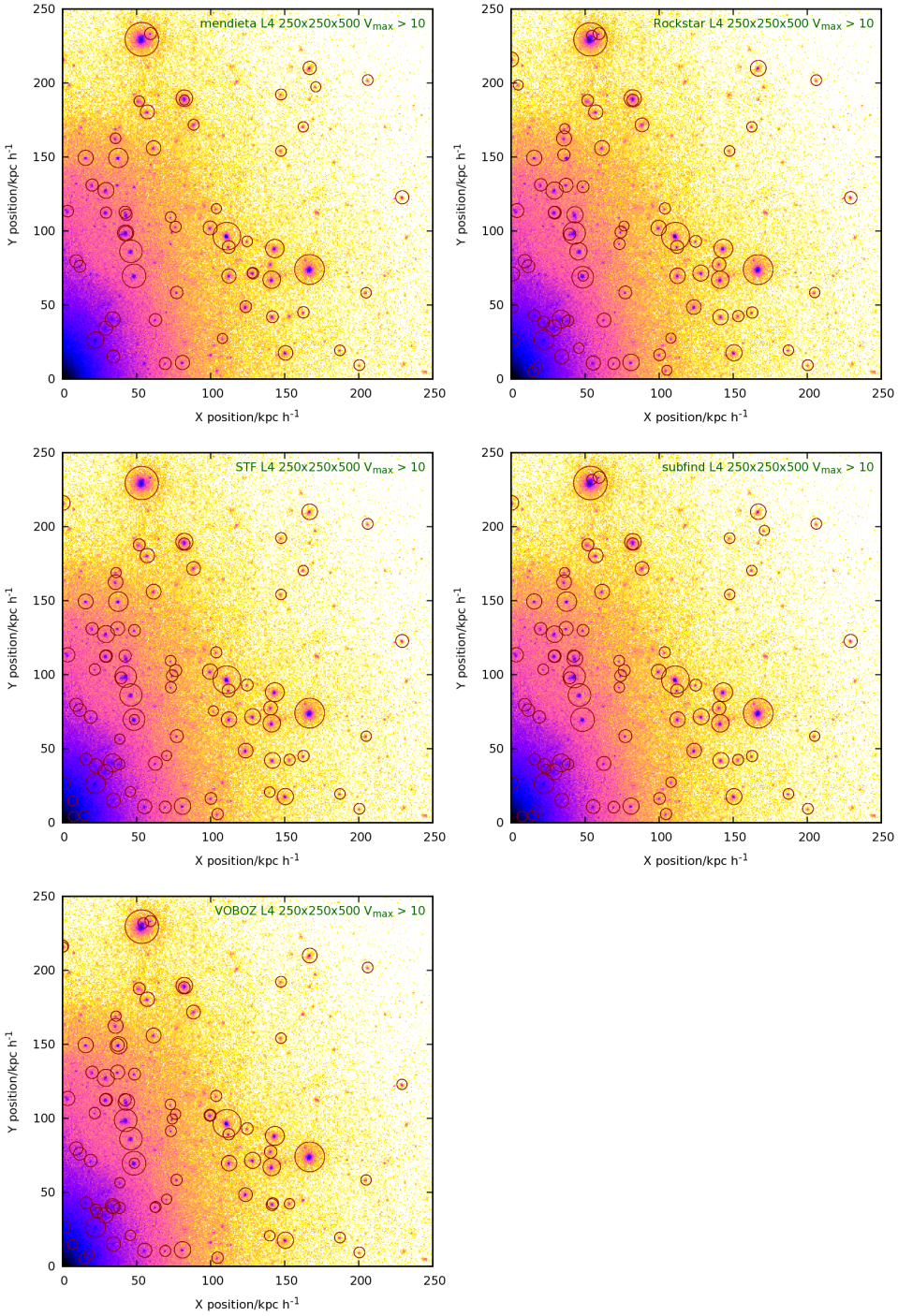

Figure 3.1: (continued) Recovered subhalo locations and $v_{\max}$ scale by labelled finder.

subhaloes with a low number of particles. The details of the algorithm are in Aubert, Pichon & Colombi (2004) and Tweed *et al.* (2009).

3.2.2 AHF (Knollmann & Knebe)

The halo finder AHF[1] (AMIGA Halo Finder) is a spherical overdensity finder that identifies (isolated and sub-)haloes as described in Gill, Knebe & Gibson (2004) as well as Knollmann & Knebe (2009). The initial particle lists are obtained by a rather elaborate scheme: for each subhalo the distance to its nearest more massive (sub-)halo is calculated and all particles within a sphere of radius half this distance are considered prospective subhalo constituents. This list is then pruned by an iterative unbinding procedure using the (fixed) subhalo centre as given by the local density peak determined from an adaptive mesh refinement hierarchy. For more details we refer the reader to aforementioned code description papers as well as the online documentation[1].

3.2.3 Hierarchical Bound-Tracing (HBT) (Han)

HBT (Han *et al.*, 2012) is a tracing algorithm working in the time domain of each subhaloes' evolution. Haloes are identified with a Friends-of-Friends algorithm and halo merger trees are constructed. HBT then traverses the halo merger trees from the earliest to the latest time and identifies a self-bound remnant for every halo at every snapshot after infall. Care has been taken to ensure that subhaloes are robustly traced over long periods. The merging hierarchy of progenitor haloes are recorded to efficiently allow satellite-satellite mergers or satellite accretion.[2]

3.2.4 HOT+FiEstAS (HOT3D & HOT6D) (Ascasibar)

HOT+FiEstAS is a general-purpose clustering analysis tool, still under development, that performs the unsupervised classification of a multidimensional data set by computing its Hierarchical Overdensity Tree (HOT), analogous to the Minimal Spanning

[1] AHF is freely available from http://www.popia.ft.uam.es/AMIGA

[2] It should be noted that HBT had access to the full snapshot data for Aquarius-A.

Tree (MST) in Euclidean spaces, based on the density field returned by the Field Estimator for Arbitrary Spaces (FiEstAS Ascasibar & Binney, 2005; Ascasibar, 2010). As explained in Knebe *et al.* (2011) in the context of halo finding, HOT+FiEstAS identifies objects with density maxima, either in configuration space (considering particle positions alone, HOT3D) or in the full, six-dimensional phase-space of particle positions and velocities (HOT6D). In both cases, the boundary of an object is always set by the isodensity contour crossing a saddle point, and its centre is defined as the density-weighted average of its constituent particles.

The main difference with respect to the version used in Knebe *et al.* (2011) is that there is now a post-processing stage, akin to a 'hard' expectation-maximization that is specifically tailored to the problem of halo finding, where:

1. $r_{\rm max}$ and $v_{\rm max}$ are computed for every object in the catalog.
2. Objects with more than 10 particles within $r_{\rm max}$ are labelled as (sub)-halo candidates.
3. Particles are assigned to the candidate that contributes most to the phase-space density at their location, approximating each candidate by a Hernquist (1990) sphere with the appropriate values of $r_{\rm max}$ and $v_{\rm max}$.

Candidates are only kept if they contain more than 5 particles within $r_{\rm max}$ *and* the density within that radius is higher than 100 times the critical density. Although a detailed discussion is obviously beyond the scope of this work, it is interesting to comment that some of the objects discarded by the latter criterion seem to be numerical artefacts, but others are clearly associated with filaments, streams, and other loose – yet physical, sometimes even gravitationally bound – structures. Since they are certainly not individual dark matter (sub)-haloes, they can be simply discarded for our present purposes.

3.2.5 Hierarchical Structure Finder (HSF) (Maciejewski)

The Hierarchical Structure Finder (HSF) identifies objects as connected self-bound particle sets above some density threshold. This method has two steps. Each particle

is first linked to a local dark matter phase-space density maximum by following the gradient of a particle-based estimate of the underlying dark matter phase-space density field. The particle set attached to a given maximum defines a candidate structure. In a second step, particles which are gravitationally unbound to the structure are discarded until a fully self-bound final object is obtained. For more details see Maciejewski *et al.* (2009).

3.2.6 MENDIETA (Sgró, Ruiz & Merchán)

The MENDIETA finder is a Friends-of-Friends based finder that is used to obtain a dark matter halo. This prospective host halo is subsequently refined by looking at peaks of increasing density by reducing the linking length. This approach decomposes the halo into its substructure plus other minor overdensities. In a final pass unbound particles are removed by checking their associated energies. MENDIETA is described more fully in Sgr, Ruiz & Merchn (2010).

3.2.7 ROCKSTAR (Behroozi)

ROCKSTAR (Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement) is a phase-space halo finder designed to maximize halo consistency across timesteps (Behroozi, Wechsler & Wu, 2011). The algorithm first selects particle groups with a 3D Friends-of-Friends variant with a very large linking length ($b = 0.28$). For each main FOF group, Rockstar builds a hierarchy of FOF subgroups in phase-space by progressively and adaptively reducing the linking length, so that a tunable fraction (70%, for this analysis) of particles are captured at each subgroup as compared to the immediate parent group. The metric for combining two particles in phase space, is to add together their differences divided by the FOF position and velocity dispersion respectively. When this is complete, Rockstar converts FOF subgroups into seed haloes beginning at the deepest level of the hierarchy. If a particular group has multiple subgroups, then particles are assigned to the subgroups seed haloes based on their phase-space proximity. This process is repeated at all levels of the hierarchy until all particles in the base FOF group have been assigned to haloes. Unbinding is performed

using the full particle potentials; halo centres and velocities are calculated in a small region close to the phase-space density maximum.

3.2.8 STF (Elahi)

The STructure Finder Hierarchical Structure Finder (Elahi, Thacker & Widrow, 2011, (STF)) identifies objects by utilizing the fact that dynamically distinct substructures in a halo will have a *local* velocity distribution that differs significantly from the mean, *i.e.* smooth background halo. This method consists of two main steps, identifying particles that appear dynamically distinct and linking this outlier population using a Friends-of-Friends-like approach. Since this approach is capable of not only finding subhaloes, but tidal streams surrounding subhaloes as well as tidal streams from completely disrupted subhaloes, we also ensure that a group is self-bound. Particles which are gravitationally unbound to a candidate subhalo are discarded until a fully self-bound final object is obtained or the object consists of fewer than 20 particles, at which point the group is removed entirely.

3.2.9 Subfind (Springel)

SUBFIND identifies substructures as locally overdense, gravitationally bound groups of particles. Starting with a halo identified through the Friends-of-Friends algorithm, a local density is estimated for each particle with adaptive kernel estimation using a prescribed number of smoothing neighbours. Starting from isolated density peaks, additional particles are added in sequence of decreasing density. Whenever a saddle point in the global density field is reached that connects two disjoint overdense regions, the smaller structure is treated as a substructure candidate, followed by merging the two regions. All substructure candidates are subjected to an iterative unbinding procedure with a tree-based calculation of the potential. The SUBFIND algorithm is discussed in full in Springel *et al.* (2001).

3.2.10 VOBOZ (Neyrinck)

VOBOZ (Neyrinck, Gnedin & Hamilton, 2005) was developed to have little dependence on free parameters. Density peaks are found using a Voronoi tessellation, which gives an adaptive, parameter-free estimate of each particle's density and set of neighbours. Each particle is joined to the peak that lies up the steepest density gradient from that particle. A halo associated with a high density peak (which is defined as the VOBOZ centre of the halo) will typically contain smaller density peaks. The significance of a halo is judged according to the ratio of its central density to a saddle point joining the halo to a halo with a higher central density, comparing to a Poisson point process. For this project, we impose a 4-σ significance threshold on subhaloes. Particles not gravitationally bound to each halo are iteratively removed, by comparing their potential energies (measured as sums over all other particles) to kinetic energies with respect to the velocity centroid of the halo's core (i.e. the particles that directly jump up density gradients to the peak). In the unbinding process, the least-bound particles are removed first; for each halo, the boundedness threshold reduces by a factor of $\sqrt{2}$ at each iteration, until it reaches its true value.

3.3 The Data

The simulation data used for this chapter forms part of the Aquarius project (Springel *et al.*, 2008a). It consists of multiple dark matter only re-simulations of a Milky Way like halo at a variety of resolutions performed using GADGET3 (based on GADGET2, Springel, 2005). We have used the Aquarius-A halo dataset at $z = 0$ for this project. This provides 5 levels of resolution, varying in complexity from the 2.3 million particles of the lowest resolution (i.e. level 5), up to the 4.25 billion particles of the highest resolution (i.e. level 1), as shown in Table 3.1. The underlying cosmology for the Aquarius simulations is the same as that used for the Millennium simulation (Springel *et al.*, 2005) i.e. $\Omega_M = 0.25, \Omega_\Lambda = 0.75, \sigma_8 = 0.9, n_s = 1, h = 0.73$. These parameters are consistent with the latest WMAP data (Jarosik *et al.*, 2011) although σ_8 is a little high. All the simulations were started at an initial redshift of 127. Precise details of the set-up and performance of these models can be found in Springel *et al.* (2008a).

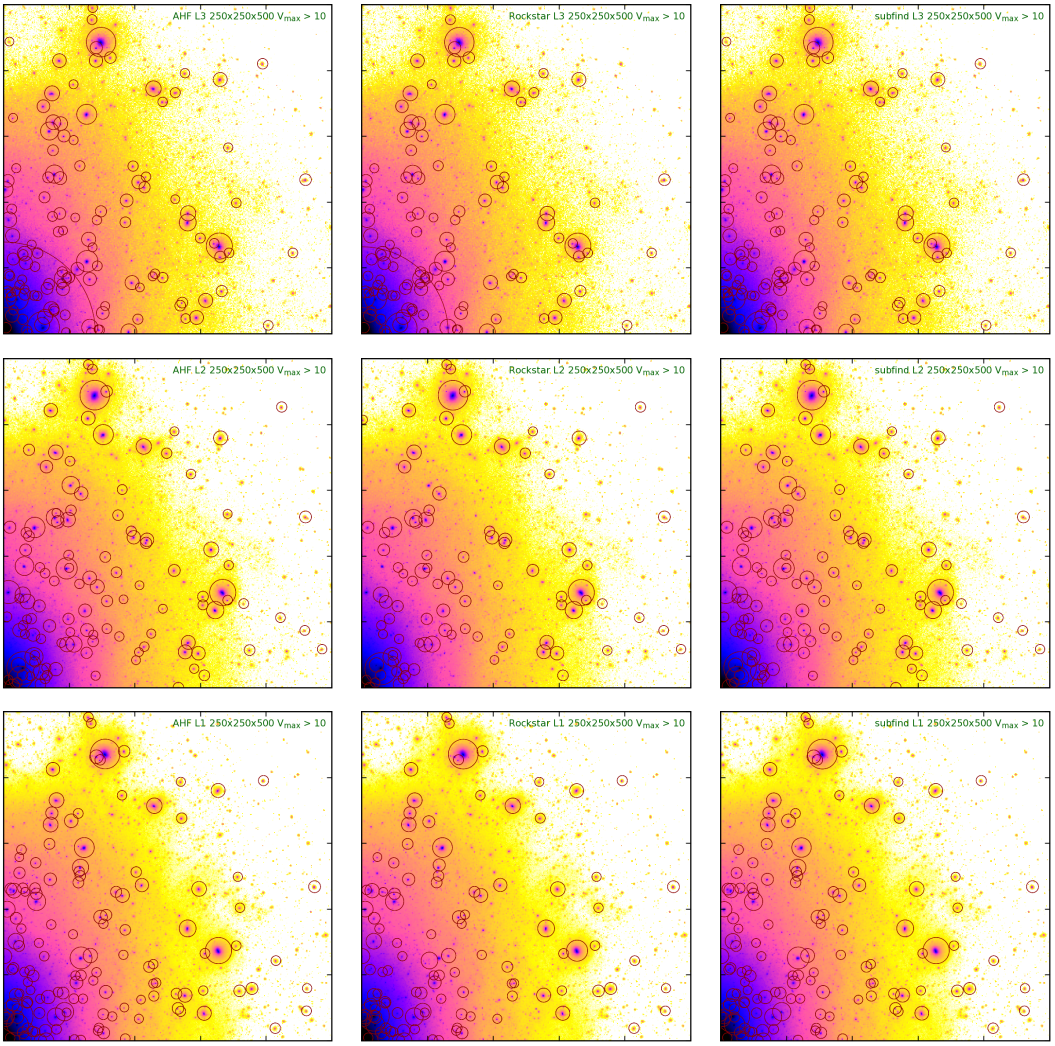

Figure 3.2: Subhalo recovery as a function of resolution. Location and size of recovered substructure from level 3 to level 1 for the three finders that reached this level. In all panels subhaloes with $v_{\rm max}>$ 10 km/s are shown, scaled by $v_{\rm max}$ as in Figure 3.1 and the background image is the smoothed dark matter density at that level. The relevant finder and level are labelled in the top right of each panel. The biggest change between levels is the additional small scale power moving the substructure locations. It shows good level of agreement in location and size, with just the occasional mismatch.

The participants were asked to run their subhalo finders on the supplied data and to return a catalogue listing the substructures they found. Specifically they were asked to return a list of uniquely identified substructures together with a list of all particles associated with each subhalo.

Finders were initially run on the smallest dataset, the Aq-A-5 data. This allowed for debugging of the common output format required by the project and some basic checks on the internal consistency of the data returned from each participant. Once this had been achieved each participant scaled up to the higher resolution datasets, continuing

Table 3.1: Summary of key numbers for each Aquarius level, the dataset used for this study. N_h is the number of particles with the highest resolution (lowest individual mass). N_l the number of low resolution particles - the sum of the remainder. N_{250} is the number of high resolution particles found within a sphere of radius 250 kpc/h from the fiducial centre at each resolution (*i.e.* those of interest for this study). M_p is the mass of one of these particles (in $\mathrm{M}_\odot/\mathrm{h}$). S is the resolution increase (mass decrease) for each level relative to level 5, and S_p is the resolution increase relative to the previous level. All particles are dark matter particles.

Data	N_h	N_l	N_{250}	M_p	S	S_p
Aq-A-5	2,316,893	634,793	712,232	2.294×10^6	1	$\times 1$
Aq-A-4	18,535,972	634,793	5,715,467	2.868×10^5	8	$\times 8$
Aq-A-3	148,285,000	20,035,279	45,150,166	3.585×10^4	64	$\times 8$
Aq-A-2	531,570,000	75,296,170	162,527,280	1.000×10^4	229	$\times 3.6$
Aq-A-1	4,252,607,000	144,979,154	1,306,256,871	1.250×10^3	1835	$\times 8$

until they reached the limits of their finder and/or the computing resources readily available to them. A summary of the number of subhaloes found by each subhalo finder at the various levels is contained in Table 3.2 as well as the size of the largest subhalo at level 4. All of the finders that participated in this study completed the analysis of the level 4 dataset which is used for the main comparison that follows and contains around 6 million particles within the region considered, a sphere of radius 250 kpc/h around a fiducial centre[3]. Three of the finders (AHF, ROCKSTAR & SUBFIND) completed the analysis of the very computationally demanding level 1 dataset. In addition to these HBT and HSF completed level 2 which contains around 160 million particles within the region examined here.

Both the halo finder catalogues (alongside the particle ID lists) and our post-processing software (to be detailed below) are publically available from the web site http://popia.ft.uam.es/SubhaloesGoingNotts under the Tab "Data".

3.4 The Comparison

We are going to primarily focus on comparing the location of subhaloes (both visually and quantitatively), the mass spectrum, and the distribution of the peak value of the rotation curve. The comparison, however, is based solely upon the provided particle lists and not the subhalo catalogues as the latter are based upon each code's own defini-

[3]We adopted a fixed and unique position for the host halo of $x = 57060.4, y = 52618.6, z = 48704.8$ kpc/h independent of the resolution.

Table 3.2: The number of subhaloes containing 20 or more particles and centres within a sphere of radius 250kpc/h from the fiducial centre found by each finder after standardised post-processing (see Section 3.4.1). Three finders (AHF, ROCKSTAR & SUBFIND) returned results from the highest resolution (level 1) within the timescale of this project. Below this we list the number of particles contained within the largest subhalo after post-processing.

Number of subhaloes within 250kpc/h of the fiducial centre after post processing.						
Name	ADAPTAHOP	AHF	HBT	HOT3D	HOT6D	HSF
Aq-A-5	353	230	228	58	136	231
Aq-A-4	2497	1599	1544	1265	1075	1544
Aq-A-3	-	11213	11693	-	-	11240
Aq-A-2	-	38441	39703	-	-	35445
Aq-A-1	-	226802	-	-	-	-
Number of particles in the largest subhalo within 250kpc/h of the fiducial centre after post processing.						
Aq-A-4	49076	77225	66470	69307	61581	73167

Name	MENDIETA	ROCKSTAR	STF	SUBFIND	VOBOZ
Aq-A-5	207	272	205	214	257
Aq-A-4	1493	1707	1521	1433	1862
Aq-A-3	10948	11797	10250	10094	13343
Aq-A-2	-	38489	-	33135	-
Aq-A-1	-	235819	-	221229	-
Aq-A-4	48387	78565	56990	50114	54685

tions and means to determine aforementioned properties and hence possibly introducing "noise" into the comparison (cf. Knebe *et al.*, 2011). In order to achieve a fair comparison between the respective finders we produced a single analysis pipeline which we used to post-process the particle lists provided by each participating group. This ensured consistency across our sample while removing differences due to the adoption of different post-processing methodologies and the particular choice of threshold criteria. The comparison detailed in this chapter is restricted to this uniform post-processed dataset, and ignores the finder derived data that was also provided. We intend to explore differences due to different methodologies in a subsequent work. However, we stress at the outset that our particular chosen post-processing methodology is not intended to be unique nor do we put it forward as the *best* way of defining a subhalo. Rather we use a single methodology so that we can first answer the most fundamental question: if we agree on a single subhalo definition do the different finders agree on the most fundamental properties they recover? Perhaps surprisingly we will see that the answer to this question is broadly yes.

We did not consider in this chapter efficiency of processing, as to make a fair comparison the codes would need to run on comparable machines with a set amount of memory and processors. In this instance the finders were run with the resources that were available to each of the participants. Some indication of the capabilities of the respective finders may be deduced from Table 3.2.

3.4.1 Post-processing pipeline

Some finders (e.g. AHF) include the mass (and particles) of a subhalo within the encompassing host halo whereas others do not (e.g. SUBFIND), preferring each particle to only be associated with a single structure. Either of these approaches has its pros and cons. For instance, keeping the subhalo mass as part of the halo mass makes it straightforward to calculate the enclosed dynamical mass of any object. However, such an approach easily leads to multiple counting of mass, particularly if there are many layers of the substructure hierarchy. In principle though it is not difficult to transform from one definition to the other given knowledge of both the halo and particle locations.

In our study, 5 of the 11 finders chose to include the mass of subhaloes whereas the other 6 did not. Following our principle of creating a uniform analysis pipeline we processed all the particle lists to ensure that a particle could only reside within a single structure.

To this end, we first sorted the returned halo catalogue into mass order. Then starting from the smallest halo we performed the centring, trimming and overdensity checks detailed below to trim the subhalo uniformly. We then tagged the particles contained within this object as being within a subhalo before continuing to the next largest subhalo and repeating the procedure ignoring particles already tagged as being used before. This preserved the maximum depth of the subhalo hierarchy while ensuring that a particle could only reside within a single subhalo. We should remark that in practice excising all the sub-subhaloes from each subhalo's particle list made little difference to any of the results presented here as at any level of the subhalo hierarchy only around 10 per cent of the material is within a subhalo of the current halo. So sub-subhaloes contribute only around 1 per cent of the halo mass, although it can affect other properties such as the centre of mass.

All the particles belonging to the list each finder identified as being associated with a subhalo were extracted from the original simulation data files to retrieve each particle's position, velocity and mass. From this data the centre of mass was first calculated, before being refined based on consideration of only the innermost 50 per cent of these particles, sorted with respect to the initial centre of mass. This procedure was repeated until a stable centre was found, i.e. until the change in the position was below the actual force resolution of the simulation. Once the centre had been defined the particles were ordered radially from this point and a rotation curve $GM(< r)/r$ and overdensity $M(< r)/(4\pi r^3/3)$ calculated until it dropped below 200 times the critical density $\rho_{\rm crit}$ defining the subhalo radius R_{200} and mass M_{200}. All particles outside R_{200} were removed which was essential in particular for the phase-space finders who also considered already stripped material as still being part of and belonging to the subhalo. Please note that our post-processing pipeline does explicitly not feature an unbinding procedure as this already formed part of most halo finding algorithms. At this point the maximum circular velocity, $v_{\rm max}$ was obtained by smoothing the rotation curve and locating its maximum by searching both inwards and outwards for a peak in the rotation curve and taking the average of these two measures, a process that stabilises the measure if the rotation curve is very flat or noisy.

We emphasise that the precise subhalo properties are somewhat sensitive to the definition of the halo centre. Various groups use the centre-of-mass as the centre of all material enclosed within the subhalo's radius (both with and without including substructure), the centre-of-mass of some smaller subset (as here for example), the location of the most bound particle, the location of the densest particle or the minimum of the gravitational potential. Additionally different groups use different methodologies for deciding whether or not a particle is bound to a halo as this involves some decisions about the global potential and can be a very time consuming process if done fully generally and iteratively.

Finally the choice of where to place the subhalo edge is also problematic. By definition the subhalo resides within some in-homogeneous background density and so at some point particles cease to belong to it and should rather be associated with the background object. Different groups split the host halo from the subhalo in different ways and

there is no *correct* method. Without a uniform choice these differences can swamp any differences due to actually finding subhaloes or not. We stress that our post-processing (where we treat each subhalo in isolation) can only remove particles from the original list of those particles associated with a subhalo. We have therefore tested whether or not our results are sensitive to our choice of 200 as an overdensity parameter by re-running our analysis with a tighter threshold of 500. Other than making all the subhalo masses smaller this has no noticeable effect on the scatter of the cumulative number counts. We therefore decided to stick to the original choice of R_{200} and M_{200}, respectively. Further, throughout the subsequent comparison only haloes with more than 20 (bound) particles within R_{200}were used, although some finders detected and returned haloes with less particles.

To summarise, our uniform post-processing pipeline involved the following steps, applied iteratively where necessary:

- The subhalo catalogues were sorted into mass order.
- Starting from the smallest subhalo, the particles associated with the current subhalo were obtained from the simulation data.
- Only particles tagged as "not used before" were considered.
- The centre-of-mass was iteratively calculated using the innermost 50 per cent of particles. (Originally we used the innermost 10 per cent but found some of the more dispersed sub structures did not converge with this value).
- A value for R_{200} was calculated based on an enclosed overdensity of 200 times the critical density.
- The subhalo mass and rotation curve peak $v_{\rm max}$ were computed based on particles inside R_{200}.
- Only substructures containing more than 20 particles were retained.

3.4.2 Visual comparison

A visual representation of the location and size (based on $v_{\rm max}$) of the recovered subhaloes at Aquarius level 4 from each of the finders is shown in Figure 3.1. A smoothed colour image of the underlying dark matter density based on all particles from the original Aquarius data are shown in one quadrant of the main halo, and this is overplotted with the recovered subhaloes from each finder indicated by circles whose size is scaled according to $v_{\rm max}$ (specifically $v_{\rm max}$ (in km/s) divided by 3). This allows a visual comparison between the finders. Only haloes with $v_{\rm max} > 10$km/s are shown. We immediately see that most of the finders are very capable of extracting the locations of the obvious overdensities in the underlying dark matter field. Wherever you would expect to find a subhalo (given the background density map) one is indeed recovered. This demonstrates that substructure finders should be expected to work well, recovering the vast majority of the substructure visible to the eye. Additionally, if our aforementioned post-processing is applied the quantitative agreement between the finders is also excellent, with the extracted structures having very similar properties between finders (see below). The majority of the finders agree very well, reliably and consistently recovering nearly all the subhaloes with maximum circular velocities above our threshold.

While Figure 3.1 illustrates the agreement between the finders at a single Aquarius level (in this case level 4, which all the participating finders have completed), in Figure 3.2 we construct a similar Figure to illustrate the agreement between levels. We show the same quadrant at level 3 to level 1 for the three finders that have completed the level 1 analysis (i.e., AHF, ROCKSTAR & SUBFIND); we deliberately omitted level 5 and level 4 as the former is not very informative and the latter has already been presented in Figure 3.1. As can be seen, the main difference between the different levels is in the exact location of the substructures. This changes because additional power was added to the Aquarius initial power spectrum to produce the additional small objects that form as the resolution is increased (fundamentally, the Nyquist frequency has changed as there are more available tracers within the higher resolution box). This extra power moves the substructure around slightly, and these differences are amplified in the, by definition, non-linear region of a collapsed object. Despite this the ready agreement between the three finders at any single level is clear to see and this is sim-

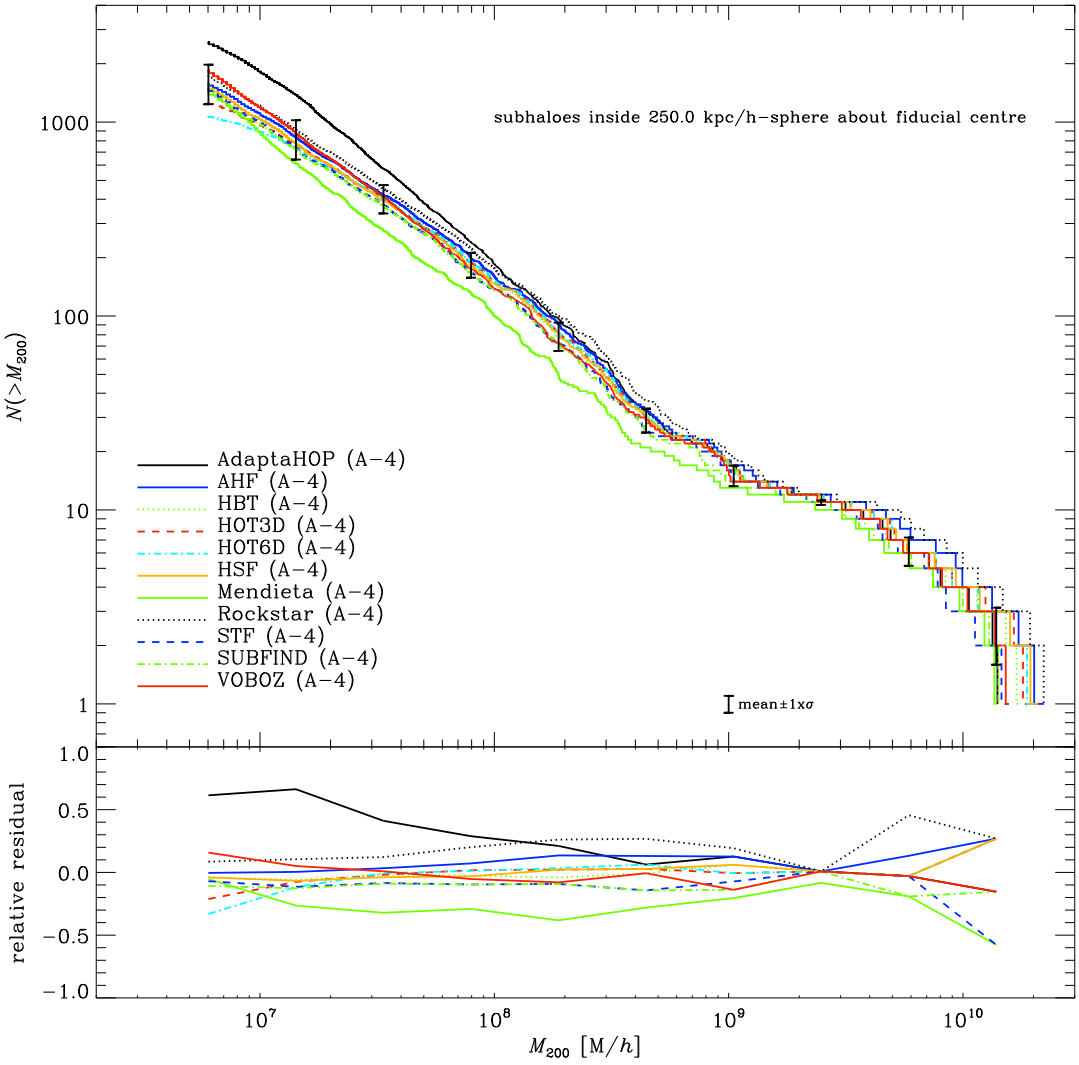

Figure 3.3: Cumulative number count of subhaloes above the indicated mass found (M_{200}) within a radius of 250 kpc/h from the fiducial halo centre after standardised post-processing at resolution level 4 (see Section 3.4.1 for details). The bottom plot shows the relative residual value offset from the mean of the cumulative mass curve.

ilarly true for both the other finders (HBT, HSF) that completed level 2. We do not explore the effect of changing the resolution on subhalo extraction in more detail here because that is not the main point of this chapter, which focuses on how well different finders extract substructure relative to each other. Also, this topic has already been well studied for SUBFIND using this same suite of models by (Springel *et al.*, 2008a).

3.4.3 Subhalo Mass Function

3.4.3.1 Level 4

Perhaps the most straightforward quantitative comparison is simply to count the number of subhaloes found above any given mass. For Aquarius level 4 this produces the cumulative mass plot (based on M_{200}) shown in Figure 3.3. Results from each participating finder are shown as a line of the indicated colour. Generally the agreement is good, with some intrinsic scatter and a couple of outliers (particularly ADAPTAHOP and MENDIETA) which do not appear to be working as well as the others, finding systematically too many or too few subhaloes of any given mass respectively. For ADAPTAHOP we like to remind the reader that this code does *not* feature a procedure where gravitationally unbound particles are removed; we therefore expect lower mass haloes stemming from Poisson noise in the background host halo to end up in the halo catalogue as well as haloes to have a higher mass in general possibly explaining the distinct behaviour of this code. But typically the scatter between codes is around the 10 per cent level except at the high mass end where it is larger as each finder systematically recovers larger or smaller masses in general. We like to remind the reader again that this scatter is neither due to the inclusion/exclusion of sub-subhaloes (which has been taken care of by our post-processing pipeline) nor the definition of the halo edge: as the 10 per cent differences still remain if choosing R_{500} as the subhalo edge.

Table 3.2 lists the number of subhaloes found that contain 20 or more particles after the uniform post-processing procedure detailed above had been performed and within 250 kpc/h of the fiducial centre of the main Aquarius halo at each level completed for all the eleven finders that participated. These number counts are generally remarkably consistent, again with a few outliers as expected from Figure 3.3. The majority of the finders are recovering the substructures remarkably well and consistent, respectively.

As an additional quantitative comparison we list the number of particles associated with the largest substructure found by each of the finders as the last row of Table 3.2. All the finders recover a structure containing $60,000$ particles ± 20 per cent. As Figure 3.3 has shown there is a lot of residual scatter for the highest mass haloes even when a uniform post-processing pipeline is used. This is most likely due to the differ-

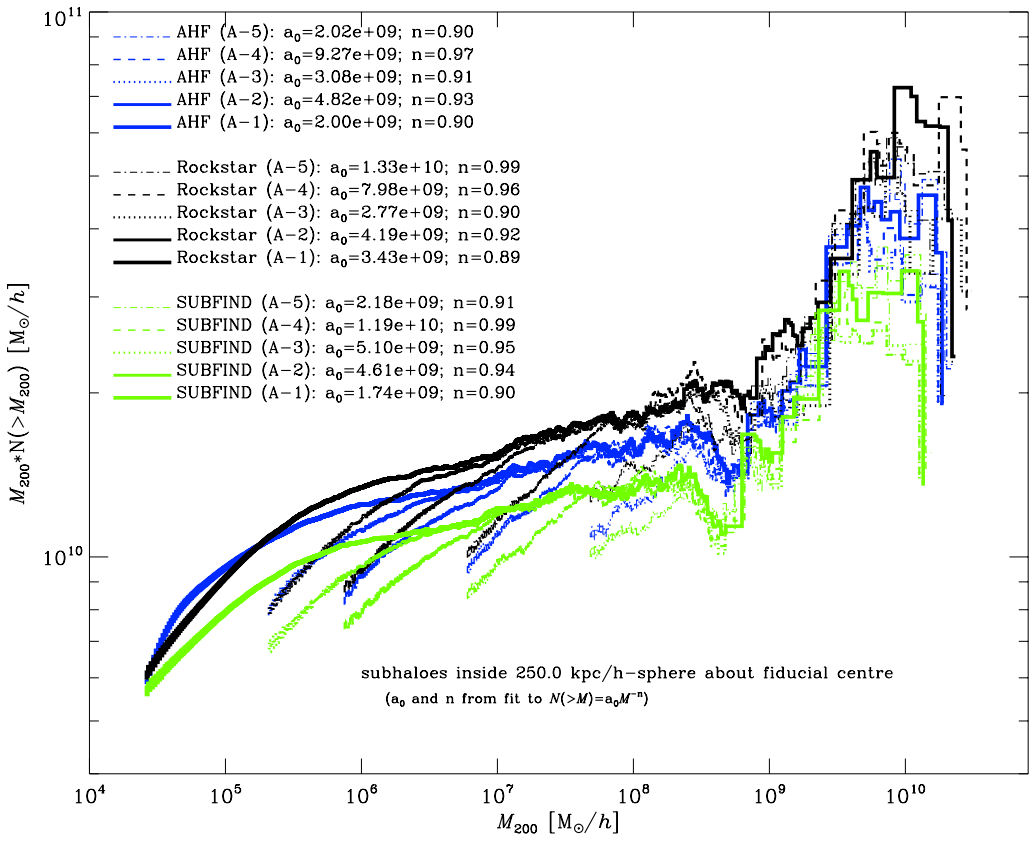

Figure 3.4: Cumulative subhalo mass function (multiplied by M to compress the vertical dynamical range) for all five Aquarius levels for the AHF, ROCKSTAR, and SUBFIND finder. We fit the function $N(>M)/N_{\rm tot} = a_0 \times M^{-n}$ between the mass equivalent to 100 particles at each level and $10^9 {\rm M}_\odot/{\rm h}$. Note: the data has *not* been shifted for clarity but is as plotted.

ent unbinding algorithms used in the initial creation of the substructure membership lists which are particularly uncertain for these large structures. At the other end of the substructure mass scale we have chosen to truncate our comparison at subhaloes containing 20 particles as this was shown to be the practical limit in Knebe *et al.* (2011). Some participants returned haloes smaller than this as this is their normal practice. They all stress that such small subhaloes should be treated with extreme caution but that there does appear to be a bound object at these locations even if its size is uncertain. We have removed them here for the purposes of a fair comparison.

Other plots that could be considered are those comparing the number of subhaloes against radial distance, or fractional mass against radial distance. Both these were produced and considered, but did not give any further insight to the comparison.

3.4.3.2 All Levels

Cumulative subhalo number counts like that shown for level 4 in Figure 3.3 can be calculated for all completed levels and compared. As shown in Figure 3.2 while increasing the resolution does not exactly reproduce the same substructures a reasonable approximation is achieved and so we expect to find a set of similar subhaloes containing more particles as we decrease the individual particle mass between levels (i.e. any specific subhalo should effectively be better resolved as the resolution increases). We show the cumulative number counts for the finders AHF, ROCKSTAR, and SUBFIND (multiplied by M to compensate for the large vertical scale) from level 5 to level 1 in Figure 3.4. We show this as an example and stress that similar plots with similar features could be produced for any of the finders that completed level 2. The curve for each level starts at 20 particles per halo and we like to stress that no artificial shifting has been applied: any differences seen in the plot are due to the different halo finding algorithms. Below about 100 particles per halo the cumulative number counts fall below the better resolved curves, indicating that subhaloes containing between 20 and 100 particles are not fully resolved and should have a slightly higher associated mass, also reported in Muldrew, Pearce & Power (2011). Above $10^9 \mathrm{M}_\odot/\mathrm{h}$ the power law slope breaks as there are less than 10 subhaloes more massive than this limit and the number of these is a property of this particular host halo. For these reasons we fit a power law of the form

$$\frac{N(>M)}{N_{\mathrm{tot}}} = a_0 M^{-n} \tag{3.1}$$

between 100 particles and $10^9 \mathrm{M}_\odot/\mathrm{h}$ where the power law breaks. Here a_0 is a normalisation (capturing the rise in the number of subhaloes due to the increase in resolution), M is the mass and n is the power law slope. The fitted values of the parameters by level are given in the legend for each finder. The subhalo cumulative number count appears to be an unbroken power law – at least in the range considered for the fitting. Similar results for SUBFIND were found by Springel *et al.* (2008a).

We extended this particular analysis of fitting a single power-law to the (cumulative) subhalo mass function to all finders at all available levels and compare the values of a_0 and n as a function of level for all participating substructure finders in Figure 3.5. There we find that at level 5 little can be said because the fitting range is so narrow. At

the lower, better resolved levels good agreement is seen between the finders (clearly ADAPTAHOP is a strong outlier on this plot, probably due to its lack of unbinding as mentioned before when discussing Figure 3.3, and HOT3D (as well as the first resolution step of MENDIETA) is inverted with respect to the main trend) and a consistent trend emerges: all agree that the power law slope, n is less than 1 and if anything decreasing with increasing resolution. Values of n less than 1 are significant because they imply that not all the mass is contained within substructures, with some material being part of the background halo. This has important ramifications for studies requiring the fraction of material within substructures such as the dark matter annihilation signal and lensing work. Although this result is robust between all high-resolution finders we remind the reader that this is for a single halo within a single cosmological model. However, it does indicate that, as perhaps expected, the most important contribution to substructure mass is from the most massive objects and that progressively smaller structures contribute less and less to the signal.

3.4.4 Distribution of $v_{\rm max}$

If, instead of quantifying the total mass of each subhalo, we rather use the maximum rotational velocity, $v_{\rm max}$ to rank order the subhaloes in size we obtain a generally much tighter relation (see below). Knebe *et al.* (2011) already found that $v_{\rm max}$ was a particularly good metric for comparing haloes and we confirm this for subhaloes. As Muldrew, Pearce & Power (2011) showed in figure 6 of their paper, this is because for an NFW profile (Navarro, Frenk & White, 1997) the maximum of the rotation curve is reached at less than 20 per cent of the virial radius for objects in this mass range so $v_{\rm max}$ is a property that depends upon only the very inner part of the subhalo and is not affected by any assumptions made about the outer edge. On the other hand, it has also been shown (Ascasibar & Gottlöber, 2008) that $v_{\rm max}$ provides a meaningful tracer of the depth of the gravitational potential (i.e. the mass scale) of the halo.

Figure 3.6 displays the cumulative $v_{\rm max}$ for all the finders for level 4 again. All the finders align incredibly well for the largest subhaloes with $v_{\rm max}>$ 20 km/s. For subhaloes smaller than this, the alignment remains tighter than the total mass comparison when extended down to rotation velocities of around 6 km/s. At level 4 haloes of this

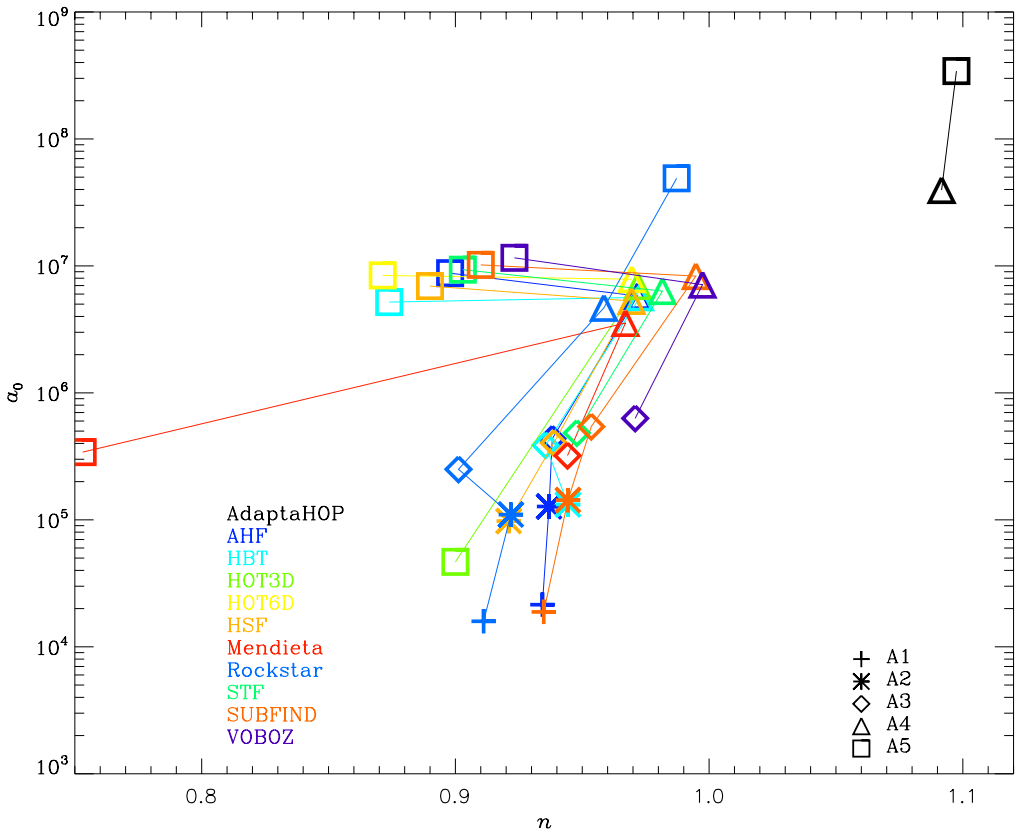

Figure 3.5: A comparison of the slope and normalisation of the fits of the mass function derived as per Figure 3.4 for all finders at all levels returned.

size contain around 80 particles in total, so $v_{\max}$ is being calculated from less than 20 particles at this point; the arrows give an indication of the number of particles inside $r_{\max}$. ADAPTAHOP, despite its missing unbinding procedure, agrees well with other finders for high rotation velocities as this particular statistic probes inner regions of the subhaloes which are less affected by unbound particles; and its deviation at the lower $v_{\max}$-end is due to the existence of (small mass) fluke objects not removed by such an unbinding step.

3.4.5 Radial Mass Distribution

The accumulated total mass of material within the subhaloes is measured by ordering the subhalo centres in radial distance from the fiducial centre of the halo and summing outwards,i.e. $\sum_{r_{\rm sat}<r} M_{\rm sat}$. We include all post-processed subhaloes above our mass threshold of 20 particles. As Figure 3.7 demonstrates at level 4 most of the finders (AHF, HBT, HOT6D, HSF, STF, VOBOZ) agree very well, finding very similar amounts

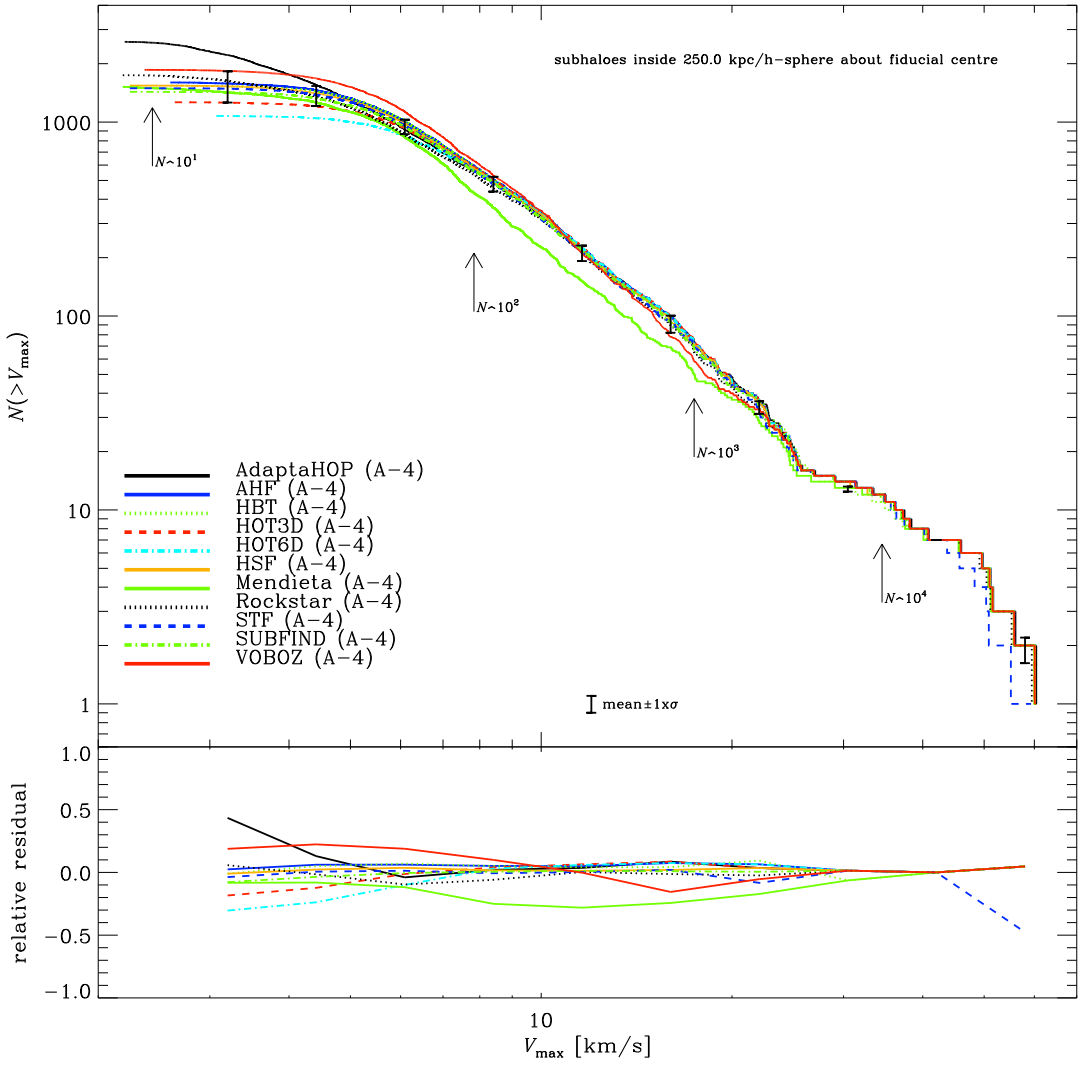

Figure 3.6: Cumulative number count of subhaloes above the indicated $v_{\rm max}$ value within a radius of 250 kpc/h from the fiducial halo centre after standardised post-processing (see Section 3.4.1). The arrows indicate the number of particles interior to $r_{\rm max}$, the position of the peak of the rotation curve. The bottom plot shows the relative offset from the mean of the cumulative count.

of substructure both in radial location and mass. ROCKSTAR finds a little more structure, particularly in the central region where its phase-space nature works to its advantage and SUBFIND finds around a factor of 25 per cent less due to its conservative subhalo mass assignment.

The MENDIETA finder appears to show significantly different results to the rest. As previously noted the ADAPTAHOP finder locates many small subhaloes and these push up the total mass found in substructure above that found by the others, particularly in the range around 50-100kpc. We note that two of the three phase-space based finders (HOT6D, & HSF) have a radial performance indistinguishable from real-spaced based

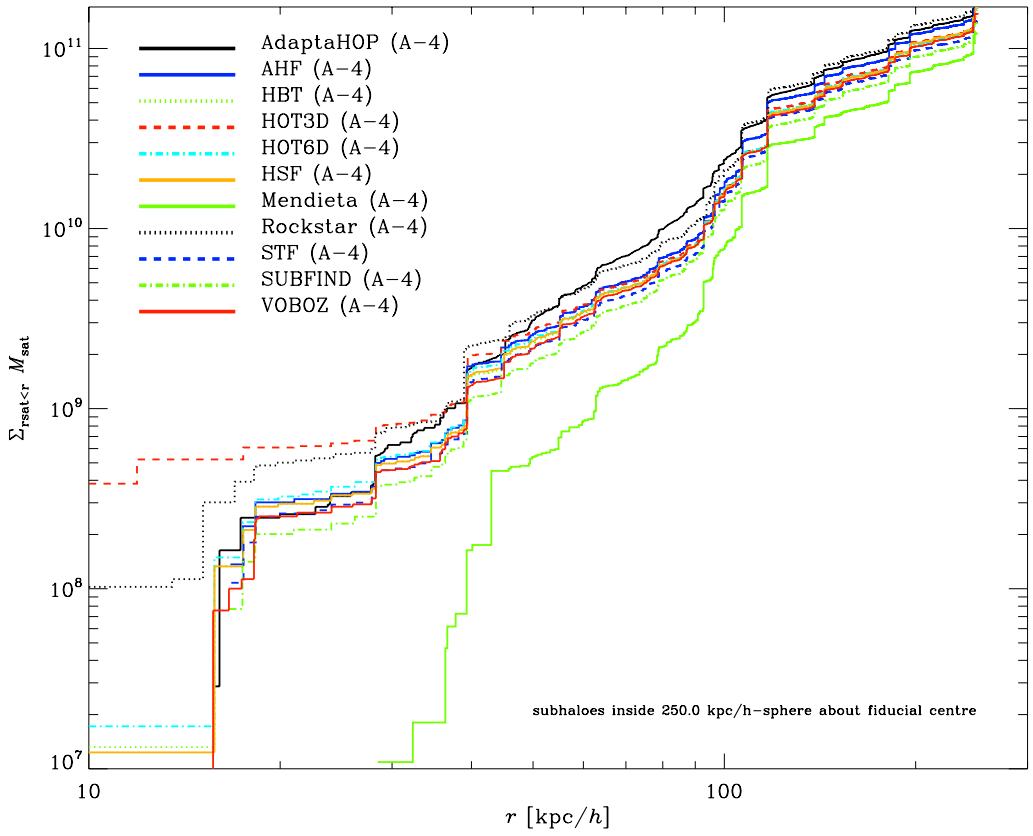

Figure 3.7: Cumulative plot of the enclosed mass within subhaloes as a function of radial distance from the fiducial centre of the host halo.

finders. The only one to show any difference is ROCKSTAR and it remains unclear whether or not this is in practice a significant improvement.

We further like to mention (though not explicitly shown here) that a visual comparison akin to Figure 3.1 but focusing on the central 20 h^{-1}kpc reveals that it is very likely that the excess mass found in that inner region by some of the finders such as HOT3D may be due to mis-identifications of the host halo as subhaloes. In the very central region it is difficult for the underlying real-space Friends-of-Friends methodology to distinguish structures from the background halo and so can show up either as multiple detections, or no structure at all.

3.5 Summary & Conclusions

We have used a suite of increasing resolution models of a single Milky Way sized halo extracted from a self-consistent cosmological simulation (i.e. the Aquarius suite

(Springel *et al.*, 2008a)) to study the accuracy of substructure recovery by a wide range of popular substructure finders. Each participating group analysed independently as many levels of the Aquarius-A dataset at redshift $z = 0$ as they could manage and returned lists of particles they associated with any subhalo they found. These lists were post-processed by a single uniform analysis pipeline. This pipeline employed a standard fixed definition of the subhalo centre and subhalo mass, and employed a standard methodology for deriving $v_{\rm max}$. This analysis was used to produce cumulative number counts of the subhaloes and examine how well each finder was able to locate substructure.

We find a remarkable agreement between the finders which are based on widely different algorithms and concepts. The finders agree very well on the presence and location of subhaloes and quantities that depend on this or the inner part of the halo are amazingly well and reliably recovered. We agree with Knebe *et al.* (2011) that $v_{\rm max}$ is a good parameter by which to rank order the haloes (in this case subhaloes). However, we also show that as $v_{\rm max}$ is only dependent upon the inner 20 per cent or less of the subhalo particles around 100 particles are required to be within the subhalo for this measure to be reliably recovered. Quantities that depend on the outer parts of the subhaloes, such as the total mass, are still recovered with a scatter of around 10 per cent but are more dependent upon the exact algorithm employed both for unbinding (intrinsic to each finder) and to define the outer edge (given by the common post-processing applied here).

The most difficult region within which to resolve substructures is the very centre of the halo which has, by definition, a very high background density. In this region real-space based finders are expected to struggle whereas the full six-dimensional phase-space based finders should do better. In practice ROCKSTAR is the only phase-space based finder that shows any indication of this (and this difference becomes *less* pronounced as the resolution is increased); but we cannot rule out mis-identifications of the host halo as subhaloes at this stage. We conclude that, as yet, none of the phase-space based finders present a significant improvement upon the best of the more traditional real-space based finders. Phase space finders are also often targeted at recovering streams and other more dynamical structure, which was not part of this comparison.

Convergence studies indicate that identified subhaloes containing less than 100 particles tend to be under-resolved and these objects grow slightly in mass if a higher resolution study is used. This could be due to the fact that particles in the outer regions of these subhaloes are stripped more readily at lower resolution or it could be an artefact of the difficulty of measuring the potential (and hence completing any unbinding satisfactorily) with this small number of particles. Several studies (Kase, Makino & Funato, 2007; Pilipenko, Doroshkevich & Gottlöber, 2009; Trenti *et al.*, 2010) have indicated the unreliability of halo properties (other than physical presence) for (sub)haloes of this size or less.

Fitting power law slopes to the convergence studies of each finder indicates that the logarithmic slope of the cumulative number count is less than 1. While this is only confirmed for a single halo within a single cosmology, and ignoring any mass in tidal streams, the result appears to be robust as it is found for all the high-resolution finders employed in this study. This indicates that the larger substructures are the most important ones and that higher levels of the (sub)subhalo hierarchy play a less significant dynamical role.

We like to close with a brief note on the removal of gravitationally unbound particles for subhaloes. We have seen that the omission of such a procedure most certainly leads to rather distinct results. However, we cannot convincingly deduce whether or not this will lead to more small mass objects (as is the case for ADAPTAHOP) or to objects more massive in general (also seen for ADAPTAHOP), likely both will occur. But we confirm that the exact differences between a subhalo catalogue based upon a halo finding method with and without unbinding depend on the actual algorithm to collect the initial set of particles to be considered part of the subhalo: we performed an analysis of the level 4 data with AHF switching the unbinding part off ending up with a subhalo mass function that was only different at the higher mass end (not shown here though) as opposed to the ADAPTAHOP results; but both these codes differ substantially in the way of assigning the primary particle set to a subhalo.

It should be noted that the Aquarius-A halo is a relatively quiescent halo (Wang *et al.*, 2011b), not having been subject to many mergers. Investigation of other haloes, and those produced by other simulation code would be interesting to compare. Therefore

more studies focusing on the actual halo catalogues returned by each finder (as opposed to the particle lists used here); other cosmological simulations and different simulated scenarios (such as disrupted galaxies); the detailed analysis of sub-substructure (which is only really practical at level 1); and other subhalo properties such as spin parameter and shape, as well as more detailed resolution studies for those codes providing an analysis of all levels will be deferred to future work.

Chapter 4

Spin across subhaloes

4.1 Introduction

Within the hierarchical galaxy formation model, dark matter haloes are thought to play the role of gravitational building blocks, within which baryonic diffuse matter collapses and becomes detectable (White & Rees, 1978; White & Frenk, 1991). Gravitational processes that determine the abundance, the internal structure and kinematics, and the formation paths of these dark haloes within the cosmological framework, can be simulated in great detail using N-body methods. However, the condensation of gas associated with these haloes, eventually leading to stars and galaxies we see today, is still at the frontier of present research efforts. A first exploration of the (cosmological) formation of disc galaxies was presented in Fall & Efstathiou (1980), where it was shown that galactic spin is linked to the surrounding larger scale structure (e.g. the parent halo). In particular, the general theory put forward by Fall & Efstathiou reproduces galactic discs with roughly the right sizes, if specific angular momentum is conserved, as baryons contract to form a disc (previously suggested by Mestel (1963)) and if baryons and dark matter initially share the same distribution of specific angular momentum.

While the theory has subsequently been refined, it always included (and still includes) such a coupling between the parent halo's angular momentum and the resulting galactic disc (cf. Dalcanton, Spergel & Summers (1997); Mo, Mao & White (1998); Navarro &

Steinmetz (2000); Abadi *et al.* (2003); Bett *et al.* (2010)). The origin of the halo's spin can now be understood in terms of tidal torque theory in which protohaloes gain angular momentum from the surrounding shear field (e.g., Peebles (1969); White (1984); Barnes & Efstathiou (1987)) as well as by the build-up of angular momentum through the cumulative transfer of angular momentum from subhalo accretion (Vitvitska *et al.*, 2002). Whichever way the halo gains its spin, it is a crucial ingredient for galaxy formation and all semi-analytical modelling of it (Kauffmann, White & Guiderdoni, 1993; Kauffmann, Nusser & Steinmetz, 1997; Frenk *et al.*, 1997; Cole *et al.*, 2000; Benson *et al.*, 2001; Croton *et al.*, 2006; De Lucia & Blaizot, 2007; Bower *et al.*, 2006; Bertone, De Lucia & Thomas, 2007; Font *et al.*, 2008; Benson, 2012).

A number of studies have been performed on the spin of haloes, in particular studies by Peebles (1969); Bullock *et al.* (2001); Hetznecker & Burkert (2006); Bett *et al.* (2007); Macci *et al.* (2007); Gottlöber & Yepes (2007); Knebe & Power (2008); Antonuccio-Delogu *et al.* (2010); Wang *et al.* (2011a); Trowland, Lewis & Bland-Hawthorn (2013); Lacerna & Padilla (2012); Bryan *et al.* (2012) but so far little has been done on subhaloes. These studies look at the spin of individual dark matter haloes found in cosmological simulations and generally do not focus on the substructure, or differences between substructure definition due to lack of resolution. Here we present a comparison of spin parameters across a number of detected subhaloes found by a variety of substructure finders. The finders use many different techniques to detect substructure within a larger host halo. This is a follow-up to a more general paper comparing the recovery of structure by different finders in Onions *et al.* (2012) included here as Chapter 3 and its predecessor Knebe *et al.* (2011).

The techniques studied here for finding substructures include real-space, phase-space, velocity-space finders, as well as finders employing a Voronoi tessellation, tracking haloes across time using snapshots, friends-of-friends techniques, and refined meshes as the starting point for locating substructure. With such a variety of mechanisms and algorithms, there is little chance of any systematic source of errors in the collection of substructure distorting the result. Subhaloes are particularly subject to distortion and evolution, more so than haloes because, by definition, they reside within a host halo with which they tidally interact. This can affect their structure and other parameters,

and in this case we are particularly interested in the spin properties. We quantify the spin with the parameter λ, a dimensionless quantity that characterises the spin properties of a halo and is explained in more detail in Section 4.2.

The rest of the paper is structured as follows. We first describe the methods used to quantify the spin of the halo in Section 4.2. The data we used is described in Section 4.3. Next we look at the overall properties of the spin in Section 4.4.1. Then we look at the correlation between the host halo and the subhaloes spin in subsection 4.4.2. Finally we look at how the spin is built up within the subhalo as a function of mass in subsection 4.4.3. We conclude in Section 4.5.

4.2 Method

4.2.1 Spin parameter

The dimensionless spin parameter gives an indication of how much a gravitationally bound collection of particles is supported in equilibrium via net rotation compared to its internal velocity dispersion. The spin parameter varies between 0, for a structure negligibly supported by rotation, to values of order 1 where it is completely rotationally supported, and in practice maximum values are usually $\lambda \approx 0.4$ (Frenk & White, 2012). Values larger than 1 are unstable structures not in equilibrium.

There are two variants of the spin parameter that are in common use. Peebles (1969) proposed to parameterise the spin using the expression given in Equation 4.1.

$$\lambda = \frac{J\sqrt{|E|}}{GM^{5/2}} \tag{4.1}$$

where J is total angular momentum, E the total energy (kinetic and potential) and M the mass of the structure. In isolated haloes, all of these quantities are conserved, which gives the definition a time independence.

Bett *et al.* (2007) measured the Peeble's spin parameter and fitted an expression to the distribution for haloes extracted from the Millennium simulation (Springel *et al.*, 2005); that is characterised by Equation 4.2

$$P(\log\lambda) = A\left(\frac{\lambda}{\lambda_0}\right)^3 \exp\left[-\alpha\left(\frac{\lambda}{\lambda_0}\right)^{3/\alpha}\right] \tag{4.2}$$

where A is

$$A = 3 \ln 10 \frac{\alpha^{\alpha-1}}{\Gamma(\alpha)} \tag{4.3}$$

The variables λ_0 and α are free parameters, and $\Gamma(\alpha)$ is the gamma function. The best fit they found for field haloes was with $\lambda_0 = 0.04326$ and $\alpha = 2.509$.

Bullock *et al.* (2001) proposed a different definition of the spin parameter, λ', expressed in Equation 4.4. As it is not dependent on measuring the energy it is somewhat faster to calculate when dealing with large numbers of haloes.

$$\lambda' = \frac{J}{\sqrt{2} M R V} \tag{4.4}$$

Here J is the angular momentum within the enclosing sphere of virial radius R and virial mass M, and V is the circular velocity at the virial radius ($V^2 = GM/R$). The Bullock spin parameter is more robust to the position of the outer radius of the structure. Bullock proposes a fitting function to the distribution as described in Equation 4.5 which was based on one from Barnes & Efstathiou (1987).

$$P(\lambda') = \frac{1}{\lambda' \sqrt{2\pi}\sigma} \exp\left(-\frac{\ln^2(\lambda'/\lambda'_0)}{2\sigma^2}\right) \tag{4.5}$$

This has free parameters λ'_0 and σ and Bullock *et al.* (2001) found a best fit for field haloes at values of $\lambda'_0 = 0.035$ and $\sigma = 0.5$.

The Peebles calculation is perhaps more well defined for a given set of particles, as it is calculated directly from the particles properties, whereas the Bullock parameter is easier to calculate from gross halo statistics, and is not dependant on the density profile. For more comparisons of the two parameters the reader is referred to Hetznecker & Burkert (2006)

4.2.2 The SubHalo Finders

In this section we briefly list the halo finders that took part in the comparison project. More details about the specific algorithms are available in Onions *et al.* (2012) and the articles referenced therein.

- ADAPTAHOP (Tweed) is a configuration space over density finder (Aubert, Pichon & Colombi, 2004; Tweed *et al.*, 2009).

- AHF (Knollmann & Knebe) is a configuration space spherical overdensity adaptive mesh finder (Gill, Knebe & Gibson, 2004; Knollmann & Knebe, 2009).

- GRASSHOPPER (GRadient ASSisted HOP) (Stadel) is a reworking of the SKID group finder (Stadel, 2001) and appears within our wider comparison for the first time here, and so is described in more detail. It takes an approach like the HOP algorithm (Eisenstein & Hut, 1998) using the density around each particle to make links between the densest of its neighbours. Density is calculated for all particles as in SKID using an 80 element smoothed particle hydrodynamic (SPH) kernel.

  After links have been created, each particle follows the chain of links until it reaches a cycle, marking oscillation about the density peak of the group. Finally since noise below a gravitational softening length causes a lot of artificial density peaks we search for particles of a cycle which are within a distance τ of any other particles in a cycle. The parameter τ is typically set to 4 times the gravitational softening length, as was the typical case for SKID. Unbinding is also performed in a nearly equivalent way to SKID, but now scales as $O(n \log n)$ as opposed to $O(n^2)$ as was the case with the original SKID.

  The group finding with GRASSHOPPER is now fast enough to allow it to be performed during a simulation but gives nearly identical results to the previous SKID algorithm.

- Hierarchical Bound-Tracing (HBT) (Han) is a tracking algorithm working in the time domain (Han *et al.*, 2012).

- HOT+FiEstAS (HOT3D & HOT6D) (Ascasibar) is a general-purpose clustering analysis tool, working either in configuration or phase space (Ascasibar & Binney, 2005; Ascasibar, 2010).

- MENDIETA (Sgró, Ruiz & Merchán) is a Friends-of-Friends based finder that works in configuration space (Sgr, Ruiz & Merchn, 2010).

- ROCKSTAR (Behroozi) is a phase-space halo finder (Behroozi, Wechsler & Wu, 2011).
- STF (Elahi) is a velocity space/phase-space finder (Elahi, Thacker & Widrow, 2011).
- SUBFIND (Springel) is a configuration space finder (Springel *et al.*, 2001).
- VOBOZ (Neyrinck) is a Voronoi tessellation based finder (Neyrinck, Gnedin & Hamilton, 2005).

4.3 The Data

4.3.1 Simulation Data

The first data set used in this paper forms part of the Aquarius project (Springel *et al.*, 2008a). It consists of multiple dark matter only re-simulations of a Milky Way-like halo at a variety of resolutions performed using GADGET3 (based on GADGET2, Springel, 2005). We have used in the main the Aquarius-A to E halo dataset at $z = 0$ for this project. This provides 5 levels of resolution, varying in complexity for which further details are available in Onions *et al.* (2012).

The underlying cosmology for the Aquarius simulations is the same as that used for the Millennium simulation (Springel *et al.*, 2005) i.e. $\Omega_M = 0.25,\ \Omega_\Lambda = 0.75,\ \sigma_8 = 0.9,\ n_s = 1,\ h = 0.73$. These parameters are close to the latest WMAP data (Jarosik *et al.*, 2011) ($\Omega_M = 0.2669$, $\Omega_\Lambda = 0.734$, $\sigma_8 = 0.801$, $n_s = 0.963$, $h = 0.71$) although σ_8 is a little high. All the simulations were started at an initial redshift of 127. Precise details on the set-up and performance of these models can be found in Springel *et al.* (2008a).

The second data set was from the GHALO simulation data (Stadel *et al.*, 2009b). GHALO uses a slightly different cosmology to Aquarius, $\Omega_M = 0.237, \Omega_\Lambda = 0.763, \sigma_8 = 0.742,\ n_s = 0.951,\ h = 0.735$ which again are reasonably close to WMAP latest results. It also uses a different gravity solver, PKDGRAV2(Stadel, 2001), to run the simulation therefore allowing comparison which is independent of gravity solver and to

Table 4.1: Summary of the key numbers in the Aquarius and GHALO simulations used in this study. N_{high} is the number of particles with the highest resolution (lowest individual mass). N_{250} is the number of high resolution particles found within a sphere of radius 250 kpc/h from the fiducial centre at each resolution (*i.e.* those of primary interest for this study).

Simulation	N_{high}	N_{250}
Aq-A-5	2,316,893	712,232
Aq-A-4	18,535,97	5,715,467
Aq-A-3	148,285,000	45,150,166
Aq-A-2	531,570,000	162,527,280
Aq-A-1	4,252,607,000	1,306,256,871
Aq-B-4	18,949,101	4,771,239
Aq-C-4	26,679,146	6,423,136
Aq-D-4	20,455,156	8,327,811
Aq-E-4	17,159,996	5,819,864
GH-4	11,254,149	1,723,372
GH-3	141,232,695	47,005,813

some extent the exact cosmology.

The details of both simulations are summarised in Table 4.1.

4.3.2 Post-processing pipeline

The participants were asked to run their subhalo finders on the supplied data and to return a catalogue listing the substructures they found. Specifically they were asked to return a list of uniquely identified substructures together with a list of all particles associated with each subhalo. The broad statistics of the haloes found are summarised in Table 4.2.

To enable a direct comparison, all the data returned was subject to a common post-processing pipeline detailed in Chapter 3. For this project we added a common unbinding procedure based on the algorithm from the AHF finder which is based on spherical unbinding from the centre. We requested data to be returned both with and without unbinding to allow a comparison of that procedure to feature in this study. Unbinding is the process where the collection of gathered particles is examined to discard those which are not gravitationally bound to the structure. This common unbinding allowed us to remove some of the sources of scatter introduced by the finders using slightly different algorithms for removing unbound particles and to find what difference this made to the results.

Table 4.2: The number of subhaloes containing 300 or more particles and centres within a sphere of radius 250kpc/h from the fiducial centre found by each finder after standardised post-processing (see Section 4.3.2).

Name	ADAPTAHOP	AHF	GRASSHOPPER	HBT	HOT3D	HOT6D
Aq-A-5	24	23	23	23	18	23
Aq-A-4	222	189	170	169	174	176
Aq-A-3	-	1259	1202	1217	-	-
Aq-A-2	-	4230	-	4036	-	-
Aq-A-1	-	30694	-	-	-	-
Aq-B-4	-	197	-	191	-	-
Aq-C-4	-	152	-	146	-	-
Aq-D-4	-	217	-	216	-	-
Aq-E-4	-	218	-	219	-	-
GH-4	-	58	58	-	-	-
GH-3	-	1172	1148	-	-	-

Name	MENDIETA	ROCKSTAR	STF	SUBFIND	VOBOZ
Aq-A-5	17	25	22	23	21
Aq-A-4	123	182	155	154	163
Aq-A-3	787	1252	1124	1117	1141
Aq-A-2	-	4161	-	3661	-
Aq-A-1	-	25009	-	26155	-
Aq-B-4	-	202	-	188	-
Aq-C-4	-	158	-	137	-
Aq-D-4	-	230	-	196	-
Aq-E-4	-	221	-	205	-
GH-4	-	60	54	54	-
GH-3	-	1148	1033	1090	-

4.4 Results

The results used were restricted to subhaloes with more than 300 particles, as these produce a relatively stable value for spin. Values below this limit tend not to converge across resolutions (Bett *et al.*, 2007).

4.4.1 Spin parameter

In general there is a proportional relationship between the Peebles and Bullock spin parameters recovered by all the finders for the same subhaloes, although there is some scatter as shown in Figure 4.1. We do not dwell on the differences between the two definitions as that has already been studied elsewhere (Hetznecker & Burkert, 2006).

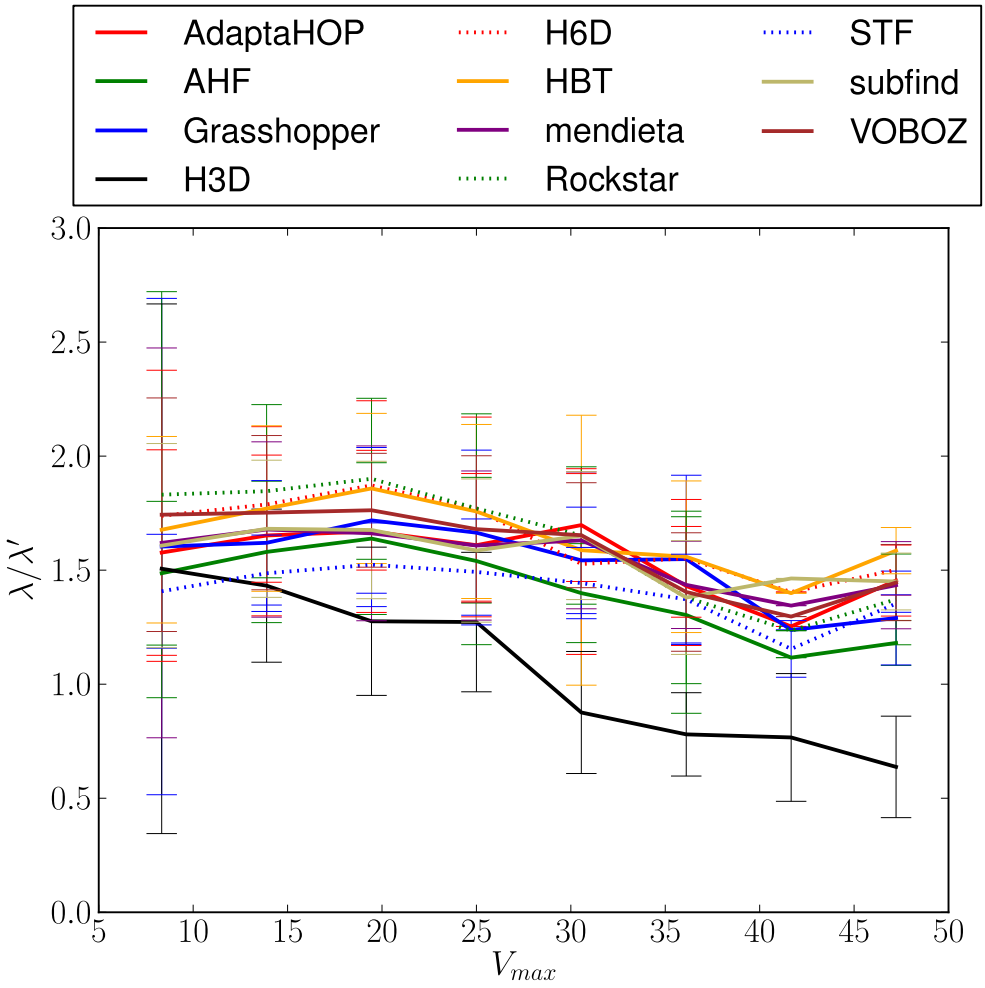

Figure 4.1: A comparison of the Peebles and Bullock spin parameters against $v_{\rm max}$ based on all finders using a common unbinding procedure from subhaloes with more than 300 particles. The mean value of λ/λ' is shown together with one standard deviation error bars. It shows there is a correlation between the two but not a one-to-one correspondence, with some scatter present. The scatter at low $v_{\rm max}$ where haloes have very few particles is particularly pronounced.

As both definitions of spin exist in the literature we consider both metrics when comparing how the spin is recovered across finders, placing particular emphasis on their application to subhaloes.

The majority of field haloes are found to cluster around a value of $\lambda_0 = 0.044$ for the Peebles spin parameter (Bett *et al.*, 2007) and $\lambda'_0 = 0.035$ for the Bullock parameter (Bullock *et al.*, 2001) with a spread of values matched by a free parameter to give the width of the distribution.

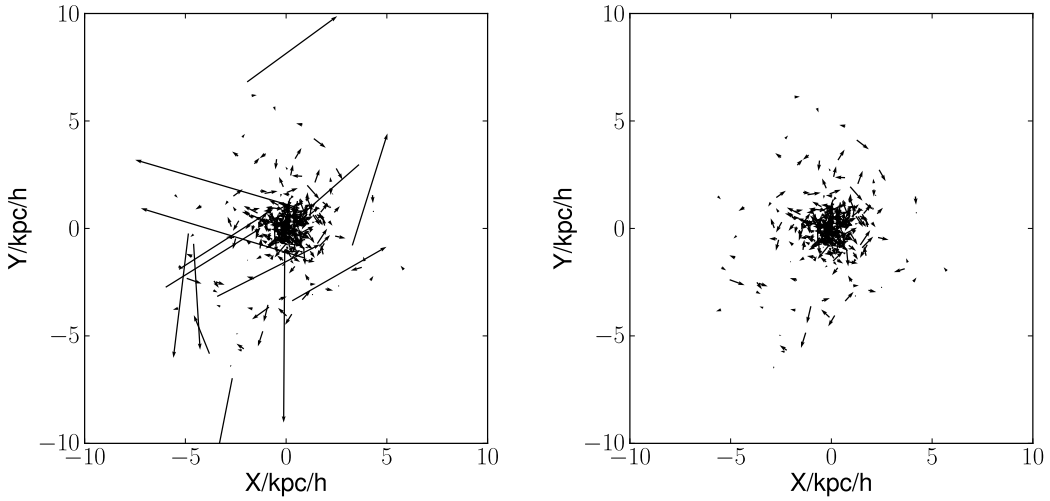

Figure 4.2: An example of the influence of unbinding. Left panel: particles in the object prior to unbinding. Right panel: particles in the object after unbinding has been performed. The vectors indicate the direction and velocity relative to the bulk velocity of the individual particles making up this example subhalo. The contribution from the background particles has only a minor influence on the mass and $v_{\max}$ of the subhalo, but a large effect on the spin parameter.

4.4.1.1 Spin for subhaloes with no unbinding performed

If unbinding has not been correctly implemented the high speed background particles can distort the spin parameter enormously.

To emphasise the type of structures that are found, an example of a subhalo without (left panel) and with (right panel) unbinding is shown in Figure 4.2. This is displayed as a vector plot of all the component particles position and velocities that make up the subhalo with the velocity vectors scaled in the same way in both panels. The bulk velocity of the subhalo has been removed and all positions and velocities are relative to the rest frame of the subhalo. Evident in the left panel of Figure 4.2 without unbinding are stray particles that are part of the background halo. Despite their small number these particles have both a large lever arm and large velocity relative to the halo, and significantly alter the derived value of the spin parameter due to their large angular momenta.

Comparing the two forms of the spin parameter in Figure 4.3 and Figure 4.4 we show how the spin parameter is quite chaotic, not matching a smooth Gaussian like profile as might be expected, and is clearly a long way removed from the idealised curve others have found for the distribution of spin. A significant number of the haloes have spin parameter values above 1, which is unphysical as these objects would be ripped

apart by this level of rotation and so clearly cannot be equilibrium systems. This result is perhaps not surprising given the contribution from unbound background particles moving with velocities far from the mean of the object being considered but clearly shows how poor unbinding methods are relatively easy to detect by looking at the spin parameter distribution. The Peebles spin parameter is more affected by the lack of unbinding than the equivalent Bullock parameter as it takes into account the kinetic energy of all the particles. Some more objective numbers for this and subsequent comparisons are given in Table 4.3.

The best fit values shown by the bold dashed lines are vastly different from the fiducial values given in Section 4.2. It is however significant that the finders HOT6D, ROCKSTAR and STF (shown by dotted lines) which all have a phase space based component in their particle collection algorithm already show a much better fit to the fiducial value than the non phase-space finders. It should be noted that when GRASSHOPPER is run without unbinding, it finds a large number of subhaloes which would normally be discarded by the unbinding procedure that is integral to the final part of the GRASSHOPPER algorithm.

4.4.1.2 Spin for subhaloes with finders own unbinding performed

Including each finder's own unbinding procedure improves the spin parameter measure considerably, as shown in Figure 4.5 and Figure 4.6. Note that as ADAPTAHOP doesn't do any unbinding in its post-processing steps it is a clear outlier on this plot. The MENDIETA finder shows a double peak, which is indicative of some of the unbinding failing, an issue that the authors of the finder are currently working on.

When fitting the best fit curves to this data obtained for the spin parameter of subhaloes, the peak of the Bullock fitting curve given in Equation 4.4 is less than that found for field haloes, by a value of about 20 percent, offsetting the mean towards smaller values of the spin parameter. For the Peebles spin parameter the best fit is again offset by about 36 percent from the field halo value, again towards a smaller value of the spin parameter.

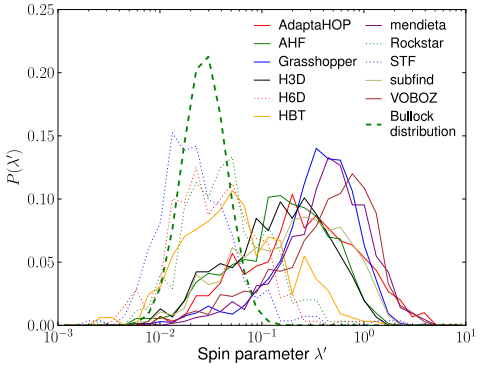

Figure 4.3: General profile of the Bullock spin parameter of all subhaloes found with more than 300 particles without unbinding performed, binned into 35 log bins. The results are normalised to give equal area under the visible curve. The dashed line is the field halo fit from Bullock *et al.* (2001). The results show a large scatter about a peak which is far distant from the fiducial fit for haloes. Dotted lines indicate finders with a phase space component of their algorithm, whereas solid lines indicate finders without a phase space component.

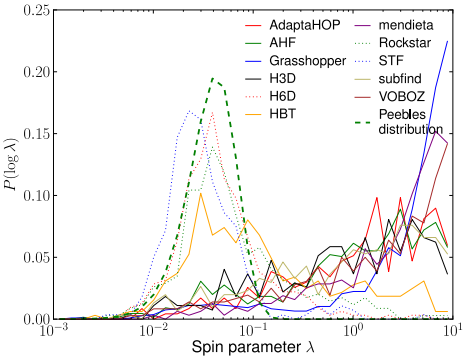

Figure 4.4: The same plot as Figure 4.3 but using the Peebles spin parameter and fitting function from Bett *et al.* (2007).

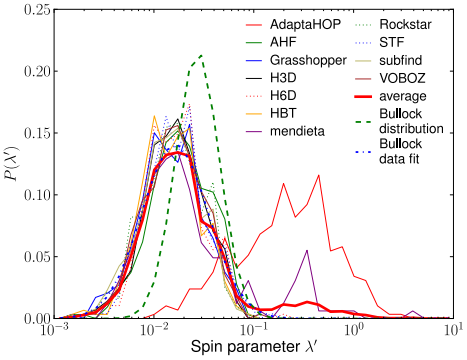

Figure 4.5: The same plot as Figure 4.3 but with the finders own unbinding processing applied to the data. This groups the spin parameters somewhat more tightly, and shows that spin is a good indicator of how well the unbinding procedure is removing spurious background particles. The ADAPTAHOP finder doesn't perform an unbinding step, and this plot also shows up a flaw in MENDIETA's unbinding procedure. The dashed line is the Bullock field halo fit curve from Bullock *et al.* (2001). The Bullock data fit is the best fit to the average using the Bullock fitting formula.

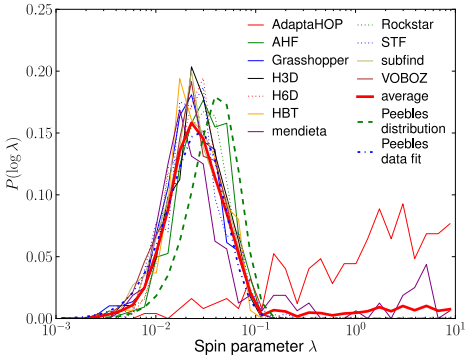

Figure 4.6: The same plot as Figure 4.5 except that this time the dashed line is the Peebles field halo fit from Bett *et al.* (2007). The Peebles best data fit is the best fit to the average of the Bett formula.

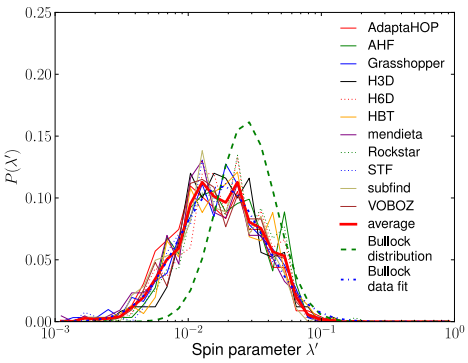

Figure 4.7: The same plot as Figure 4.3 but with a common unbinding processing applied to the data. This groups the spin parameters much more tightly, and shows that spin is a good predictor of how well the unbinding procedure performs at removing spurious background particles. The dashed line is the Bullock best fit field halo curve from Bullock *et al.* (2001).

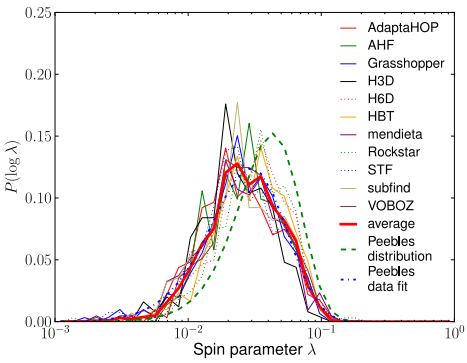

Figure 4.8: The same plot as Figure 4.4 but with a common unbinding processing applied to the data. The dashed line is the Peebles best fit curve from Bett *et al.* (2007).

4.4.1.3 Spin for subhaloes with a common unbinding performed

Once a common unbinding is done, the curves move significantly closer to the idealised curve, although there is still some separation. The plots of Figure 4.7 and Figure 4.8 compare the spin parameter distribution of the different finders using a common unbinding process. It shows the match between the best fit curve quoted in Bullock *et al.* (2001) and Bett *et al.* (2007) and the haloes found by the finders taking part in the comparison. The values are now offset by 10 percent for the Bullock fit, and 30 percent for the Peebles fit. This results in the closest fit to the data, although the subhalo spin again extends to slightly lower values for both parameters, and follows the best fit line at larger values. These results also have a similar trend for the Aquarius B-E haloes and the GHALO data sets. These inclusions show that the results are not influenced greatly by the simulation, simulation engine or small changes in the cosmology used.

4.4.1.4 Spin at higher resolutions

Going to higher resolutions afforded by the level 1 data as shown in Figure 4.9, the trend to a lower spin distribution peak continues, although only three of the finders were able to manage such a computationally intensive task.

There is a more pronounced tendency to depart from the field halo fit line at low spin

Table 4.3: Summary of the best fit parameters for the graphs shown. Shown are the values for λ_0 and the other free parameter (α or σ) used in the best fit, and their difference from the published field halo fit value. The subscripts F, N, O and C are for field haloes, no unbinding, own unbinding and common unbinding respectively. The Δ values are the difference from the field halo values, and the change is the percentage difference. All results are for level 4 data except the last which is level 1

Plot	λ_0	$\Delta\lambda_0$	change	σ/α	$\Delta\sigma/\alpha$	change
Bullock$_F$	0.035			0.5		
Peebles$_F$	0.044			2.509		
Bullock$_N$	1.646	1.611	+4600%	1.36	0.86	172%
Peebles$_N$	12.6	12.573	+29000%	41	39.2	1560%
Bullock$_O$	0.028	-0.007	-20%	0.727	0.227	45.5%
Peebles$_O$	0.028	-0.016	-36%	3.643	1.134	45.2%
Bullock$_C$	0.031	-0.004	-10.4%	0.75	0.25	50.0%
Peebles$_C$	0.03	-0.013	-30%	3.96	1.448	57.7%
Bullock-L1$_O$	0.022	-0.013	-38%	0.693	0.193	38.6%

part of the distribution, with the peak and bulk of the distribution moving towards lower spin parameter values. The finders also show more scatter with each of them identifying the peak of the distribution in slightly different places. The agreement particularly at the low end of the spin distribution is good but with slightly lesser agreement at the high end.

Although AHF appears to find slightly more higher spin haloes, this is a result of the spherical unbinding algorithm it uses, which tends to also increase the spin distribution of the other finders slightly when used as the common unbinding procedure.

The dashed line representing the level 4 data is included to allow a direct comparison between the level 4 and level 1 average fits. It shows the continued movement of the distribution towards lower spin values with higher resolution and an increase in data.

4.4.1.5 Spin distribution summary

The best fit curve figures for all these plots are summarised in Table 4.3. Even after cleaning the catalogues significantly by utilising a common unbinding procedure (subsection 4.3.2) for all finders there remains a definite trend for substructure spin to be less than that found for field haloes. We investigate the reason for this in the next sections.

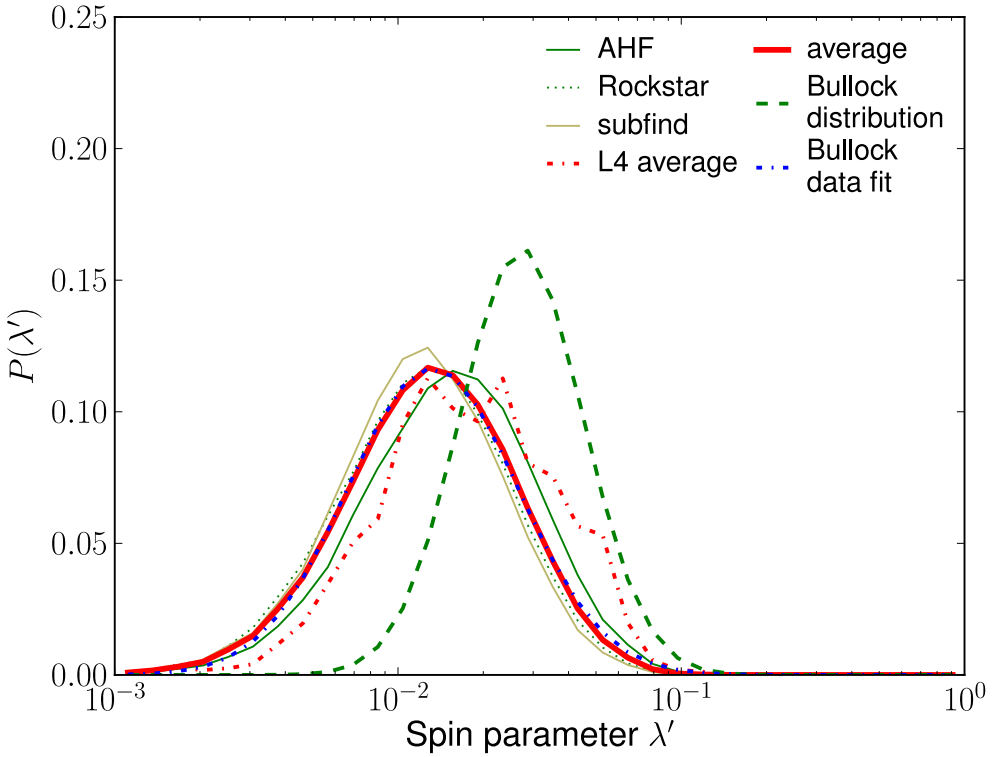

Figure 4.9: The same plot as Figure 4.5 but using the level 1 data which has much higher resolution. The lower spin haloes are more obvious in this plot, as is the difference between finders. The level 4 average is included for comparison.

4.4.2 Host halo radial comparison

Next we consider whether the location of a subhalo within a host halo has any effect on the recovered spin parameter. First we demonstrate in Figure 4.10 that any effect is not an artefact of the finding process. Substructures closer into the centre of the host halo are more difficult to detect particularly by some finders, and therefore subject to a loss of constituent particles that could be attached to the subhalo as shown in Muldrew, Pearce & Power (2011). To test this supposition we took a subhalo found in the outskirts of the Aquarius-A main halo, and repositioning it at points closer to the location of the centre of the halo. Then two of the finders (AHF and ROCKSTAR) were rerun on the new data and the spin value calculated anew. The results shown in Figure 4.10 indicate that there is little change in the value of the spin parameter with radius despite some variation in the recovered number of particles.

Next we look at whether the mean value of the measured spin parameters changes with respect to the distance from the centre of the host halo. Figure 4.11 displays this radial

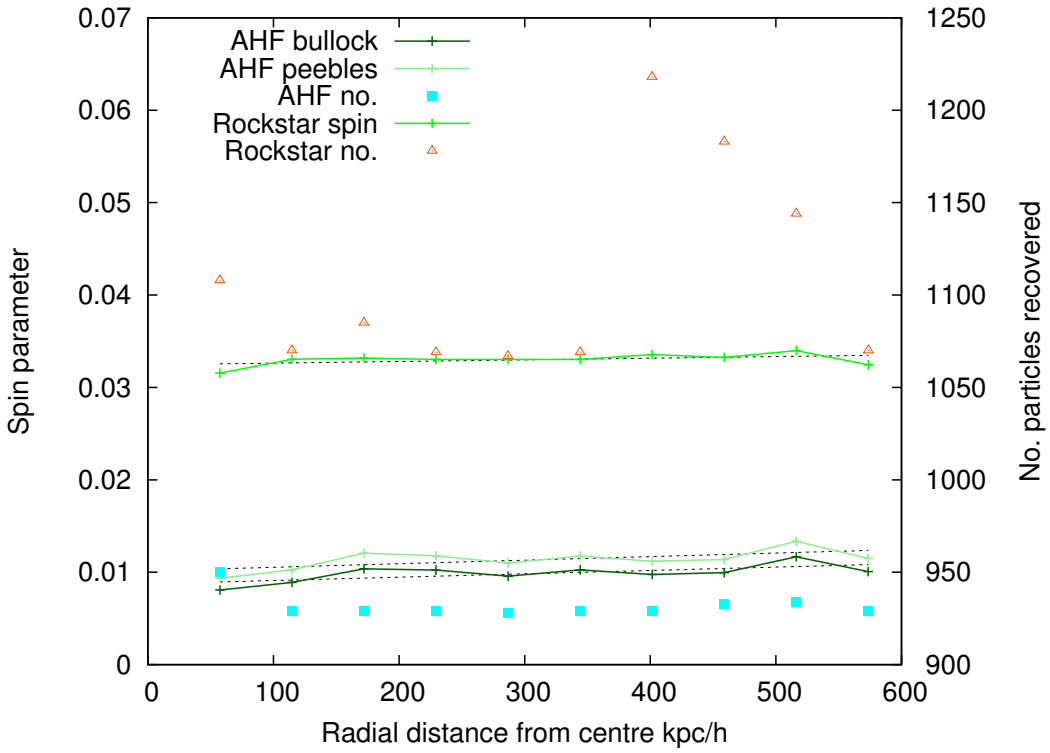

Figure 4.10: The spin parameter as recovered by AHF (Peebles and Bullock) and ROCKSTAR (Peebles) of an outer subhalo repositioned progressively closer to the centre. The finders own spin calculations were used in this case rather than the full pipeline. The spin is seen to be approximately unchanging across the radius.

dependence for the indicated finders after a common unbinding step has been applied. The background points indicate the scatter in the spin parameter for any individual halo, as seen in the previous section. This shows a small trend for a lower mean spin as the subhaloes get closer to the centre of the host halo. This confirms the result that were found in Reed *et al.* (2004) but is shown here at higher resolution and across more finders than this earlier paper.

Equivalent results are found when we compare 6 different simulations generated by two different N-body codes and aggregate the average of the different finders across multiple haloes in Figure 4.12. This effect of reduced spin for subhaloes (as noted in Reed *et al.* 2004) is difficult to detect observationally, as most substructure will form galaxies before falling in so will have its spin detectable from observations of galactic rotation curve already fixed (Kauffmann, White & Guiderdoni, 1993). The possible exception to this are galaxies forming at high redshift where the infalling substructure has not yet formed stars, such as gas-rich dark galaxies (Cantalupo, Lilly & Haehnelt, 2012), made entirely of dark matter and gas, which may form structure after falling

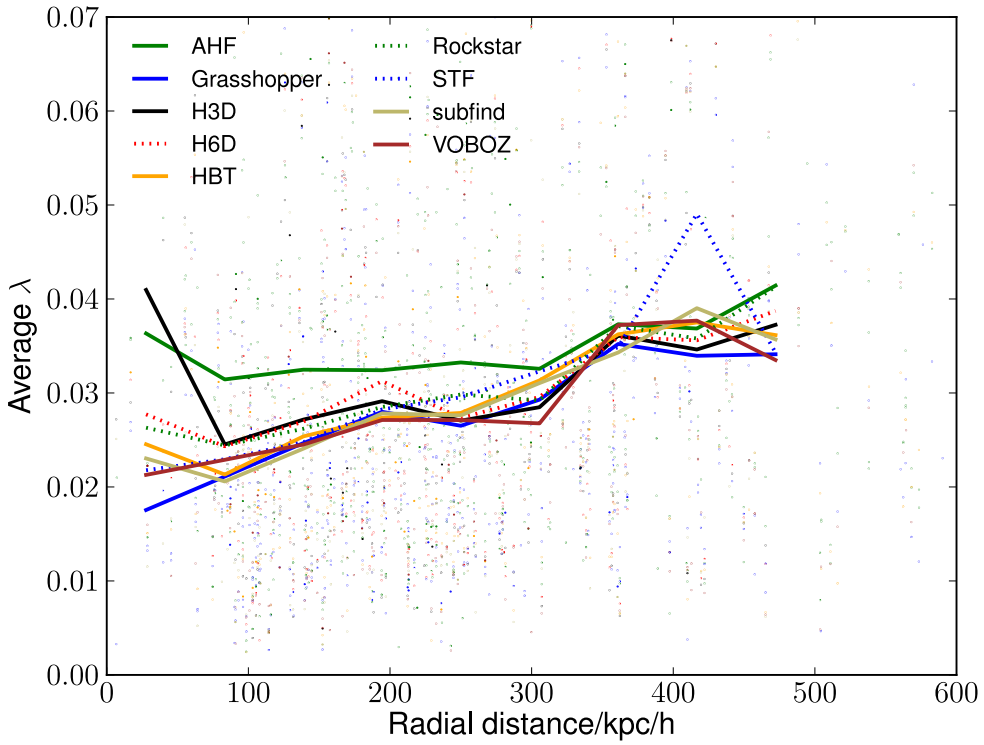

Figure 4.11: Comparison of mean spin parameter against radius from the centre of the host halo. Common unbinding was applied in the pipeline in this case. There is some additional scatter at low radial values as few haloes above 300 particles are found there. The background points indicate the measured spin parameter for individual subhaloes.

into a parent halo.

4.4.3 Build up of the spin parameter within a subhalo

This leads to the question of what causes the drop in the measured spin parameter with proximity to the centre of the host halo. Figure 4.13 shows the average change in the measured spin parameter as the detected subhalo is analysed from the centre outwards to its radius. This procedure is computed after the common processing and unbinding steps have been done. The subhaloes analysed in this way are then further binned into radial bins determined from the centre of the host halo. The outermost subhaloes, which are the least disrupted, show an initial decrease in measured spin parameter as particles are removed from their outer edges. Subhaloes extracted from nearer the centre of the host halo do not show this initial decrease but instead have a monotonically rising spin parameter as material is removed.

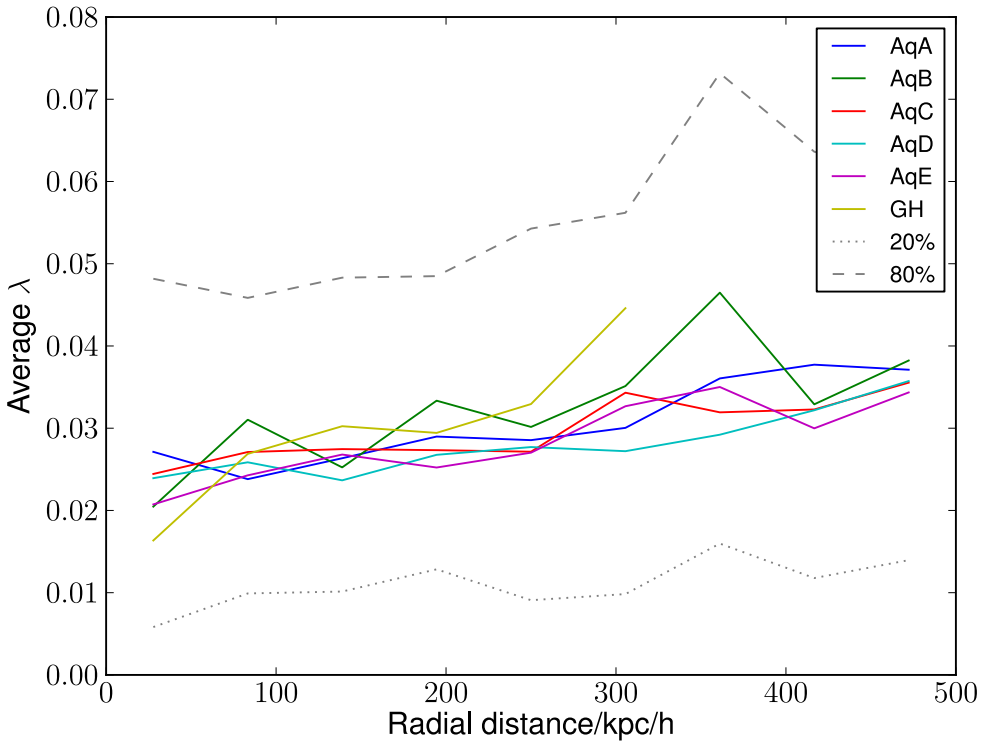

Figure 4.12: Comparison of mean spin parameter against radius from the centre of the host halo for several different haloes. The finders own unbinding procedure was used in the pipeline in this case. Each line is the average of the spin parameter binned into 10 bins across all finders partaking (AHF, GRASSHOPPER, ROCKSTAR, SUBFIND and STF). The haloes used were the Aquarius-A to E and GHALO all at level 4 of the resolution. The dashed/dotted lines indicate 20 and 80 maximum percentiles across all data.

This trend suggests that subhaloes are preferentially stripped of high angular momentum particles which are likely to be the most weakly bound particles, leading to a decrease in the spin parameter as they enter the host halo. The outermost particles are usually those least bound so are the most likely to be removed on infall.

We can also examine how the spin parameter is built up as mass is added to a subhalo. In Figure 4.14 we look at how the spin parameter changes at various mass cuts of the subhalo, $M(< M_{tot})$. This shows how the spin is built up across the structure of the subhalo. For each halo we calculate the spin parameter at fractions of 0.25, 0.5, 0.75, 0.95 of the subhalo's total mass for all the contributing halo finders. We plot the mean and the standard deviation at each mass cut.

As expected from Figure 4.13 all finders agree that the calculated spin increases as the fraction of the subhalo mass that is used to calculate the spin parameter is reduced. Note that haloes have steeply rising density profiles and so the inner 50% of the mass

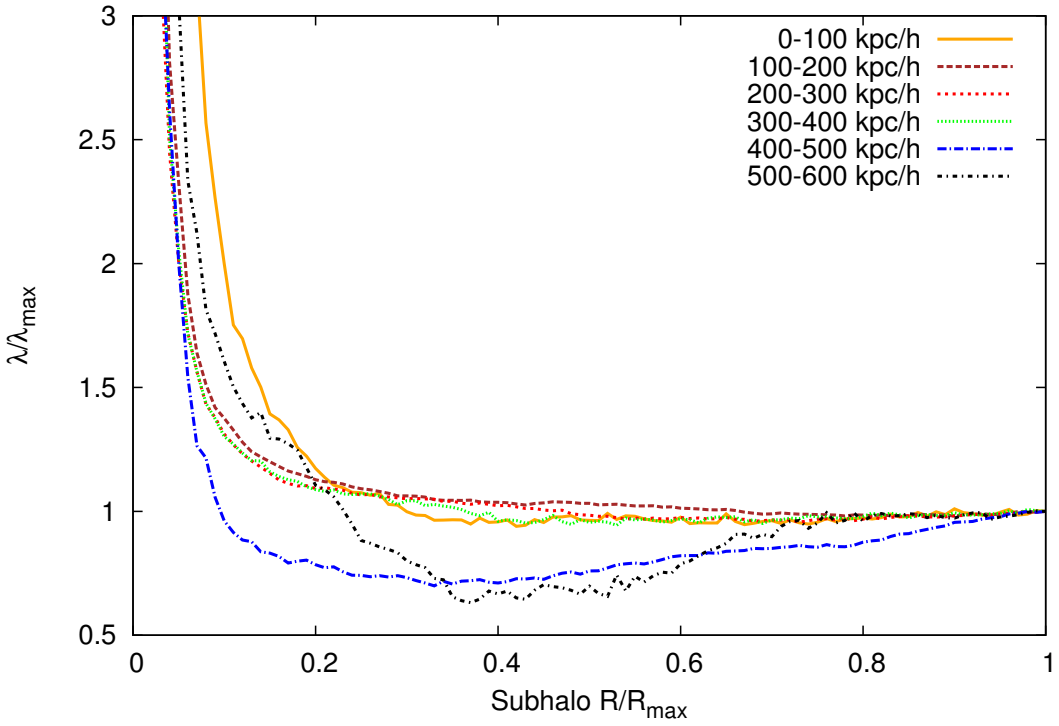

Figure 4.13: The radial profile of the spin parameter across the subhalo (λ_{max} is the spin parameter at the R_{max}). This shows the change in the measured spin parameter as spin is analysed from the centre to the radius of the subhalo. Here R_{max} is the subhaloes maximum radius. Each line represents a different host halo radial bin. Subhaloes near the centre of the host halo show monotonically rising spin parameter values spin, whereas further out the spin parameter initially drops before rising.

is contained within a much smaller fraction of the radius and that this result is averaged over all the recovered haloes and not split in radial bins.

4.5 Summary & Conclusions

There is a good level of agreement amongst the finders on the recovery of the distribution of the spin of subhaloes, although differences are still evident, causing scatter in some of the comparisons. Undoubtedly some of the scatter is due to different types of subhalo that are being recovered by the finders, some finders focusing on stream like structures and some on simple overdensities. There is still some room for improvement of the finders as the common unbinding test shows. Some of the possible improvements and sources of error will be outlined in Knebe et al. (in prep).

The distribution of spin provides a very good indicator of the finders unbinding ability and seems broadly unaffected by the cosmology and simulation engine in use. As

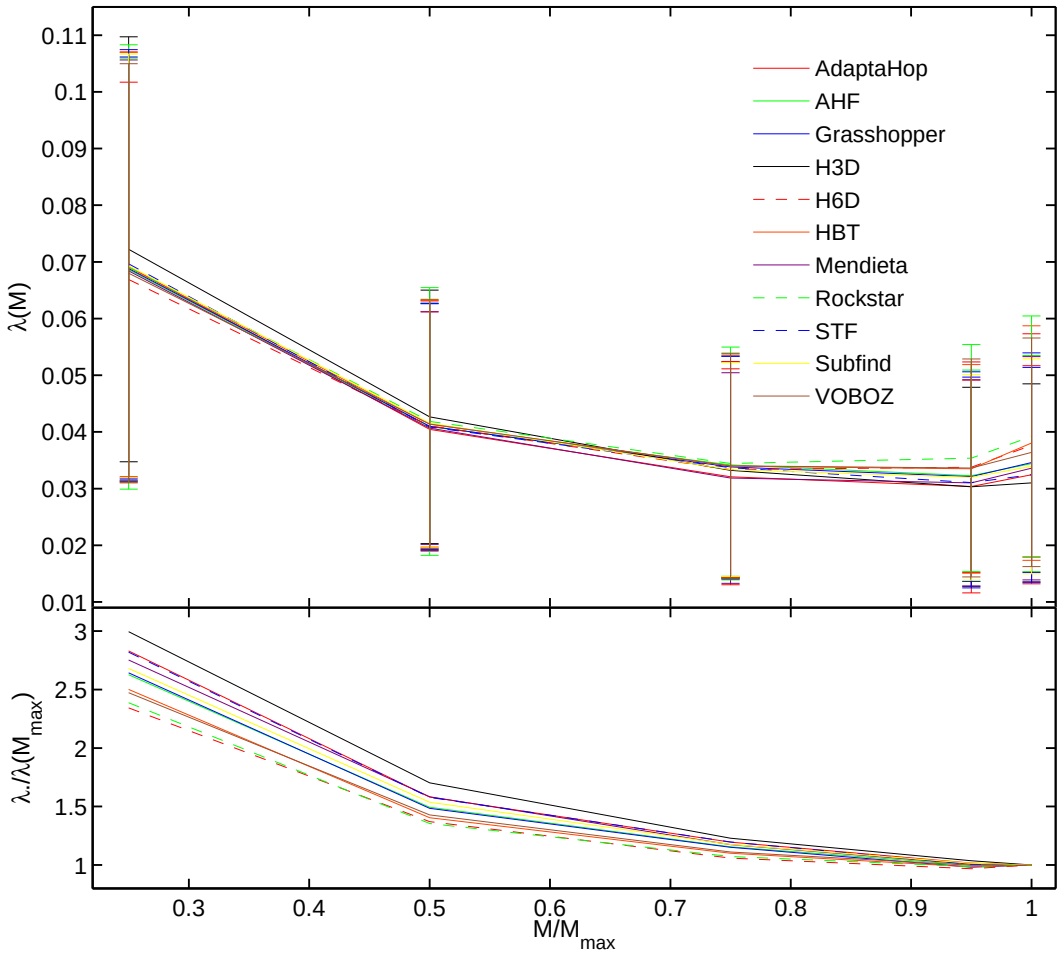

Figure 4.14: Comparison of the normalised mean Peebles spin at different mass shells of all subhaloes. The cuts were taken at 0.25, 0.5, 0.75, 0.95 and the complete mass of the subhalo. A common unbinding procedure was run on the results. There is a clear decrease in spin with increasing contained mass, and about a 3-fold drop is evident. The top plot shows the value of the spin parameter, and the bottom plot the spin parameter normalised to the value of λ at the subhaloes R_{max}. Error bars are one standard deviation.

such, the spin distribution serves as a mechanism to detect if substructure finders are performing the unbinding correctly. The unbinding errors can be masked in other comparisons such as $v_{\rm max}$ and mass plots but show up in an obvious way when the spin distribution is examined. Phase-space finders are less sensitive to poor unbinding as they have some implicit unbinding in their selection criteria when looking at velocity components. Indeed Hetznecker & Burkert (2006) and D'Onghia & Navarro (2007) both show there is a good correlation between the virialisation of haloes and the spin parameter, thus indicating its use for the determination of how relaxed the halo is, which is not unrelated to the unbinding process.

The mean spin parameter of subhaloes decreases as they approach the host halo's cen-

tre. This is a real effect and not an artefact of any difficulty in recovering structure as the subhalo approaches the centre of the main halo. This effect is apparent in the spin parameter distribution which matches that of field haloes at larger radii but has a broader width than other published fits, extending to lower spin values. This difference between the spin properties of subhaloes and field haloes needs to be taken into account if precise measurement of the spin parameter distribution are to be made.

The recovered spin parameter goes through a minimum for subhaloes near the edge of the host at about half the $r_{\rm max}$ value. Here, if outer particles are stripped tidally as a substructure falls into a host halo, the result will be a decrease in the spin. This implies a radial dependant factor needs to be taken into account when compiling substructure catalogues, as the infalling haloes tend to have their outer particles removed. Once the outer layer has been lost the spin parameter generally increases to smaller radii as less and less mass is considered.

The value of the spin parameter measured is dependent upon the choice of where to place the outer edge and precisely which material is included in the calculation. As we have shown here and elsewhere these choices are very halo finder dependent and so care should be taken when inter-comparing spin parameter measurements from different codes.

In a future project we looked more closely at the difference between field and substructure haloes. This work is reported in the work Wang *et al.* (2015a) where temporal evolution is also investigated.

Chapter 5

Merger Trees with JMERGE

The building of merger trees appears to be a fairly straightforward process. It involves taking a halo at timestep T_N and linking it to a halo at timestep T_{N+1}. The merger part of the name comes from the fact that there may be more than one progenitor for a given halo at the next timestep, when two or more haloes merge. There is also the remote possibility of fragmentation of a halo into two or more parts as the result of an interaction. This does require a rather convoluted set of multi-body encounters to fragment a halo. For instance a close encounter strips a tail of matter that is then split from the halo by encounter with a third halo. These are not normally considered except in a theoretical sense.

There is, as ever with simple procedures, much that has to be taken into consideration when the fine details are considered. Some of the important considerations are:

- In general, with a few exceptions, the merger tree process is only as good as the data provided. It has been shown that even the most sophisticated finders struggle to recover substructure deep within a host halo, even when the substructures have a significant difference in velocity (Muldrew, Pearce & Power, 2011; Knebe *et al.*, 2011; Onions *et al.*, 2012).

- When two haloes are approaching a merger situation, finders can struggle to allot the particles between the two overlapping haloes. This can result in large jumps of mass, or central positions as new particles are incorporated into what the finder decides is the major halo. In some cases this leads to two adjacent halos

"flip-flopping" in mass and position in subsequent snapshots as they struggle to separate the two. A number of these issues are investigated in Behroozi *et al.* (2015).

- Small haloes can often hover around the cut off density that the finder uses to differentiate a halo from the background density field. This results in small haloes popping into existence, and then vanishing in the next timestep, possibly to reappear in a subsequent snapshot.

- Timesteps may not be linear between snapshots. Often there is little happening in the early stages of the simulation, and large time steps are taken between snapshots. As the more interesting phases approach, then the time between snapshots can be reduced to improve the detailed story. Not all merger tree algorithms take this information as input, but this can result in some interesting edge cases. This variation of timestepping was investigated more thoroughly in Wang *et al.* (2015b).

5.1 The JMERGE Algorithm

The algorithm used by JMERGE is constructed in 3 parts, which are termed tracking, major merging, and minor merging. These are described in detail here.

JMERGE was written to compare the approach of not using particle identifiers for matching, which means a program with less memory requirements, faster matching and an overall improvement in speed. If the results were comparable with existing merger tree builders, then this might show an interesting development. It would also mean much less data needed to be stored to build a merger tree.

5.1.1 Tracking

For the majority of haloes, especially at early times, they remain in isolation and drift through the universe. Therefore at this stage, most of what is required is to match up haloes at T_N with haloes at T_{N+1}. This can be done by predicting where they will have moved from based on their peculiar velocity and the interval between snapshots.

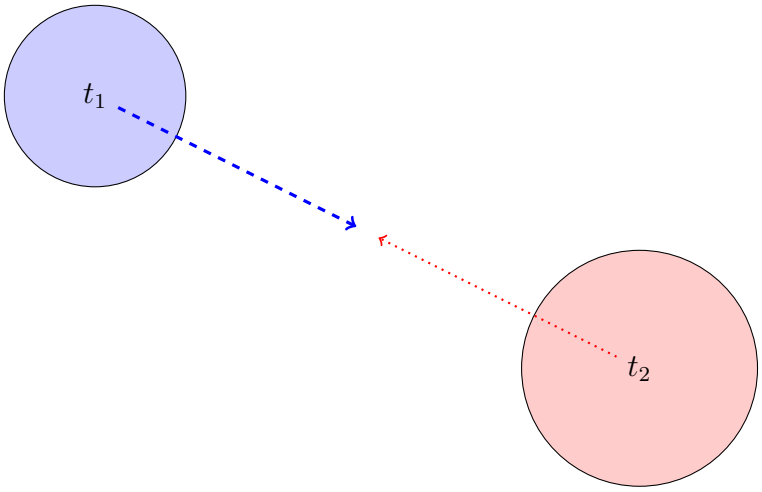

Figure 5.1: The halo tracking step of JMERGE, matching halo at time T_N (T_1) with that at T_{N+1} (T_2). The T_1 halo is projected forward in time, and the T_2 backwards in time to attempt a best match.

A more complete analysis could be done, which would amount to an N-body integration of the halo masses, taking into account all forces that are interacting with the objects. This would allow the paths of the objects to be more closely predicted, although it would still be an approximation as growth and stripping of the halo would be difficult to account for. However in practice the haloes typically don't move a large distance between snapshots, and there are other issues that tend to dominate the matching. A good approximation can therefore be made by advancing the T_N halo along its velocity vector for half the time interval, and likewise projecting haloes at T_{N+1} backwards along their negative velocity vector for a similar time.

Then it is a case of trying to find a match of two haloes based on a combination of constraints. The first constraint being a distance from the predicted point. The search distance was fixed based on a value proportional to the peculiar velocity of the halo at time T_N and the virial radius of this halo. Thus more latitude would be given to fast moving haloes, and also those with large volumes.

A smaller distance is obviously a good constraint to place here, but distance alone is not enough to ensure a good match. In particular small haloes with as few as 20 particles can often spontaneously appear leading to bad matches. A reasonable secondary match is therefore mass, or its proxy such as number of particles.

Under normal circumstances haloes can be expected to grow by smooth accretion of dark matter over time, therefore in an idealised scenario a good match for a halo at time

T_N would be one with slightly more mass than it currently has. There are also cases though, where haloes may be stripped of particles particularly when orbiting larger haloes or undergoing other interactions. Thus it is possible in some cases for the mass difference to be negative. This is further amplified by deficiencies in the finders that locate the haloes losing haloes as they pass through each other. By empirical testing it was found that a factor of 4 increase, and 0.7 decrease gave reasonable results. The first such match, which is naturally ordered based on the k-d tree (a method for fast spatial look up) used for look up was accepted.

Other possibilities besides mass can be considered, in particular some serious consideration was given to other parameters:

- V_{max}, the maximum circular velocity of the particles, was initially used as a further constraint, based on its success at comparing haloes in previous projects. This proved to be no better than mass in most cases however.

- Another option tested was σ, the velocity dispersion of the particles. This can be quite a sensitive metric, and is well used in observational astronomy for its powers to derive relationships and reconstructive abilities (Old *et al.*, 2014). Whilst this does give a tight relationship for many cases, it does not match the outliers very well.

A number of other parameters were investigated, as shown in Figure 5.2 but the original choice of mass seems to be at least as good as any of the other options, and leads to some obvious arithmetic options when applied to the following steps (mass addition being the primary one), which are more difficult to assess with some of the derived parameters. For example what happens to the V_{max} parameter when two halos combine?

This phase is complete when as many haloes at T_N have been matched directly with an acceptable counterpart at T_{N+1}. This leaves potentially a set of T_N haloes that have found no suitable match. For haloes of small size this may be acceptable as finders can detect loose connections of particles as haloes, which then subsequently disperse. However it is more serious with larger haloes, and this leads to the next step.

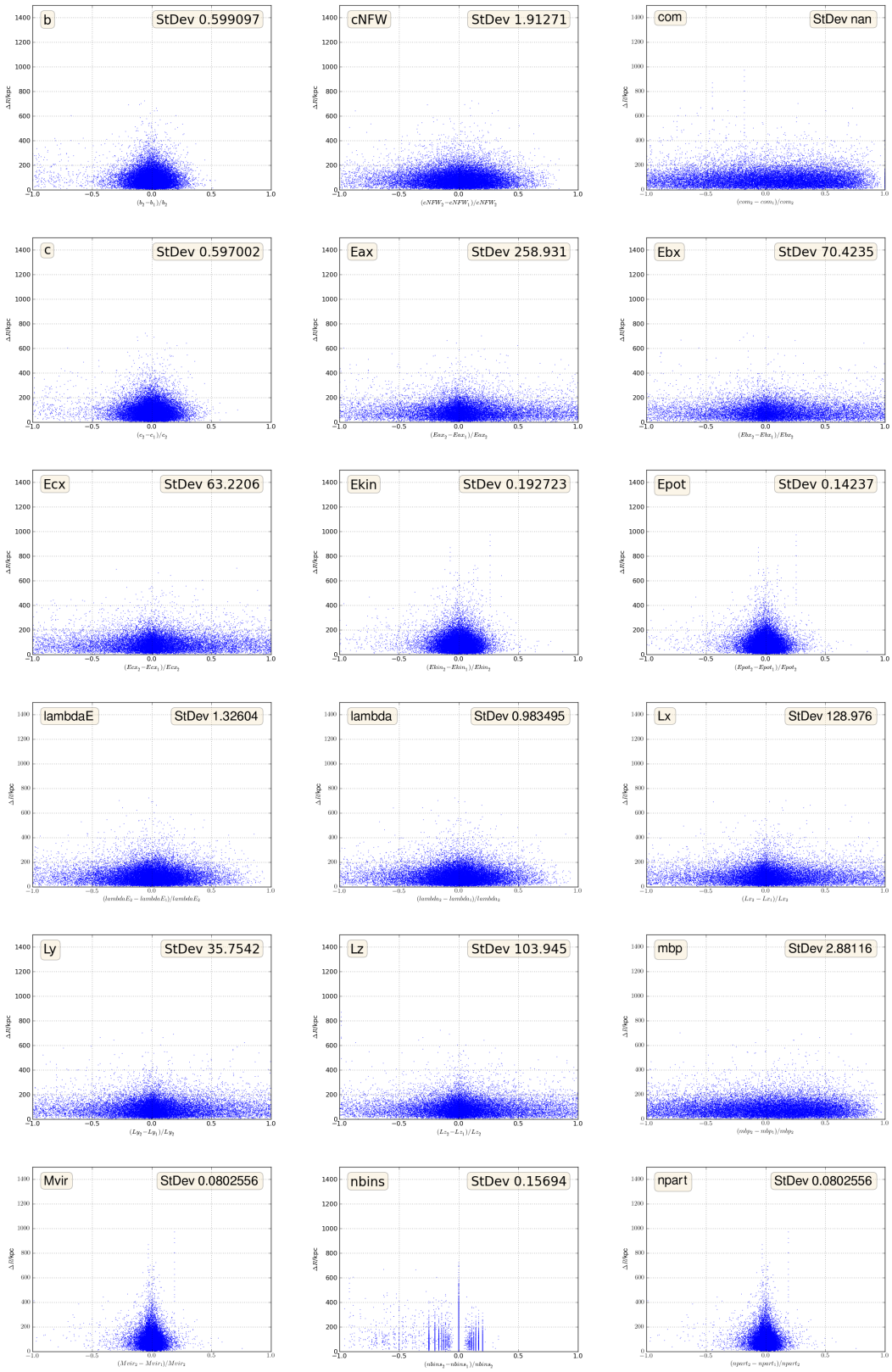

Figure 5.2: A comparison of a number of parameters used to match haloes taken from MERGERTREE results allowing the tightness of comparison to be seen. Each plot shows the relative difference in the parameter between the two matched halos (x-axis) and the difference in two haloes position. The standard deviation figure is inset into each figure. The tightest clustering around the mid point shows the best parameters. The parameters are described in Table 5.1

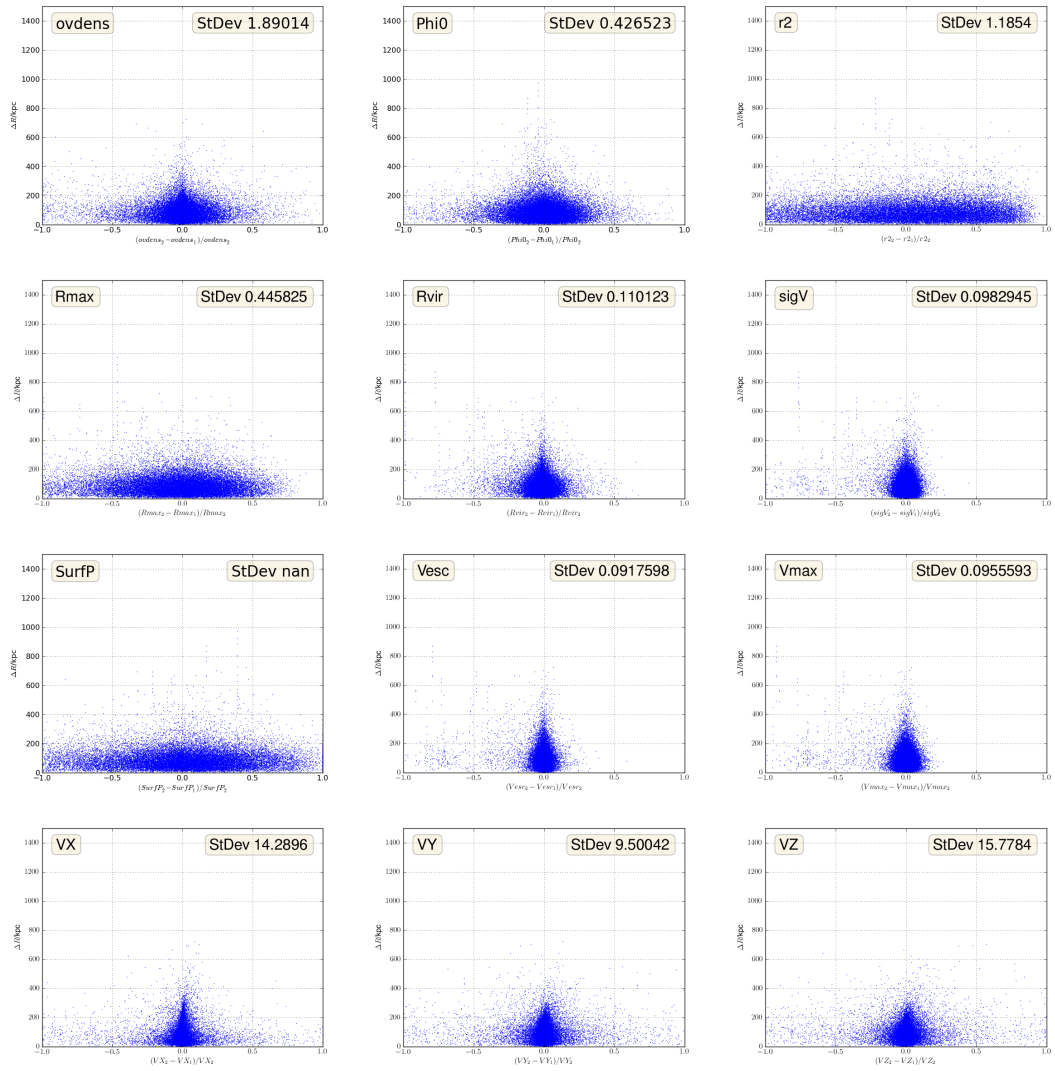

Figure 5.2 (cont.)

5.1.2 Major Mergers

Given a set of unmatched haloes, the next task is to try and match up combinations of haloes at T_N to form a single halo at T_{N+1}. This allows for two or more haloes to be combined to form a single combined halo. In this way two small haloes can combine to match a single larger halo which would be rejected for either of them, by the mass matching constraints in the first pass. Major mergers aren't a particularly common occurrence across cosmic time, but works such as Toomre (1977); Conselice *et al.* (2003); Conselice, Rajgor & Myers (2008); Lotz *et al.* (2011) indicate they do occur for most large galaxies - despite the exact frequency still being an area of investigation, so they should be expected for dark matter haloes too.

The operations for this phase mirror the first pass to some extent. A number of can-

Table 5.1: List of parameters plotted in Figure 5.2

Parameter	Stdev	Description
b	0.599	second largest axis to main ratio b/a.
cNFX	1.913	The NFW concentration parameter
com	–	Centre of mass to halo centre offset
c	0.597	the c axis relative to the a axis
Eax	258.931	Orientation of the X axis of a
Ebx	63.221	Orientation of the X axis of b
Ecx	63.221	Orientation of the X axis of c
Ekin	0.193	The kinetic energy
Epot	0.142	The potential energy
lambdaE	1.326	The Peebles dimensionless spin parameter
lambda	0.983	The Bullock dimensionless spin parameter
Lx	1287.976	Angular momentum x vector
Ly	35.754	Angular momentum y vector
Lz	103.945	Angular momentum z vector
mbp	2.881	offset between centre and most bound particle
Mvir	0.080	virial mass
nbins	0.157	number of bins used for profile
npart	0.090	number of particles
ovdens	1.890	overdensity at virial radius
Phi0	0.427	Unbinding energy parameter
r2	1.185	peak of ρr^2
Rmax	0.446	Radius at Vmax
Rvir	0.110	Virial radius
sigV	0.098	velocity dispersion
SurfP	–	surface profile parameter
Vesc	0.0912	Escape velocity
Vmax	0.096	maximum circular velocity
VX	14.290	Velocity in X direction
VY	9.500	velocity in Y direction
VZ	15.778	velocity in Z direction

didates are assembled based on proximity match, and then mass constraints applied by totalling up the T_N candidates mass, and constraining it to the T_{N+1} mass with appropriate adjustment for smooth accretion or tidal stripping.

In this phase using other matching criteria is more difficult. For instance a major merger would probably increase the velocity dispersion to a greater or lesser extent depending on how the interaction progresses and the incident angles relative to the major spin axis.

Once this phase is complete, again there may be a set of unmatched T_N haloes. These feed into the final stage.

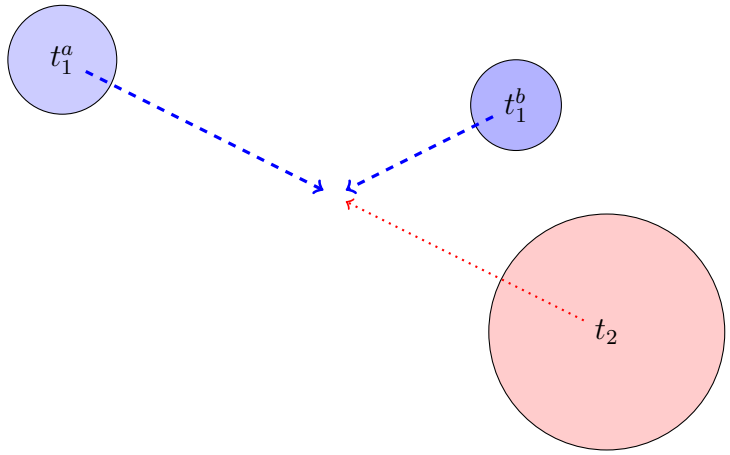

Figure 5.3: The JMERGE major merger step. Matching two or more haloes at T_N with a single halo at T_{N+1}. The combination of t_1^a and t_1^b together come to approximately the mass of t_2 and with roughly coincident locations when the $t_1^{a,b}$ are played forwards and t_2 played backwards in time, can make a good match and a major merger.

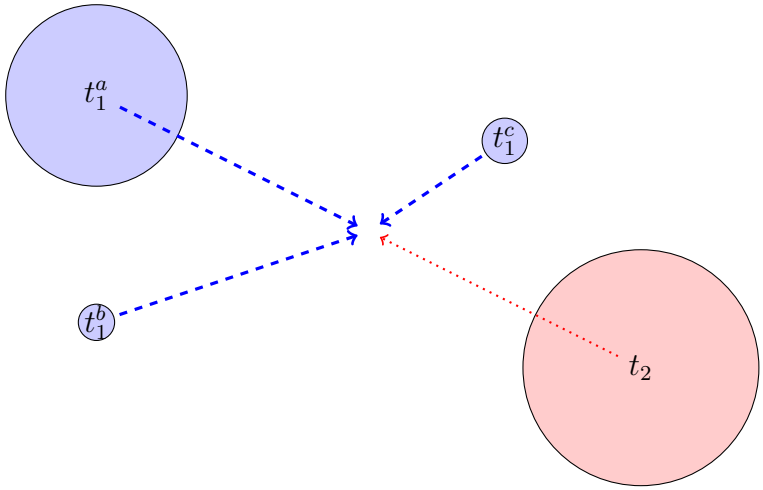

Figure 5.4: The JMERGE step of assigning minor mergers. In this case the tracking of halo t_1^a into subsequent t_2 has already been identified in step one, and no appropriate match found for t_1^b and t_1^c. However the location and the effect on the mass of t_2 are such that it is reasonable to assume that they have undergone a minor merger with t_1^a.

5.1.3 Minor Mergers

In this stage unmatched haloes are examined to see if they can be attached to already matched combinations. This is typically small sized haloes that are merged into larger haloes as minor mergers.

Similar metrics are used to evaluate these matches taking distance and combined mass into account when considering accreting these haloes.

5.1.4 Finalisation

This process works between two snapshots and produces as output a list of haloes at T_N that match those found at T_{N+1}. A subsequent process can then read this collection of outputs and assemble them into a tree in some suitable format for further processing.

5.2 JMERGE Results

The preliminary results of this algorithm are documented and published in Srisawat *et al.* (2013). They show a moderate success compared to the more traditional approaches but slower method of using particle ID's to match up haloes. However it is clear from some of the plots that the trees produced are often truncated or broken, not extending back across the snapshots.

Looking in more detail at this, some of the plots from the Srisawat *et al.* (2013) are given here to show the strengths and weaknesses of the approach.

5.2.1 Main branch depth

As in the Srisawat *et al.* (2013) paper, we can start by looking at how well the tree is constructed. This is done by tracing the main branch back through time and snapshots to indicate what sort of tree lengths are recovered. This shows how well the tree builder is doing on average at reassembling the haloes history. The results are shown in Figure 5.5.

The main branch depth shows the general trend of all the tree builders that were involved in the workshop. In most cases the tree length is expected to be reasonably long, having grown from an initial seed halo at early red shift, so the expected result is most haloes having long main branch lengths. In this case, the JMERGE algorithm holds up fairly well, fitting in with other tree builders following the detection of mostly long chain trees, with perhaps a few more shorter length trees evident in the haloes with mass above $10^{11}h^{-1}\mathrm{M}_{\odot}$.

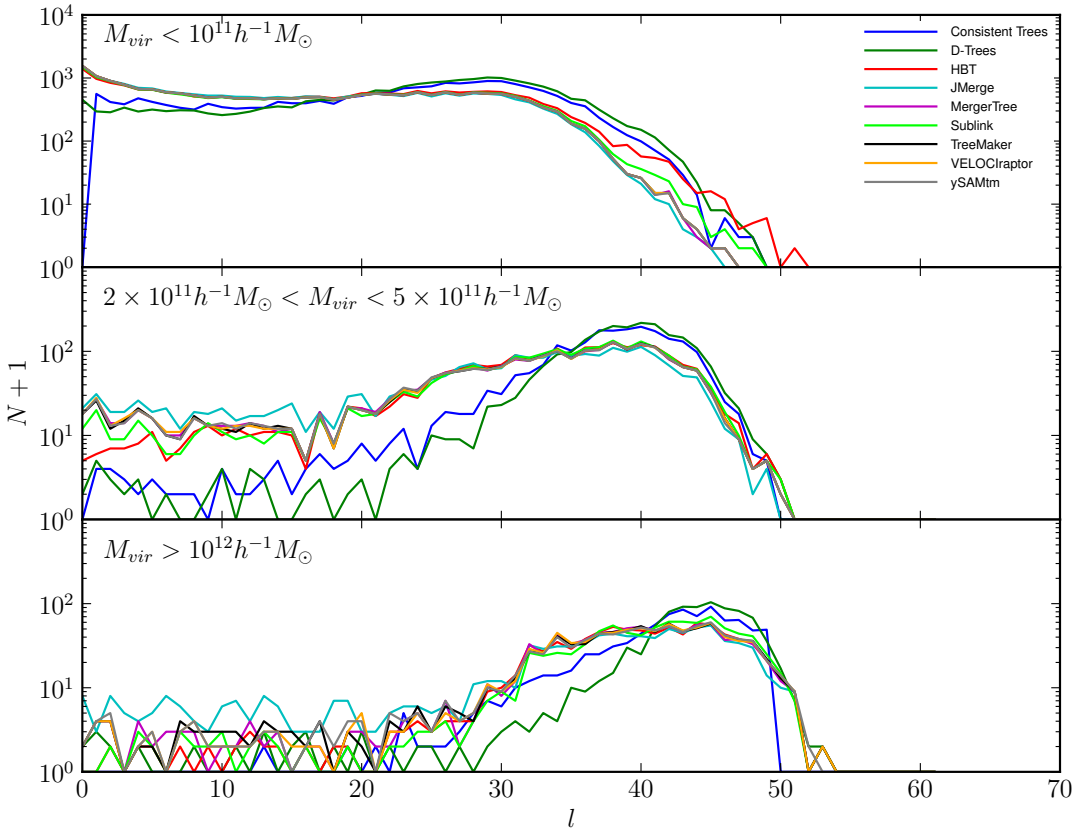

Figure 5.5: The length of the main branch for haloes identified at $z = 0$ (Snapshot 61). The ordinate is $l = 61 - S$, where S is the snapshot number at the high-redshift end of the main branch. The upper, middle and lower panels show the halo mass ranges at $z = 0$, as indicated in the panel, which correspond to roughly < 100, 200-500 and > 1000 particles respectively. Srisawat *et al.* (2013)

5.2.2 Branching ratio

Another metric to investigate is the branching ratio. This is defined as the number of direct progenitors making up the halo at each node in the merger tree. As might be expected the normal case is for one halo to have a single progenitor. This is the most common case as shown in Figure 5.6. Following on from this, the next most common case is for a halo to have zero progenitors, so therefore detected for the first time as a new halo.

In this test, JMERGE agrees closely with the other tree builders for the zero and single progenitor case. It is evident though that it does not find as many progenitors as some other algorithms, which can find up to twenty or more progenitors for a given halo in some cases. This is probably attributable to the multi-stage approach the JMERGE takes, preferring to find single matches in the first case, and looking for mergers in a separate step. This leads to it detecting matches of, for example, a small snapshot

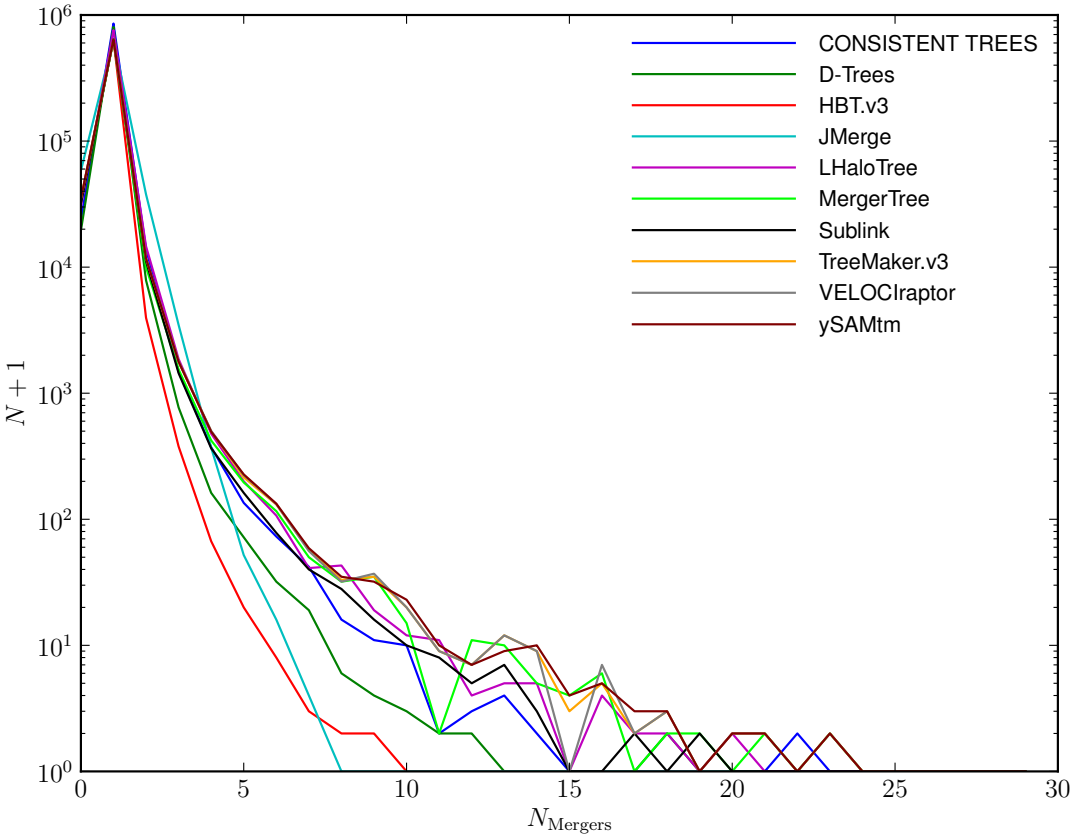

Figure 5.6: Histograms of the number of haloes with N_{dprog} direct progenitors, using all halos from $z = 0$ to $z = 2$. Srisawat *et al.* (2013)

$N-1$ halo with a newly formed snapshot N halo, and so then not looking for a merger event.

5.2.3 Misidentified haloes

The next metric of comparison shown in the paper was the misidentification of haloes. This was done by choosing a metric that gave a value for how far a halo had moved from its expected location. This is encapsulated in Equation 5.1. The value of Δ_r stays reasonably small provided the linked haloes do not show a non-uniform acceleration or there are errors in the linkage of the two haloes.

$$\Delta_r = \frac{|\mathbf{r}_B - \mathbf{r}_A - 0.5(\mathbf{v}_A + \mathbf{v}_B)\,(t_B - t_A)|}{0.5\,(R_{200A} + R_{200B} + |\mathbf{v_A} + \mathbf{v_B}|(t_B - t_A))} \tag{5.1}$$

Here $\mathbf{r}$ and $\mathbf{v}$ are the location and associated velocity of the two haloes under consideration. R_{200} is the radius of the halo enclosing 200 times the critical density, and t is

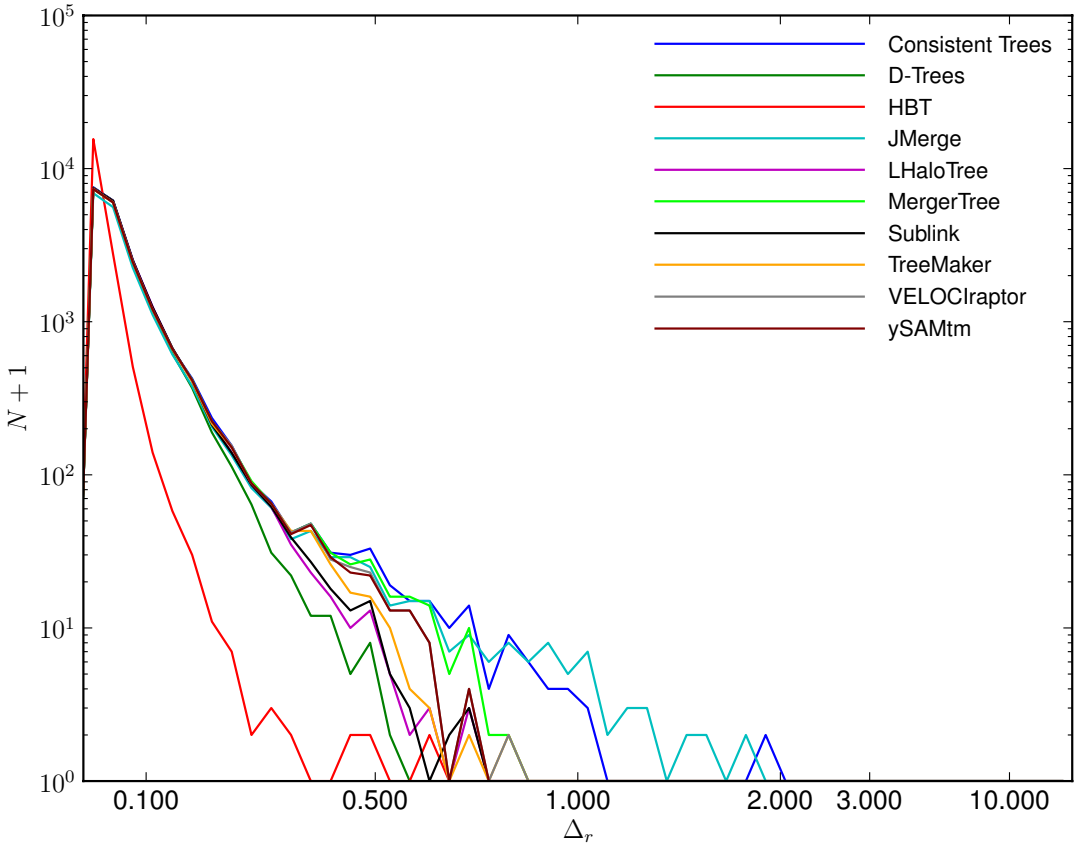

Figure 5.7: Histograms of the displacement statistic, Δ_r, for main haloes and their main progenitor for which both of them have $M_{200} > 10^{12}\,h^{-1}M_{\odot}$. The vertical lines show the 90^{th} and 99^{th} percentiles for MERGERTREE (but are approximately the same for all algorithms except HBT). Srisawat *et al.* (2013)

the time of the snapshots from which the halo instances were taken.

The results of this are shown in Figure 5.7. Whilst again JMERGE shows a good agreement for the majority of the haloes, it shows considerable deviation for a number, showing that it is probably not linking up the correct haloes in all cases. Values for this parameter approach 2.0 which puts it as a clear outlier when compared to the majority of the other tree builders.

5.2.4 Influence on Semi-Analytic Models

The use of JMERGE shows some agreement with the other algorithms in the previous analysis, it fairs much less well when the merger tree is actually used to produce results.

The following are some of the results where the combination of merger trees were used as input into a semi-analytic model, ySAM. The full details are given in Lee

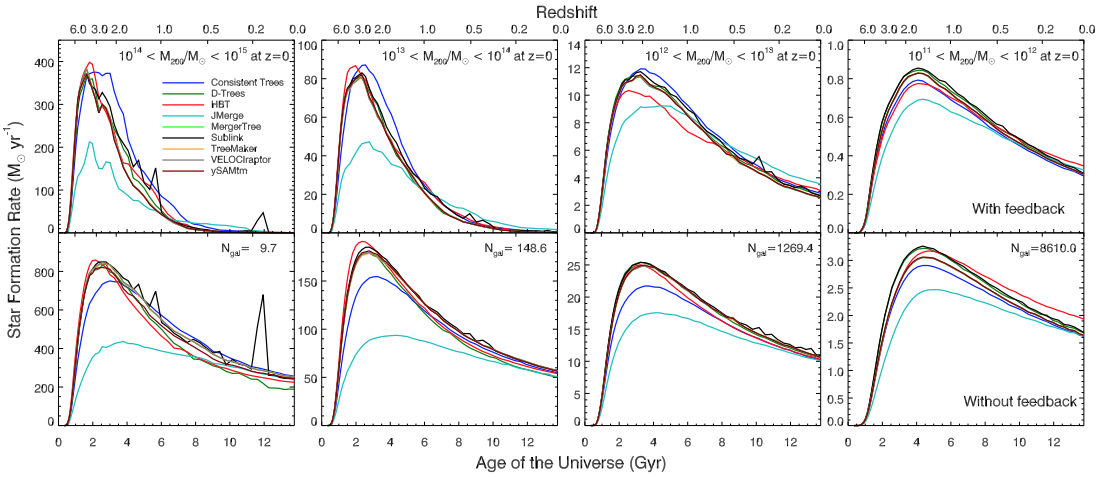

Figure 5.8: Star formation histories of main galaxies with respect to the M_{200} of haloes. The upper and lower panels show models with and without feedback, respectively. The colour coding, consistent throughout Srisawat *et al.* (2013) and this chapter, represents the nine algorithms. $\mathrm{N_{gal}}$ shows the mean number of main galaxies, averaged over different trees in each halo mass range. The upper panels have the same number of galaxies. (Lee *et al.*, 2014)

et al. (2014) but some of the illustrative plots from that paper are included below with commentary.

The first plot, Figure 5.8 shows the star formation of the main galaxies derived from the various merger trees, partitioned into four different mass ranges. The growth of the halo, and thus the resulting star formation rate is strongly correlated. Star formation is triggered in particular by mergers which are derived from the merger tree data.

As can be seen, the line for JMERGE is conspicuously below all the other tree builders. This probably relates back to the number of progenitors of a given halo, which is shown in Figure 5.6.

Another significant plot shows the number density evolution of galaxies. Figure 5.9 shows both the evolution in number count of central galaxies and satellite galaxies, and again it is very evident that JMERGE is a quite severe outlier in this case for central galaxies, and even worse for satellites. This is attributed to the shorter tree lengths produced by JMERGE.

5.3 Improvements

After the comparison paper (Srisawat *et al.*, 2013) was finished, the results were released and it was possible to use the data from other merger tree programs as a training

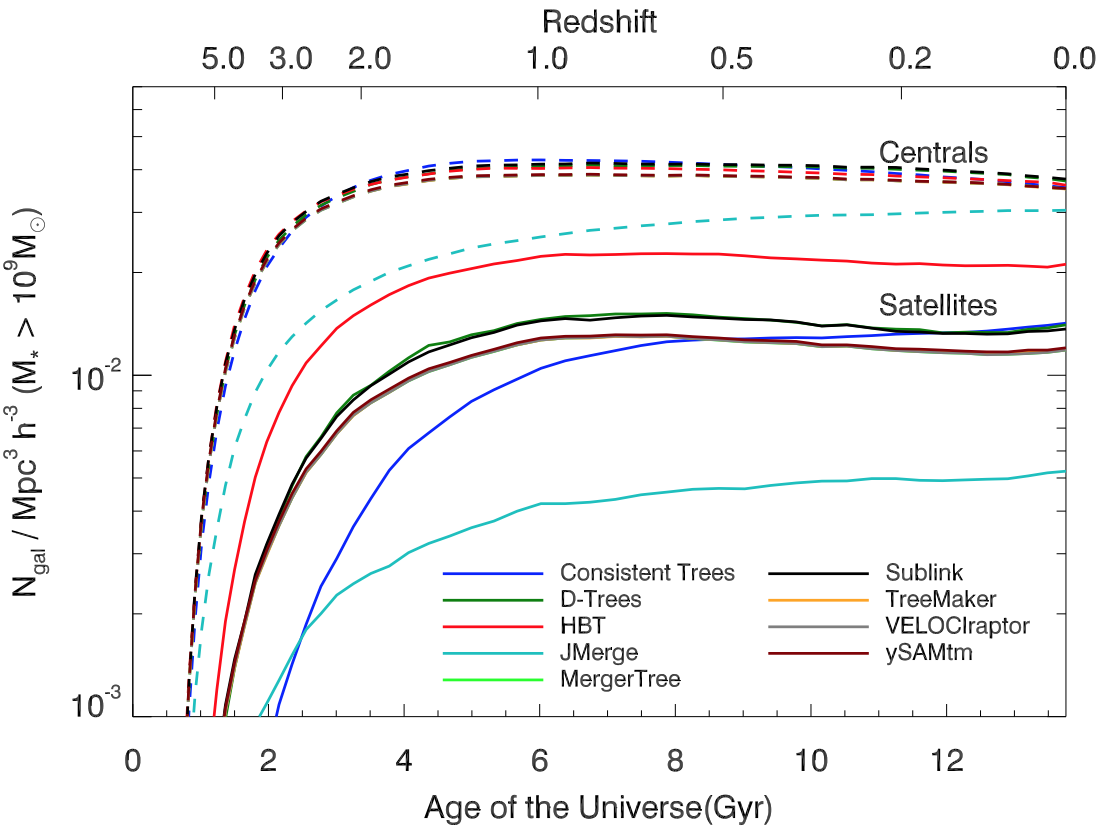

Figure 5.9: The number density evolution of galaxies more massive than $10^9 M_\odot$ in a co-moving volume. The dashed lines represent the number density of the central galaxies. The solid lines display the evolutionary history of the satellite number density. (Lee *et al.*, 2014)

set to try and improve JMERGE.

Swapping from a distance ordered list matched by mass to a metric that minimised the combination of mass and distance together, improved the results somewhat, but the results were still not comparable with the particle based tree builders.

5.3.1 Multi-snapshot spanning

Some possible ideas for improvement include searching across multiple snapshots to try and locate haloes. This caters for the case where haloes "disappear" in a given snapshot, when they are traversing a larger halo. While this could be levelled as a deficiency in the halo finders, which fail to track some haloes when conditions get noisy, in practice it is a deficiency that all current halo finders have to one degree or another. Therefore a strategy which attempts to locate subhaloes that do not match up in a suitable way, by skipping a snapshot and attempting to locate it in a subsequent one may counter the issue where long chains of haloes are not preserved in the tree.

This has been shown to work well in other particle based halo finders such as HBT (Jiang *et al.*, 2014) and Consistent-Trees (Behroozi *et al.*, 2013). The best scheme would be one using temporal information, and to scan ahead with sufficient time to try and relocate the halo, taking into account the typical crossing time of a halo. This would enable the rediscovery of the halo in a snapshot that would not necessarily be the next one taken, but might be several timesteps on.

5.3.2 N-body calculations

The current algorithm assumes that a linear prediction of the haloes final destination is acceptable, together with a fuzzy match of position. The accuracy of the position could be improved by doing a stepwise integration of forces based on the haloes gravitational interaction to improve the location. This would be at the cost of much increased computational complexity. If the timescale of the snapshots is relatively short the increased accuracy may not be significant and other factors may dominate the matching accuracy.

5.3.3 Void finding

An opportunity arose to test the algorithm for tracking voids rather than haloes. Voids are the inverse of haloes, being areas of underdensity rather than overdensity. There are a few issues with tracking voids, but the algorithm is basically the same. Voids are identified by a list of particles that are present in the void. They tend to congregate on the edge of the void, making an outline of it, with fewer particles in the centre. This is a natural result of void finders, which typically accumulate areas of underdensity up to some limit.

Another issue with voids is that they can be extremely large. As they grow across time, they tend to join together as matter increasingly congregates into the gravity wells. In some cases a major void can encompass many mega-parsecs. This conflicts with some of the strategies employed by JMERGE. In particular the scanning within twice or some other factor of radius of a halo when employed to large voids can mean considering a large fraction of the cosmological box in an attempt to match. This results in a very large number of potential matches to be considered which can make the algorithm

very slow. In these cases switching from the radius of the halo/void to locate matching structures, to a strategy just based on tracking centres but with a margin of error helps somewhat.

The results of applying the technique to voids are examined in more detail in Sutter *et al.* (2014).

5.4 Summary and Conclusions

The purpose of writing JMERGE was to compare it against particle based tree construction codes. The results can be summarised in the following points, and of course in more detail in Srisawat *et al.* (2013).

+ The algorithm is very fast compared to particle comparison methods. This is because there is obviously no matching of particle ID's and therefore runs much faster and requires very little memory in comparison, only needing the halo catalogs.

+ The algorithm can run in cases where particle data is not preserved, or generated in the first case. This is or has been the case for some simulation codes such as Harnois-Déraps *et al.* (2013).

- The accuracy compared to particle based tree codes is much reduced. The algorithm only matches gross features of haloes and predicts a consistent growth. Therefore it is not resilient to haloes missing for a time and then reappearing. If there is an ambiguous match between two or more candidates, there is little that can be done to break the tie other than pick the most likely.

- It tends to make much shorter trees than the particle based tree codes. This is largely down to losing track of haloes between snapshots.

This leads to the conclusion that preserving the particle data, and having access to it, despite vastly increasing the size of the saved data, is essential to constructing well formed merger trees.

Chapter 6

niFTy - Effects of the Mass Definition on Semi-Analytic Models

6.1 Introduction

Semi analytic models are a well known technique in simulations to allow the population of dark matter haloes with galaxies based on numerical recipes. They use this to approximate the baryonic physics that takes place. Whilst N-body simulations involving gas such as Vogelsberger *et al.* (2014); Schaye *et al.* (2014) are becoming more main stream, they are still considerably more compute time intensive to run, often by several orders of magnitude. Thus application of semi-analytic models to form representative galaxies in dark matter haloes are still an important tool.

Semi-analytic models fall into a number of categories, although most rely on an analysis of an underlying merger tree. The merger tree construction itself can therefore influence the final form of the analysis if not well defined (Srisawat *et al.*, 2013; Avila *et al.*, 2014) and therefore can result in changes to the results of the SAM model as shown in Lee *et al.* (2014). The merger tree in turn relies on the finder code to locate and characterise dark matter haloes for which the following investigations have already been carried out Onions *et al.* (2012); Knebe *et al.* (2013).

There are other techniques that rely on halo occupation techniques which work from similar inputs but use different techniques.

In this study, we look at what effect using different definitions of the mass might make on the outcome. We use the data that came out of the "niFTy cosmology" [1] workshop held in Madrid.

The data is described in Section 6.2, the results are shown in Section 6.3. We summarise and conclude in Section 6.4.

6.2 The Data and Models

6.2.1 The Data

The data used was a 62.5 h^{-1}Mpc box generated by the N-body code GADGET3 Springel (2005). The cosmology used the parameters from WMAP-7 data (Jarosik *et al.*, 2011), and used 270^3 dark matter particles each of mass $9.31 \times 10^8 h^{-1} M_\odot$.

This was then analysed to produce halo catalogues with SUBFIND (Springel *et al.*, 2001) and assembled into merger trees using MERGERTREE from the publicly available AHF package (Knollmann & Knebe, 2009). Both the finder and mergertree packages have been found to be robust in recovering structure and hierarchies (Onions *et al.*, 2012; Srisawat *et al.*, 2013).

The mass definitions used in the halo data were the following:

- FOF Mass - this is mass based on the friends-of-friends grouping, taking all particles in the group without further refinement.
- Bound mass - this is the FOF mass given above, but refined by removing all particles that are not gravitationally bound to the halo.
- 200c - this is the definition of the halo based on a BOUNDmass that is restricted to a limiting density of 200 times the critical density of the universe. Thus particles may be removed once the halo reaches 200 times the critical density to truncate the halo.
- 200b - this is the definition of the halo based on a density of 200 times the

[1] `http://popia.ft.uam.es/nIFTyCosmology`

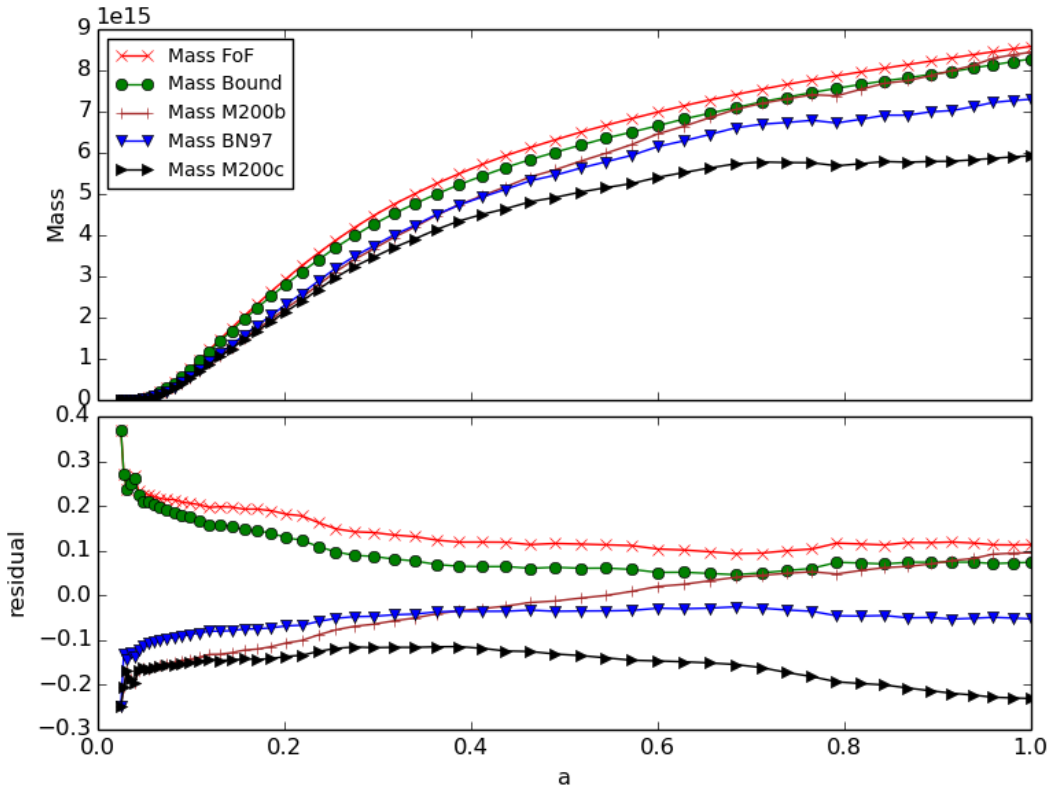

Figure 6.1: The mass growth over time for the haloes using the five different mass definitions for the finder. The time is from redshift 49 ($a = 0.02$ to the present day. The residual shows the difference from the average of the masses.

average (or background) density of the universe. Note this value changes with the expansion of the universe.

- BN97 - this uses the definition of mass based on (Bryan & Norman, 1998). This reflects the commonly used tophat collapse density.

The relative growth of these different masses is show in Figure 6.1. The analysis is done at redshift zero to allow the maximum effect and also there is more data available at redshift zero, as some of the HOD SAM models were only run on the final snapshot.

6.2.2 The SAM Models

The semi analytic models chose a mass to calibrate to, then ran their models. They then ran the models again with every other mass definition without re-calibrating the models. In this way we can see the effect of the different mass definitions on the output of the semi analytic models.

Name	Calibration Mass	Reference and Notes
GALACTICUS	BN98	Benson (2012)
GALFORM	BOUND	Durham (Gonzalez-Perez *et al.*, 2014)
LGALAXY	M200c	Munich (Henriques *et al.*, 2013)
MORGANA	FOF	Monaco, Fontanot & Taffoni (2007)
SAGE	M200c	Durham based (Croton *et al.*, 2006)
SAG	BOUND	(Lagos, Cora & Padilla, 2008)
SKIBBA	M200c	HOD based (Skibba & Sheth, 2009)
YSAM	M200c	Lee & Yi (2013)

Table 6.1: The participating codes and their mass definitions that they calibrate to.

The SAM models are described more in Knebe *et al.* (2015) and the references described there in, but fall broadly into the categories shown in Table 6.1

The initial data runs took place at the nIFTy workshop in Madrid, comprising fourteen different codes, however not all were able to complete the data comparison, so only those that did all the different mass data runs are analysed here.

6.3 Results

The following results were found.

In Figure 6.1 we show the underlying mass growth of the different mass definitions plotted against expansion factor. It is noted that the FOF mass as expected is the largest of the mass definitions, with the others appearing in the order shown. Note that the 200c definition is the smallest of the mass definitions, and that the remaining three definitions cross one another over expansion time. At redshift zero they appear in the order of mass high to low as:

1. FOF
2. 200b
3. BOUND
4. BN98
5. 200c

The simplest effect expected of the mass difference on the resulting SAM models would be for the change in mass to be reflected directly in the various mass quantities calculated in the SAM models. There may be more complexity going on within the model that makes this simple prediction invalid, but it makes a good starting point for comparison.

Most of the graphs shown from here on show the relative difference for each of the SAM models compared to their FOF mass findings. In addition, the expected mass offset is taken into account for the mass plots, by removing what the expected difference would be using the difference in mass definitions at redshift zero taken from 6.1. Thus the simplistic expected result would be all lines would lie on top of each other if there was a direct relationship between input halo mass and output derived mass quantities. The points are connected with dotted lines to guide the eye in following a particular quantity, and do not imply a connection.

6.3.1 Stellar mass

First we look at the total stellar mass across the SAM models at redshift zero in Figure 6.2. Given a change in halo mass, the expected output might be that a reduction in dark matter halo mass would result in fewer stars in the occupying galaxies constructed.

One would expect the stellar mass quantities to follow those shown in Figure 6.1 with all mass definitions being less than the FOF Mass. Whilst this is broadly true, GALFORM finds a larger mass than the FOF mass for the 200c case, and GALACTICUS finds nearly the same for both 200c (the smallest mass definition) and BOUND mass as for the FOF mass. LGALAXY and SAG follow more closely the original mass definitions. SAGE meanwhile finds almost no difference in stellar mass no matter which definition is used.

6.3.2 Hot gas mass

Turning to the mass recovered in the hot phase (heated gas), this is shown in Figure 6.3 In this case GALACTICUS finds very little difference in the mass of hot gas found, no

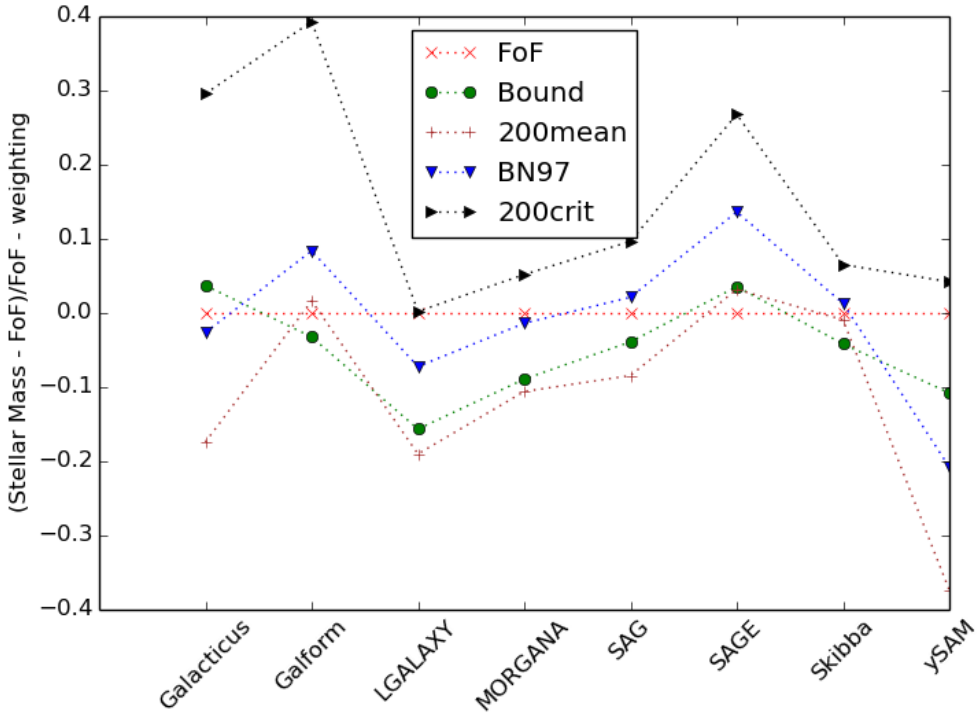

Figure 6.2: The stellar mass at redshift zero for the different SAM models each using the different mass definitions. The y axis shows the relative difference from the FoF mass in terms of $(\mathrm{stellarmass} - \mathrm{mass}_{\mathrm{FoF}})/\mathrm{mass}_{\mathrm{FoF}} - \mathrm{mass}_{\mathrm{expectedincrease}}$ which takes into account a simplistic dependence on the difference shown in Figure 6.1.

matter which mass definition it starts with. The same is largely true for GALFORM, although the bound mass (normally second in rank) is anomalous. SAGE again shows to be little influenced by the mass in use. The remaining SAM models find results broadly in line with the mass ranking of the input.

6.3.3 Cold gas mass

Now looking at the total material held in cold gas mass as shown in Figure 6.4, we again see perhaps unexpectedly high results for both GALACTICUS and GALFORM. GALACTICUS shows either little change in total cold mass for the different definitions, or slightly above the FOF mass, whilst GALFORM shows a small increase in cold mass for the smaller mass definitions. SAGE again shows little influence by the different mass definitions.

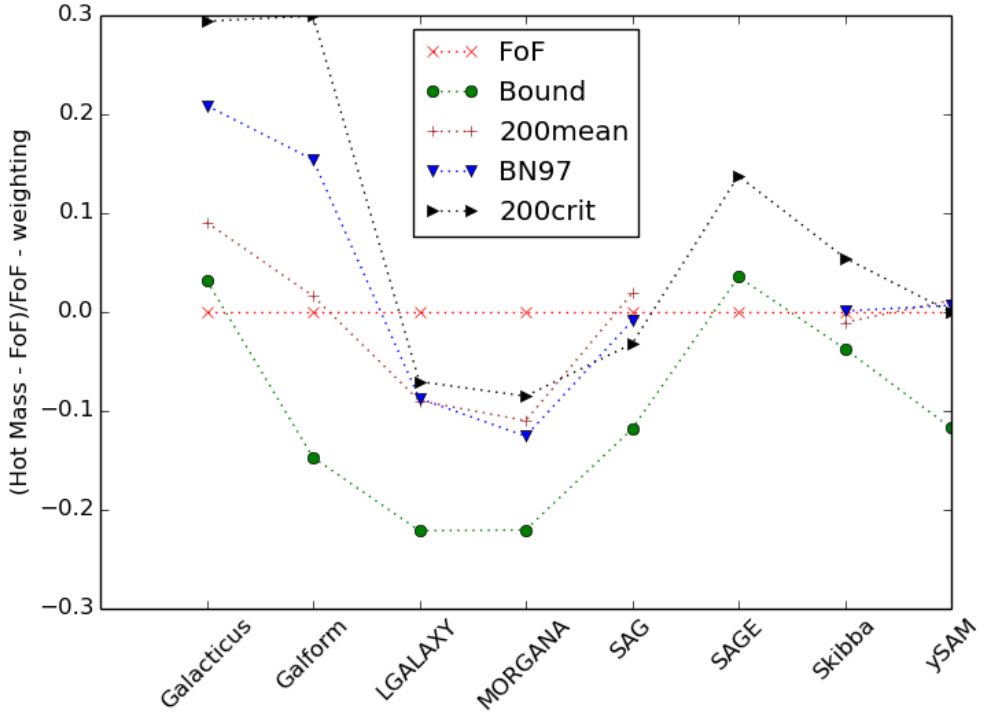

Figure 6.3: The difference in the mass of hot material at redshift zero relative to the FoF mass definition for different SAM models.

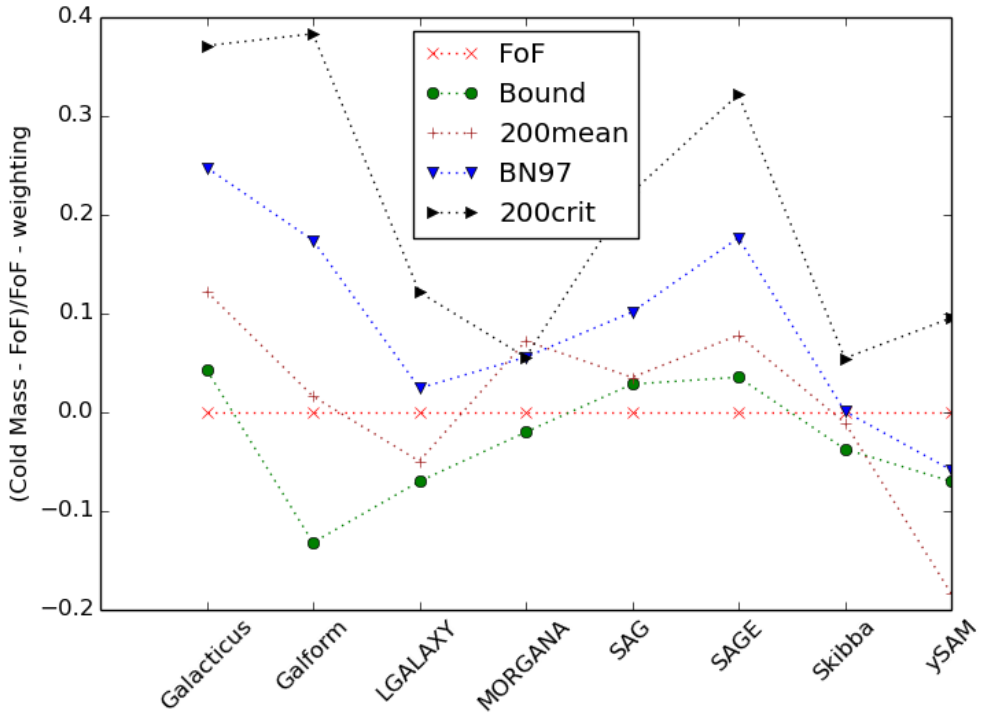

Figure 6.4: The relative difference in total mass at redshift zero for matter in the cold phase.

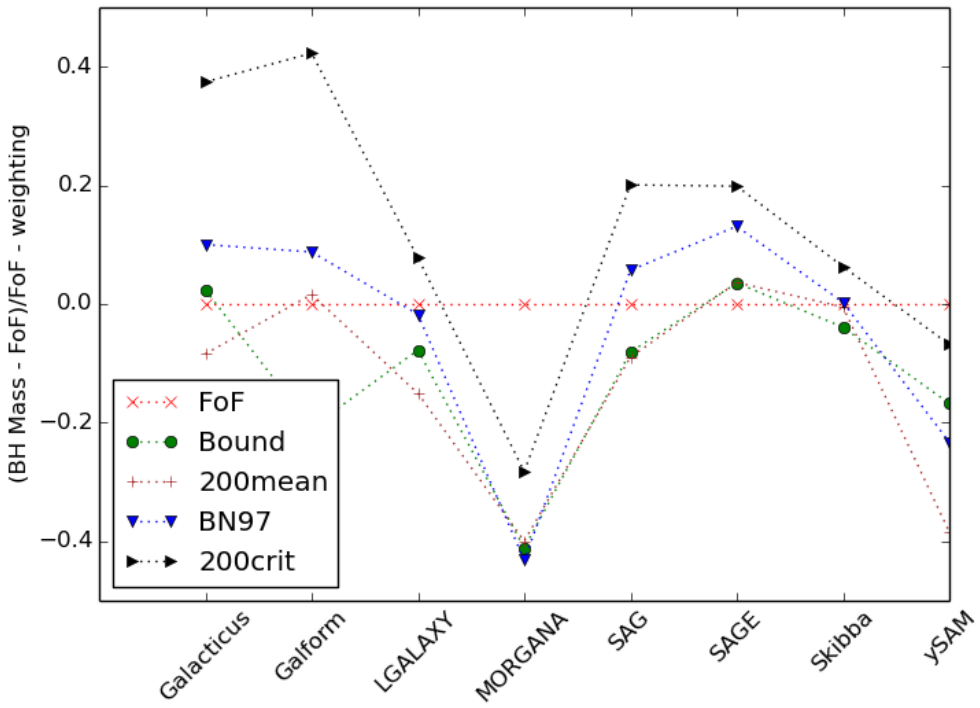

Figure 6.5: The relative quantity of mass found in central black holes relative to the FoF definition at redshift zero.

6.3.4 Central Black hole mass

Finally we show the mass concentrated in the central black holes in Figure 6.5. Again both GALACTICUS and GALFORM show unexpectedly high masses for the 200c mass definition. MORGANA has a substantial decrease in the central black hole mass for all the different mass definitions. The others seem to follow the input mass more closely.

6.3.5 Star formation rate

Looking now to non-mass quantities, we start with the star formation rate as shown in Figure 6.6. These graphs do not have an adjustment for the expected difference in mass that was shown in Figure 6.1.

Star formation is obviously dependent on many quantities, but in particular relates somewhat to the quantity of cold gas available(Robert C. Kennicutt, 1998; Schmidt, 1959) at any one time, although strongly influenced by AGN and supernova feedback. However again, simplistically increasing the overall total mass one would expect to

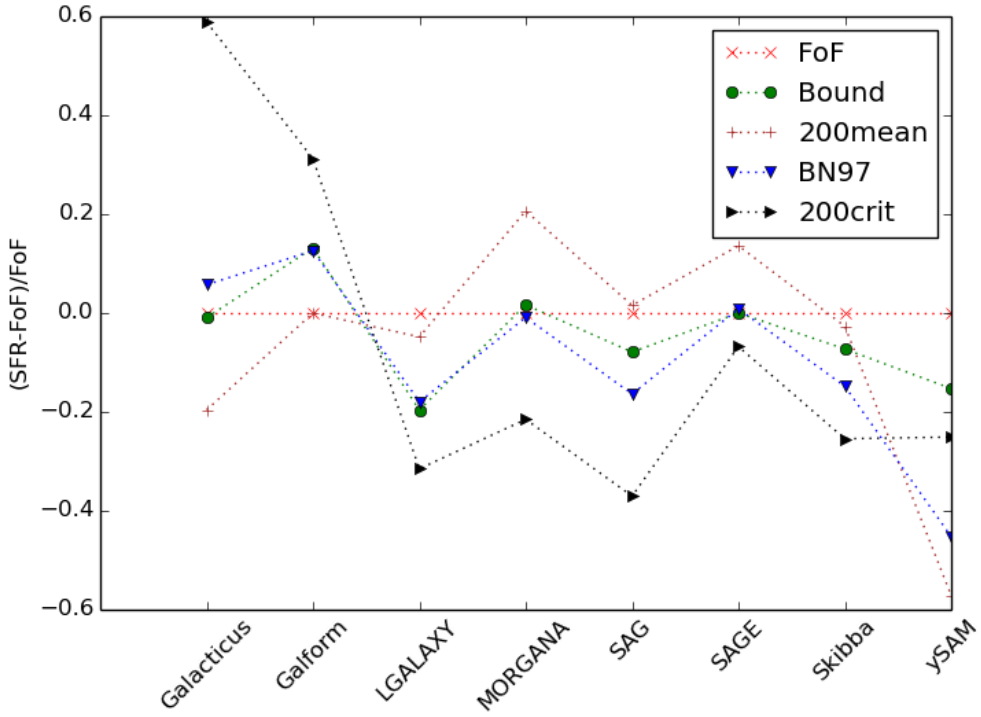

Figure 6.6: The difference in median star formation rate at redshift zero compared to the FoF mass definition for the different SAM models.

feed through in some cases to more available mass for star formation.

It is evident that the 200c mass definition leads to a much larger star formation rate for GALACTICUS and a similar though lesser effect for GALFORM. The other SAM models match the more simplistic model of mass decrease although YSAM shows more mixed results.

If we compare with the changes in cold mass shown in Figure 6.4 there is a overall broad trend of the SFR following the cold mass changes. However GALACTICUS, GALFORM and YSAM all vary from the changes related to cold mass, particularly in the 200c case.

6.3.6 Metallicity

Finally we look at the effect on the metallicity of stars at redshift zero shown in Figure 6.7.

For YSAM the metals decrease in a similar way to mass, which is also the case for

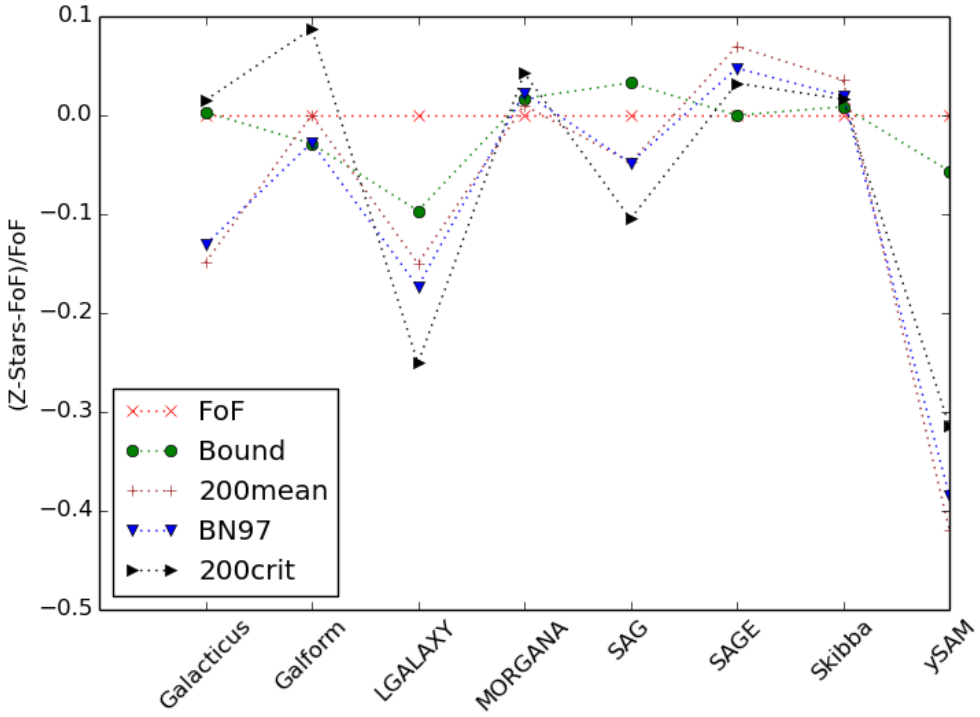

Figure 6.7: The difference in median metallicity compared to the FoF Mass input for the different SAM models.

LGALAXY. MORGANA and SKIBBA are almost unaffected by the change in mass, as is SAGE. GALFORM shows a increase in metals for the 200c case, but other masses are unaffected.

6.4 Summary and Conclusions

This chapter has shown that a change in the mass definition can have quite wide ranging results in the output of semi-analytic models. It can change the results by a factor of $\pm 40\%$. In some cases these are perhaps as might be predicted, but in other cases the results are quite at odds with what might be naively expected from a simple change in mass.

It should be noted that each SAM model general expects to work with a particular mass definition, and so having an unexpected mass definition may well confuse some of the internal workings of algorithm, leading to incorrect virial radii and perhaps circular velocities, in turn unsettling some assumptions made. So this is not an unexpected

issue.

This doesn't argue that any one of these models is right or wrong, in how they approach the change in mass, as it is not clear exactly what the correct answer might be. It does illustrate that SAM models can be extremely sensitive to the mass definition input, and that it is therefore important to be sure of the definition fed into the model. In some cases this can influence the output by more than 50 percent of the expected values.

Chapter 7

Tracing the possible progenitors of ultra-compact dwarf galaxies

7.1 Introduction

The current paradigm of ΛCDM structure formation revolves around the dark-matter-led hierarchical construction of gravity wells, which merge under the influence of gravity to form the larger structures observed today (White & Rees, 1978). The density of these haloes is somewhat set at the time of their initial formation as it is dependent on the density of material at the time.

When observable galaxies form, as the baryons cool and condense within these dark matter gravity wells (White & Frenk, 1991), smaller structures fall into the main halo undergoing merging and tidal interactions as the components reach a stable formation (Ghigna *et al.*, 1998).

Aside from the regular collection of elliptical, spiral, lenticular and dwarf galaxies, and their smaller counterparts the globular clusters, ultra-compact dwarf galaxies are a recent discovery (Phillipps *et al.*, 2001; Drinkwater *et al.*, 2003). They are characterised by a small galactic mass, of the order of $10^8 M_\odot$ and a half light radius of around 20 pc (Brodie *et al.*, 2011). They typically have a luminosity comparable to faint dwarf galaxies - around $10^7 L_\odot$, but with a morphology that is very different. They have an old stellar population, but are larger and brighter than globular clusters. Their mass-to-

light ratio is higher than for globular clusters (Dabringhausen, Hilker & Kroupa, 2008) which indicates either an anomalous mass function, or a small scale dark matter concentration. It is possible that at least some of them contain super massive black holes (Seth *et al.*, 2014), which would indicate they share a formation history somewhat similar to regular galaxies, as opposed to globular clusters.

Such compact structures are most easily formed early on in the universe when it was much smaller and phase-space density could be large within such a structure. As the universe expands and galaxies accrete more mass, such compact structures are harder to form. The possibilities for UCD construction are broadly split into three types:

1. The first is from galaxies that evolved as normal, but end up through tidal interaction with neighbours getting stripped of their outermost stars, leaving a condensed nucleus (Bekki, Couch & Drinkwater, 2001; Thomas, Drinkwater & Evstigneeva, 2008). We refer to this as a type one UCD (UCD-I).

2. Another option is that the galaxies evolve early on, and grow rapidly to begin with, but then are starved of further material and so stay as a compact object that doesn't grow further, a possibility discussed briefly in Drinkwater *et al.* (2004). These objects have to form early and stay compact. This is termed the type two UCD (UCD-II) in this chapter.

3. They may also form from the merger of stellar super-clusters where young massive star clusters form in collision between gas rich galaxies (Fellhauer & Kroupa, 2002). We don't consider this population in this chapter as these are not easily tracked in this simulation.

In all cases however the core has to be compact, and this is achieved most easily by being formed when the universe was considerably denser than it is now.

7.2 Method

We used an N-Body simulation running with GADGET3 configured with the PLANCK cosmology ($h = 0.6777$, $\Omega_m = 0.307115$, $\Omega_\Lambda = 0.692885$, $\Omega_{Baryon} = 0.022032$)

initial conditions, using a 125 h^{-1}Mpc box size. This is the first step in a process that we plan to expand to far bigger box sizes once the analysis techniques are tried and tested.

The haloes were then recovered with the ROCKSTAR halo finder (Behroozi, Wechsler & Wu, 2011), and a merger tree built with consistent-trees (Behroozi *et al.*, 2013), both of which have proved robust tools for such analysis (Onions *et al.*, 2012; Srisawat *et al.*, 2013).

The halo and merger tree catalog was then processed through a semi-analytic model GALACTICUS (Benson, 2012). A sample of the largest haloes were chosen at a redshift of six, matching the early formers, and then traced through to the present day using the merger tree structures. With the semi analytic model catalogs noted above, it is then possible to see what type of galaxies inhabit these haloes both at early and late times, and to select on galaxy size as well as halo size. It can also be done the other way around, by selecting the largest galaxies at redshift six, finding their associated haloes and then following them through to the present day.

7.3 Results

We start by looking at how haloes evolve over time. A point in the past is chosen, in this case redshift six, as this is around the epoch where the first galaxies are generally thought to form. We then select all dark matter haloes above $10^{11}\mathrm{M}_{\odot}$ at this redshift, as these are those that have formed large and early. These haloes are traced through the merger tree to redshift zero to see how they evolve. When we compare the mass of the halo at redshift six with its corresponding final mass we find a few anomalies. We would expect these haloes to be the progenitors of the largest present day haloes, as they have a head start on all the others in terms of mass growth. Some of the redshift six haloes in Figure 7.1 appear not to grow at all staying at or even below the line of constant mass. The environment these haloes end up in is also included, with the points being colour coded with the average density of halo mass in the surrounding 8 h^{-1}Mpc sphere.

The general trend is for the more massive redshift zero haloes to be in a denser envi-

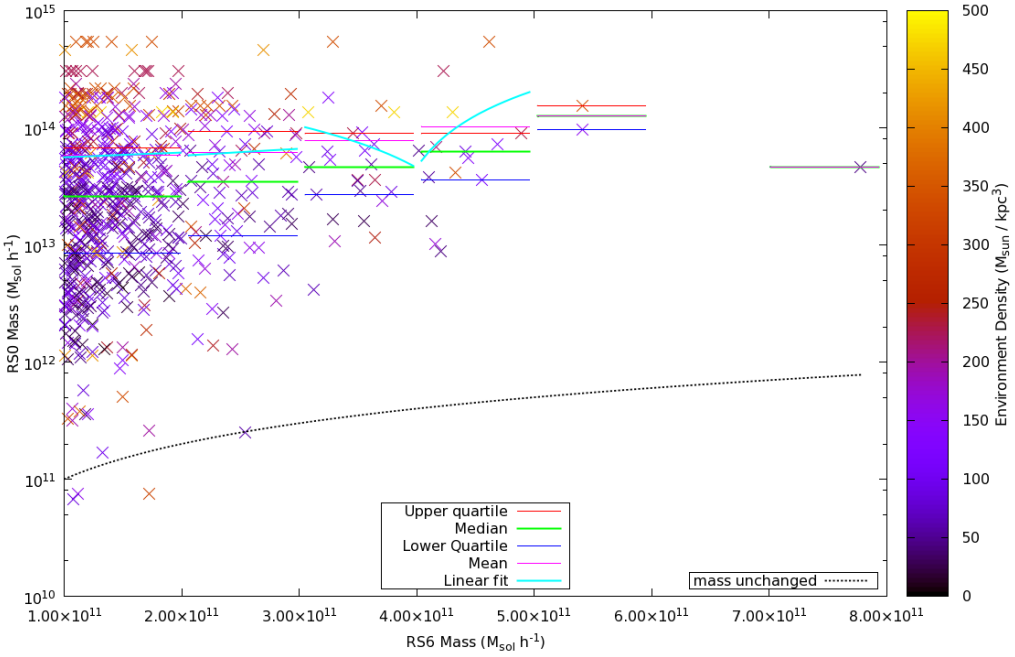

Figure 7.1: A comparison of the halo mass as it evolves from redshift six to redshift zero, showing the growth of haloes. The points are coloured by environmental density which is defined as the average mass in haloes within an 8 Mpc sphere centred on the given halo.

ronment, and for the median mass to increase with the initial redshift six mass. This is an expected trend, with the larger starting masses generally accreting more mass over time than smaller starting masses, and those in higher density environment having more chance to grow by mergers. There are some inconsistent points however, with some of the lower redshift zero mass haloes also being in high density environments.

It also appears that there are some haloes that do not grow at all, and some in fact shrink, indicated by a few points at or below the line of zero mass growth. However if we look at how some of these smallest haloes evolve over time, as shown in Figure 7.2 we can see that this is an artefact of their environment. The haloes do grow to a reasonable size, but then as a result of infall into a bigger halo, the original is gradually stripped of particles. As the dark matter particles are stripped and the radius shrinks, eventually there will not be a large enough dark matter halo to keep hold of the outer stars, where the rotation curve is mostly flat. This does require a substantial part of the halo to be stripped, but will lead to stars being stripped both by loss of mass of the halo/galaxy and so being flung out, and by gravitational attraction of the larger host halo stealing away the looser bound stars and dark matter material.

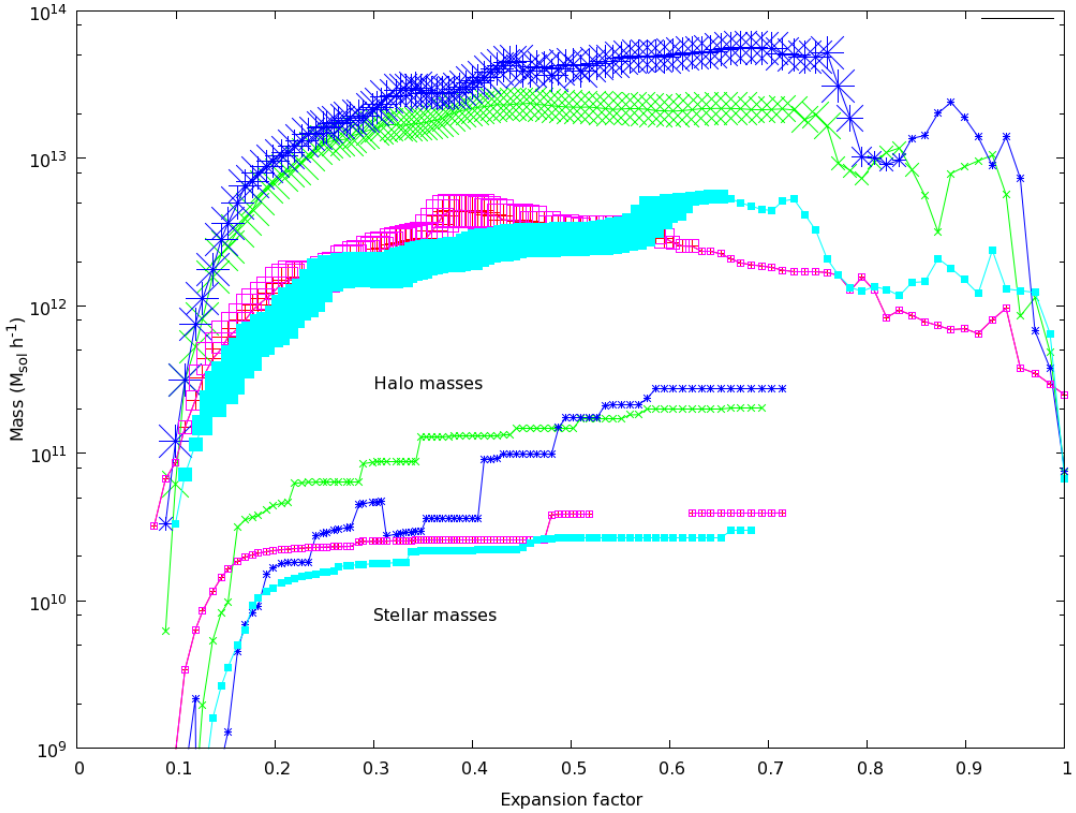

Figure 7.2: The evolution in mass for the haloes in Figure 7.1 that are below $10^{11}M_{\odot}$ ayt $z = 6$, showing their growth between redshift six and zero, and the rise and fall in mass as they get stripped. The size of the point on the line is indicative of the width of the merger tree at that point. Plotted below is the equivalent stellar mass of the largest galaxy contained within the halo. The semi-analytic model loses the connection between the halo and the galaxy at about expansion factor of 0.7.

Indeed if we follow the stellar mass of such haloes, as shown plotted in the lower curves in Figure 7.2, we can see a rise, and then flattening of the stellar mass, up to the point they start getting stripped, at which point the semi-analytic model loses the association between the halo and the galaxy.

Such haloes are good candidates for the UCD-I category. They grow rapidly at early times, but then shrink as they are stripped of much of their mass. Once the dark matter mass is stripped sufficiently, it will start to have an effect on the baryons and stars. It is well known that stars in most galaxies are moving too fast to be held by the gravity of the baryonic matter alone, as explained in the introduction. Therefore, once the dark matter haloes get to be small, the outer stars will be shed. The inner stars, especially of those galaxies formed early on have high phase space density, so will be held as a tight core. The resultant galaxy will just be the core component of the original galaxy.

An example of one of these haloes is shown in Figure 7.3, where it grows early on in a

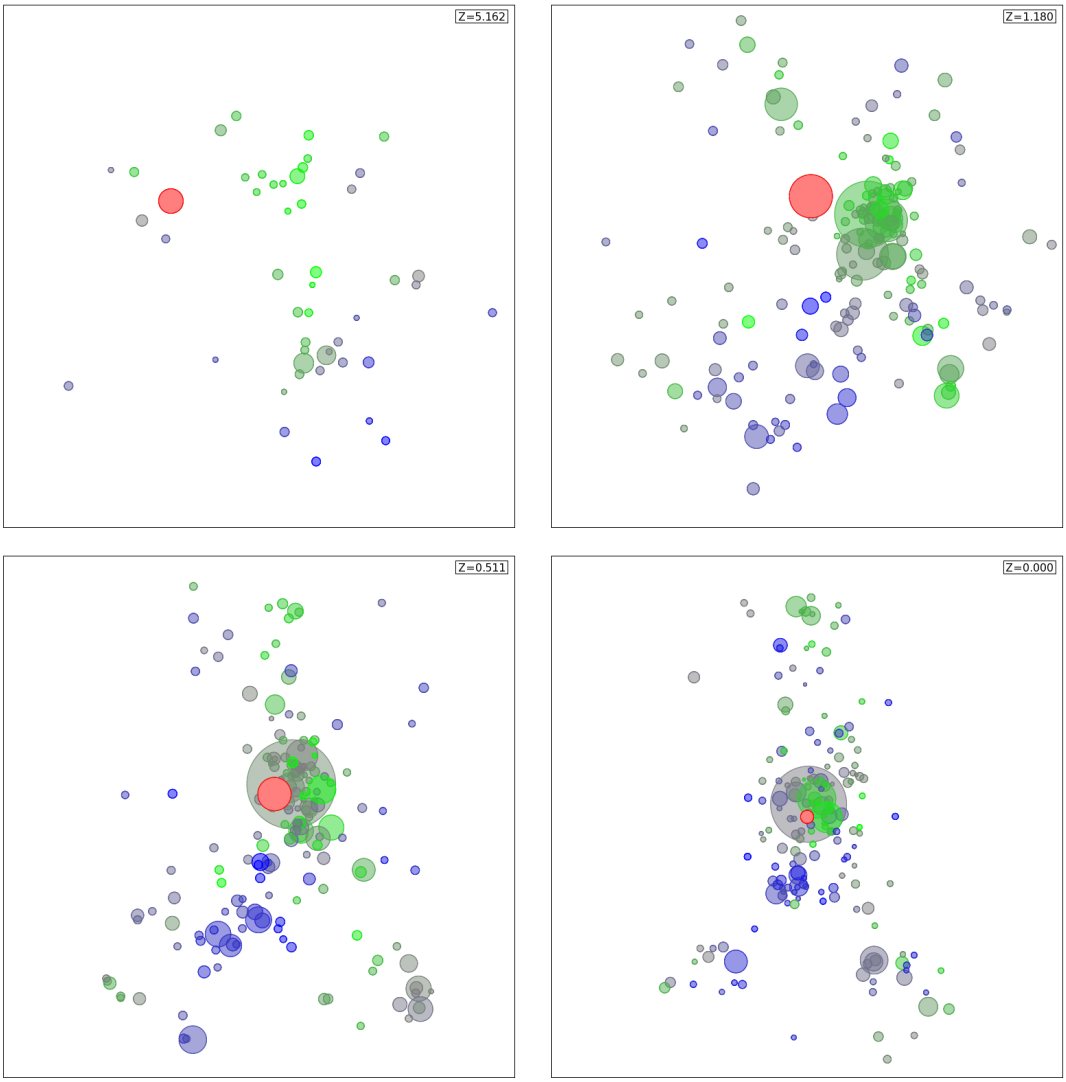

Figure 7.3: Snapshots at a sequence of indicated redshifts showing the growth and stripping of one of the haloes (in red) in Figure 7.2. Plotted as X and Y coordinates, with radius, and Z coordinate encoded in the colour. It shows the progression of the halo in red $z = 5.2$ to $z = 0$ growing in the first two panels and then being stripped on infall in the latter two.

more isolated environment, but is then drawn in to a bigger halo and stripped of much of its mass, ending up smaller than its redshift six mass. This dark matter halo will soon disappear merging into the main halo, but any attached galaxy may survive for a considerable time as a tightly bound structure making it a reasonable UCD-I body. In a study, Moore, Lake & Katz (1998) showed for simulated galaxies undergoing harrasment, approximately 50% of the stars of harrased galaxies were lost to the interstellar medium, which agrees with production of UCD-I processes considered here. These stars would naturally be those outermost ones, dragged away in tidal tails as their study shows.

In contrast the haloes that are in more isolated environments grow quickly when the universe is small, but then stagnate. This is shown in the mass plots in Figure 7.4.

In these cases, the stellar mass essentially flat-lines despite a very small growth in the dark matter halo.

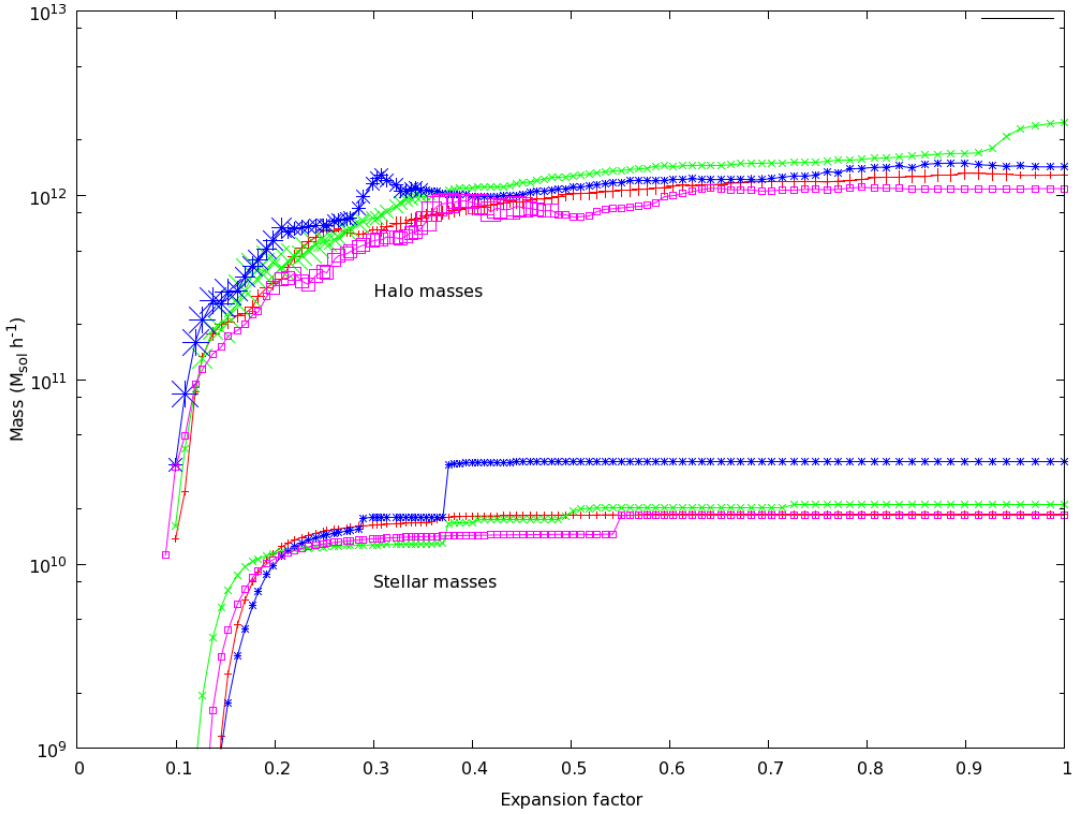

Figure 7.4: The evolution in mass for the haloes in Figure 7.1 that are low in bushiness. The size of the graph symbol is indicative of how many mergers (of any size) this halo undergoes. A few cases are obvious early on, but they soon settle down to just gradual growth.

If instead we consider the maximum mass that the halo reaches, rather than its final mass as shown in Figure 7.5 we see that all haloes now grow by construction. All haloes are above the dotted zero mass change line. Almost all haloes are now a factor of at least 10 bigger, at a minimum of 10^{12}M$_{\odot}$, and there is also a reasonable trend with density. The haloes that have grown the most largely end up in the densest environment. There is also a trend with the median mass increasing slightly with each bin shown. The biggest haloes tend to get bigger than the smallest, although the very largest haloes at redshift zero are spread across the original range.

Another way of looking at the same data is to colour it by the "bushiness" of the merger

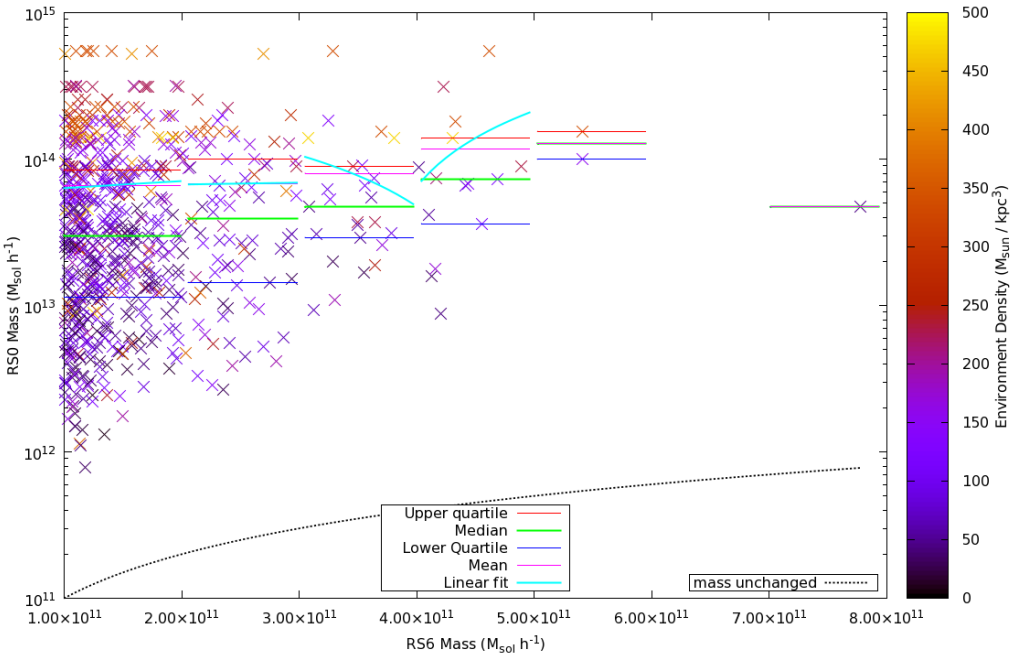

Figure 7.5: A comparison of the halo mass as it evolves from redshift six, showing the growth of haloes to their maximum subsequent mass. The points are coloured by environmental density which is defined as the average mass in haloes within an 8 Mpc sphere centred on this halo.

tree. Bushiness (B) is defined in Wang *et al.* (2015b) as

$$B = \frac{\text{No. of progenitors}}{\text{maximum tree length}} \tag{7.1}$$

which encapsulates how many mergers the halo has undergone, although it doesn't differentiate between minor and major mergers. In this case, with a fixed distance between redshift six and zero, all merger trees have the same length, as any haloes that do not survive to redshift zero would be dropped, although in practice there are none found here.

The correlation between environmental density and bushiness is a rather weak one as shown in Figure 7.6, that is a little unexpected. In dense environments it would be normal for more mergers to happen, as there are more haloes around to be merged with, but this appears not to be strongly indicated in this observation.

In contrast there is a much stronger relationship between mass and bushiness as shown in Figure 7.7. Here the correlation is almost linear on the log-log graph, showing a power law. Both the largest stellar mass selected and largest halo mass selected objects are shown here, and there is a lot of similarity as might be expected between the two data sets, and both show the same good correlation.

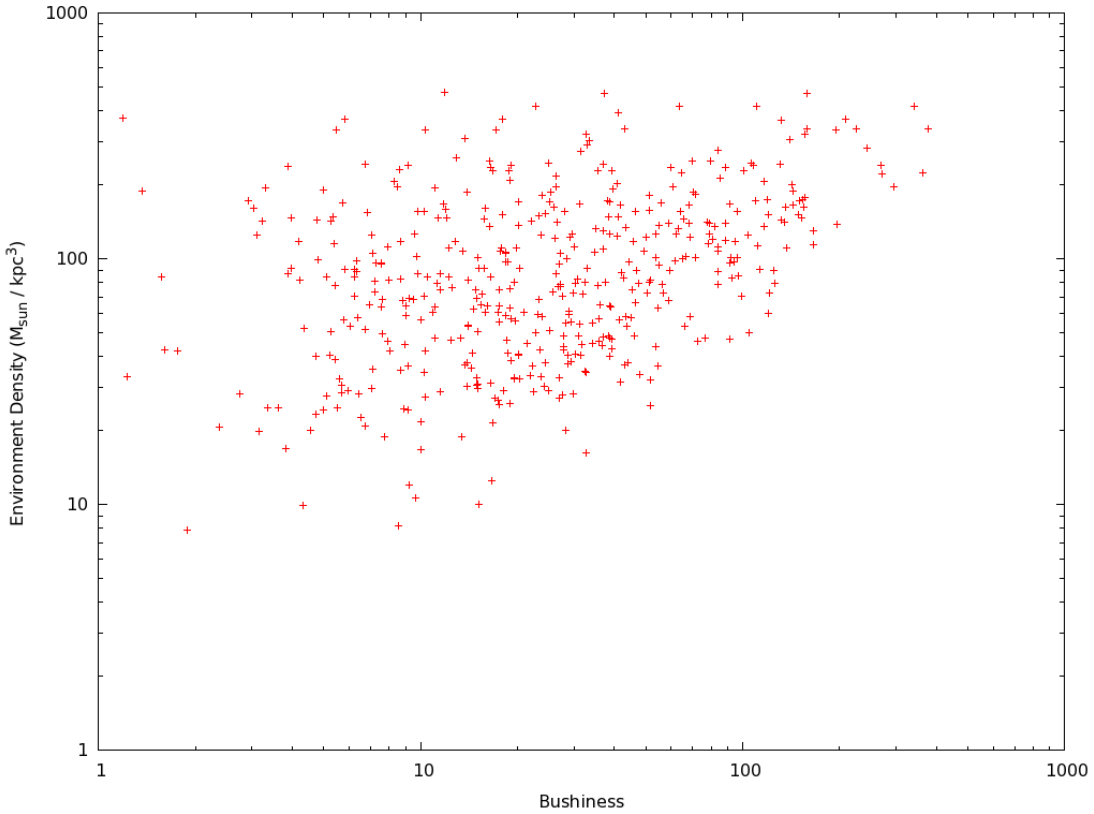

Figure 7.6: The correlation of environmental density and tree bushiness has a very weak agreement for these haloes.

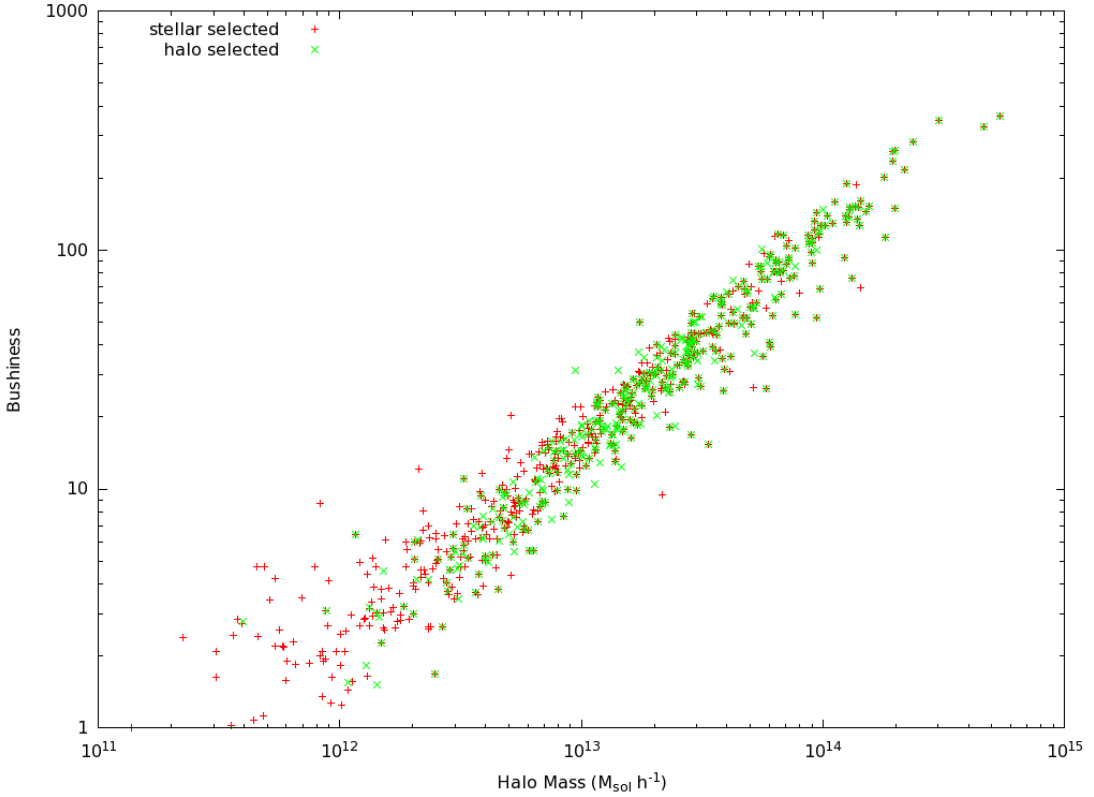

Figure 7.7: The halo mass correlation to bushiness is much stronger for the largest haloes at redshift six evolving to redshift zero.

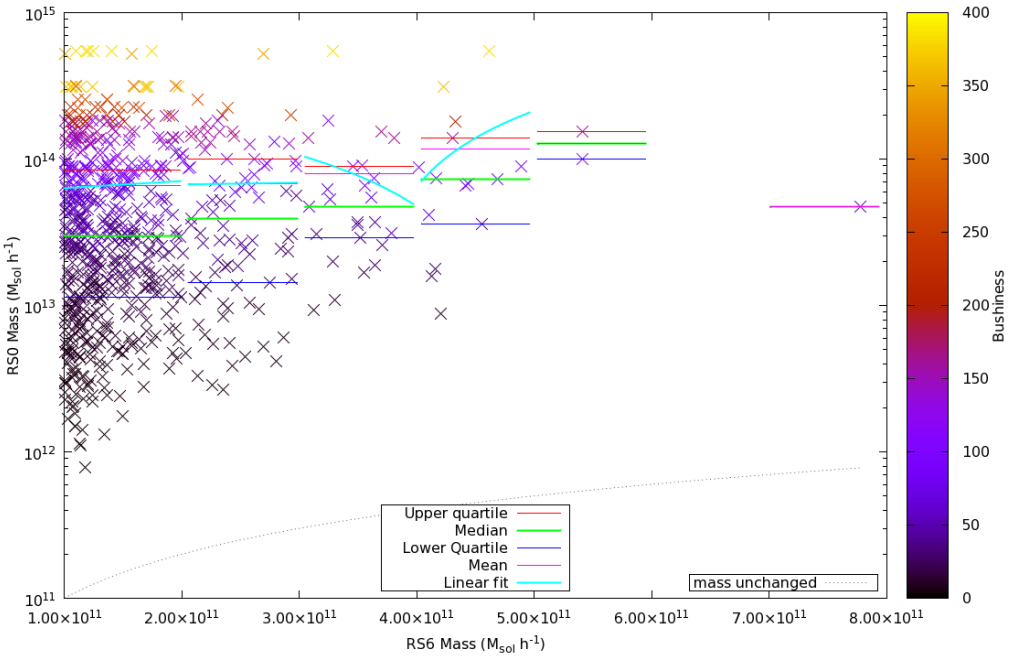

Figure 7.8: A comparison of the halo mass as it evolves from redshift six to redshift zero, showing the growth of haloes. The mass is the largest mass the halo has ever reached by redshift zero. The points are coloured by "bushiness" which is defined in (Wang *et al.*, 2015b) as the total number of branches of the tree divided by the maximum tree length.

This results in the plot shown in Figure 7.8, also indicate a good correlation between the bushiness of the halo and its final redshift zero mass. The most massive haloes clearly have a much higher degree of merging and the smaller ones almost no history of mergers involved in their growth as expected.

If we look at the end state of the haloes evolved from redshift six by their overall contribution, we can see that the very biggest haloes end up in the densest environments. In Figure 7.9 we can see the gradual decline in those initially large haloes found in low density environments. By the time the $10^{14.75}$ bin is reached, all are in very dense environments corresponding to galaxy clusters. However the type UCD-II's that have started large and have then not grown very much are more prevalent in the low density environments.

Using all these indicators and data, it is therefore possible to split apart the UCD-I and UCD-II candidate haloes when an appropriate selection criteria is chosen. A selection of most massive haloes at redshift six, that end up as not having grown much, but have a final mass much smaller than their maximum mass matches the UCD-I reasonably well. Then a filter that picks those with low mass, with low bushiness and a small

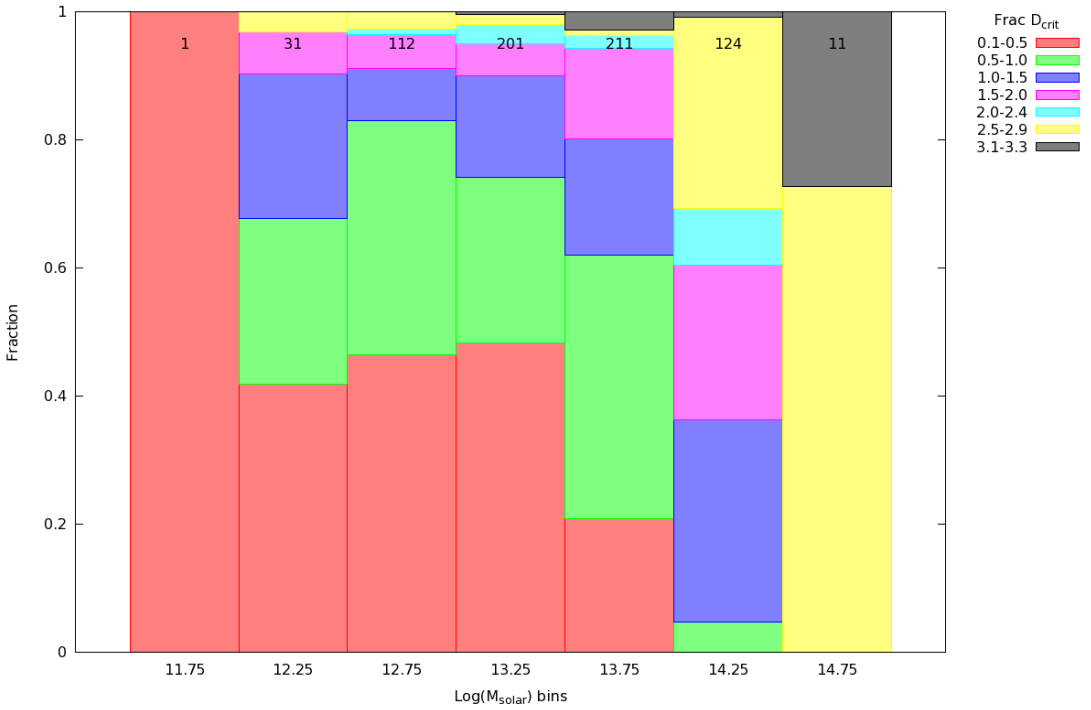

Figure 7.9: The distribution of halo masses placed in logarithmic bins, and colour coded by their density compared to the critical density, after evolving from redshift six to the current day. The numbers across the top are the total haloes in each bin.

difference between maximum mass and final pass picks out the UCD-II category. The results of such a filter is shown in Figure 7.10.

Figure 7.11 shows the total population of haloes, as selected by stellar mass at redshift six, broken into the two groups at an arbitrary cut of $10^9 M_\odot$. These are then evolved to redshift zero, and their destination over plotted. Also included are all the haloes below the initial cut. Many of these later haloes merge during growth and so end up in the same final halo, and are therefore displayed as a combined symbol. Just a few large haloes are not merged with the smaller ones. Again there is an obvious correlation with bushiness in the evolution towards the most massive haloes.

If instead the haloes are selected by halo mass, at a cut of $10^{11} M_\odot$ we see the data in Figure 7.12. This shows broadly the same story.

When now split into 4 arbitrary bins based on final halo mass, we can see how many of the haloes end up in each bin. The smallest bin is just for reference. Then the $10^{12} M_\odot$ bin represents galaxy sized haloes, and shows that about 3% of the haloes that are large at redshift six will end up here. These are possible UCD-II candidates, that have grown very little.

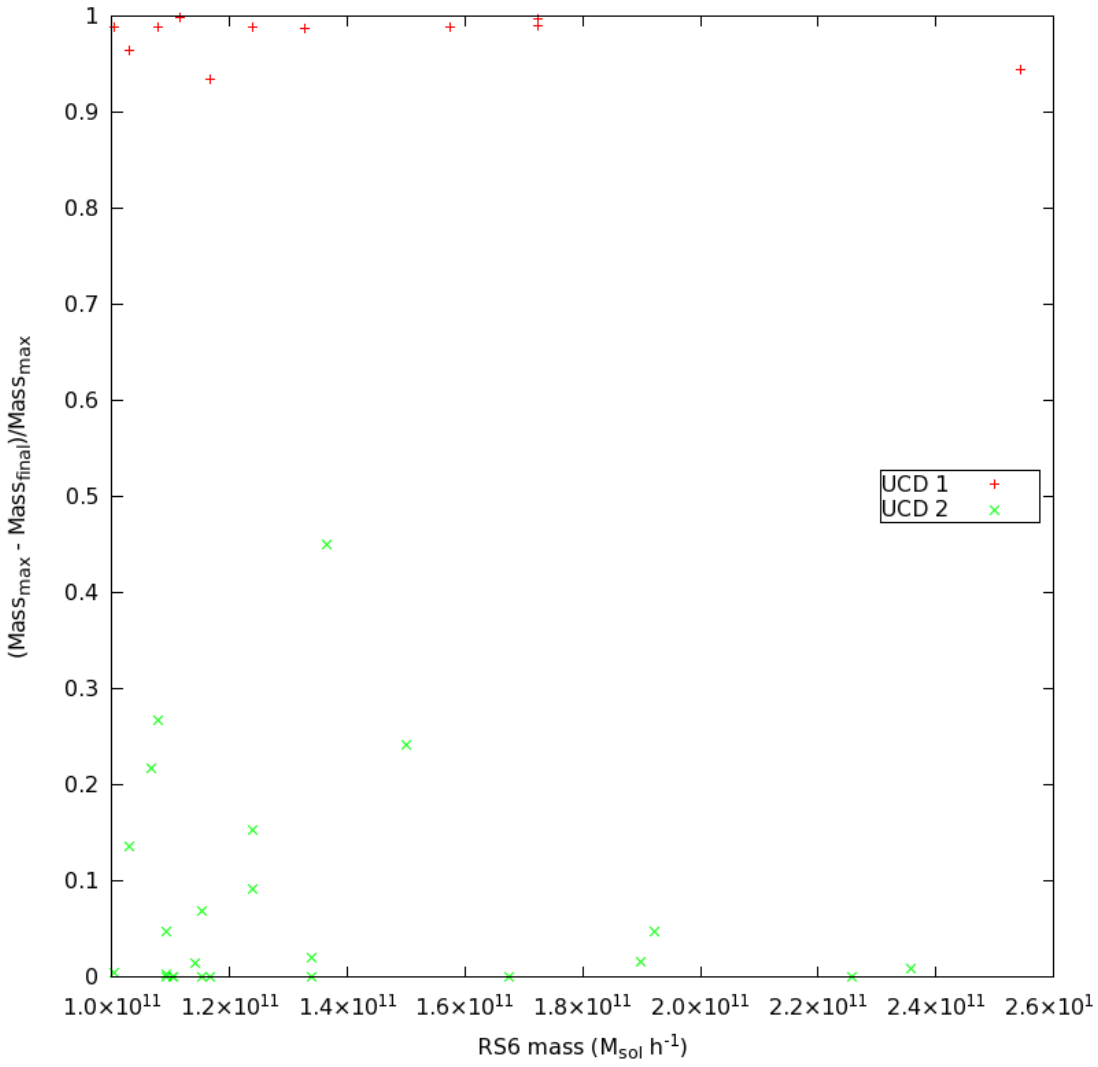

Figure 7.10: Selecting the two populations. The UCD-I are matched with low final mass, but much smaller than their maximum mass (1/10th in this case). The UCD-II are matched with low final mass, low bushiness and a small difference in final mass to maximum mass ratio.

In the next bin, which might represent galaxy groups, about $1/3$ of the haloes in this range were large at redshift six. In the final group representing galaxy clusters, the vast majority were large haloes at redshift six, that have grown by accretion and merger.

Looking first at the selection of haloes chosen by stellar mass in Figure 7.13, most of the large haloes at redshift six go on to become the largest haloes at redshift zero, as might be expected. Only about 6 percent of the biggest haloes at redshift zero are not formed from large progenitors. Meanwhile at the lower end on the $10^{12}M_{\odot}$ bin, there are about 4 percent of the haloes that show little growth.

Finally if we look at where the haloes end up selected by initial halo mass as shown in Figure 7.14, there is a similar picture, but reduced percentage in each case. This time

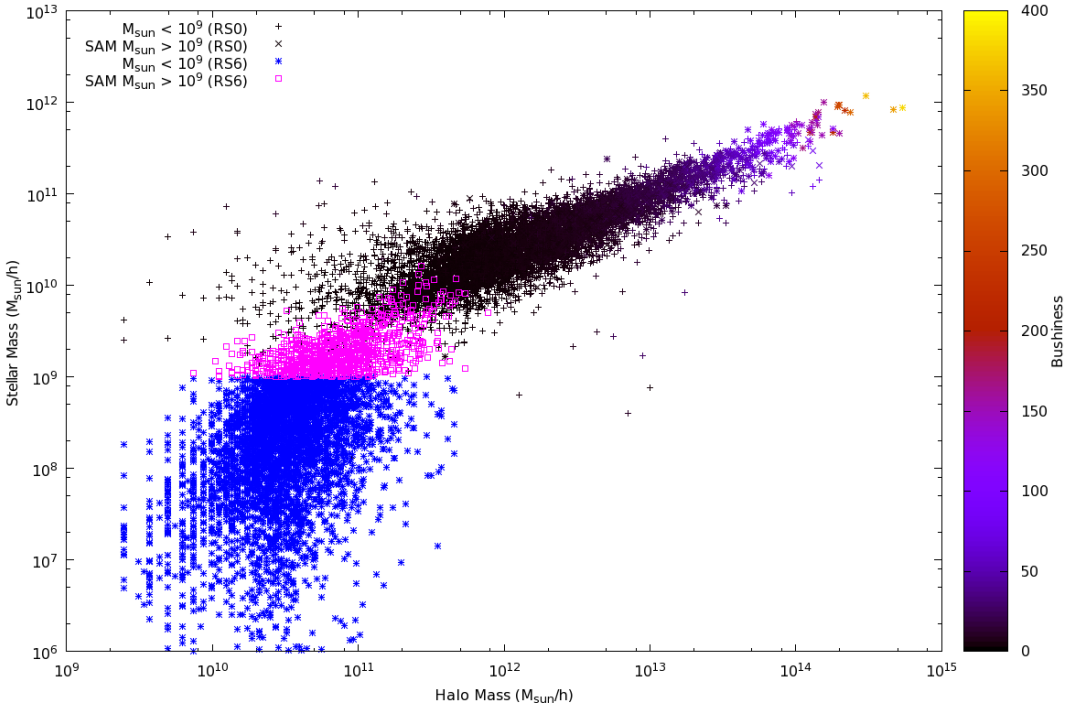

Figure 7.11: Haloes at redshift six selected by the stellar mass of the largest galaxy in purple, and those below the cut in blue, evolved to redshift zero and colour coded by bushiness.

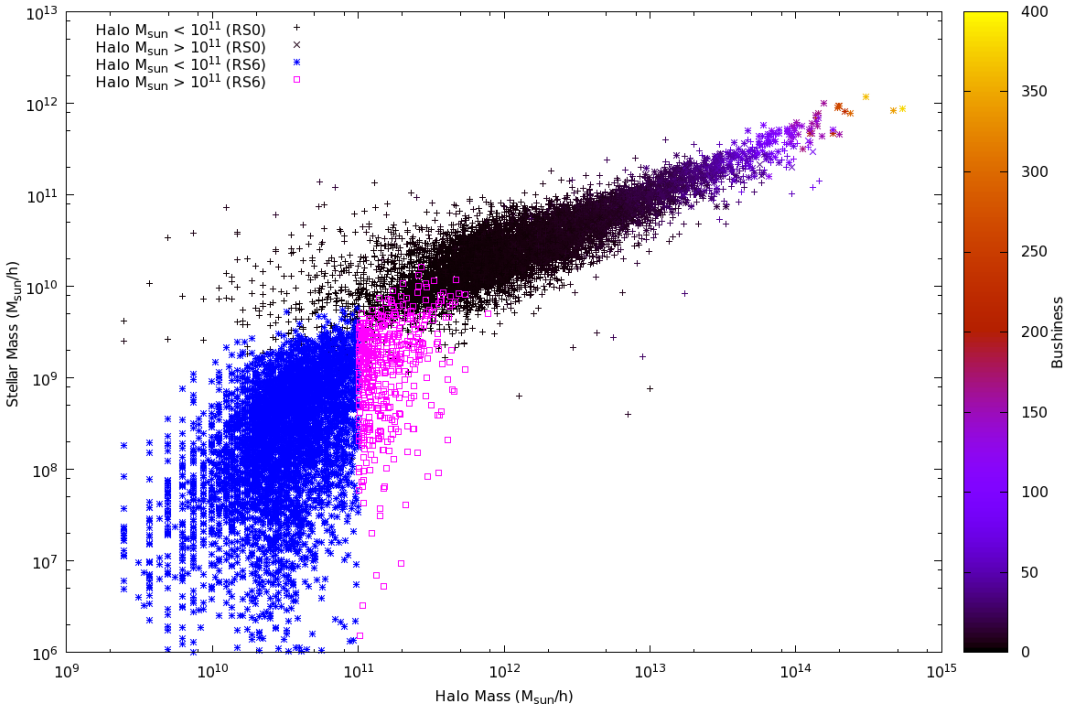

Figure 7.12: Haloes at redshift six selected by their halo mass in purple, and those below the cut in blue, evolved to redshift 0 and colour coded by bushiness.

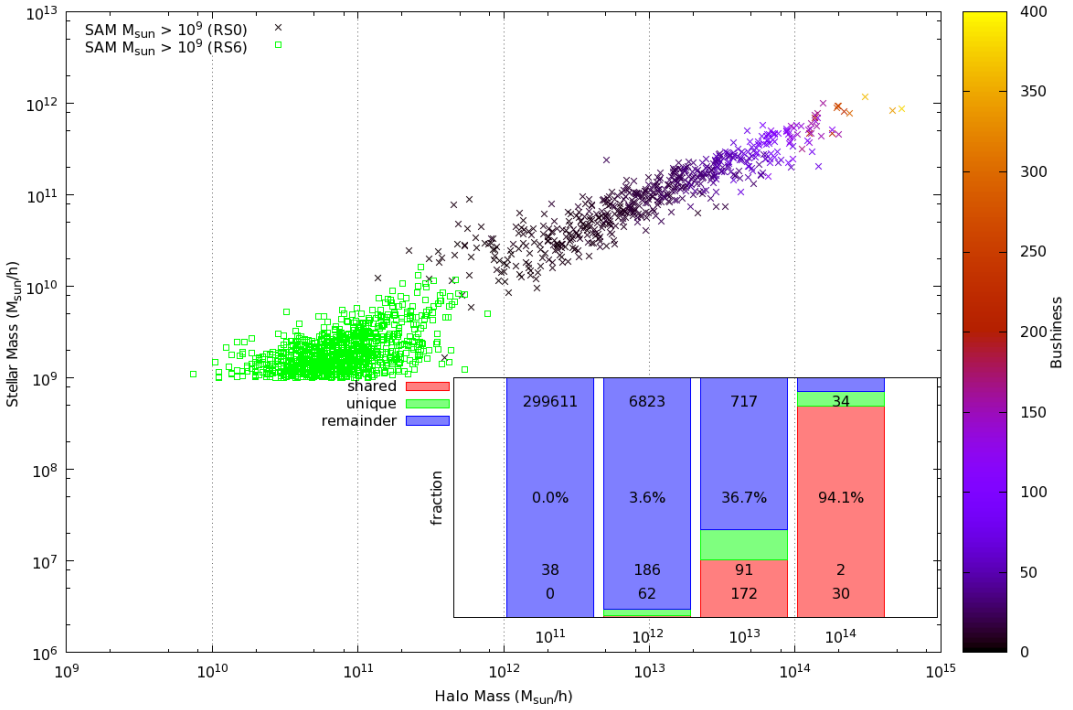

Figure 7.13: Haloes selected by stellar mass at redshift six traced to descendants at redshift zero. Inset is the number of haloes present in different mass bins, roughly shown as small, large, group and cluster sized haloes. The shared portion is those haloes present in both halo mass and stellar mass cuts, the unique portion is those only appearing from the stellar mass sample, and the remainder is all other haloes in this bin. The numbers on the histogram are as follows. The top number is the total number of haloes in that bin, the percentage is the number that are shared in the selections. The bottom two numbers are the number of unique haloes to this selection, and the final the number of haloes that are present in both selections.

only about 2 percent of the haloes show little growth, and 91 percent of the largest haloes start off as the largest haloes at redshift six.

For further work we intend to scale this up to use one of the multidark simulations that we have access to. This is a 1 h^{-1}Gpc box, so is much less susceptible to cosmic variance, compared to the 125 h^{-1}Mpc box that was produced for this preliminary work. The two mass functions appear to give very similar results, at least up to very small scales as show in Figure 7.15, but give much better large mass scales because of the bigger box.

7.4 Summary & Conclusions

This chapter has shown two possible mechanisms for production of UCD's traced through their dark matter halo and associated SAM galaxies.

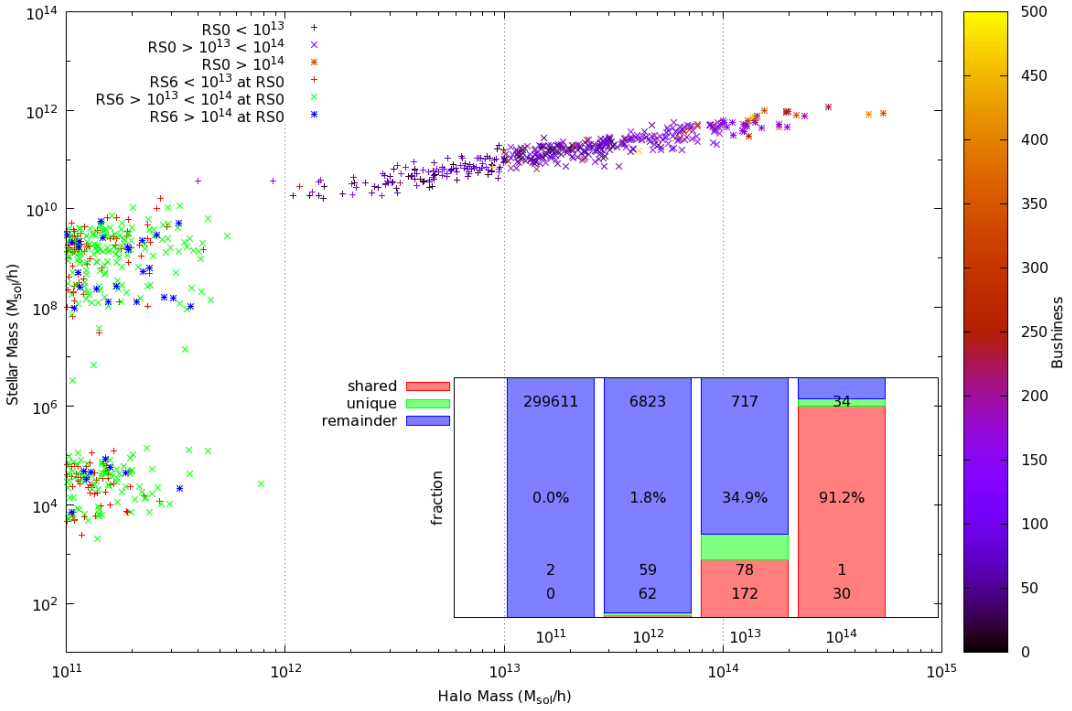

Figure 7.14: Haloes selected by halo mass at redshift six traced to redshift zero, and colour coded by bushiness. Inset is the number of haloes present in the mass bins that have evolved to these bins from redshift six.

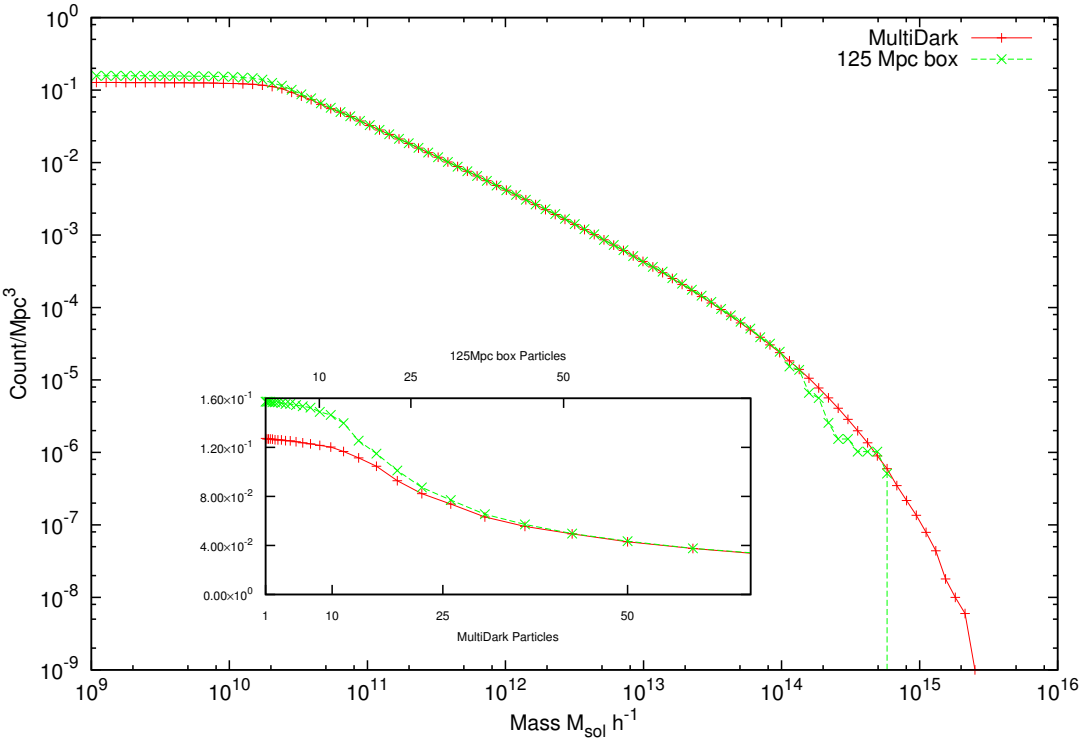

Figure 7.15: A comparison of the mass function of the large scale multidark simulation (1 h^{-1}Gpc box) with the mass function of the 125 h^{-1}Mpc box used to produce the results herein.

The first, UCD-I, are the more traditional route which others have looked at, where they are the cores of high phase space density early formation galaxies that have been harassed and stripped of much of their halo in a cluster environment. In this investigation a number of these have turned up. In this simulation such haloes tend to be very transitory, as once they start to get stripped, they can quickly evaporate and so vanish from the catalogs. The galaxies that were resident within them should continue to exist after much of the dark matter is stripped, but are sometimes missed by the semi analytic models. Without a host halo, they can sometimes be lost in the models, and in this case it seems GALACTICUS is unable to track such stripped haloes on infall. The galaxies should have a higher metallicity than galaxies of comparable size, and be of small mass $10^8 M_\odot$ and have a half light radius of around 20 pc. The fidelity of the Semi analytic models is probably not capable of reproducing these structures at this resolution.

The second class, UCD-II, are ones that form in sparsely occupied areas, that quickly accrete all available materials around, and are then starved of resources subsequently, and in particular have a very poor merger history. Although a number of these were detected in this simulation, these are much less prevalent than the first case.

There is also shown a method of selecting the two, the UCD-I class show little to no growth when traced from early on to present day, although they gain in mass only to have it removed. The UCD-II class show up in the maximum mass plot as having low mass growth, and low bushiness, which are also related.

Of the two classes, the UCD-II class show up as more common in this small simulation, by about a factor of two.

Whilst there is good reason to believe the UCD-I class of haloes exist, and would appear visually much like the observed UCDs do, the evidence that the UCD-II would appear similar is lacking. These objects would be harder to find, as they would probably be within voids and so not the subject of a targeted search. They would also probably not be as compact, as they will have little harassment and so will keep most of their stars, relying instead on being formed compact early on.

This simulation is relatively small, and we hope to repeat this work with a much larger simulation. However so far this has not been possible because of the much larger data

size. We plan on following up this work with the multidark simulation data, together with some data we have prepared but not managed to process so far.

Chapter 8

Conclusions

8.1 Overall Summary

In this thesis I have tried to show how effective various common analysis techniques are at processing and analysing data running across the pipeline of processes.

This is important, not only to validate how well current analysis programs manage to recover key data and parameters, but also to act as a benchmark for future versions and new variations of the analysis. For instance if a new halo finder, or merger tree builder is written, it should at least compare against the results here.

8.2 Accuracy

The accuracy with which simulations can reproduce large scale structure is reasonably good, particularly on the Mpc scale. However smaller scale structure is more difficult to reproduce. I have shown that even small changes in the initial definitions, the parameters used in the simulation or the time steps used can account for significant changes in the small scale structure.

This gives a substantial limit on the fidelity of the simulation and what it can reproduce, when looking for a fine accuracy such as the 1% target that EUCLID and others wish for weak lensing maps. This topic is explored in more detail in the paper Schneider *et al.* (2015).

8.3 Structure Finders

It is apparent that most of the well known structure finders give very similar results. They all struggle to recover structure within the middle of a host halo, no matter what detection scheme they make use of.

The unbinding step of removing gravitationally unbound particles is particularly important if some of the basic parameters of a halo are to be recovered.

Naturally what is recovered also depends on usage. When trying to determine lensing mass for instance, the instantaneous mass is required, in which case unbinding is not required - and can be detrimental.

This leads on to the definition of what a subhalo is. Despite extensive discussions and proposals at some of the workshops, there is no clear definition. Everyone generally agrees on them when they see them, but being dynamic and sometimes ephemeral structures, the definition often depends on the use to which they are being built for.

As light propagates far faster than the motion of the halo particles, a snapshot in time of all mass within a specific radius is what is required for lensing signals. The same is true for any other gravitational interaction parameter that needs to be found.

If the structure needs to be examined in more detail, to recover parameters of the subhalo such as spin or shape, then more care is needed to remove the non-gravitationally bound particles. These particles can include high speed background particles traversing the subhalo giving spurious results for spin and velocity dispersion. Other particles move with slower velocities on the outskirts of the halo, and are more difficult to remove but have less impact on the overall parameters. These are often either particles in the process of leaving or joining the substructure. Thus if temporal existence of the substructure is required, these particles may be important in the future if they are part of the accretion of the structure, or if they are undergoing stripping through gravitational interactions with other structures.

In the longer term, another definition of the substructure can ask how long the structure will exist for. Ultimately all substructures will usually be merged in the gravity well given long enough, so it depends often on the timescale that the existence of the

structure is around for exactly what counts as substructure.

8.4 Spin of subhaloes

The spin of subhaloes tends to increase as they fall into a main halo. This is primarily the result of the out lying particles being stripped away from the substructure, as tidal forces and stronger gravitational binding of the main and other substructures all conspire to remove the most weakly bound particles.

One of the useful side effects of measuring the spin was found to be that it was an excellent indicator of how well the unbinding steps managed in their removal of particles that are not gravitationally bound. Very high anomalous spin rates were seen, placing the spin parameter far outside the normal range and distribution that would be expected. Although there is a small shift in the peak of the spin distribution because of the interaction with the main halo, this is very small compared to the often wildly unphysical results that are found when unbinding is partially or wholly failing.

8.5 Merger trees

The merger tree work, and in particular JMERGE shows that care is needed in constructing merger trees, and in particular that breaks in the hierarchy can cause issues when the data is used for further analysis. In particular semi-analytic models do not respond well to merger trees with short branches.

8.6 Semi Analytic Model mass definition

Semi-analytic models are shown in most cases to be very sensitive to which mass definition is in use. It is important that the catalog and the SAM agree on what mass definition is in use, and that different mass definitions vary against one another over the life of the universe.

8.7 Ultra Compact Dwarf Galaxy Progenitors

Haloes that UCDs may appear in, possibly exist in (at least) two types. The most common is a early formed halo that is then subsequently stripped of most of its mass as it falls into a bigger halo. The dark matter can be completely stripped away leaving a tight core of a galaxy that will persist past the demise of the halo.

The other type is the halo that forms in a void, grows quickly initially, and is then being subsequently starved of material.

8.8 Future Work

One of the areas not looked at here is the generation of initial conditions. There are a number of commonly used tools to generate initial conditions, and it could prove fruitful to compare the results of running the same simulations on initial conditions generated by different methods. A comparison of the results after running to redshift zero might indicate what issues might be expected.

There are a lot of possibilities around the progression of SAM models and comparing one with another and consider them in relation to observational data. This is the subject of an upcoming workshop, and hopefully data from this will allow more of the aspects of SAM models and their fitting.

Another interesting exercise would be perturbing the input and/or parameters of the SAM models, to see how resilient they are to change. Whether this is because the physics is unstable or the models are unstable to small changes, or if the general results hold across a wide range of changes.

Meanwhile there is a wealth of data to be gleaned from the bigger dark matter runs that have been completed. Looking to the accuracy of these, it would be interesting to explore how small changes in input parameters affect the hydrodynamic gas models and their subsequent structures. Investigation of unusual structures across these models could be interesting, as the cosmic variance may show low probability structures evident in the output.

Bibliography

Aarseth S. J., Hoyle F., 1963. *MNRAS*, **126(3)**, 223. *Dynamical Evolution of Clusters of Galaxies, I*.

Abadi M. G., Navarro J. F., Steinmetz M., Eke V. R., 2003. *The Astrophysical Journal*, **591(2)**, 499. *Simulations of Galaxy Formation in a Cold Dark Matter Universe. I. Dynamical and Photometric Properties of a Simulated Disk Galaxy*.

Alam S., Albareti F. D., Prieto C. A., Anders F., Anderson S. F., Andrews B. H., Armengaud E., Aubourg É., Bailey S., Bautista J. E. *et al.*, 2015. *arXiv preprint arXiv:1501.00963*. *The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III*.

Alpher R. A., Bethe H., Gamow G., 1948. *Phys. Rev.*, **73**, 803. *The Origin of Chemical Elements*.

Angulo R. E., Lacey C. G., Baugh C. M., Frenk C. S., 2009. *MNRAS*, **399**, 983. *The fate of substructures in cold dark matter haloes*.

Antonuccio-Delogu V., Dobrotka A., Becciani U., Cielo S., Giocoli C., Macciò A. V., Romeo-Veloná A., 2010. *MNRAS*, **407**, 1338. *Dissecting the spin distribution of dark matter haloes*.

Ascasibar Y., Binney J., 2005. *MNRAS*, **356(3)**, 872. *Numerical estimation of densities*.

Ascasibar Y., Gottlöber S., 2008. *MNRAS*, **386**, 2022. *The dynamical structure of dark matter haloes*.

Ascasibar Y., 2010. *Comput. Phys. Commun.*, **181**, 1438. *Estimating multidimensional probability fields using the Field Estimator for Arbitrary Spaces (FiEstAS) with applications to astrophysics*.

Aubert D., Pichon C., Colombi S., 2004. *MNRAS*, **352(2)**, 376. *The origin and implications of dark matter anisotropic cosmic infall on ?L? haloes*.

Avila S., Knebe A., Pearce F. R., Schneider A., Srisawat C., Thomas P. A., Behroozi P., Elahi P. J., Han J., Mao Y.-Y., Onions J., Rodriguez-Gomez V., Tweed D., 2014. *MNRAS*, **441(4)**, 3488. *SUSSING MERGER TREES: the influence of the halo finder*.

Babcock H. W., 1939. *Lick Observatory Bulletin*, **19**, 41. *The rotation of the Andromeda Nebula*.

Barnes J., Efstathiou G., 1987. *ApJ*, **319**, 575. *Angular momentum from tidal torques*.

Barnes J., Hut P., 1986. *Nat*, **324**, 446. *A hierarchical O (N log N) force-calculation algorithm*.

Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013. *ApJ*, **763**, 18. *Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology*.

Behroozi P., Knebe A., Pearce F. R., Elahi P., Han J., Lux H., Mao Y.-Y., Muldrew S. I., Potter D., Srisawat C., 2015. *arXiv preprint arXiv:1506.01405*. *Major Mergers Going Notts: Challenges for Modern Halo Finders*.

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2011. *ArXiv preprint arXiv:1110.4372*. *The Rockstar Phase-Space Temporal Halo Finder and the Velocity Offsets of Cluster Cores*.

Bekki K., Couch W. J., Drinkwater M. J., 2001. *ApJ*, **552(2)**, L105. *Galaxy threshing and the formation of ultracompact dwarf galaxies*.

Benson A. J., Pearce F. R., Frenk C. S., Baugh C. M., Jenkins A., 2001. *MNRAS*, **320(2)**, 261. *A comparison of semi-analytic and smoothed particle hydrodynamics galaxy formation*.

Benson A. J., 2012. *New Astronomy*, **17(2)**, 175 . *Galacticus: A semi-analytic model of galaxy formation*.

Bentley J. L., 1975. *Communications of the ACM*, **18(9)**, 509. *Multidimensional binary search trees used for associative searching*.

Bertone S., De Lucia G., Thomas P. A., 2007. *MNRAS*, **379(3)**, 1143. *The recycling of gas and metals in galaxy formation: predictions of a dynamical feedback model*.

Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007. *MNRAS*, **376(1)**, 215. *The spin and shape of dark matter haloes in the Millennium simulation of a cold dark matter universe*.

Bett P., Eke V., Frenk C. S., Jenkins A., Okamoto T., 2010. *MNRAS*, **404(3)**, 1137. *The angular momentum of cold dark matter haloes with and without baryons*.

Binney J., 1987. *Galactic dynamics*, Princeton university press.

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006. *MNRAS*, **370(2)**, 645. *Breaking the hierarchy of galaxy formation*.

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011a. *ArXiv preprint arXiv:1111.2048. The Milky Way's bright satellites as an apparent failure of LCDM*.

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011b. *MNRAS*, **415**, L40. *Too big to fail? The puzzling darkness of massive Milky Way subhaloes*.

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009. *MNRAS*, **398**, 1150. *Resolving cosmic structure formation with the Millennium-II Simulation*.

Brodie J. P., Romanowsky A. J., Strader J., Forbes D. A., 2011. *ApJ*, **142(6)**, 199. *The Relationships among Compact Stellar Systems: A Fresh View of Ultracompact Dwarfs*.

Bryan G. L., Norman M. L., 1998. *ApJ*, **495(1)**, 80. *Statistical properties of x-ray clusters: Analytic and numerical comparisons*.

Bryan S. E., Kay S. T., Duffy A. R., Schaye J., Dalla Vecchia C., Booth C. M., 2012. *ArXiv preprint arXiv:1207.4555*. *The impact of baryons on the spins and shapes of dark matter haloes*.

Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001. *ApJ*, **555(1)**, 240. *A Universal Angular Momentum Profile for Galactic Halos*.

Bundy K., Bershady M. A., Law D. R., Yan R., Drory N., MacDonald N., Wake D. A., Cherinka B., Snchez-Gallego J. R., Weijmans A.-M., Thomas D., Tremonti C., Masters K., Coccato L., Diamond-Stanic A. M., Aragn-Salamanca A., Avila-Reese V., Badenes C., Falcn-Barroso J., Belfiore F., Bizyaev D., Blanc G. A., Bland-Hawthorn J., Blanton M. R., Brownstein J. R., Byler N., Cappellari M., Conroy C., Dutton A. A., Emsellem E., Etherington J., Frinchaboy P. M., Fu H., Gunn J. E., Harding P., Johnston E. J., Kauffmann G., Kinemuchi K., Klaene M. A., Knapen J. H., Leauthaud A., Li C., Lin L., Maiolino R., Malanushenko V., Malanushenko E., Mao S., Maraston C., McDermid R. M., Merrifield M. R., Nichol R. C., Oravetz D., Pan K., Parejko J. K., Sanchez S. F., Schlegel D., Simmons A., Steele O., Steinmetz M., Thanjavur K., Thompson B. A., Tinker J. L., van den Bosch R. C. E., Westfall K. B., Wilkinson D., Wright S., Xiao T., Zhang K., 2015. *ApJ*, **798(1)**, 7. *Overview of the SDSS-IV MaNGA Survey: Mapping nearby Galaxies at Apache Point Observatory*.

Cantalupo S., Lilly S. J., Haehnelt M. G., 2012. *MNRAS*, **425(3)**, 1992. *Detection of dark galaxies and circum-galactic filaments fluorescently illuminated by a quasar at $z = 2.4$*.

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000. *MNRAS*, **319**, 168. *Hierarchical galaxy formation*.

Colless M., Dalton G., Maddox S., Sutherland W., Norberg P., Cole S., Bland-Hawthorn J., Bridges T., Cannon R., Collins C., Couch W., Cross N., Deeley K., de Propris R., Driver S. P., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Madgwick D., Peacock J. A., Peterson B. A., Price I., Seaborne M., Taylor K., 2001. *MNRAS*, **328(4)**, 1039. *The 2dF Galaxy Redshift Survey: spectra and redshifts*.

Conselice C. J., Bershady M. A., Dickinson M., Papovich C., 2003. *ApJ*, **126**, 1183. *A Direct Measurement of Major Galaxy Mergers at z¡3*.

Conselice C. J., Rajgor S., Myers R., 2008. *MNRAS*, **386(2)**, 909. *The structures of distant galaxies–I. Galaxy structures and the merger rate to z 3 in the Hubble Ultra-Deep Field*.

Contini E., De Lucia G., Borgani S., 2011. *ArXiv preprint arXiv:1111.1911*. *Statistics of Substructures in Dark Matter Haloes*.

Cooper A. P., Cole S., Frenk C. S., White S. D. M., Helly J., Benson A. J., De Lucia G., Helmi A., Jenkins A., Navarro J. F., Springel V., Wang J., 2010. *MNRAS*, **406**, 744. *Galactic stellar haloes in the CDM model*.

Croton D. J., Springel V., White S. D. M., De Lucia G., Frenk C. S., Gao L., Jenkins A., Kauffmann G., Navarro J. F., Yoshida N., 2006. *MNRAS*, **365(1)**, 11. *The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies*.

Dabringhausen J., Hilker M., Kroupa P., 2008. *MNRAS*, **386**, 864. *From star clusters to dwarf galaxies: the properties of dynamically hot stellar systems*.

Dalcanton J. J., Spergel D. N., Summers F. J., 1997. *ApJ*, **482(2)**, 659. *The Formation of Disk Galaxies*.

Davis M., Efstathiou G., Frenk C. S., White S. D., 1985. *ApJ*, **292**, 371. *The evolution of large-scale structure in a universe dominated by cold dark matter*.

De Lucia G., Blaizot J., 2007. *MNRAS*, **375(1)**, 2. *The hierarchical formation of the brightest cluster galaxies*.

De Lucia G., Kauffmann G., Springel V., White S. D. M., Lanzoni B., Stoehr F., Tormen G., Yoshida N., 2004. *MNRAS*, **348**, 333. *Substructures in cold dark matter haloes*.

di Cintio A., Knebe A., Libeskind N. I., Yepes G., Gottlöber S., Hoffman Y., 2011. *MNRAS*, **417**, L74. *Too small to succeed? Lighting up massive dark matter subhaloes of the Milky Way*.

Diemand J., Kuhlen M., Madau P., Zemp M., Moore B., Potter D., Stadel J., 2008. *Nat*, **454**, 735. *Clumps and streams in the local dark matter distribution*.

D'Onghia E., Navarro J. F., 2007. *MNRAS*, **380**, L58. *Do mergers spin-up dark matter haloes?*

Drinkwater M. J., Gregg M. D., Hilker M., Bekki K., Couch W. J., Ferguson H. C., Jones J. B., Phillipps S., 2003. *Nat*, **423(6939)**, 519. *A class of compact dwarf galaxies from disruptive processes in galaxy clusters*.

Drinkwater M. J., Gregg M. D., Couch W. J., Ferguson H. C., Hilker M., Jones J. B., Karick A., Phillipps S., 2004. *Publications of the Astronomical Society of Australia*, **21(4)**, 375. *Ultra-compact dwarf galaxies in galaxy clusters*.

Dyer C. C., Ip P. S. S., 1993. *ApJ*, **409**, 60. *Softening in N-body simulations of collisionless systems*.

Efstathiou G., Eastwood J. W., 1981. *MNRAS*, **194**, 503. *On the clustering of particles in an expanding universe*.

Eisenstein D. J., Hut P., 1998. *ApJ*, **498(1)**, 137. *HOP: A New Group-finding Algorithm for N-Body Simulations*.

Eke V. R., Baugh C. M., Cole S., Frenk C. S., Norberg P., Peacock J. A., Baldry I. K., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Dalton G., Propris R. D., Driver S. P., Efstathiou G., Ellis R. S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D., Peterson B. A., Sutherland W., Taylor K., (the 2dFGRS Team), 2004. *MNRAS*, **348(3)**, 866. *Galaxy groups in the 2dFGRS: the group-finding algorithm and the 2PIGG catalogue*.

Elahi P. J., Thacker R. J., Widrow L. M., 2011. *MNRAS*, **418(1)**, 320. *Peaks above the Maxwellian Sea: a new approach to finding substructures in N-body haloes*.

Fall S., Efstathiou G., 1980. *MNRAS*, **193**, 189. *Formation and rotation of disc galaxies with haloes*.

Fellhauer M., Kroupa P., 2002. *MNRAS*, **330**, 642. *The formation of ultracompact dwarf galaxies*.

Ferguson H. C., Binggeli B., 1994. *A&AR*, **6(1-2)**, 67. *Dwarf elliptical galaxies*.

Ferrero I., Abadi M. G., Navarro J. F., Sales L. V., Gurovich S., 2011. *ArXiv preprint arXiv:1111.6609. The dark matter halos of dwarf galaxies: a challenge for the LCDM paradigm?*

Fields B. D., Molaro P., Sarkar S., 2014. *Chin.Phys.*, **C38**. *Big-Bang Nucleosynthesis*.

Font A. S., Bower R. G., McCarthy I. G., Benson A. J., Frenk C. S., Helly J. C., Lacey C. G., Baugh C. M., Cole S., 2008. *MNRAS*, **389(4)**, 1619. *The colours of satellite galaxies in groups and clusters*.

Freeman K., Bland-Hawthorn J., 2002. *ARA&A*, **40**, 487. *The New Galaxy: Signatures of Its Formation*.

Frenk C., White S., 2012. *Annalen der Physik*, **524(9-10)**, 507. *Dark matter and cosmic structure*.

Frenk C. S., Baugh C. M., Cole S., Lacey S., 1997. *Numerical and Analytical Modelling of Galaxy Formation and Evolution*, In: *Dark and Visible Matter in Galaxies and Cosmological Implications*, 335, eds Persic M., Salucci P.

Frieman J., 2011. *Dark Energy Survey*, In: *Bulletin of the American Astronomical Society*, 20501.

Gao L., White S. D. M., Jenkins A., Stoehr F., Springel V., 2004. *MNRAS*, **355**, 819. *The subhalo populations of ΛCDM dark haloes*.

Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998. *MNRAS*, **300(1)**, 146. *Dark matter haloes within clusters*.

Gill S. P. D., Knebe A., Gibson B. K., Dopita M. A., 2004. *MNRAS*, **351**, 410. *The evolution of substructure - II. Linking dynamics to environment*.

Gill S. P., Knebe A., Gibson B. K., 2004. *MNRAS*, **351(2)**, 399. *The evolution of substructureI. A new identification method*.

Gonzalez-Perez V., Lacey C., Baugh C., Lagos C., Helly J., Campbell D., Mitchell P., 2014. *MNRAS. How sensitive are predicted galaxy luminosities to the choice of stellar population synthesis model?*

Gottlöber S., Yepes G., 2007. *ApJ*, **664**, 117. *Shape, Spin, and Baryon Fraction of Clusters in the MareNostrum Universe*.

Guillet T., Teyssier R., Colombi S., 2010. *Monthly Notices of the Royal Astronomical Society*, **405(1)**, 525. *The effect of baryons on the variance and the skewness of the mass distribution in the Universe at small scales*.

Han J., Jing Y. P., Wang H., Wang W., 2012. *MNRAS*, **427(3)**, 2437. *Resolving Subhalos' Lives with the Hierarchical Bound-Tracing Algorithm*.

Harnois-Déraps J., Pen U.-L., Iliev I. T., Merz H., Emberson J. D., Desjacques V., 2013. *MNRAS*, **436(1)**, 540. *High-performance P3M N-body code: cubep3m*.

Heitmann K., White M., Wagner C., Habib S., Higdon D., 2010. *ApJ*, **715(1)**, 104. *The Coyote Universe. I. Precision Determination of the Nonlinear Matter Power Spectrum*.

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013. *MNRAS*, **431(4)**, 3373. *Simulations of the galaxy population constrained by observations from z = 3 to the present day: implications for galactic winds and the fate of their ejecta*.

Hernquist L., 1990. *ApJ*, **356**, 359. *An analytical model for spherical galaxies and bulges*.

Hetznecker H., Burkert A., 2006. *MNRAS*, **370**, 1905. *The evolution of the dark halo spin parameters λ and λ' in a ΛCDM universe: the role of minor and major mergers*.

Holmberg E., 1941. *ApJ*, **94**, 385. *On the Clustering Tendencies among the Nebulae. II. a Study of Encounters Between Laboratory Models of Stellar Systems by a New Integration Procedure.*

Huchra J. P., Geller M. J., 1982. *ApJ*, **257**, 423. *Groups of galaxies. I - Nearby groups*.

Huterer D., Takada M., Bernstein G., Jain B., 2006. *Monthly Notices of the Royal Astronomical Society*, **366(1)**, 101. *Systematic errors in future weak-lensing surveys: requirements and prospects for self-calibration*.

Ivezic Z., Tyson J., Acosta E., Allsman R., Anderson S., Andrew J., Angel R., Axelrod T., Barr J., Becker A. *et al.*, 2008. *arXiv preprint arXiv:0805.2366. LSST: from science drivers to reference design and anticipated data products*.

Jarosik N., Bennett C. L., Dunkley J., Gold B., Greason M. R., Halpern M., Hill R. S., Hinshaw G., Kogut A., Komatsu E., Larson D., Limon M., Meyer S. S., Nolta M. R., Odegard N., Page L., Smith K. M., Spergel D. N., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2011. *ApJS*, **192(2)**, 14. *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results*.

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001. *MNRAS*, **321(2)**, 372. *The mass function of dark matter haloes*.

Jiang L., Helly J. C., Cole S., Frenk C. S., 2014. *MNRAS*, **440(3)**, 2115. *N-body dark matter haloes with simple hierarchical histories*.

Kase H., Makino J., Funato Y., 2007. *PASJ*, **59**, 1071. *Missing Dwarf Problem in Galaxy Clusters*.

Kauffmann G., Nusser A., Steinmetz M., 1997. *MNRAS*, **286**, 795. *Galaxy formation and large-scale bias*.

Kauffmann G., White S. D. M., Guiderdoni B., 1993. *MNRAS*, **264**, 201. *The Formation and Evolution of Galaxies Within Merging Dark Matter Haloes*.

Klypin A., Gottlöber S., Kravtsov A. V., Khokhlov A. M., 1999a. *ApJ*, **516**, 530. *Galaxies in N-Body Simulations: Overcoming the Overmerging Problem*.

Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999b. *ApJ*, **522(1)**, 82. *Where Are the Missing Galactic Satellites?*

Klypin A. A., Trujillo-Gomez S., Primack J., 2011. *ApJ*, **740**, 102. *Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation*.

Knebe A., Power C., 2008. *ApJ*, **678(2)**, 621. *On the Correlation between Spin Parameter and Halo Mass*.

Knebe A., Gill S. P. D., Kawata D., Gibson B. K., 2005. *MNRAS*, **357**, L35. *Mapping substructures in dark matter haloes*.

Knebe A., Knollmann S. R., Muldrew S. I., Pearce F. R., Aragon-Calvo M. A., Ascasibar Y., Behroozi P. S., Ceverino D., Colombi S., Diemand J., 2011. *MNRAS*. *Haloes gone MAD: The Halo-Finder Comparison Project*.

Knebe A., Pearce F. R., Lux H., Ascasibar Y., Behroozi P., Casado J., Moran C. C., Diemand J., Dolag K., Dominguez-Tenreiro R., Elahi P., Falck B., Gottlber S., 2013. *MNRAS*, **435(2)**, 1618. *Structure finding in cosmological simulations: the state of affairs*.

Knebe A., Pearce F. R., Thomas P. A., Benson A., Blaizot J., Bower R., Carretero J., Castander F. J., Cattaneo A., Cora S. A., Croton D. J., Cui W., Cunnama D., De Lucia G., Devriendt J. E., Elahi P. J., Font A., Fontanot F., Garcia-Bellido J., Gargiulo I. D., Gonzalez-Perez V., Helly J., Henriques B., Hirschmann M., Lee J., Mamon G. A., Monaco P., Onions J., Padilla N. D., Power C., Pujol A., Skibba R. A., Somerville R. S., Srisawat C., Vega-Martnez C. A., Yi S. K., 2015. *MNRAS*, **451(4)**, 4029. *nIFTy cosmology: comparison of galaxy formation models*.

Knollmann S. R., Knebe A., 2009. *ApJS*, **182**, 608. *AHF: Amiga's Halo Finder*.

Kuhlen M., Diemand J., Madau P., 2008. *ApJ*, **686**, 262. *The dark matter annihilation signal from galactic substructure: predictions for GLAST*.

Lacerna I., Padilla N., 2012. *MNRAS*, **426(1)**, L26. *The nature of assembly bias II. Halo spin*.

Lacey C., Cole S., 1994. *MNRAS*, **271(3)**, 676. *Merger rates in hierarchical models of galaxy formation II. Comparison with N-body simulations*.

Lagos C. d. P., Cora S. A., Padilla N. D., 2008. *MNRAS*, **388(2)**, 587. *Effects of AGN feedback on CDM galaxies*.

Laureijs R., Racca G., Stagnaro L., Salvignol J.-C., Alvarez J. L., Criado G. S., Venancio L. G., Short A., Strada P., Colombo C. *et al.*, 2014. *Euclid mission status*, In: *SPIE Astronomical Telescopes+ Instrumentation*, p. 91430H, International Society for Optics and Photonics.

Lee J., Yi S. K., 2013. *ApJ*, **766(1)**, 38. *On the Assembly History of Stellar Components in Massive Galaxies*.

Lee J., Yi S. K., Elahi P. J., Thomas P. A., Pearce F. R., Behroozi P., Han J., Helly J., Jung I., Knebe A., Mao Y.-Y., Onions J., Rodriguez-Gomez V., Schneider A., Srisawat C., Tweed D., 2014. *ArXiv e-prints*. *Sussing Merger Trees : The Impact of Halo Merger Trees on Galaxy Properties in a Semi-Analytic Model*.

Libeskind N. I., Knebe A., Hoffman Y., Gottlöber S., Yepes G., 2011. *MNRAS*, **418**, 336. *Disentangling the dark matter halo from the stellar halo*.

Lokas E. L., Mamon G. A., 2003. *MNRAS*, **343(2)**, 401. *Dark matter distribution in the Coma cluster from galaxy kinematics: breaking the massanisotropy degeneracy*.

Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011. *ApJ*, **742(2)**, 103. *The Major and Minor Galaxy Merger Rates at z ¡ 1.5*.

LHuillier B., Park C., Kim J., 2014. *New Astronomy*, **30**, 79 . *Effects of the initial conditions on cosmological N-body simulations*.

Macci A. V., Dutton A. A., Van Den Bosch F. C., Moore B., Potter D., Stadel J., 2007. *MNRAS*, **378(1)**, 55. *Concentration, spin and shape of dark matter haloes: scatter and the dependence on mass and environment*.

Maciejewski M., Colombi S., Springel V., Alard C., Bouchet F. R., 2009. *MNRAS*, **396(3)**, 1329. *Phasespace structures II. Hierarchical Structure Finder*.

Maciejewski M., Vogelsberger M., White S. D. M., Springel V., 2011. *MNRAS*, **415**, 2475. *Bound and unbound substructures in Galaxy-scale dark matter haloes*.

Mestel L., 1963. *MNRAS*, **126**, 553. *On the galactic law of rotation*.

Mo H. J., Mao S., White S. D. M., 1998. *MNRAS*, **295**, 319. *The formation of galactic discs*.

Monaco P., Fontanot F., Taffoni G., 2007. *MNRAS*, **375(4)**, 1189. *The morgana model for the rise of galaxies and active nuclei*.

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999. *ApJ*, **524(1)**, L19. *Dark Matter Substructure within Galactic Halos*.

Moore B., Lake G., Katz N., 1998. *ApJ*, **495(1)**, 139. *Morphological Transformation from Galaxy Harassment*.

Moore G. E., 1965. *Electronics*, **38(8)**, 114. *Cramming more components onto integrated circuits*.

More S., Kravtsov A. V., Dalal N., Gottlöber S., 2011. *ApJS*, **195**, 4. *The Overdensity and Masses of the Friends-of-friends Halos and Universality of Halo Mass Function*.

Muldrew S. I., Pearce F. R., Power C., 2011. *MNRAS*, **410(4)**, 2617. *The accuracy of subhalo detection*.

Navarro J. F., Steinmetz M., 2000. *ApJ*, **538(2)**, 477. *Dark Halo and Disk Galaxy Scaling Laws in Hierarchical Universes*.

Navarro J. F., Frenk C. S., White S. D. M., 1997. *ApJ*, **490(2)**, 493. *A Universal Density Profile from Hierarchical Clustering*.

Neyrinck M. C., Gnedin N. Y., Hamilton A. J. S., 2005. *MNRAS*, **356(4)**, 1222. *voboz: an almostparameterfree halofinding algorithm*.

Oh S.-H., Brook C., Governato F., Brinks E., Mayer L., de Blok W. J. G., Brooks A., Walter F., 2011. *ApJ*, **142**, 24. *The Central Slope of Dark Matter Cores in Dwarf Galaxies: Simulations versus THINGS*.

Old L., Skibba R. A., Pearce F. R., Croton D., Muldrew S. I., Muoz-Cuartas J. C., Gifford D., Gray M. E., der Linden A. v., Mamon G. A., Merrifield M. R., Mller V., Pearson R. J., Ponman T. J., Saro A., Sepp T., Sifn C., Tempel E., Tundo E., Wang Y. O., Wojtak R., 2014. *MNRAS*, **441(2)**, 1513. *Galaxy cluster mass reconstruction project I. Methods and first results on galaxy-based techniques*.

Onions J., Knebe A., Pearce F. R., Muldrew S. I., Lux H., Knollmann S. R., Ascasibar Y., Behroozi P., Elahi P., Han J., Maciejewski M., Merchn M. E., Neyrinck M., Ruiz A. N., Sgr M. A., Springel V., Tweed D., 2012. *MNRAS*, **423(2)**, 1200. *Subhaloes going Notts: the subhalo-finder comparison project*.

Oort J. H., 1932. *Bull. Astron. Inst. Netherlands*, **6**, 249. *The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems*.

Peebles P., 1969. *ApJ*, **155**, 393. *Origin of the angular momentum of galaxies*.

Phillipps S., Drinkwater M. J., Gregg M. D., Jones J. B., 2001. *ApJ*, **560(1)**, 201. *Ultracompact Dwarf Galaxies in the Fornax Cluster*.

Pilipenko S., Doroshkevich A., Gottlöber S., 2009. *Astronomy Reports*, **53**, 976. *Evolution of dark-matter halos in numerical models*. 10.1134/S106377290911002X.

Planck Collaboration, Ade P. A. R., Aghanim N., Arnaud M., Ashdown M., Aumont J., Baccigalupi C., Banday A. J., Barreiro R. B., Bartlett J. G., et al., 2015. *ArXiv e-prints. Planck 2015 results. XIII. Cosmological parameters*.

Pontzen A., Governato F., 2011. *ArXiv preprint arXiv:1106.0499*. *How supernova feedback turns dark matter cusps into cores*.

Press W., Davis M., 1982. *ApJ*, **259(2)**. *How to identify and weigh virialized clusters of galaxies in a complete redshift catalog*.

Primack J. R., 2005. *New Astronomy Reviews*, **49(2)**, 25. *Precision cosmology*.

Reed D., Governato F., Quinn T., Gardner J., Stadel J., Lake G., 2004. *MNRAS*, **359(4)**, 1537. *Dark Matter Subhaloes in Numerical Simulations*.

Robert C. Kennicutt J., 1998. *The Astrophysical Journal*, **498(2)**, 541. *The Global Schmidt Law in Star-forming Galaxies*.

Roberts M. S., Whitehurst R. N., 1975. *ApJ*, **201**, 327. *The rotation curve and geometry of M31 at large galactocentric distances.*

Rubin V. C., Ford, Jr. W. K., 1970. *ApJ*, **159**, 379. *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*.

Schaye J., Crain R. A., Bower R. G., Furlong M., Schaller M., Theuns T., Dalla Vecchia C., Frenk C. S., 2014. *ArXiv e-prints*. *The EAGLE project: Simulating the evolution and assembly of galaxies and their environments*.

Schmidt M., 1959. *ApJ*, **129**, 243. *The Rate of Star Formation.*

Schneider A., Teyssier R., Potter D., Stadel J., Onions J., Reed D. S., Smith R. E., Springel V., Pearce F. R., 2015. *ArXiv e-prints*. *Matter power spectrum and the challenge of percent accuracy*.

Seth A. C., van den Bosch R., Mieske S., Baumgardt H., den Brok M., Strader J., Neumayer N., Chilingarian I., Hilker M., McDermid R. *et al.*, 2014. *Nature*, **513(7518)**, 398. *A supermassive black hole in an ultra-compact dwarf galaxy*.

Sgr M. A., Ruiz A. N., Merchn M. E., 2010. *BAAA*, **53**, 43. *Hierarchical Friend-of-Friend algorithm to extract substructures from dark matter halos*.

Skibba R. A., Sheth R. K., 2009. *MNRAS*, **392(3)**, 1080. *A halo model of galaxy colours and clustering in the Sloan Digital Sky Survey*.

Skrutskie M. F., Cutri R. M., Stiening R., Weinberg M. D., Schneider S., Carpenter J. M., Beichman C., Capps R., Chester T., Elias J., Huchra J., Liebert J., Lonsdale C., Monet D. G., Price S., Seitzer P., Jarrett T., Kirkpatrick J. D., Gizis J. E., Howard E., Evans T., Fowler J., Fullmer L., Hurt R., Light R., Kopan E. L., Marsh K. A., McCallon H. L., Tam R., Dyk S. V., Wheelock S., 2006. *The Astronomical Journal*, **131(2)**, 1163. *The Two Micron All Sky Survey (2MASS).*

Smoot G. F., Bennett C. L., Kogut A., Wright E. L., Aymon J., Boggess N. W., Cheng E. S., de Amici G., Gulkis S., Hauser M. G., Hinshaw G., Jackson P. D., Janssen M., Kaita E., Kelsall T., Keegstra P., Lineweaver C., Loewenstein K., Lubin P., Mather J., Meyer S. S., Moseley S. H., Murdock T., Rokke L., Silverberg R. F., Tenorio L., Weiss R., Wilkinson D. T., 1992. *ApJ*, **396**, L1. *Structure in the COBE differential microwave radiometer first-year maps*.

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001. *MNRAS*, **328(3)**, 726. *Populating a cluster of galaxies I. Results at*.

Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005. *Nat*, **435**, 629. *Simulations of the formation, evolution and clustering of galaxies and quasars*.

Springel V., Wang J., Vogelsberger M., Ludlow A., Jenkins A., Helmi A., Navarro J. F., Frenk C. S., White S. D., 2008a. *MNRAS*, **391(4)**, 1685. *The Aquarius Project: the subhaloes of galactic haloes*.

Springel V., White S. D. M., Frenk C. S., Navarro J. F., Jenkins A., Vogelsberger M., Wang J., Ludlow A., Helmi A., 2008b. *Nat*, **456(7218)**, 73. *Prospects for detecting supersymmetric dark matter in the Galactic halo*.

Springel V., Yoshida N., White S. D., 2001. *New Astronomy*, **6(2)**, 79. *GADGET: a code for collisionless and gasdynamical cosmological simulations*.

Springel V., 2005. *MNRAS*, **364(4)**, 1105. *The cosmological simulation code gadget2*.

Srisawat C., Knebe A., Pearce F. R., Schneider A., Thomas P. A., Behroozi P., Dolag K., Elahi P. J., Han J., Helly J. *et al.*, 2013. *MNRAS*, **436(1)**, 150. *Sussing Merger Trees: The Merger Trees Comparison Project*.

Stadel J., Potter D., Moore B., Diemand J., Madau P., Zemp M., Kuhlen M., Quilis V., 2009a. *MNRAS*, **398**, L21. *Quantifying the heart of darkness with GHALO - a multibillion particle simulation of a galactic halo*.

Stadel J., Potter D., Moore B., Diemand J., Madau P., Zemp M., Kuhlen M., Quilis V., 2009b. *MNRAS*, **398(1)**, L21. *Quantifying the heart of darkness with GHALO a multibillion particle simulation of a galactic halo*.

Stadel J. G., 2001. *PhD thesis*, University of Washington.

Sutter P. M., Elahi P., Falck B., Onions J., Hamaus N., Knebe A., Srisawat C., Schneider A., 2014. *MNRAS*, **445(2)**, 1235. *The life and death of cosmic voids*.

Teyssier R., 2002. *A&A*, **385(1)**, 337. *Cosmological hydrodynamics with adaptive mesh refinement*.

Thomas P., Drinkwater M., Evstigneeva E., 2008. *MNRAS*, **389(1)**, 102. *Formation of ultra-compact dwarf galaxies: tests of the galaxy threshing scenario in Fornax*.

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlber S., Holz D. E., 2008. *ApJ*, **688(2)**, 709. *Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality*.

Toomre A., 1977. *Mergers and Some Consequences*, In: *Evolution of Galaxies and Stellar Populations*, 401, eds Tinsley B. M., Larson, D. Campbell R. B. G.

Trenti M., Smith B. D., Hallman E. J., Skillman S. W., Shull J. M., 2010. *ApJ*, **711**, 1198. *How Well do Cosmological Simulations Reproduce Individual Halo Properties?*

Trowland H. E., Lewis G. F., Bland-Hawthorn J., 2013. *ApJ*, **762(2)**, 72. *The Cosmic History of the Spin of Dark Matter Halos within the Large-scale Structure*.

Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009. *A&A*, **506**, 647. *Building merger trees from cosmological N-body simulations. Towards improving galaxy formation models using subhaloes*.

Vitvitska M., Klypin A. A., Kravtsov A. V., Wechsler R. H., Primack J. R., Bullock J. S., 2002. *The Astrophysical Journal*, **581(2)**, 799. *The Origin of Angular Momentum in Dark Matter Halos*.

Vogelsberger M., Helmi A., Springel V., White S., Wang J., Frenk C., Jenkins A., Ludlow A., Navarro J., 2009. *MNRAS*, **395(2)**, 797. *Phase-space structure in the local dark matter distribution and its signature in direct detection experiments*.

Vogelsberger M., Genel S., Springel V., Torrey P., Sijacki D., Xu D., Snyder G., Bird S., Nelson D., Hernquist L., 2014. *Nature*, **509(7499)**, 177. *Properties of galaxies reproduced by a hydrodynamic simulation*.

Wang H., Mo H. J., Jing Y. P., Yang X., Wang Y., 2011a. *MNRAS*, **413**, 1973. *Internal properties and environments of dark matter haloes*.

Wang J., Navarro J. F., Frenk C. S., White S. D. M., Springel V., Jenkins A., Helmi A., Ludlow A., Vogelsberger M., 2011b. *MNRAS*, **413**, 1373. *Assembly history and structure of galactic cold dark matter haloes*.

Wang Y., Lin W., Pearce F. R., Lux H., Muldrew S. I., Onions J., 2015a. *ApJ*, **801(2)**, 93. *Solving the Puzzle of Subhalo Spins*.

Wang Y., Pearce F. R., Knebe A., Schneider A., Srisawat C., Tweed D., Jung I., Han J., Helly J., Onions J., Elahi P. J., Thomas P., Behroozi P., Yi S. K., Rodriguez-Gomez V., Mao Y.-Y., Jing Y., Lin W., 2015b. *ArXiv e-prints*. *Sussing Merger Trees: Stability and Convergance*.

Warnick K., Knebe A., Power C., 2008. *MNRAS*, **385**, 1859. *The tidal streams of disrupting subhaloes in cosmological dark matter haloes*.

White S., Frenk C., 1991. *ApJ*, **379**, 52. *Galaxy formation through hierarchical clustering*.

White S. D. M., Rees M. J., 1978. *MNRAS*, **183**, 341. *Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering*.

White S. D. M., 1984. *ApJ*, **286**, 38. *Angular momentum growth in protogalaxies*.

White M., 2004. *Astroparticle Physics*, **22(2)**, 211. *Baryons and weak lensing power spectra*.

Zavala J., Springel V., Boylan-Kolchin M., 2010. *MNRAS*, **405(1)**, 593. *Extragalactic gamma-ray background radiation from dark matter annihilation*.

Zwicky F., 1933. *Helvetica Physica Acta*, **6**, 110. *Die Rotverschiebung von extragalaktischen Nebeln*.

Zwicky F., 1937. *ApJ*, **86**, 217. *On the Masses of Nebulae and of Clusters of Nebulae*.