University of Waterloo

Faculty of Science

Automatic Identification and Analysis of Dark Matter Halo Growth in N-Body Simulations

Rishita Gudapati

4B Mathematical Physics

April 15 2019

This report was submitted for partial credit in PHYS 437B
under the supervision of Dr. James Taylor

Abstract

This report outlines an automated process developed to locate dark matter halos and track their growth in an N-body cosmological simulation. Halos are identified using a modified friends-of-friends algorithm and their growth rates are given by their mass accretion histories (MAHs), determined by the program through a largest progenitor search. The properties of halos, including their mass, concentration, and NFW Profile parameters, are tracked through time along each largest progenitor chain. The concentration-age relation derived from the MAHs is examined for two cosmologies with differing values of Ω_m and σ_8.

Contents

List of Figures

1 Introduction

1.1 Background

The prevailing cosmological Lambda-Cold Dark Matter model (ΛCDM) is based on several parameters, including the matter density parameter Ω_m, the fluctuation in matter density σ_8 (averaged over spheres of radius 8 Mpc/h), and the age of the universe [3]. A consequence of this model is that large scale structure evolved hierarchically: matter was distributed uniformly in the early universe with the exception of small fluctuations which clustered together over time under the influence of gravity [4]. The result of this clustering is the configuration of dark matter halos, filaments, and voids shown on the next page. Halos are dense pockets of gravitationally-bound dark matter, and filaments are the trails of particles connecting halos. Spaces that are evacuated by dark matter particles during this clustering are called voids. Halos grow by merging with other halos nearby. Galaxies are found in the centre of dark matter halos [5].

This evolution of complex structure in the real universe can be replicated in a computer using N-body simulations. A popular simulation code is GADGET-2, developed by Volker Springel [6]. The image below is taken from a snapshot of a GADGET-2 simulation. A snapshot is a description of the positions, mass, and velocities of all the particles in the simulation at a given moment in time. A halo is circled in the image, and an arrow points to a filament. Voids are represented in white. The image shown is only a small fraction of a single snapshot.

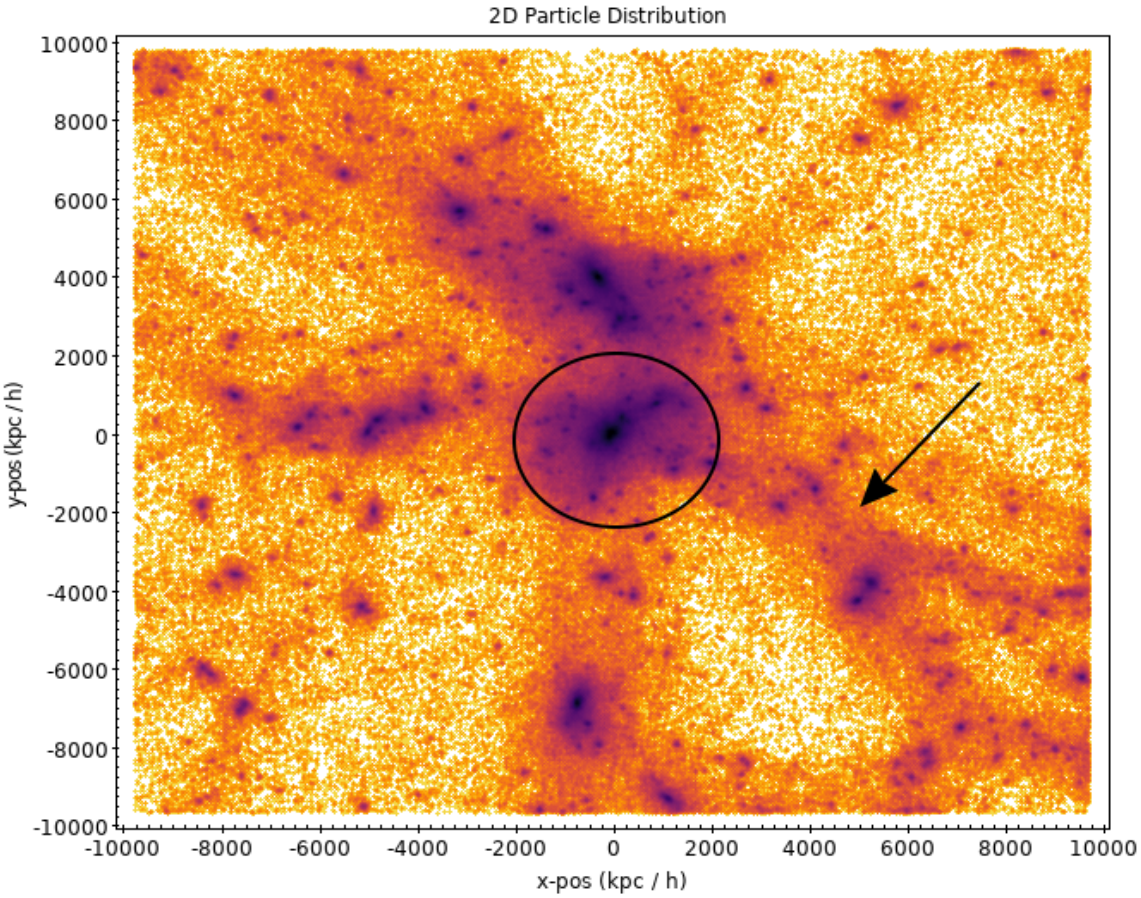

Figure 1: 2-dimensional particle distribution showing large scale structure in a GADGET-2 snapshot

Dark matter halos were observed to fit a simple universal density profile described by Navarro, Frenk, and White in 1997. The NFW Profile is given by

$$\rho(r) = \frac{\rho_s}{\frac{r}{R_s}\left(1 + \frac{r}{R_s}\right)^2} \tag{1}$$

R_s is the halo's scale radius, where its spherically averaged density is proportional to r^{-2}, and ρ_s is a density amplitude [1]. The profile is shown below for various cosmologies in a reproduction from the original paper.

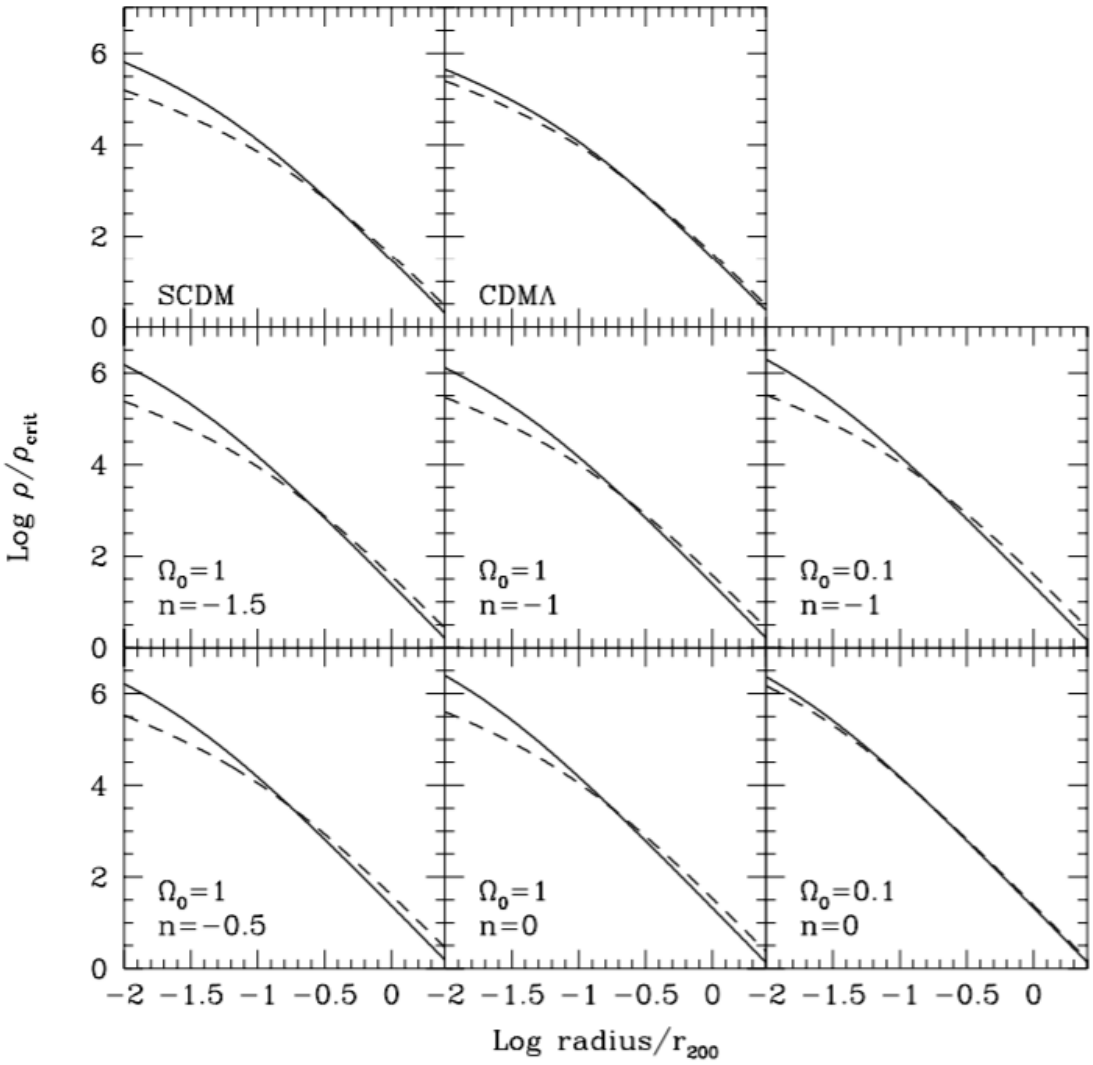

Figure 2: scaled and spherically averaged density profiles of a low mass (solid) and high-mass (dashed) halo taken from [1]

The virial radius of a halo, R_{vir} or r_{200}, is taken to be the radius of a sphere centered around the halo with a mean interior density of $200\rho_c$ [1]. Here, ρ_c is the critical density of the universe, or the density required for the universe to be flat. A halo's virial radius can be determined through its relationship to a virial mass [7]

$$M_{\text{vir}} = \Big(\frac{4\pi R_{\text{vir}}^3}{3}\Big)200\rho_c \tag{2}$$

In this report, M_{vir} is taken to be the FOF Mass defined in Section 2.2 and ρ_c is $1\text{E}{-4} \times h^2$. The virial radius and the scale radius can be used to find the halo's concentration c

$$R_{\text{vir}} = cR_s \tag{3}$$

The Mass Accretion History (MAH) of a halo can be found by tracking its mass as the halo undergoes hierarchical merging.

1.2 Purpose

We want to know the values of Ω_m and σ_8 precisely. The matter density parameter $\Omega_m = \rho_m/\rho_c$ is the ratio of the real matter density of the universe to the critical density, and σ_8 is the root-mean-square fluctuation in the matter density as described in Section 1.1 [5]. The values of these two parameters have been determined separately by space-based and ground based experiments.

The space based WMAP and Planck missions calculated the values of Ω_m and σ_8 by mapping the cosmic microwave background (CMB) [8] [9], whereas the ground based Canada France Hawaii Lensing Survey (CFHTLenS) determined these val-

ues by looking at effects from gravitational lensing [10]. CFHTLenS constrains Ω_m and σ_8 together without determining either of them separately, and although Planck+WMAP were able to provide more precise values, they disagreed somewhat from the CFHTLenS results.

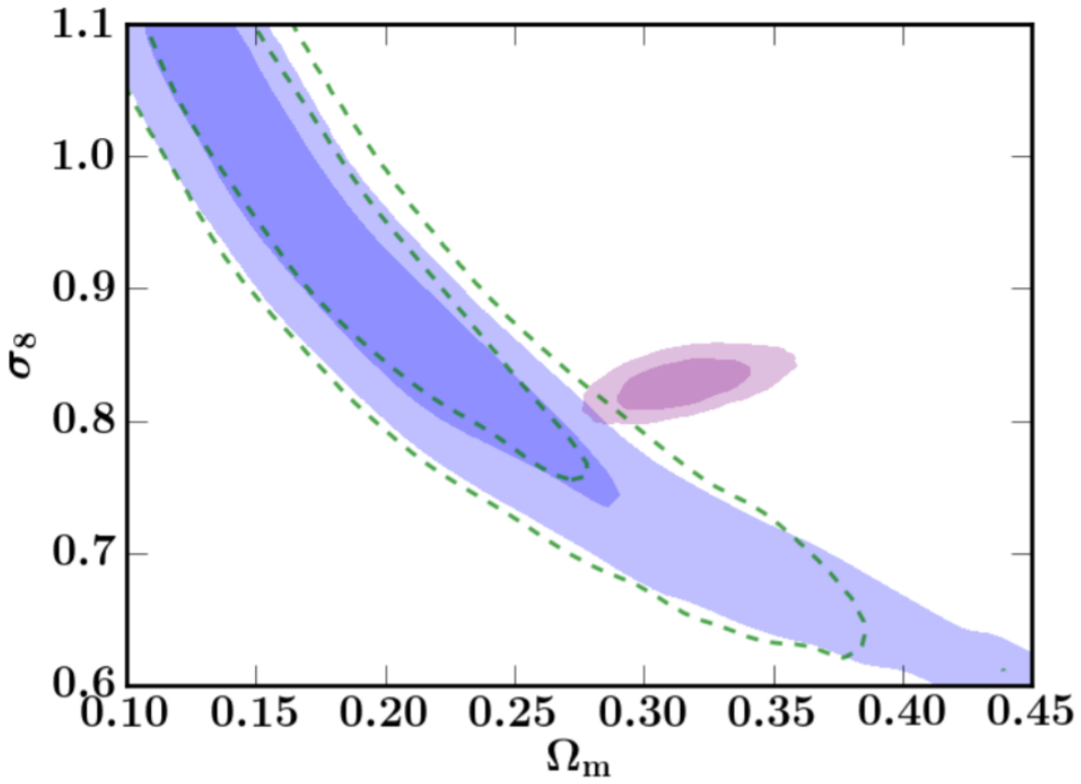

Figure 3: constraints on Ω_m and σ_8 from Planck+WMAP (filled pink) and CFHTLenS (filled purple includes small scales and dashed green excludes them). Inner contours are to 1σ error and outer contours are to 2σ error [2]

To improve the estimates of Ω_m and σ_8 we can try to distinguish between a high Ω_m, low σ_8 universe and a low Ω_m, high σ_8 universe. In a high Ω_m, low σ_8 universe, halo formation is relatively recent. In a low Ω_m, high σ_8 universe, halos are older on average: they stop growing at low redshift but continue to grow at the same rate as the high Ω_m cosmology at higher redshifts [11].

Navarro, Frenk, and White hypothesized that the concentration parameter c of a halo is related to its formation time: halos that form earlier are more concentrated [1]. However, the exact relationship between concentration and age is not known, and neither is the relationship between Ω_m, σ_8, and age.

By tracking the concentration of a halo throughout its MAH we will be able to get a better understanding of the concentration-age relationship. Furthermore, by completing this analysis on different cosmologies we can understand how the concentration-age relation changes with the values of Ω_m and σ_8.

The relationship between concentration c and average growth history of a halo $M(z)/M(0)$ has been explored in Wong & Taylor 2012 for $\Omega_m = 0.27$ and $\sigma_8 = 0.8$ [7]. In this report we first develop some automated techniques to extract the MAHs of halos from GADGET-2 simulations and then discuss the concentration-age relationship for extracted MAHs in $\Omega_m = 0.3, \sigma_8 = 0.8$ and $\Omega_m = 0.25, \sigma_9 = 0.9$ cosmologies.

2 Methods

2.1 Code Structure

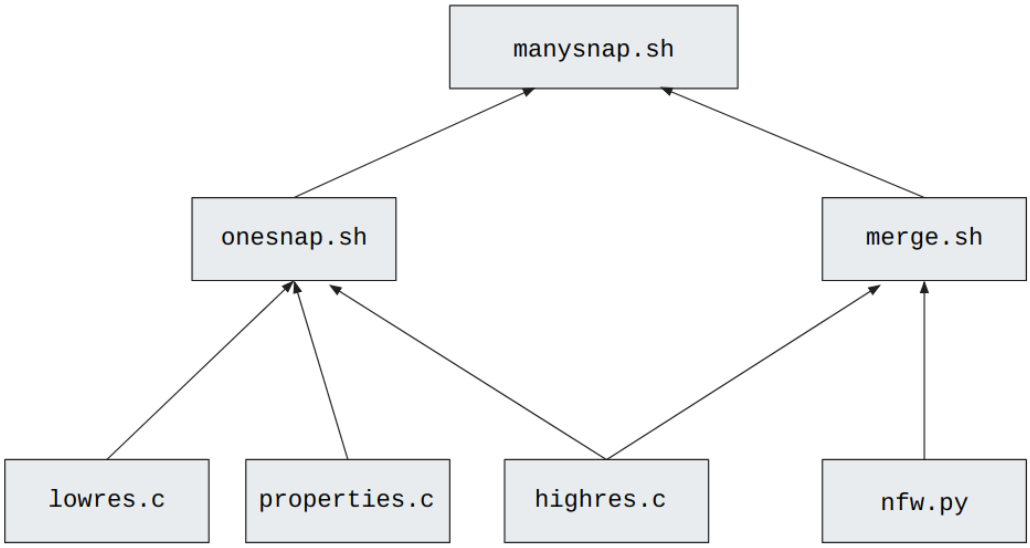

Figure 4: layout of modules in the Automated Simulation Processing Program (ASPP)

The automated processing program contains multiple modules working together to identify halos, build mass accretion histories, and fit density profiles to extract halo concentrations. The entire process is fairy hands off for the user; once paths to script, simulation, and output directories are set up, the user only has to run `manysnap.sh` and all relevant output is automatically generated. The modularity of the Automated Simulation Processing Program (ASPP) also allows for efficient testing and modifications to be made to the code without the need to run a full analysis for every test. A brief overview of each module is below, and a more detailed discussion of important algorithms is given in the following subsections.

Top Layer: ManySnap

The wrapper module `manysnap.sh` is found in the user-defined `$OUTPUT` directory. It requires five parameters, in order:

- `[BEG]` - integer number of the first snapshot to be processed
- `[END]` - integer number of the last snapshot to be processed
- `[NUMHALOS]` - integer number of halos in each snapshot to be processed
  - If `NUMHALOS ==` N the program processes the N largest halos in each snapshot with `cutout1` containing the largest halo
  - this is also the number of progenitor chains that will be produced by the end of the program
- `[LINKLR]` - integer or float friends-of-friends linking length used at low resolution to processes full snapshots
- `[LINKHR]` - integer or float friends-of-friends linking length used at high resolutions to process cutouts around halos

The modified friends-of-friends algorithm used for halo identification by ASPP is expanded on in Section 2.2 which includes information on how to calculate `LINKLR` and `LINKHR` for a given simulation – these values are *not* to be defined by user choice.

example: to run this module `./manysnap.sh 47 51 100 124 78.12`

note: `manysnap.sh` rings a bell once it has finished running, so the user does not have to wait by their computer.

Middle Layer: Management Modules

All code for management modules is found in the user-defined `$OUTPUT` directory. Modules in the middle or bottom layer of the program tree should not be run by the user for any purpose other than testing.

- `onesnap.sh` identifies halos in a single snapshot at low particle resolution and then repopulates cutouts of the snapshot around each halo at high resolution. This is where the modified friends-of-friends algorithm described in Section 2.2 is run.

- `merge.sh` builds largest progenitor chains for N halos using the method described in Section 2.3 and produces the important `evol.merge` file described in Section 2.6. It can only be run after `onesnap.sh` has finished running for all snapshots.

Bottom Layer: Mathematical Modules

All code for for the mathematical modules layer is found in the user-defined `$SCRIPTS` directory.

- `lowres.c` processes a single snapshot in its entirety by returning the positions of a *representative sample* of particles across the snapshot (hereby refered to as low particle resolution). No area of the snapshot is over or under sampled.

- `properties.c` calculates the centers of mass (COM) and diameters d of the N largest halos at low resolution

- `highres.c` using the COM and radial calculations from `properties.c`, this script produces a fully sampled cutout of particles in a 5×diameter box around

the previously determined COM. It is also used by `merge.sh` to return the redshift of each snapshot that is processed.

2.2 Halo Identification

2.2.1 Friends of Friends

The friends-of-friends algorithm described in [12] can be used to identify halos in N-body simulations. FOF groups particles together if they are separated by a distance smaller than the linking length L, which is usually taken to be a function of the mean interparticle separation $\bar{l}$:

$$L = 0.2 \times \bar{l}$$

The units of L should be simulation length units kpc / h.
The "FOF Mass" is then the total mass of particles in a group identified by the FOF algorithm.

This report uses the University of Washington FOF algorithm which takes particle input in TIPSY binary format; TIPSY is another UWashington program that can turn human-readable particle positions into a binary file that is quicker for FOF to process. Both of these resources can be found in the Appendix.

2.2.2 Modified Two-Pass Algorithm

The modified two-pass friends-of-friends (FOF) algorithm used by ASPP is described below for a single snapshot. It involves running FOF once at low particle resolution and then again at complete resolution within a smaller cutout. This process requires different linking lengths at high and low resolution.

- `LINKHR` $= 0.2 \times (b/P)$ where b is the boxsize of the full simulation and P is the total number of particles in the simulation (ie. full resolution). It is dimensionless.
- `LINKLR` $=$ `LINKHR` $\times S^{(1/3)}$ where S is the "skip size" or the decrease in resolution. If 1/4 of the total number of particles P is represented at low resolution, then $S = 4$.

In summary, the full algorithm is:

1. process the entire snapshot at low particle resolution
2. run FOF with `LINKLR` on the entire snapshot
3. choose the N largest halos as determined by FOF. For each halo:
   (a) find its diameter d (described below)
   (b) calculate its center of mass
   (c) cutout a box of size $5 \times d$ around the center of mass
   (d) re-process the cutout box at high particle resolution
   (e) run FOF again on this cutout with `LINKHR`
   (f) the largest group determined by FOF after this second run is the halo of interest in this cutout

The purpose for this two-pass algorithm is to better resolve the edges of each halo and to have a fully sampled box for density calculations.

To find d, the preliminary estimate of a halo's diameter, the program finds the particles at the very outer edges of the halo identified at low resolution. It then calculates the maximum displacement between particles on opposite sides of the halo in each of the three dimensions x, y, and z. It chooses the greatest of these three maximum displacements to be d for the halo.

2.3 Largest Progenitor Search

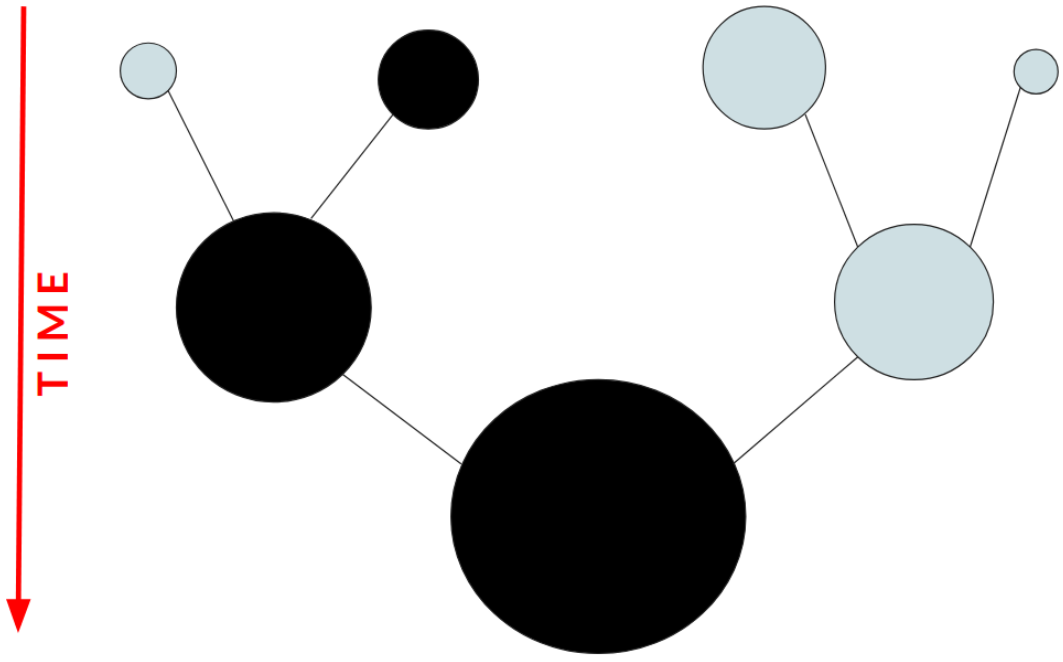

Figure 5: visualization of a dark halo merger tree with the largest progenitor chain shown in black

By looking at a halo's growth through time, we can extract a largest-progenitor chain that describes the MAH of a halo. The process is described below, where "cutout"

refers to the high resolution box around a halo that is produced by the two-pass FOF algorithm.

1. find the halo of interest in the "last" snapshot (the one that is most recent in time or has the lowest redshift z).

2. make a list of all halos in the immediately preceding snapshot which lose at least 50% of their particles to the halo of interest.

3. of all halos in the previous list, chose the one with the greatest mass to be the largest progenitor of the halo of interest

Since we want this process to be done at full resolution, the program does Steps 1 and 2 inside cutouts of the snapshot and repeats for N cutouts, thereby producing N progenitor chains. The program decides which cutouts to use by comparing the center of mass of all cutouts in Step 2's snapshot to the center of mass of the cutout in Step 1. Among all cutouts in Step 2, it chooses the cutout that is minimally separated from the cutout in Step 1 to do the largest progenitor search in.

An important note here is that since high Ω_m simulations generally accrue most of their mass later in time, the progenitor chains for these simulations are relatively short; very quickly, no progenitor can be found because halos become too small.

2.4 Density Fitting to Calculate Halo Concentration

The module `merge.sh` finds halos that appear in a progenitor chain and runs the script `nfw.py` on each of them; `nfw.py` fits a halo to an NFW-density profile in order to determine its concentration. The process is as follows:

1. determine the halo's virial radius R_{vir} using its second-pass FOF mass and equation (2)

2. calculate the center of mass of the halo at full resolution using only the particles that appear in the second-pass FOF group for the halo

3. find all particles in the cutout in an R_{vir} region around the COM regardless of whether or not they are part of the second-pass FOF group from Steps 1 and 2

4. fit the spatial distribution of these particles to a spherically averaged NFW-density profile using equation (1) and `scipy.optimize.curve_fit` to return the halo's scale radius R_s and density amplitude ρ_s; the lower range for the fit is the softening length and the upper range is the virial radius

5. calculate the halo's concentration using equation (3)

This is the least reliable part of the program because a poor fit can be caused by a variety of things. The first is the unreliability of the fit near the virial radius if the center of mass is off or the edge of the halo is ill-defined. This is partially fixed by including all particles in the cutout within the virial range (as we did in Step 3) in case our FOF algorithm missed a few particles during its grouping. We can see the improvement in fit obtained from this step on the next page.

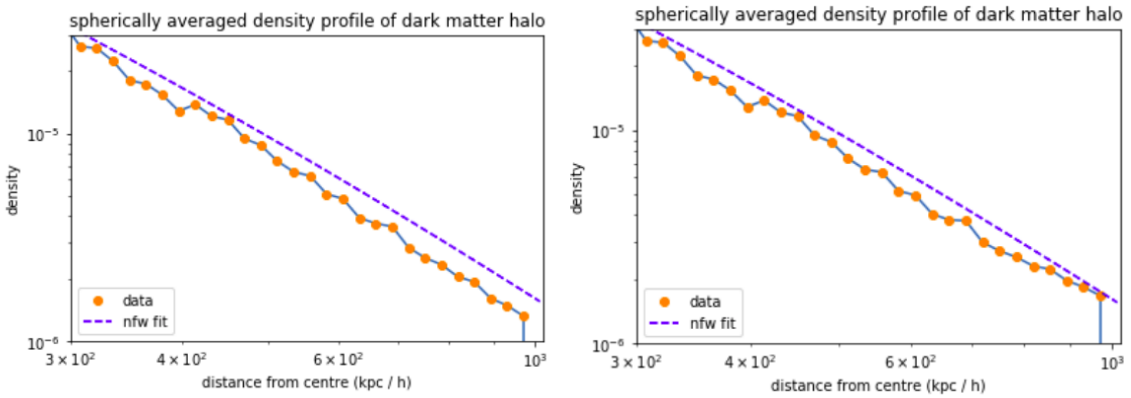

Figure 6: improvements of fit near virial radius for a low concentration halo

The NFW fit is generally poor for low concentration halos. It is worth it to enforce a monotonically increasing requirement on the concentration when analyzing progenitor chains because fits at very high redshift (how high depends on the cosmology) are generally nonesense. As an example, below is the density profile of a halo identified by FOF at the high-redshift/low-concentration end of a progenitor chain; it is extremely irregular and doesn't follow an NFW profile.

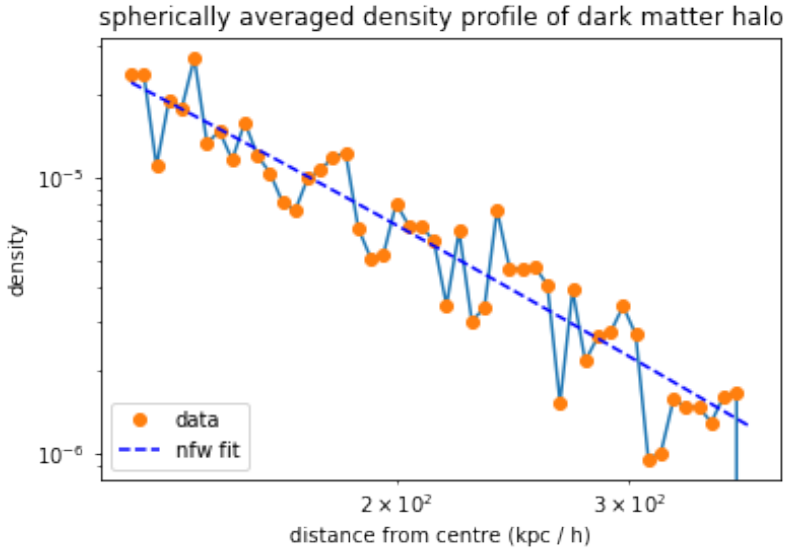

Figure 7: NFW profile fit to a very low concentration halo

Another reason the program might return a bad fit is that the halo is irregularly shaped and/or not isolated enough. As as an example of this, see the cutout below which contains a central halo of interest with overlayed contours. Due to the irregular shape of the halo and the proximity of its neighbour, the virial radius was over-estimated and the fit was done over a radial range that was too large.

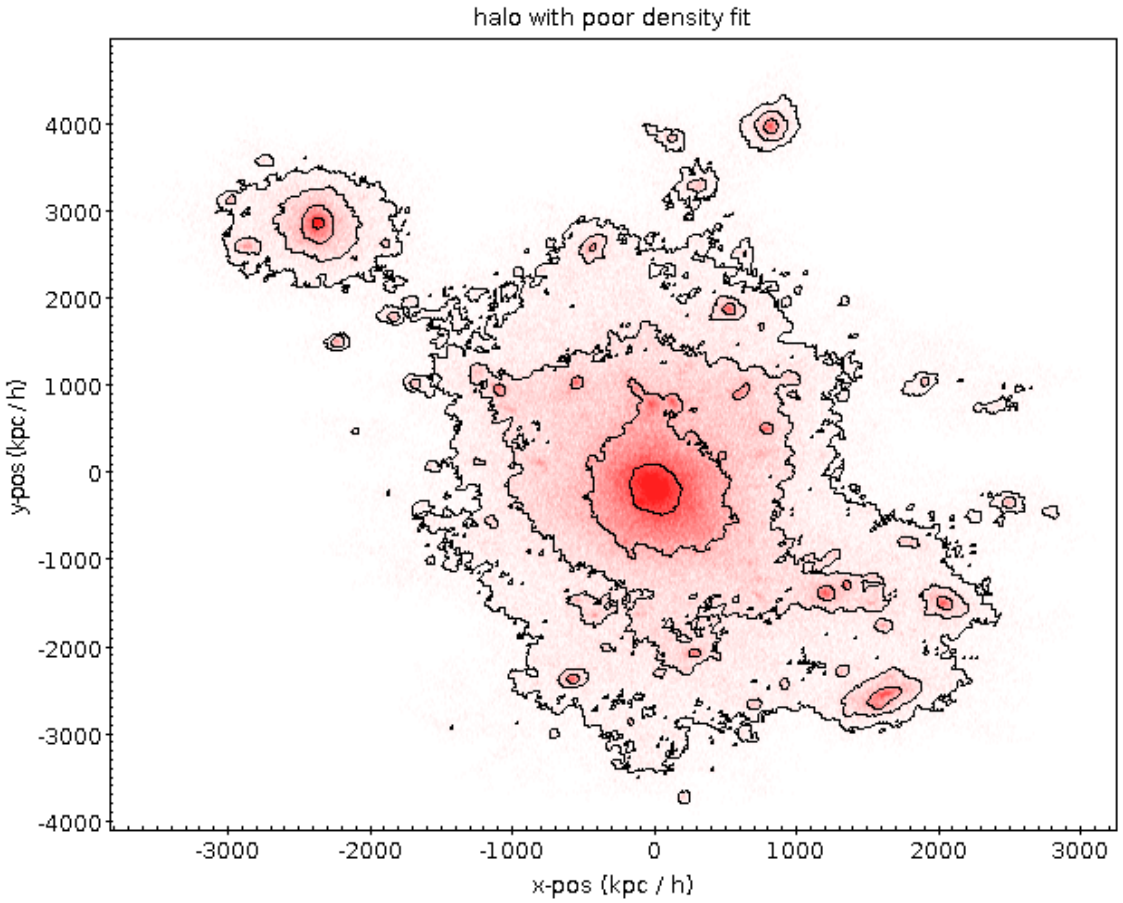

Figure 8: a halo which results in a bad NFW profile fit

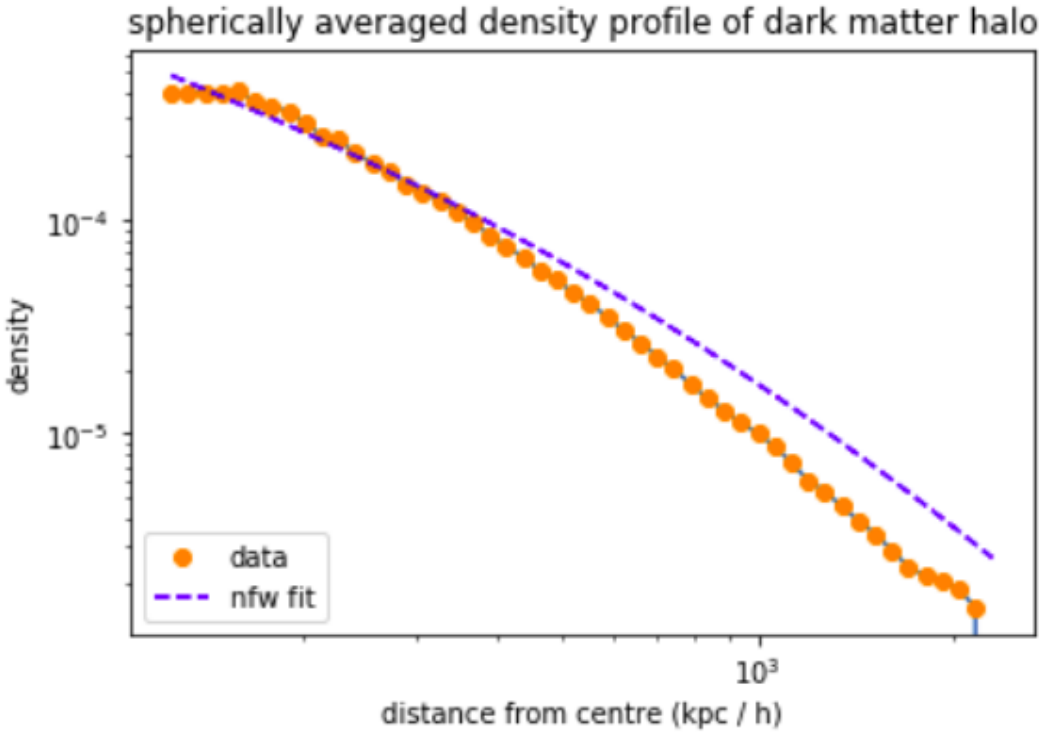

Figure 9: typical NFW profile fit of an irregular and/or non-isolated halo

It is difficult to tackle these problems without manual intervention, so the best option is to increase the number of progenitor chains N and select only those chains with well-behaving fits throughout.

Unfortunately, these problems mean that high Ω_m simulations will more frequently result in bad fits since the shape of halos in these simulations are more irregular and they are less concentrated.

2.5 Time & Space Costs

The upper time limit to process a single halo is 160 real-time seconds on a 16GB RAM machine. This figure increases linearly for the *total* number of halos. For example, if the user wanted to process 5 snapshots with $N = 10$ halos each, she would be processing a total of 50 snapshots which would take her around 2 hours. Generally, a single halo can take anywhere from 90 seconds to 160 seconds to process depending on its size, so this is only an upper estimate.

Sys-time has consistently been around 1/10th real-time on my machine, so 50 halos would take around 15 sys-minutes. Similarly, user-time has consistently been 6/10ths the real time, so 50 total halos would take about 1.5 hours in user-time.

The space-cost is more complicated. Each snapshot produces a 1.5GB file and each halo within a snapshot produces a file that is anywhere from 35MB to 200MB depending on the size of the halo (with 85MB being about average). If we take the upper limit of this requirement, we need $[0.2(s \times N) + 1.5 \times s]$ GB of space for output files where s is the number of snapshots processed and N is as usual the number of halos per snapshot.

2.6 List of Output

- file containing the IDs, positions, and first-pass FOF group numbers of the entire snapshot (one per snapshot)
- summary text file containing the low-resolution center of mass and diameter calculations for each halo (one per snapshot)
- text file containing the number of particles in each of the N halos (one per snapshot)
- file containing IDs, positions, and second-pass FOF group numbers of a cutout surrounding a single large halo (one per N halos)
- an `evol.merge` text file containing N largest progenitor chains with columns in order: snapshot, cutout, and second-pass group number that halo is found in, halo mass, redshift, virial radius, scale radius, density amplitude, and concentration (just one total)

3 Results

The entire simulation box was a 200 Mpc / h cube with 512^3 particles inside it.

3.1 Mass Accretion Histories & Density Profiles

The high Ω_m, low σ_8 cosmology used in this analysis was $\Omega_m = 0.3$ and $\sigma_8 = 0.8$.

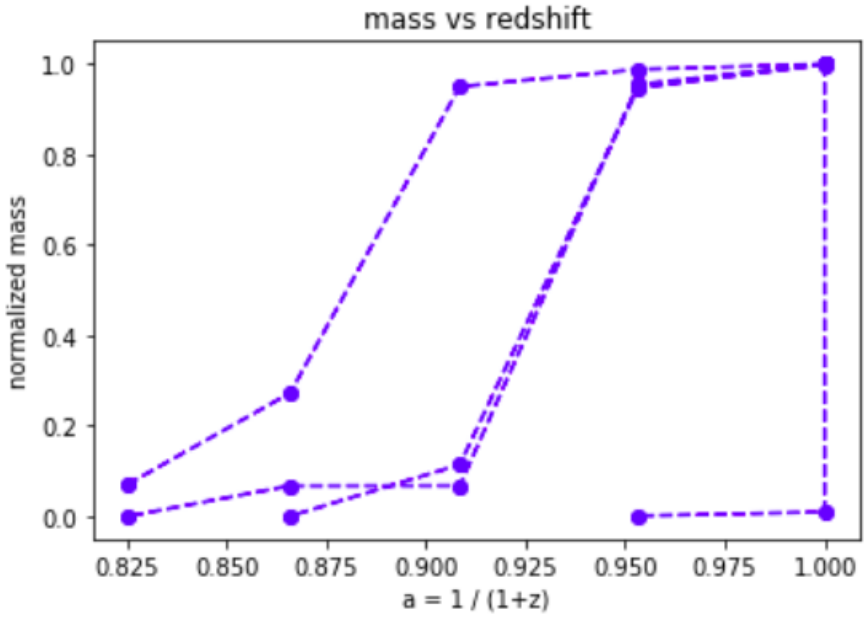

Figure 10: MAH of halos in a high Ω_m low σ_8 cosmology

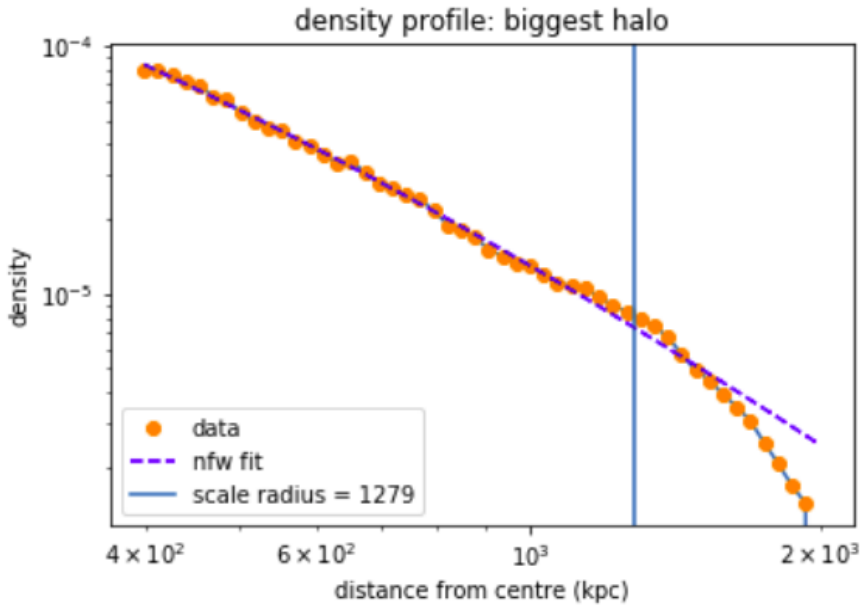

Figure 11: NFW profile fit to halo in a high Ω_m low σ_8 cosmology

The low Ω_m, high σ_8 cosmology used in this analysis was $\Omega_m = 0.25$ and $\sigma_8 = 0.9$.

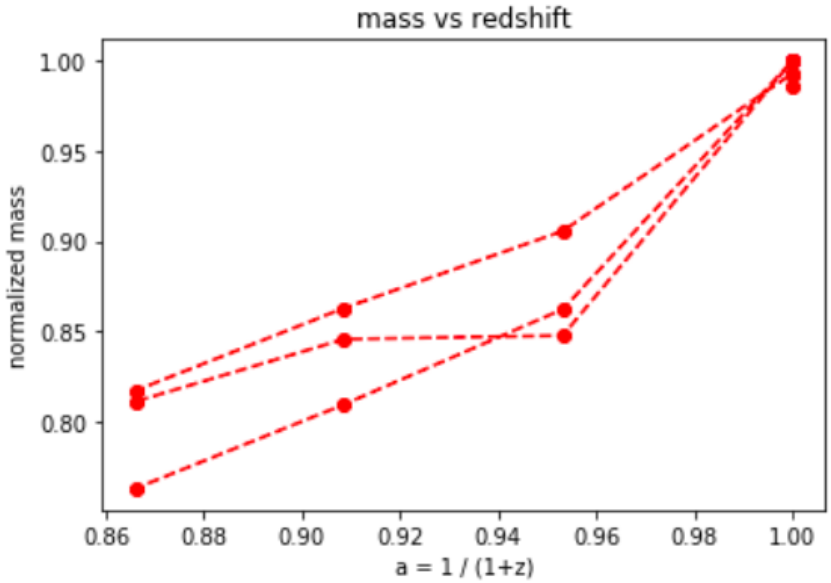

Figure 12: MAH of halos in a low Ω_m high σ_8 cosmology

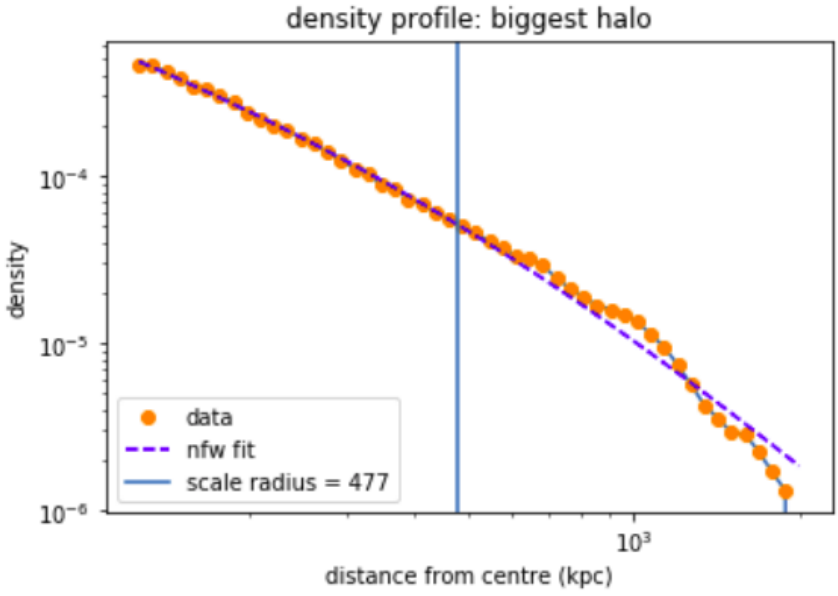

Figure 13: NFW profile fit to halo in a low Ω_m high σ_8 cosmology

It is clear that halos in the high Ω_m universe form later in time when compared to the low Ω_m universe. This is expected.

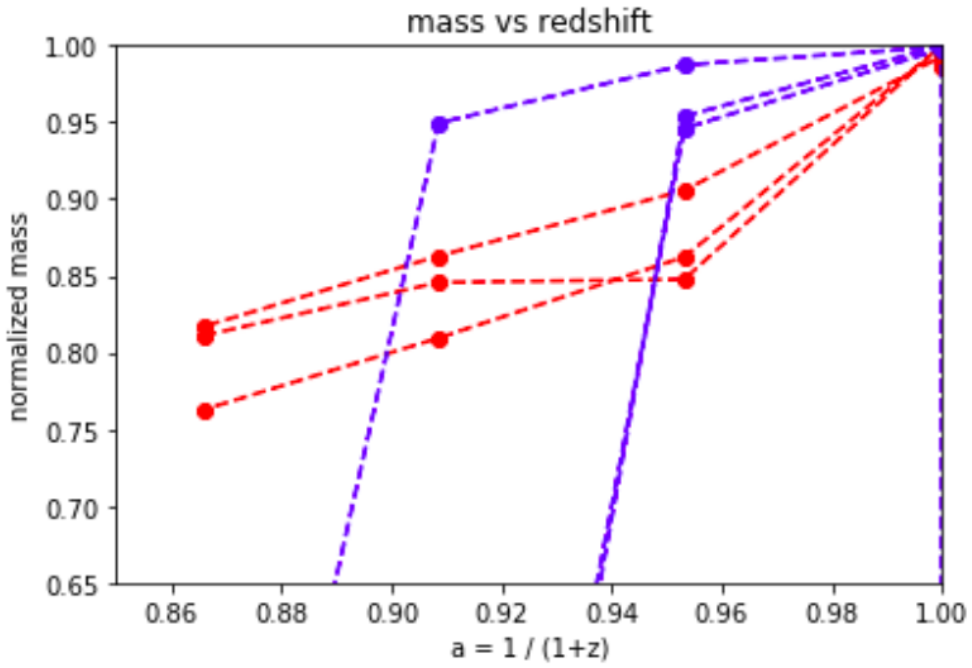

Figure 14: MAH overlay of high Ω_m halos (purple) and low Ω_m halos (red)

3.2 Concentration-Age Relationship

Although there were many good fits in both cosmologies, it was rare that I was able to find a progenitor chain in which every halo was fit well. In this section I will show the result of one consistently well-fit progenitor chain per cosmology. These fits were double-checked by hand. Since I only had a small number of snapshots for the low Ω_m cosmology available, the value of the comparison below is limited because of the small redshift range.

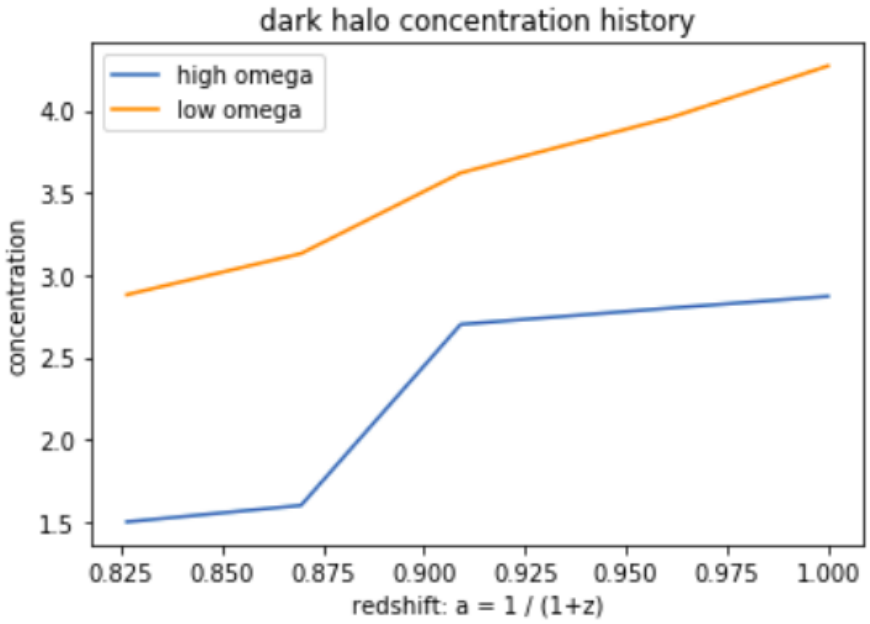

Figure 15: concentration history of a dark halo from its largest-progenitor chain

Regardless, we do see some expected results: the high Ω_m halo is always less concentrated than the low Ω_m one because it forms later; it has less time by the end of the simulation to collect into a uniform shape, and there is less matter available nearby for it to merge with as the universe expands. There is also a jump near the higher a end of the graph as the high Ω_m halo quickly accrues mass. By contrast, the concentration of the low Ω_m halo increases at a steadier rate.

4 Conclusions

In this report, we discussed automating the process of halo identification and growth tracking in an N-body simulation. We also briefly touched on the concentration-age relationship of halos with respect to different cosmologies. We found that the Automated Simulation Processing Program (ASPP) was able to return expected Mass Accretion Histories (MAHs) for two different cosmologies. In the high Ω_m, low σ_8 cosmology, halos accrued their mass slowly at first and then very quickly at lower redshifts. By contrast, the MAHs of low Ω_m, high σ_8 halos were more consistent. The rate of concentration increase in low Ω_m cosmologies was also relatively constant when compared to the large jump in concentration that was evident toward the low redshift end of halos growing in a high Ω_m cosmology.

The ASPP still needs to be developed further, particularly in its ability to make consistently good fits to NFW-density profiles. If this is accomplished, it will be possible to get more reliable concentration calculations more often. Particularly, fits for high Ω_m cosmologies need to be improved because halo concentrations in these cosmologies tend to be smaller. Fits for all low-concentration halos, particularly at lower redshifts in any cosmology, must be improved significantly before there can be any further analysis of concentration-age relationships.

Finally, once the concentration calculations become reliable, ASPP needs to be run for an entire simulation with at least $N = 1000$ halos processed per snapshot. Although one full simulation analysis was run for this report, it was for a high Ω_m cosmology and therefore produced short progenitor-chains regardless of the number of snapshots processed. This is because halos in high Ω_m cosmologies accrue mass

relatively slowly; halos earlier in the simulation will not have grown much and therefore will be too small to have a clear largest-progenitor.

Although we have found some results in this report, once full simulation analyses are run for multiple cosmologies, the dependence of the concentration-age relation on the values of Ω_m and σ_8 will become even more clear.

5 Acknowledgements

I would like to thank my supervisor James Taylor for overseeing my project and providing guidance throughout the process. I'd also like to thank the Department of Physics & Astronomy for the research opportunity and for evaluating my results. Finally, I'd like to thank Michael Vansteenkiste for proofreading my report.

References

[1] J. F. Navarro, C. S. Frenk, and S. D. White, "A Universal Density Profile from Hierarchical Clustering," *The Astrophysical Journal*, vol. 490, no. 2, p. 493, 1997.

[2] N. MacCrann, J. Zuntz, S. Bridle, B. Jain, and M. R. Becker, "Cosmic discordance: are planck cmb and cfhtlens weak lensing measurements out of tune?," *Monthly Notices of the Royal Astronomical Society*, vol. 451, no. 3, pp. 2877–2888, 2015.

[3] D. N. Spergel, "The dark side of cosmology: Dark matter and dark energy," *Science*, vol. 347, no. 6226, pp. 1100–1102, 2015.

[4] G. R. Blumenthal, S. Faber, J. R. Primack, and M. J. Rees, "Formation of Galaxies and Large-Scale Structure with Cold Dark Matter," *Nature*, vol. 311, no. 5986, p. 517, 1984.

[5] P. Schneider, *Extragalactic Astronomy and Cosmology: An Introduction*. Springer Berlin, 2016.

[6] V. Springel, "User Guide for GADGET-2," 2005.

[7] A. W. Wong and J. E. Taylor, "What do Dark Matter Halo Properties Tell Us about Their Mass Assembly Histories?," *The Astrophysical Journal*, vol. 757, no. 1, p. 102, 2012.

[8] P. A. Ade, N. Aghanim, M. Alves, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, H. Aussel, C. Baccigalupi, *et al.*, "Planck 2013 results. i. overview of products and scientific results," *Astronomy & Astrophysics*, vol. 571, p. A1, 2014.

[9] G. Hinshaw, D. Larson, E. Komatsu, D. Spergel, C. Bennett, J. Dunkley, M. Nolta, M. Halpern, R. Hill, N. Odegard, *et al.*, "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," *The Astrophysical Journal Supplement Series*, vol. 208, no. 2, p. 19, 2013.

[10] C. Heymans, E. Grocutt, A. Heavens, M. Kilbinger, T. D. Kitching, F. Simpson, J. Benjamin, T. Erben, H. Hildebrandt, H. Hoekstra, *et al.*, "Cfhtlens tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments," *Monthly Notices of the Royal Astronomical Society*, vol. 432, no. 3, pp. 2433–2453, 2013.

[11] Q. Guo, S. White, R. E. Angulo, B. Henriques, G. Lemson, M. Boylan-Kolchin, P. Thomas, and C. Short, "Galaxy Formation in WMAP 1 and WMAP 7 Cosmologies," *Monthly Notices of the Royal Astronomical Society*, vol. 428, no. 2, pp. 1351–1365, 2012.

[12] A. Knebe, S. R. Knollmann, S. I. Muldrew, F. R. Pearce, M. A. Aragon-Calvo, Y. Ascasibar, P. S. Behroozi, D. Ceverino, S. Colombi, J. Diemand, *et al.*, "Haloes gone MAD: the halo-finder comparison project," *Monthly Notices of the Royal Astronomical Society*, vol. 415, no. 3, pp. 2293–2318, 2011.

Appendix

The University of Washington FOF algorithm can be found here:
`https://drive.google.com/drive/folders/0B826B9MKlmILc0toeTc2akFsX3M`.
The version used in this report is `fof-1.1`

The above link is also where the latest version of TIPSY can be found and installed.

Figures 1 and 8 were produced by TOPCAT. Its documentation can be found here:
`http://www.star.bris.ac.uk/~mbt/topcat/`

GADGET-2 is found along with its documentation here: `https://wwwmpa.mpa-garching.mpg.de/gadget/`

All code referred to in this report can be found at the following GitHub Repository
`https://github.com/r2gudapa/darkhalosims`