GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Fakultät für Physik

Master Forschungspraktikum

Laboratory Exercise AG.DMH

Numerical Analysis and Data Visualisation for Dark Matter Halos

supervised by

Dr. Benedikt Eggemeier

updated version of the previous AG.DMH course by Dr. K. Clough

Last updated March 2023

Summary

The main aim of this lab is to get some experience in visualisation of large data sets, and the packages that researchers commonly use for these tasks. We will be doing this in the context of exploring simulations of cosmological dark matter, and specifically the analysis of dark matter (DM) halos. This is a fascinating and rich topic in modern physics that brings together theory, observation and computational research to address a number of unanswered questions about our Universe.

We will all start with the same publicly available data files, the outputs from an N-body dark matter simulation produced using the code Enzo[1]. The tasks will be a mixture of visualisation and analysis, with some opportunities for drawing simple conclusions about the data being studied. You may optionally look at how to parallelise the tasks (to split the work between several computer processors simultaneously), which is important for larger data sets.

Please note the following:

All levels of computing experience are welcome. I hope that those without a strong background in computing and programming will be encouraged to find out more about how computers can facillitate research - it really is never too late to learn. For those with a stronger computing background, I hope that you will challenge yourselves by going beyond the standard exercises suggested.

There are a lot of things that you can do in this lab - part of the exercise is using your judgement to decide what is interesting to investigate and what is not, given the time available.

You may work in a pair during this lab, and as a result your plots may be the same, but you should both write your own report independently, especially the research sections. Any form of copying beyond sharing the data and plots created in the Lab will not be accpeted.

We will be using various programs and packages, as detailed in the Introduction section. Please use your own laptop and install the required packages prior to the lab.

[1] If you are interested in generating the files yourself, details on running some basic simulations with Enzo are provided in the Appendix, see Section 1.6, but this does not form part of the lab.

Installation instructions are provided in the Appendix, see Section 1.6.

The report should be typeset in English and as brief and clear as possible, while still addressing all the tasks. It also needs to include brief explanations of all equations used, including proper citations for all sources that you consulted. Please note for the research tasks that Wikipedia is NOT an acceptable reference, although it may be a useful starting point to direct you to published, peer reviewed sources. The written report is to be handed in within 14 days of completion of the lab.

I suggest using LaTex to write up your report, which supports proper referencing, and is the current standard for most scientific publications. The online editor Sharelatex (sharelatex.gwdg.de) is a very good option. I recommend using the `cleveref` package for consistent referencing to equations, figures and tables.

Please feel free to contact me by email (benedikt.eggemeier@phys.uni-goettingen.de) with any questions or comments on the material.

Figure 1.1: The visible matter in the Andromeda galaxy - stars and gas. The DM halo is (currently) only visible from its gravitational effects. (Commons license image).

1.1 Introduction

In this lab we are going to visualise the data from N-body simulations, and identify and analyse the dark matter halos formed.

When we look at a picture of a galaxy, like the one in Fig. 1.1, what we are seeing is the visible galactic disc, which is made up of "normal" matter like stars and gas. A dark matter halo is a component of a galaxy that permeates and extends beyond the visible galactic disc, and, although invisible, it dominates the total mass of most galaxies.

When we say that dark matter is invisible, what we mean is that it interacts very weakly (if at all) with normal matter via the electromagnetic interaction - much like neutrinos, which we are familiar with, but remain difficult to detect "directly". The existence of dark matter may be inferred through its gravitational effects on the motion of the visible stars and gas in galaxies, and also by its effect on the Cosmic Microwave Background (CMB) radiation.

Dark matter halos play a key role in current models of galaxy formation and evolution, and one of the main research approaches consists in N-body simulations of the evolution of halos from initial density perturbations in the early Universe. When we are able to

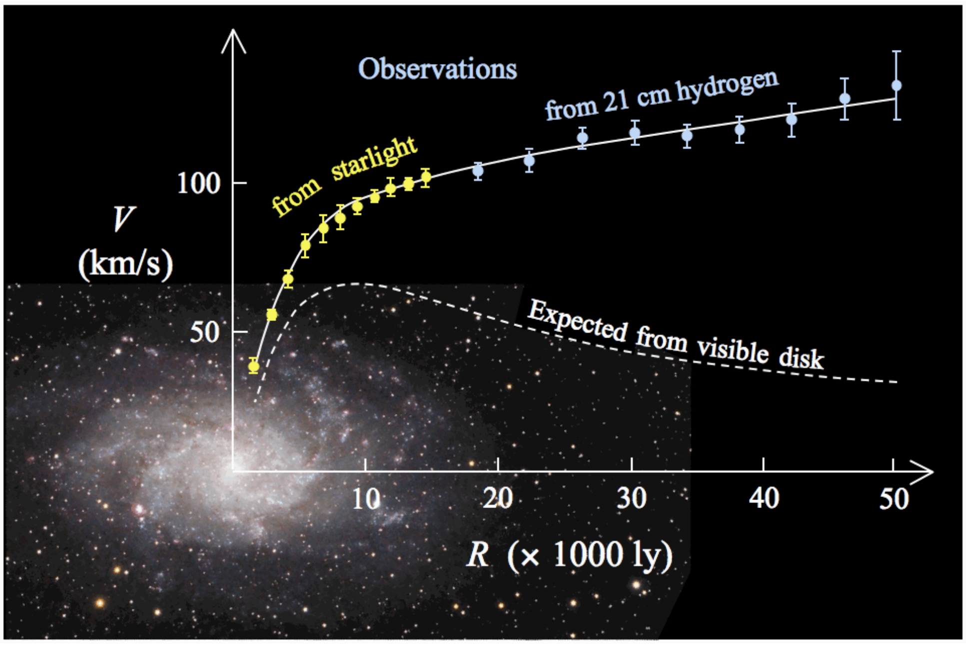

Figure 1.2: Rotation curve of spiral galaxy M33 (yellow and blue points with error bars), and a predicted one from distribution of the visible matter (white line). The discrepancy between the two curves can be accounted for by adding a dark matter halo surrounding the galaxy. (Commons license image).

accurately simulate galaxies with properties that correspond to those we observe in the Universe, this gives us confidence that we understand the physics behind dark matter. There are currently several discrepancies between simulations and observations, and reconcilling these differences is an active area of research. In this lab, we will be studying the output of some such simulations, generated using the Enzo code.

1.1.1 Background theory

The presence of a dark matter (DM) halo is inferred from its gravitational effect on a galaxy's rotation curve - this is a plot of the orbital velocity of the stars versus their radial distance from the galactic centre. A typical example is shown in Fig. 1.2.

Based on the distribution of visible matter in galaxies, the rotational velocity of the stars should decrease at large distances from the galactic center. However, observations of spiral galaxies, particularly radio observations of line emission from neutral atomic hydrogen (HI), show that the rotation curve of most spiral galaxies flattens out to a fairly constant value of orbital velocity. One can also estimate the mass of visible matter in a galaxy from its brightness, but this is found to be insufficient to maintain the velocities observed. This implies either that unobserved ("dark") matter exists, or that the theory of motion under gravity (General Relativity) is incomplete.

Dark matter halos are believed to have played a major role in the early formation of galaxies. When galaxies were first being formed, the baryonic matter should not have been able to form gravitationally self-bound objects. The prior formation of dark matter structure in the early universe (assuming that the dark matter is "cold") overcomes this problem.

The current hypothesis for DM structure formation begins with density perturbations in the Universe arising from quantum fluctuations during Inflation. These grow until they reach a critical density, after which they stop expanding and collapse to form gravitationally bound dark matter halos. These halos continue to grow, either through accretion of material from their immediate neighborhood, or by merging with other halos.

Numerical "N-body" simulations support this view. In a standard simulation, a volume containing small density perturbations initially expands with the expansion of the Universe. As time proceeds, small-scale perturbations grow and collapse to form small halos. At a later stage, these small halos merge to form a single virialized dark matter halo with an ellipsoidal shape, often with some substructure in the form of dark matter sub-halos. Once these subhalos have formed, the gravitational interaction with baryons allow the latter to collapse into the first stars and galaxies.

Typical N-body simulation data is shown in Fig. 1.3, from which you can see that it models the movement of many (N) dark matter particles (bodies). Whilst simulations of early galaxy formation roughly match the structures observed in galactic surveys, several differences have been noted, which imply that we do not yet fully understand the physics of dark matter structure formation.

1.1.2 Summary of packages used

The packages we will use are briefly described in this section, and information and links to the download instructions are provided in the Appendix, see Section 1.6.

We will use the Python 3 based packages Jupyter (formerly called iPython notebooks)

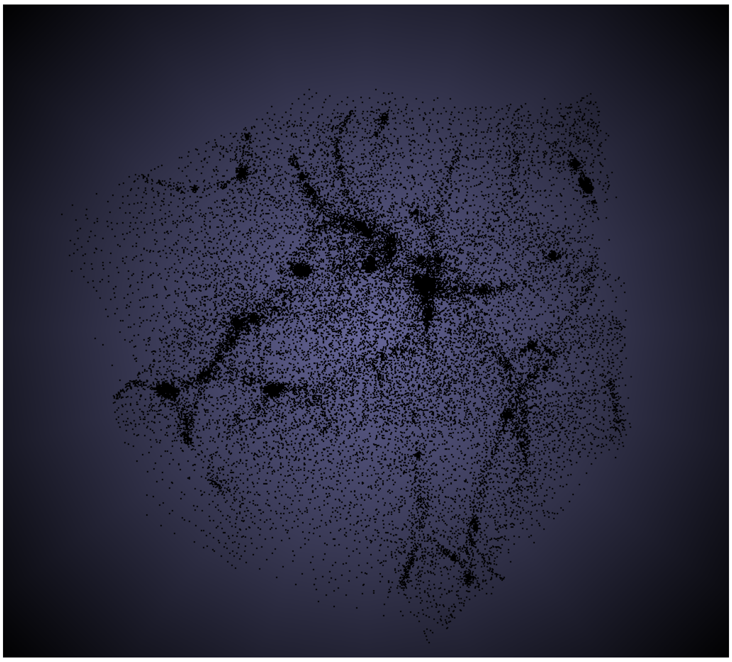

Figure 1.3: Visualisation (in VisIt) of Enzo particle data after structure formation.

and yt. These are developed in Python under the open-source model, and thus a wealth of user documentation and community forums exist.

Jupyter allows us to create “notebooks” in which we can execute snipets of Python code “on the fly”, without having to run an executable on the command line. It provides an incentive to present your code in a readable and ordered fashion - you should aim to produce something that you could share with someone else, such that they could understand it without further explanation. Two template Jupyter notebooks will be provided which you will update during the lab, and a print out of your notebooks will form part of your report. Another useful feature is that they allow tab-completion of commands, their arguments and file locations.

yt is a community-developed analysis and visualization toolkit for volumetric data. yt has been applied mostly to astrophysical simulation data, but it can be applied to many different types of data including seismology, radio telescope data, weather simulations, and nuclear engineering simulations. There is a strong collaboration between yt and Enzo, which makes it an ideal package for viewing the data. The yt website gives comprehensive instructions for using the code in a variety of situations, and I will borrow heavily from its demo sections for the work below.

You can check Jupyter works by opening a terminal window and typing:

```
jupyter notebook
```

which should bring up the current directory in an internet browser window. Click on "new" to create a new Python 3 notebook, and test it works by typing something simple like:

```
a = 4
b = 2
print(a*b)
```

The command "Shift-Enter" executes the code block, and the result of the print operation should be displayed. The variables a and b are now stored in the kernel (the memory) and can be used in the next box for further operations.

Note: The notebook equivalent of "turning it off and on again" when everything goes wrong is to restart the kernel, clearing all your data and hopefully any bugs you have introduced. You should usually clear the kernel before starting a new notebook.

1.2 Pre lab research tasks

You will gain more from the lab if you understand some of the physics behind the simulations in advance. You should therefore try to perform the following tasks before the lab. A concise summary (3-4 pages maximum) should be part of the theory section of your report.

TASK 1.2.1 Galaxy rotation curves and the NFW profile

1. Derive from first principles (Newtonian gravity) and plot the expected orbital velocity profiles (orbital velocity versus radius) for stars in a galaxy, assuming a circular orbit, and the following density profiles:

a) Constant density $\rho(r) = \rho_0$ within a radius R_0 and zero density for $r > R_0$

b) Density which decreases as $\rho(r) = \rho_0/(1 + (r/R_0)^2)$

c) The NFW profile (see below) $\rho(r) = \frac{\rho_0}{\frac{r}{R_s}\left(1 + \frac{r}{R_s}\right)^2}$

2. Astronomers studying galaxy UGC 128 have measured its brightness and calculated that the mass of visible matter within a radius of 1.30×10^{21} m is 3.34×10^{40} kg. Stars orbiting at this radius has been measured travelling at a speed of 1.30×10^5 m/s. What percentage of the mass within this radius is dark matter?

3. The Navarro Frenk White (NFW) profile is a spatial mass distribution of dark matter fitted to dark matter haloes identified in N-body simulations by Julio Navarro, Carlos Frenk and Simon White (read the abstract of their paper at `https://arxiv.org/abs/astro-ph/9508025`). The NFW profile is one of the most commonly used model profiles for dark matter halos, although it is not consistent with observations of low surface brightness galaxies which have less central mass than predicted. This is known as the cusp-core problem.

The NFW profile is an approximation to the equilibrium configuration of dark matter in a simulation, ie, it assumes that it has virialised. Explain what “virialised” means (in the context of dark matter structure formation). What is the virial radius?

TASK 1.2.2 Enzo simulations

Look at the Enzo code paper `https://arxiv.org/abs/1307.2265v1` to understand what it does. This is a very long paper and I do not expect you to read it all. You should at least read the introduction, and look at the key equations in section 2.1. Then try searching for "dark matter" and focus on understanding the N-body aspects of the code.

Summarize in your own words what Enzo does and how it models Dark Matter. What interactions does the Dark Matter have with the other components in the simulation? What assumptions are made? A basic simulation involves the collapse of dark matter particles under their own gravity. What other physics can be added to the simulations beyond this?

TASK 1.2.3 Dark matter theory

Research and summarise your findings on AT LEAST ONE of the following theory topics:

1. Research different dark matter candidates and briefly summarise your findings. What is are the strengths and weaknesses of each model from a theoretical and observational perspective?

2. Research current direct detection experiments. What limits have been put on the mass and interaction cross section of the candidate dark matter particles? What are the key backgrounds and how do detector experiments compensate for these?

3. Galaxy rotation curves are not the only evidence for dark matter in our universe. Research and summarise the other key pieces of evidence for dark matter.

4. Some researchers believe that the need for Dark Matter can be removed by modifying how gravity acts on larger scales. Provide arguments for and against this view, and provide brief details of a modified gravity model of your choice.

5. Summarise some of the discrepancies between simulations and observations, which occur mainly on small scales. For example, the "too big to fail", and "cusp-core" problems.

1.3 Visualisation of Enzo data with yt

Enzo is an adaptive mesh refinement (AMR), grid-based hybrid code (hydro + N-Body) which is designed to perform simulations of cosmological structure formation. **yt** is one of the most suitable packages for detailed analysis of Enzo data, and in particular a number of tools have been developed for use with Enzo by the yt community. However, yt can also be used for datasets generated by most other modern simulation tools.

The learning objectives are:

- Understand what Enzo is simulating.
- Understand how to load data into yt, view it, do analysis and produce plots.
- Understand how to run yt in parallel, why this is useful, and what the limitations are.

1.3.1 Enzo simulations and data

Download the `enzo_tiny_cosmology` and `Enzo_64` data from the yt site here http://yt-project.org/data/. Look at the file `32Mpc_32.inits` and, if you can, describe some of the parameters that have been set for the simulation.

1.3.2 Visualisation with yt

We will now use yt to plot and analyse the data in a Jupyter notebook. Make your own copy of the example `DMH_Visualisation_Example.ipynb` notebook that you find in the Lab folder somewhere in your home directory. Open it using Jupyter notebook, and use shift-enter to execute each section, checking you understand what is being done at each stage. You should generate some images like the one in Fig. 1.4.

Note that you may have to amend the path to the data files, depending on where you have saved your notebook.

One of the main learning outcomes should be that if you want to do some data analysis or visualisation task, you almost certainly aren't the first person who has needed to do something similar. Therefore what you want to do probably exists somewhere on the

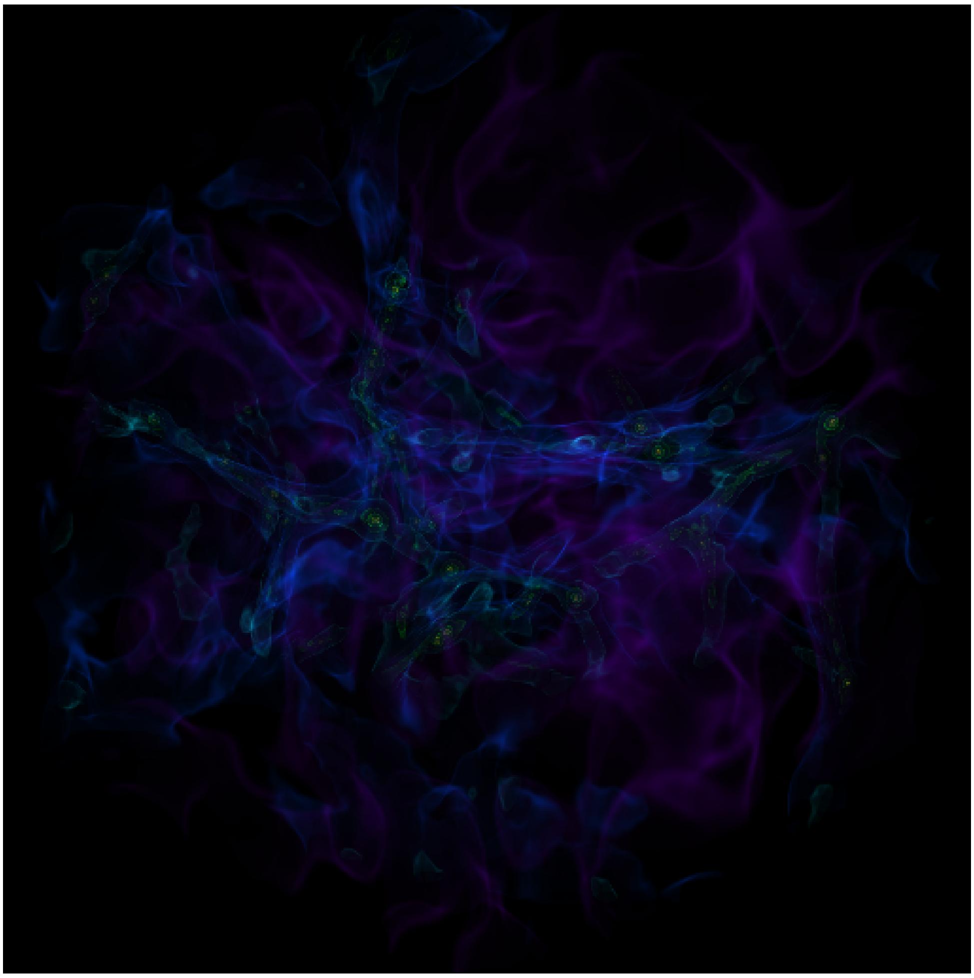

Figure 1.4: Volume rendering of density in yt, using Enzo simulation data.

internet. Don't reinvent the wheel. Having said that, the tools can sometimes make the tasks so easy that one does not understand the assumptions that have gone into the data extraction. Therefore one should not mindlessly copy, but rather take some time to understand the way in which the analysis is being conducted, and assess if the method is suited to the task at hand.

Having executed the example notebook, amend and/or add to it to complete the following tasks:

TASK 1.3.1 yt visualisation tasks

Take some time to explore other features of yt. If you get stuck, a yt cheatsheet is provided in the Lab folder, or you can search the online documentation for other examples and hints.

- Recreate the particle and temperature plots for an earlier output dataset from the simulation (maybe the second or third). What are the key differences?

- Create three plots which you find interesting. In this context, interesting means that you either learned a new yt technique, or that you can derive a sensible physical/cosmological result from your plot. If you use the same plots as in the example, change something about them - for example, change the colours and/or zoom in to show a region of interest, add isocontours, overlay the simulation grid to check when resolution is added, or label key features. Provide details of the plotting tools applied, and draw some relevant conclusions about the physics that each plot shows.

- Find the maximum and minimum densities of dark matter. Make sure you give the correct units taking into account the dimensions of the simulation box (Note: comoving units Mpccm/h, are Mpc / (1+z) / h, with h the hubble paramter $h \sim 0.7$). Assuming dark matter is a WIMP (Weakly Interacting Massive Particle) with a mass of 10 GeV, what is the maximum number density of particles implied by the simulation?

- Create a Vector plot of the baryon gas velocities in a slice through the point of maximum dark matter density (likely the center of a halo). Does it appear virialised? You may wish to slice at several angles to build up a better picture of this.

- Improve the plot of the max density versus redshift for the time series data (use the matplotlib cheat sheet, adding labels, scales etc) and plot a time series of a second quantity, (e.g. min/min temperature) in the same way. What do these plots show?
- OPTIONAL: You can also repeat some of the tasks with the larger Enzo64 data set.

TASK 1.3.2 (OPTIONAL) Parallelisation of visualisation tasks

Parallelisation shares the work yt does between computer processors, which means you can do more work in less time. This is useful for very large data sets. Parallelisation is in principle possible in the Jupyter notebook, but realistically if you are working with a large data set, you will want to create an executable python file like the one we will use now, perhaps to run it on a computer cluster, rather than having an interactive notebook.

Open the file `DMH_parallel_script.py` and try to understand what it does. The code enables parallelism in yt, which is described in detail here: `http://yt-project.org/doc/analyzing/parallel_computation.html`. However, all you really need to do is add `yt.enable_parallelism()` at the top and yt does the rest. You can run this code on the command line using:

```
mpirun -np 2 python script.py
```

which will run it on 2 cores (if you have a dual core machine). Most laptops have a dual core processor, and so you should be able to get an approximately 2x speedup in this case, compared to running it in serial, which you do by typing

```
python script.py
```

Here we use the time package in Python to measure the time take to excecute the code, which is very useful for measuring code performance and identifying "bottlenecks".

Note down how much of a speed up you get from 2 cores versus one. Then change the code so that it executes several different tasks (e.g. doing a slice plot and then a projection plot), and again note what speed up you get. (NOTE: Don't worry too much if you don't get any speed up, but make sure you can see that two processors are doing the work - you should get double the amount of output.)

Explain why one can never achieve exactly 2x the speed up on two cores for this kind of visualisation task. For some tasks it may even take longer in parallel - suggest why this may be the case.

1.4 Halo finders in yt

Having run simulations which generate dark matter halos, the halos need to be identified. In a large simulation, doing this "by eye" is going to be inconsistent and inefficient. We want an automated way of identifying where halos are. This, however, requires us to define what we mean by a dark matter halo.

The basic idea of any Halo finder is to identify clusters of particles which are dense relative to their surroundings, and this is usually done by identifying chains of particles which are within a certain distance of each other. One of the main Halo finders is HOP, for which the algorithmic details can be found here: `https://arxiv.org/abs/astro-ph/9712200v1`

Learning objectives in this section are:

- Understand how Halo finder algorithms work.
- Learn how to apply them to Enzo data sets in yt and visualise the results.
- Investigate the density profiles of the halos and compare them to theoretical expectations.

1.4.1 Halo finding

In your report you should summarize (briefly) in your own words what HOP does and how it identifies DM halos.

We will now use yt to find and plot halos. Make your own copy of the example `DMH_Halo_Example.ipynb` notebook that you find in the Lab folder. Open it using Jupyter notebook, and use shift-enter to execute each section, checking you understand what is being done at each stage. You should generate some images like the one in Fig. 1.5.

Having executed the example notebook, amend and/or add to it to complete the following tasks:

TASK 1.4.1 Halo finding tasks

- Add the NFW density profile from Section 1.2 onto the plot of the density profile for the 2nd halo (you can pick some values of the constants R_S and ρ_0 to fit the

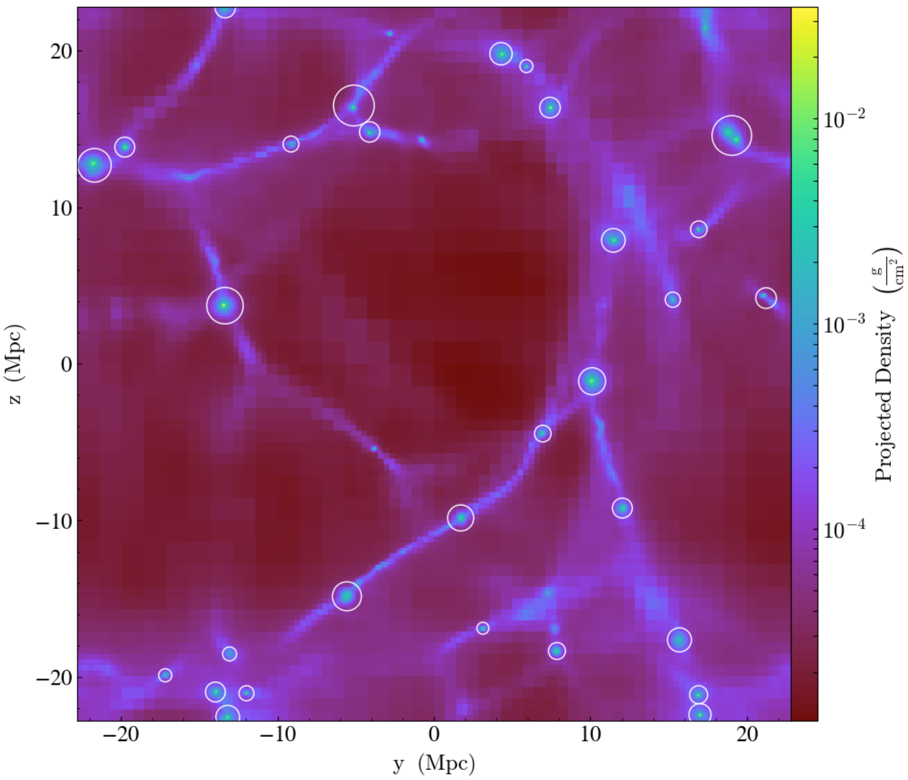

Figure 1.5: Identification of halos with yt Halo finder - projected density plot with DM Halos marked.

data as best you can). Is it a good fit? Plot the density profile versus radius for another halo for which the fit of the NFW profile is not so good, and suggest why this may be the case.

- Now plot the profiles for the Temperature versus radius for two of the halos. Comment on your result.

- The `finder_method` options can be given as "fof" or "hop" (or "rockstar", but only in parallel, so using this is optional). Try each option and compare the differences. If you have time in the lab, try to find some information about the FOF algorithm, and explain the resulting difference in the results.

- OPTIONAL TASK: There are parallel versions of HOP, for which details can be found at `https://arxiv.org/abs/1001.3411`. Summarise briefly the method used for parallelisation of the Halo finding task for this algorithm, and try running it in parallel to see whether you can achieve a speed up in the Halo finding process.

1.5 References and sources of further information

Most of the resources used are online, and the URLs have been given in the relevant parts of the text above. The key papers and references are:

- **The Structure of Cold Dark Matter Halos**
  Julio F. Navarro, Carlos S. Frenk, Simon D.M. White
  `https://arxiv.org/abs/astro-ph/9508025`

- **ENZO: An Adaptive Mesh Refinement Code for Astrophysics**
  The Enzo Collaboration
  `https://arxiv.org/abs/1307.2265v1`

- **HOP: A New Group Finding Algorithm for N-body Simulations**
  Daniel J. Eisenstein and Piet Hut
  `https://arxiv.org/abs/astro-ph/9712200v1`

- **An Introduction to Galaxies and Cosmology**
  Mark H. Jones and Robert J.A. Lambourne

1.6 Appendix: Other useful information

This appendix gives details of how to download the packages used, and very brief instructions on getting started using Enzo and VisIt. These are resources which you would certainly use if you chose to do more research in this area. These do not form part of the lab, and are simply included for your own interest and reference.

1.6.1 Download instructions for packages used

We will use the Python 3 based packages Jupyter (formerly called iPython notebooks) and yt. These are developed in Python under the open-source model, and thus a wealth of user documentation and community forums exist.

To use these packages, you will need Python 3, plus the packages themselves. I highly recommend installing Python using Anaconda. Anaconda conveniently installs Python and other commonly used packages for scientific computing and data science. Instructions can be found here:
`https://www.continuum.io/downloads`
Note that you should install the latest version of Python 3 (currently 3.6) and NOT Python 2[2]. If you do it this way Jupyter will be automatically installed, which you can check by opening a terminal window and typing:

```
jupyter notebook
```

Having installed Anaconda, yt can be installed by simply typing

```
conda install yt
```

in your terminal, as detailed here:
`http://yt-project.org/doc/installing.html#installing-yt-using-anaconda`.

To use yt in parallel, one must also obtain an MPI (Message Passing Interface) package, by typing

[2]There is an idealogical split in the Python community over which is better out of Python 3 and Python 2. For now Python 3 seems to be winning. The main thing is that one must be consistent with what version one uses for all packages.

```
conda install mpi4py
```

in the terminal.

1.7 Using Enzo

Follow the "Boot Camp" instructions on Enzo here:
`http://enzo-project.org/BootCamp.html`
to install and run an example simulation on your desktop. More detailed instructions can be found at:
`https://enzo.readthedocs.io/en/enzo-2.5/`
Use the example Unigrid and AMR dark matter-only cosmology simulation parameter files at:
`https://enzo.readthedocs.io/en/enzo-2.5/tutorials/SampleParameterFiles.html`
to run a dark matter N-body simulation.

1.8 Using VisIt

VisIt is a GUI (Graphical User Interface) based application for visualisation of large data sets, althought it also permits a Python based interface. It is thus an alternative to yt. The advantage of VisIt is that it has a GUI and thus is in some respects more "user friendly". It is particularly well suited to just "seeing what has happened" in the simulation, and moving around the data in 3D.

Details can be found here:
`https://wci.llnl.gov/simulation/computer-codes/visit`
You can download the latest version of Visit to your machine from:
`https://wci.llnl.gov/simulation/computer-codes/visit/executables`
For Mac and Windows there are installers, for Linux you should download the tar file, plus the "Visit Install Script" (in the bullets above the executable) and follow the instructions in "Visit Install Notes". It may take a little while to compile. A user friendly wiki for VisIt can be found at `http://www.visitusers.org/index.php?title=Main_Page`.

Open the VisIt GUI by loading the application. On Linux you can type:

```
./bin/visit
```

in the VisIt directory you have created (or you can add the VisIt bin directory to your path).

Go into the folder containing the data from the first timestep, and look at the data. You can do this by selecting `Open`, navigating to the directory containing the files, and opening the `DD0001.hierarchy` file with `Open file as type -> Enzo`.

One way to look at the particles is to `Add -> Mesh -> Particles`. You can make your output look nicer with the options available under `Controls -> Annotation` and `Plot Attributes -> Mesh`. Note there are different tabs and you need to click `Apply` to apply them to your plot. You can also double click on the plot identifiers under the `Add/Operators/Delete` buttons. You should get something like Fig. 1.3 if you choose a dataset in which collapse has occurred.

Explore other features of VisIt. In particular, try out the Volume plot for the Dark Matter Density, adjusting the Opacity in the 1D transfer function so that you can see the areas of high density. You could also do a Vector plot of the particle velocity, or explore the correlation between quantities using a Scatter plot. Try applying an operator to a plot, such as an Isosurface or Spherical Slice, to see what they do. Options for saving your plots are in the `File` menu.