ML4Science

First Meeting - Week 1

Michele Bianco
Riccardo Brioschi
Gabriele D'Angeli
Federico Di Gennaro

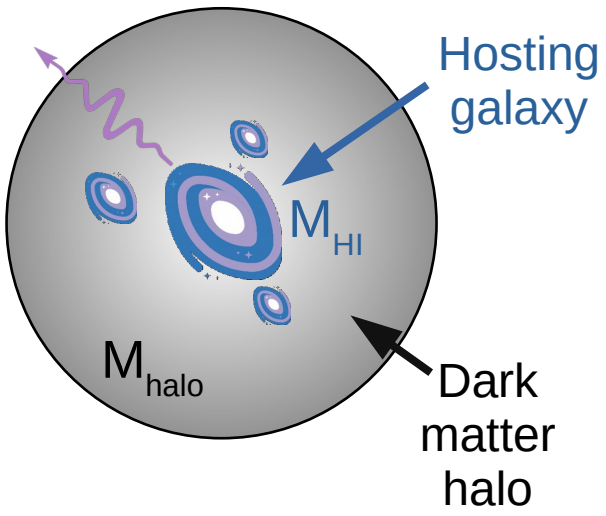
Today Meeting

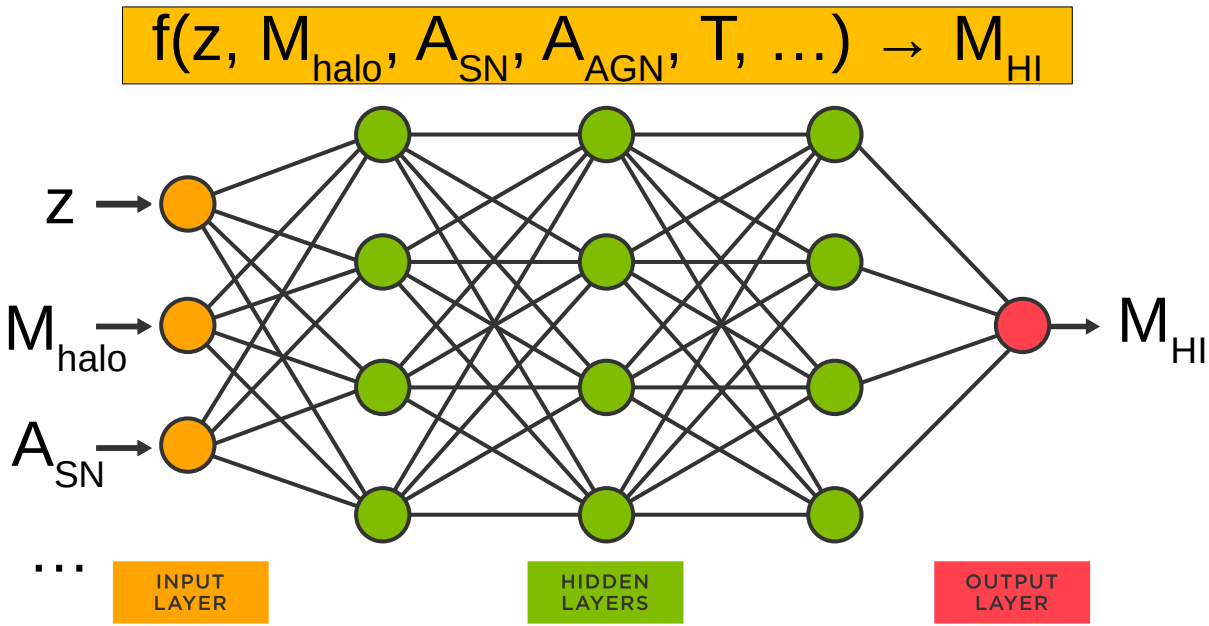
- Quick overview of the project goal
  - Review from previous meeting
  - Review of the proposed network and analysis
- General information on the future meetings
  - Organisation of the meetings
  - Request access to EPFL cluster
- Overview of the data and python test script
  - Walk trough python script to access the data
  - Iris-like analysis to get familiar with the data
- Discussion & hands-on

The ML4Science Project

Use CAMEL dataset to model the HI mass in galaxies:

- K-means clustering algorithm (data analysis)
- Emulate the M_{HI} vs M_{halo} with Fully-Connected Neural Networks

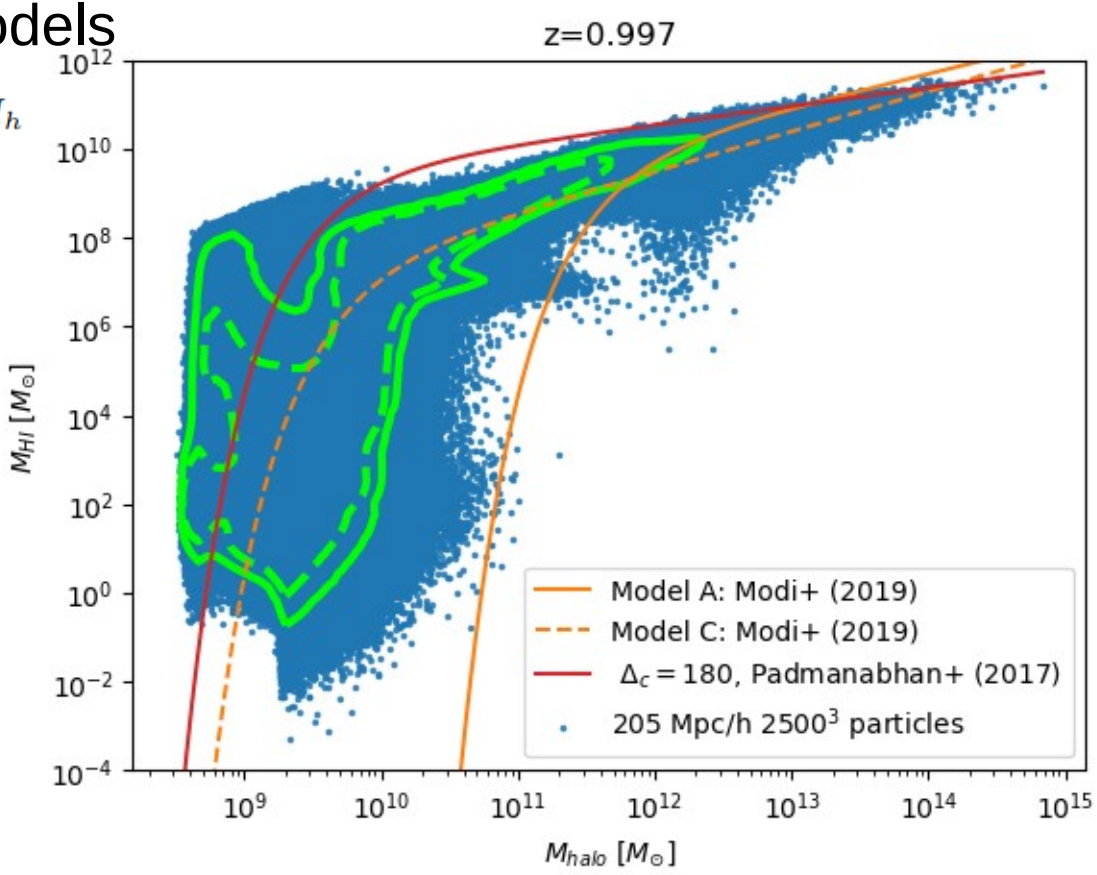




Comparison Simulation vs Models

hydro-dynamic simulations show a complex **M_{HI} vs M_{halo}** relation when compared to standard one-to-one models

$$M_{HI}(M_h; z) = A(z)(M_h/M_{\rm cut})^{\alpha(z)} e^{-M_{\rm cut}(z)/M_h}$$



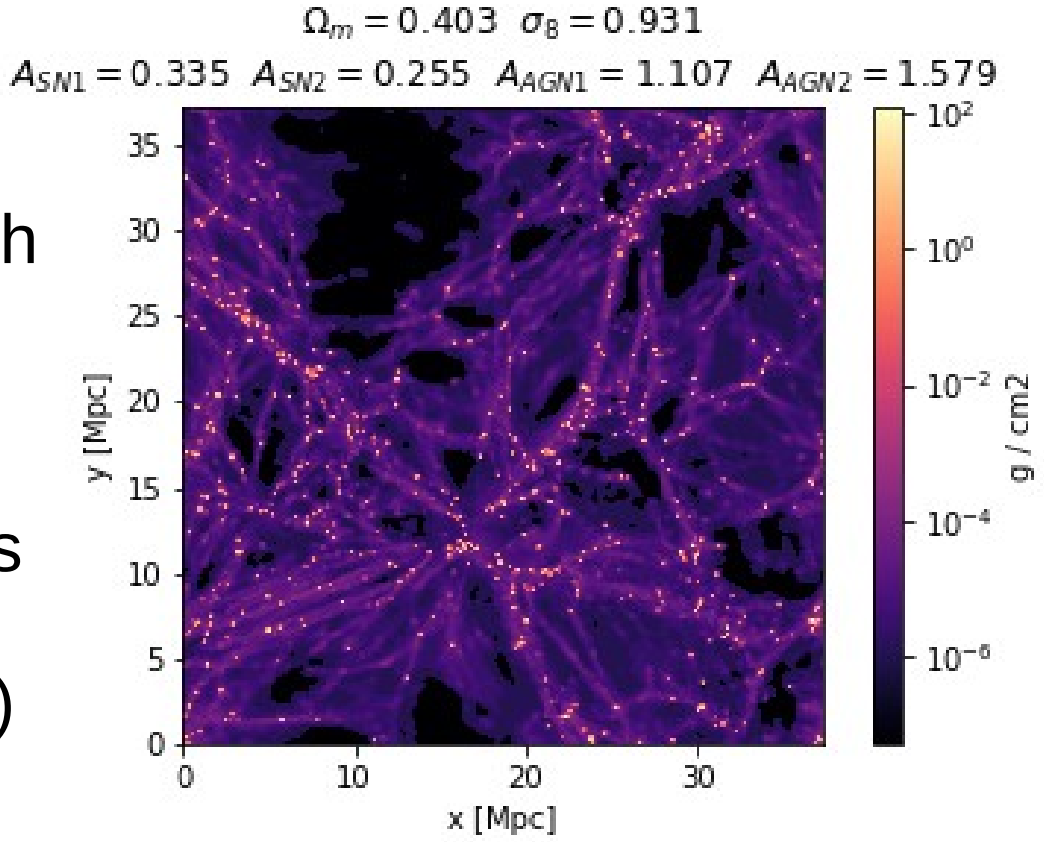
We do not know:

- The relation between M_{HI} and the halo environmental variables (temperature, time, mass, etc…)
- The non-linear relation of neutral hydrogen content with the halo

The Available Dataset

Use CAMEL dataset to model the HI mass in galaxies:

- 1’000 simulations
- Small volume size: 25 Mpc/h
- Different astrophysics and cosmological parameters
- List of Halo and its variables (black hole mass, number of galaxies, halo mass, etc.)



ML4Science Guidelines

Group of 3 students need to: ML4Science guidalines

- Written report: max 4 pages (example ML4Science 2021)
- Code: in Tensorflow / Keras or Pytorch
  - Results reproducibility
  - External libraries citations

The hosting lab

- Grading the domain-specific merit of your contribution
- Provide support to the lab project

Weekly Meeting

The organisation of the future meetings: 6 weeks 1 + 3 + 1 + 1

- Week 1: **(14 Nov)**
  Intro and first data analysis
- Week 2-4: **(21, 28 Nov and 5 Dec)**
  Weekly updates presented by one student
- Week 5: **(12 Dec)**
  Finalise the results start writing the report
- Week 6: **(18 Dec)**
  Finalise the report (plots, text, review report, etc.)

For this Week

- Set up a GitHub page (example ML4Science 2021)
  - Create a folder “Weekly Meetings” and store the student and host supervisor presentations
  - Git push your scripts with instructions
- Request access to EPFL cluster (link)
- Script for reading the dataset
- Start analyse the dataset and K-mean algorithm (example IRIS)

A quick look at what we will do next week:

- a look at Fully-Connected Neural Networks FNN, (tutorial)