
Development and Validation of Physics Modules for the New Reionization Simulation Code DYABLO

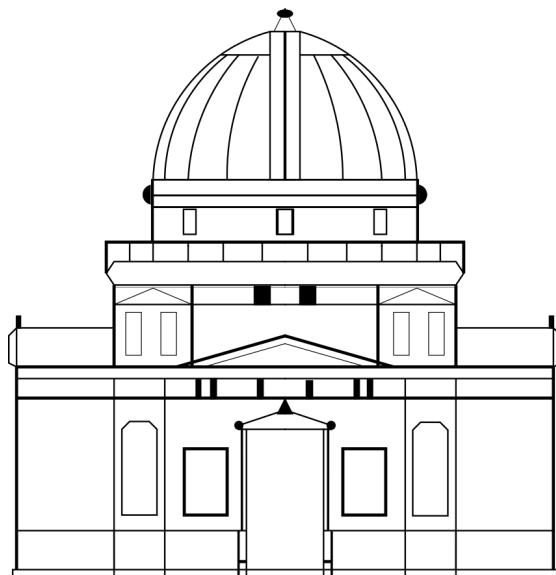
SAMUEL MESQUITA

TÉLÉCOM PHYSIQUE STRASBOURG

GENERAL ENGINEERING DEGREE

PROMOTION 2025

17/02/2025 - 25/07/2025



*Observatoire Astronomique de Strasbourg
11 Rue de l'Université,
67000 Strasbourg, France*

Supervised by :
DOMINIQUE AUBERT
dominique.aubert@astro.unistra.fr

Acknowledgments

First, I want to express my deepest gratitude to Dominique AUBERT for granting me this wonderful internship opportunity. His valuable insights, daily visits, and thoughtful pedagogical approach played a central role in making this internship a truly rewarding educational experience.

I also want to give my thanks Olivier MARCHAL, our talks were very enjoyable and significantly enriched my understanding of the DYABLO code and philosophy.

Special thanks to Maxime GRESSIER, Camille HAVARD, Baptiste MORIN, Nicolas OBRIER and Lucile PETRONIN, my colleagues and friends, for their collaboration and feedback. Being with them during the whole internship was a true pleasure.

Finally, I am especially thankful for my family and their constant support and encouragement throughout this journey.

Acronyms

- **ObAS** Observatoire Astronomique de Strasbourg
- **AMR** Adaptive Mesh Refinement
- **HPC** High Performance Computing
- **Λ CDM** Lambda Cold Dark Matter
- **EoR** Epoch of Reionization
- **HMF** Halo Mass Function
- **CMF** Conditional Mass Function
- **PS** Press–Schechter
- **EPS** Extended Press–Schechter
- **ST** Sheth–Tormen
- **HOP** Hierarchical Overdensity Peak

List of Figures

2.1	Logo of the Astronomical Observatory of Strasbourg	3
2.2	Simplified organizational chart of the ObAS	4
3.1	Original velocity-distance graph from Hubble (1929)	8
3.2	Redshift and scale factor plotted against the age of the Universe	10
3.3	Schematic diagram of the history of the Universe	11
3.4	Ionization bubbles & ionization fraction as a function of redshift	11
3.5	CNN reconstruction of ionization fields from SKA-like noisy 21cm maps . .	12
3.6	Eulerian & Lagrangian descriptions	13
3.7	Validation roadmap diagram	16
3.8	Development roadmap diagram	17
4.1	Position of one hundredth of the particles at the end of the simulation ($z = 0$)	20
4.2	Halos found among the the particles forming them at $z = 0$	21
4.3	Matter power spectrum at $z = 0$ inferred from cosmological probes	22
4.4	Schematic representation of a overdensity trajectory with halo formation .	24
4.5	HMF reproduction from Meriot & Semelin (2024)	26
4.6	CMF reproduction from Meriot & Semelin (2024)	27
4.7	HMF at $z = 0$ from the ST formalism and empirical from the simulation with their 3σ dispersion	28
4.8	Schematic representation of the concept of Lagrangian radius	30
4.9	CMF for different classes of δ_{NL} at $z = 0$	31
4.10	Comparing the empirical and theoretical collapsed mass fraction at $z = 0$.	32
5.1	Plot of the f_{coll} fit found by <i>PySr</i>	36
5.2	Star Formation Rate (SFR) as a function of redshift.	36
5.3	Ionizing fraction as a function of redshift	37
5.4	Snapshots of the evolution density of ionized gas as a function of redshift .	38
6.1	Results at redshift $z \simeq 0.04$ ($\simeq 13.29$ Gyr)	xvi
6.2	Results at redshift $z \simeq 0.14$ ($\simeq 11.99$ Gyr)	xvii
6.3	Results at redshift $z \simeq 0.24$ ($\simeq 10.88$ Gyr)	xviii
6.4	Results at redshift $z \simeq 0.36$ ($\simeq 9.73$ Gyr)	xix
6.5	Results at redshift $z \simeq 0.51$ ($\simeq 8.54$ Gyr)	xx
6.6	Results at redshift $z \simeq 0.66$ ($\simeq 7.54$ Gyr)	xxi
6.7	Results at redshift $z \simeq 0.84$ ($\simeq 6.57$ Gyr)	xxii
6.8	Results at redshift $z \simeq 1.05$ ($\simeq 5.67$ Gyr)	xxiii

ACRONYMS

6.9	Results at redshift $z \simeq 1.30$ ($\simeq 4.81$ Gyr)	xxiv
6.10	Results at redshift $z \simeq 1.66$ ($\simeq 3.91$ Gyr)	xxv
6.11	Results at redshift $z \simeq 2.48$ ($\simeq 2.64$ Gyr)	xxvi

Contents

Acknowledgments	i
Acronyms	iii
List of Figures	vi
1 Introduction	1
2 Host Organization	3
2.1 Overview	3
2.2 Structure of the ObAS	4
2.2.1 Formation department	4
2.2.2 Research departments	5
3 Context & Objectives	7
3.1 Cosmological background	7
3.2 Epoch of Reionization	10
3.2.1 Characterization	10
3.2.2 Observations	12
3.3 Numerical methods	13
3.3.1 Lagrangian & Eulerian descriptions	13
3.3.2 DYABLO	13
3.3.3 Radiative transfer	14
3.4 Internship outline	16
4 Dark matter halos mass functions	19
4.1 Simulation	19
4.1.1 Cosmological and numerical parameters	19
4.1.2 Detecting dark matter halos	20
4.2 Density field and power spectrum	21
4.3 Mass functions	23
4.3.1 Halo Mass Function	23
4.3.2 Conditional Mass Function	25
4.3.3 Implementation of the theoretical HMF & CMF	26
4.3.4 Validation of the abundance of halos	27
4.3.4.1 Halo Mass Function	27

CONTENTS

4.3.4.2	Conditional Mass Function	28
4.3.4.2.a	Segmenting the domain	28
4.3.4.2.b	Lagrangian radius	29
4.3.4.2.c	Results	30
4.4	Collapsed mass fraction	31
5	Improving DYABLO's ionization module	35
5.1	Computing the fraction of collapsed mass	35
5.2	Implementation of the photon source term	37
6	Conclusion	41
	Bibliography	xii
	Appendix	xiii
	Appendix A - Metrics	xiii
	Appendix B - Friedmann's derivation of the density parameters	xiv
	Appendix C - Results on DYABLO's structure formation at any redshift	xv
	Résumé & Abstract	xxviii

Chapter **1**

Introduction

Astrophysical phenomena happen not only over vast spatial scales but also temporal ones from the tiniest particles to entire galaxies and over millions to billions of years. These characteristics make them impossible to study through direct experimentation, observations or purely analytical solutions. Therefore, astrophysics increasingly relies on numerical simulations since they enable astrophysicists to test theories, visualize complex processes as well as explore new ideas without the need for expensive experimentation.

DYABLO ([Aubert & Durocher, 2021](#)) is a new high-performance hydrodynamics simulation code for modeling astrophysical fluids on adaptive mesh refinement (AMR) grids. It is designed for exascale supercomputers (10^{18} floating point operations per second) thanks to its use of massively parallel hardware and hardware independent framework which allows it to efficiently function across thousands of processing units, including GPUs. The code solves equations of fluid dynamics, gravity and radiation, enabling accurate simulations of a wide range of astrophysical phenomena from the formation of large-scale structures to the internal dynamics of stars.

While the foundations of DYABLO have already been laid out and allow it to produce satisfying results, some of the physics modules still lack accuracy. This is the case with the photon source formation module which not only needs to be validated but also improved to account for phenomena that occur on spatial scales smaller than a simulation grid cell.

Indeed, during the Epoch of Reionization (EoR), small galaxies, which played a major role in ionizing the intergalactic medium, are often poorly resolved or totally absent in cosmological simulations due to the limited spatial resolution. Therefore, the contribution of unresolved sources must be approximated which results in uncertainties into the timing and topology of the EoR. Thus, improving the current source model in DYABLO is critical to obtain a physically reliable reionization process which is the aim of my internship.

The first chapter of this report will provide an in-depth presentation of the Astronomical Observatory of Strasbourg. The following chapter will give the context and key elements required to fully understand the rest of the report. The subsequent chapter dives into the methods and results of the validation of the structure formation module of DYABLO. Then, the method to implement a physically motivated model to the ionization module of DYABLO will be explained. The report ends with an overall conclusion.

Chapter 2

Host Organization

Contents

2.1	Overview	3
2.2	Structure of the ObAS	4
2.2.1	Formation department	4
2.2.2	Research departments	5

2.1 Overview

The Astronomical Observatory of Strasbourg (ObAS) is a french research center dedicated to the astrophysical sciences. It is part of the University of Strasbourg and hosts the Strasbourg Astronomical Data Center (CDS). Built in 1881 while Strasbourg was a german city, the observatory consists of three buildings : the large dome and two others, one of which was originally used as the director's residence.

It is one of the thirty-five faculties, schools and institutes of the University of Strasbourg. Moreover, like any other institute that deals with the Earth and Universe Sciences in France, it is co-managed by the CNRS (National Center for Scientific Research) that supports the ObAS by providing funding, resources and research infrastructure.

According to the report from [Hcéres \(2023\)](#) which evaluates higher education and public research in France, there is ~ 55 permanent workers whether it is astrophysicists, engineers, technicians or support staff members while a smaller portion (approximately ~ 35) consists of interns, PhD students and postdoctoral researchers.

Over a period of six years, the total budget received by the ObAS is 11,86 million euros. This amount is divided as follows : 21.9% from regular funding (excluding salaries) provided by supervising institutions, 2.8% from regional project calls, 26.2% from national project calls, 39.4% from international project calls and 9.7% from contracts, patents,

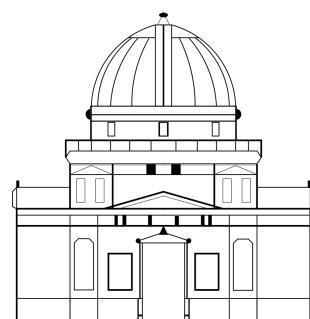


Figure 2.1: ObAS logo

2.2. STRUCTURE OF THE OBAS

services or partnerships with industry.

2.2 Structure of the ObAS

The observatory's current director is Pierre-Alain Duc who orchestrates strategic planning, operational management, resource allocation and communication among the various teams that make up the ObAS.

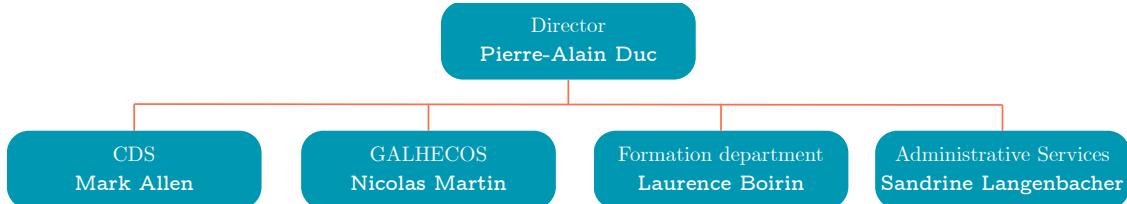


Figure 2.2: Simplified organizational chart of the ObAS

The Administrative and Technical Services department make sure the day-to-day operations of the institution are running without any problems. They also manage logistics, information systems and infrastructure to ensure smooth and efficient support for research activities.

2.2.1 Formation department

The formation department not only aims to educate students to astrophysics but also to keep researchers up to date with the latest advances in astronomy. Multiple components make that department :

- Education of masters student : The ObAS offers a Master 2 named Astrophysics & Data Science in partnership with the University of Strasbourg. Pierre Maggi is at the head of this program and is responsible for coordinating the curriculum and making sure that the classes aligns with the needs from research and industry.
- Link with the doctoral school : The doctoral school of Strasbourg (ED18) organizes the training of PhD students and provide ethical and scientific support throughout the doctoral journey. Rodrigo Ibata, a member of the ObAS, is part of the council of the ED18.
- Internship : Every year, the ObAS offers multiple internships for physics and computer science post-baccalaureate students. Those internships help students gain experience in research environments. Christian Boily and Andre Schaaff are responsible of internships offered at the ObAS.
- Seminars : In order to keep researchers up to date with the latest advances in astronomy or to introduce them to new research topics, a weekly seminar is held

featuring a guest speaker (often an external researcher) who shows their recent work and engages in discussions. Benoit Famaey and Giacomo Monari are supervising the schedule and talks.

2.2.2 Research departments

In 1972, the ObAS founded its first team : the astronomical Data Center of Strasbourg (CDS) whose goal is to collect, improve and share astronomical data worldwide. Since 2012, the CDS has been recognized as a national research infrastructure by the French Ministry of Higher Education, Research, and Innovation (MESRI).

Nicolas Martin is the current director of this team. The CDS runs multiple services like :

- Simbad (<https://simbad.cds.unistra.fr>) : a database used to identify and study astronomical objects. Its goal is to give detailed information through its records of object, a bibliography of related publications and a name resolver to match object names.
- VizieR (<https://vizier.cds.unistra.fr/>) : a service to explore astronomical catalogues. Its goal is to make published astronomical data easily searchable and usable for research.
- Aladin (<https://aladin.cds.unistra.fr>) : an interactive sky atlas used to visualize and analyze astronomical images and data. It has various surveys and allows multi-resolution sky mapping. Its goal is to help astronomers explore the sky, overlay data and study objects in their astronomical context.
- X-Match (<http://cdsxmatch.u-strasbg.fr>) : a service that helps astronomers quickly find and match objects between very large catalogues or compare a list of their own positions to a catalogue.

For the careful development and maintenance of these services, there are ~ 30 workers including astronomers, computer engineers and documentalists. Mark Allen is the current director of this team.

This work takes place in the other team of the observatory named GALHECOS (Galaxies, High Energy, Cosmology, Compact Objects & Stars) which studies the formation and evolution of galaxies in a cosmological context. The team plays a role in the SSC-XMM consortium who is responsible for delivering XMM-Newton source catalogs to the international community. It is also part of multiple national (ANR) and European projects aimed at understanding dark matter, the evolution and structure of galaxies, pulsar physics, etc.

During this internship I was able to use two computers that were bought thanks to the ANR ExaSKA project involving Pierre Ocvirk, Olivier Marchal and Dominique Aubert :

2.2. STRUCTURE OF THE OBAS

- The first one is Apu. It has 64 GB of RAM and an 8-core processor with 1 thread per core meaning 8 threads total. I used it to do some post-processing of a simulation that I was given.
- The second one, and much more powerful, is Flanders. It has 1008 GB of DDR5 RAM and 48-core processor with 2 threads per core meaning 96 threads total. The GPU is an NVIDIA H100 NVL GPU which is designed for large-scale AI and High Performance Computing (HPC) workloads. It has costed ~ 20 k€ while the whole computer costed roughly twice that amount. Fully compiling and running a simulation took ~ 1 hour on Flanders while it takes ~ 10 hours to do on Apu. I used it to run demanding simulations and post-process them on the spot.

Flanders is only being used by people who work on DYABLO, for researchers at the ObAS, there is a more powerful computer that is being shared by more users.

Chapter **3**

Context & Objectives : toward a model for unresolved sources in cosmological simulations

Contents

3.1	Cosmological background	7
3.2	Epoch of Reionization	10
3.2.1	Characterization	10
3.2.2	Observations	12
3.3	Numerical methods	13
3.3.1	Lagrangian & Eulerian descriptions	13
3.3.2	DYABLO	13
3.3.3	Radiative transfer	14
3.4	Internship outline	16

3.1 Cosmological background

This internship focuses on the study of the *Epoch of Reionization* using high performance numerical simulations. More precisely, we would like to implement a source model in DYABLO in order to accurately simulate the generation and propagation of photons created by the first galaxies throughout the grid cells regardless of the simulation's spatial resolution.

A few notions of cosmology and an overview of the cosmological context are necessary. Lemaître (1927) and then later Hubble (1929) have found a relationship between the distance of galaxies and their radial velocity. When individual stars could be resolved in

3.1. COSMOLOGICAL BACKGROUND

images, astronomers used Cepheid variables to estimate distances. Cepheids have a known average intrinsic brightness that can be determined from the period of their brightness variations. By comparing this to the observed brightness, the distance can be calculated¹. Using that, astronomers were able to accurately determine distances which could then be compared with the galaxy's radial velocity (measured from Doppler shifts of spectral lines) to establish the velocity-distance relation.

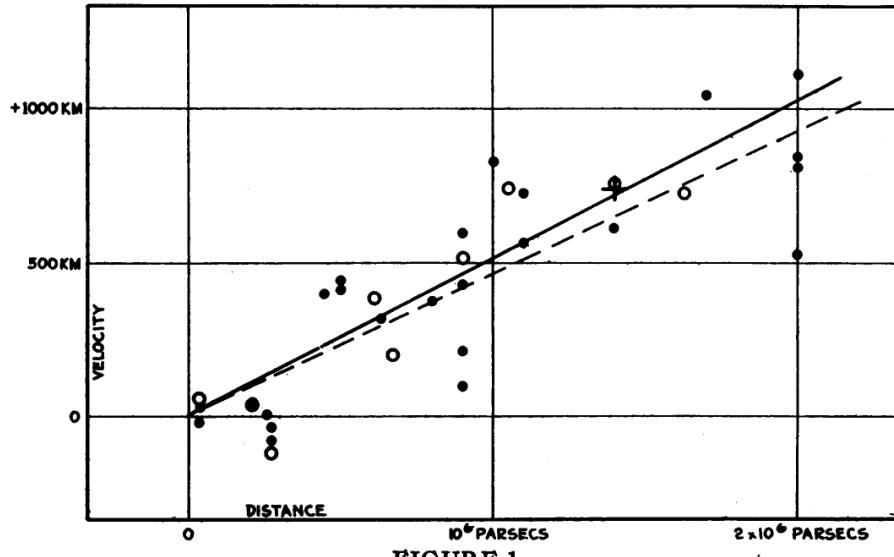


FIGURE 1
Velocity-Distance Relation among Extra-Galactic Nebulae.

Figure 3.1: Original velocity-distance graph by [Hubble \(1929\)](#)

Lemaître and Hubble (Figure 3.1) realized that galaxies are receding from us with a radial velocity proportional to their distance. This lead to the Hubble-Lemaître's law :

$$v = H_0 d, \quad (3.1)$$

with v and d the aforementioned radial velocity and distance respectively. H_0 is called the Hubble constant, it has a value of $\simeq 70$ km/s/Mpc. For example, a galaxy that is 10 Mpc away would recede from us at a velocity of approximately 700 km/s.

Galaxies receding from us was the first major observational fact that lead to the argument of an expanding Universe. It has changed the way the Universe's structure, evolution, origin is studied and ultimately lead to the development of cosmology as a science.

This expansion is not due to galaxies moving through space but rather space itself that is stretching over time. To describe it mathematically, cosmologists introduced the *scale factor* $a(t)$ that characterizes how distances in the Universe expand over time. This quantity is dimensionless and normalized such that $a(t = 0) = a(t_0) = 1$ today.

¹Once the intrinsic brightness is known, the distance d can be calculated using $F = L/(4\pi d^2)$ where F is the observed flux (apparent brightness) and L the intrinsic brightness.

The physical distance² $d(t)$ between two comoving points³ grows proportionally to $a(t)$ according to $d(t) = a(t)d_0$ where $d_0 = d(t_0)$ is called the comoving distance which is constant and is equal to the physical distance at the present time.

The *Hubble parameter* $H(t)$ (which generalizes H_0) is defined as :

$$H(t) = \frac{\dot{a}(t)}{a(t)}, \quad (3.2)$$

with $\dot{a}(t)$ the derivative of the scale factor. The current Hubble constant is simply the value of this function today $H_0 = H(t_0)$. It indicates how fast the Universe is expanding at a given epoch t .

One of the most important notion in cosmology is the *redshift* (noted z), it quantifies how much the observed wavelength (λ_{obs}) of an object is 'shifted to the red' compared to the wavelength of the same object emitted at rest (λ_{emit}).

$$z = \frac{\lambda_{obs} - \lambda_{emit}}{\lambda_{emit}}. \quad (3.3)$$

While this might look like a Doppler effect (with $z = v/c$) since the galaxies are receding from each other because of the expansion of the Universe, it would not be the right interpretation. Indeed, the distortion is not created by the motion of the source or the observer but rather by the *stretching of space itself as light travels through it*. In fact, it can also be expressed as $1 + z(t) = 1/a(t)$. This expansion causes the wavelength of light to stretch when propagating, making astrophysical objects appear redder than they are. Redshift measures how far is an object and how far in the past its light was emitted. Thus, the redshift and the scale factor have effectively replaced time as the variable to study the evolution of the Universe.

Figure 3.2 shows the relation between the redshift (z), the scale factor ($a(t)$) and time for given *cosmological parameters*. Here, they follow the Λ CDM cosmology. The latter is most successful cosmological model at this date due to its ability to explain a vast number of cosmological observations (e.g. the Cosmic Microwave Background ; see [Planck Collaboration et al. \(2020\)](#) & [Spergel et al. \(2007\)](#)) with a relatively simple set of six key parameters which are presented in Table 3.1.

H_0 km/s/Mpc	h	Ω_m	Ω_b	Ω_Λ	σ_8	n_s
67.0	0.67	0.3175	0.049	0.6825	0.8344	0.96

Table 3.1: Λ CDM cosmological parameters (used in this work)

In Table 3.1, $h = H_0/(100 \text{ km/s/Mpc})$ is the (dimensionless) reduced Hubble constant and used to account for the uncertainty on H_0 precise value. n_s is the scalar spectrum

²also known as *proper distance*.

³comoving points are locations in space that only move relatively to each other due to expansion.

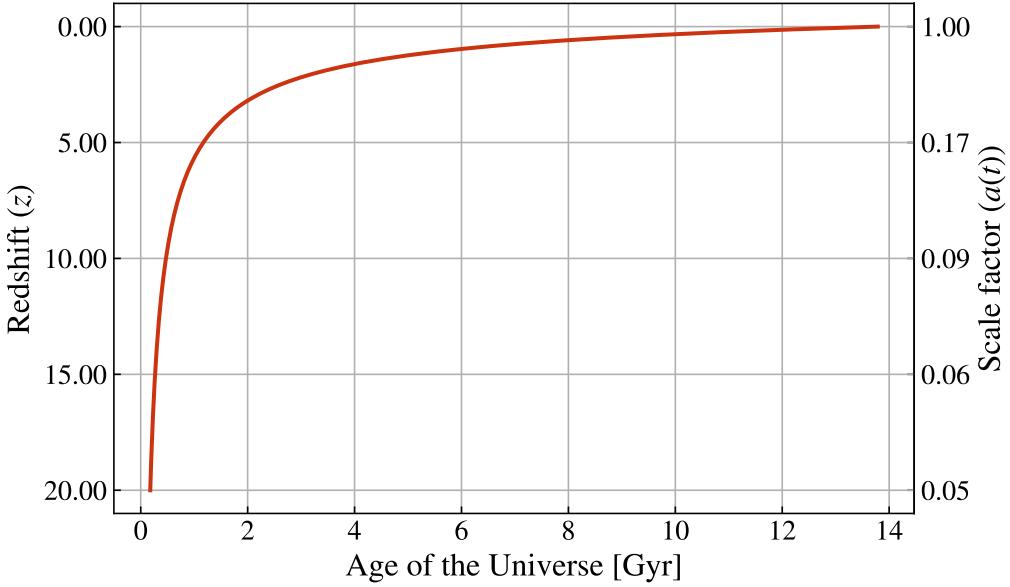


Figure 3.2: Redshift and scale factor plotted against the age of the Universe in a Λ CDM cosmology.

power-law index, σ_8 corresponds to the amplitude of matter density fluctuations in the Universe today on a scale of $8 h^{-1}\text{Mpc}$ (further details of the amplitude of matter density fluctuations in section 4.2) while the *density parameters*⁴ $\Omega_{m,b,\Lambda}$ describe the ratio of the actual density of a component (like matter (m), baryons (b) or dark energy (Λ)) to a critical density.

A Universe that follows this cosmology has baryons, radiation, cold dark matter (CDM) and an unknown energy called Λ . This Universe also follows the Cosmological Principle which states that : '*On sufficiently large scales, the Universe is spatially homogeneous and isotropic.*' Homogeneous meaning in this context that there is the same density of mass-energy everywhere at large scales and isotropic meaning that the properties are the same in every direction. The Cosmological principle ensures that observations made from a single place represents the whole Universe and can be used to challenge cosmological models. If this principle is someday proven to be unreliable then cosmology would 'die' as a science since observations and measurements would not apply to the entire Universe.

3.2 Epoch of Reionization

3.2.1 Characterization

The *Epoch of Reionization* (hereafter EoR) is characterized by the transformation of cosmic hydrogen from a mostly neutral state to a mostly ionized one thanks to the emergence of the first luminous sources. It occurred roughly within the range $6 \lesssim z \lesssim 15$ (Planck Collaboration et al., 2020).

⁴More information on their derivation in Appendix (A) and (B).

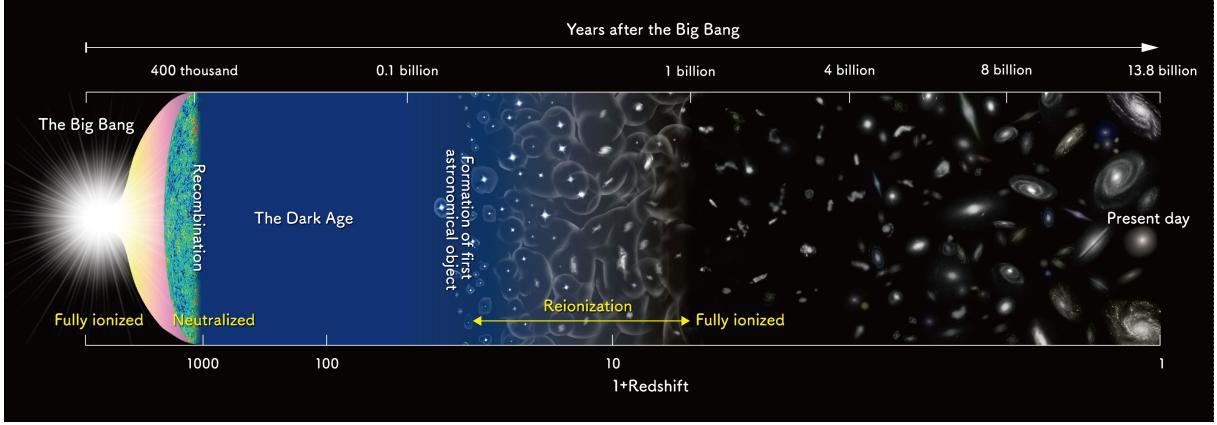


Figure 3.3: Schematic diagram of the history of the Universe. Credit : NOAJ

The Figure 3.3 shows the history of the Universe : the first few hundred thousand years after the Big Bang ends with the *Recombination Epoch* ($z \simeq 1100$) when electrons and protons combined to form neutral hydrogen and the cosmic microwave background (CMB) was released. Then arrived the *Dark Ages* when the Universe was neutral and without any luminous source. This epoch ended at $z \simeq 30$ when matter collapsed under gravity to form dark matter halos, the first stars and galaxies starting the *Cosmic Dawn*.

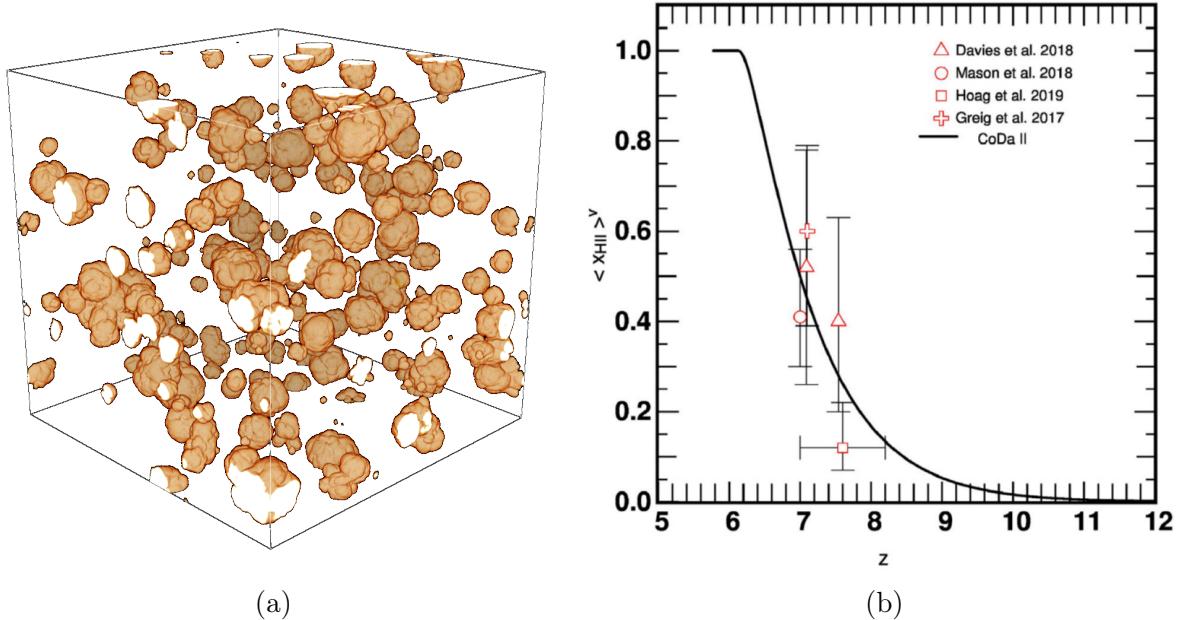


Figure 3.4: (a) Ionization bubbles in the IGM from a DYABLO simulation (b) Ionization fraction as a function of redshift from [Ocvirk et al. \(2020\)](#)

Indeed, those first objects were surrounded by neutral hydrogen in the intergalactic medium (IGM) during their formation. Then, because of their ultraviolet (ionizing) radiation, they started to form ionized bubbles (Furlanetto et al., 2004) in the IGM (Figure 3.4a). Those grew and merged over time until most of the hydrogen in the Universe was no longer neutral. This transition characterizes the EoR ; it can be seen on the plot representing the evolution of the ionization fraction as a function of redshift (Figure 3.4b)

3.2. EPOCH OF REIONIZATION

which can be constrained by observational data (Davies et al. (2018), Mason et al. (2018)), Hoag et al. (2019) & Greig et al. (2017)). The ionization fraction is given by :

$$x = \frac{\rho_{\text{ionized gas}}}{\rho_{\text{gas}}} = \frac{\rho_{HII}}{\rho_{\text{gas}}}. \quad (3.4)$$

The exact timing and length of the EoR are still being studied. Understanding them is critical to build models like the ionization modules discussed later in this work.

A key point about the EoR is that only galaxies with stellar masses above roughly $10^8 M_\odot$ were able to participate in reionization. This is because only in those galaxies the gas could cool down enough to form stars. In less massive dark matter halos, gravity is too weak to increase the temperature at levels that would trigger atomic cooling, meaning the gas stays hot and cannot collapse to form stars. Consequently, these low mass galaxies did not produce enough ionizing radiation to have a major impact on the reionization of the Universe.

3.2.2 Observations

Some telescopes like the Square Kilometer Array (SKA), a radio interferometer currently under construction in Australia and South Africa, will be a useful tool to probe the EoR and the Cosmic Dawn. It involves observing the 21cm hyperfine transition of neutral hydrogen (see Figure 3.5). This transition happens when the electron in the neutral hydrogen atom flips its spin to be reversed to the spin of the proton.

The photon released during this spin-flip transition has a very specific energy which corresponds to a frequency of $\simeq 1420$ MHz which is equivalent to a wavelength of approximately 21 cm. It can be observed using instruments studying the $\simeq 50 - 200$ MHz range due to the redshift $z \in [6, 30]$ of the epochs of interest.

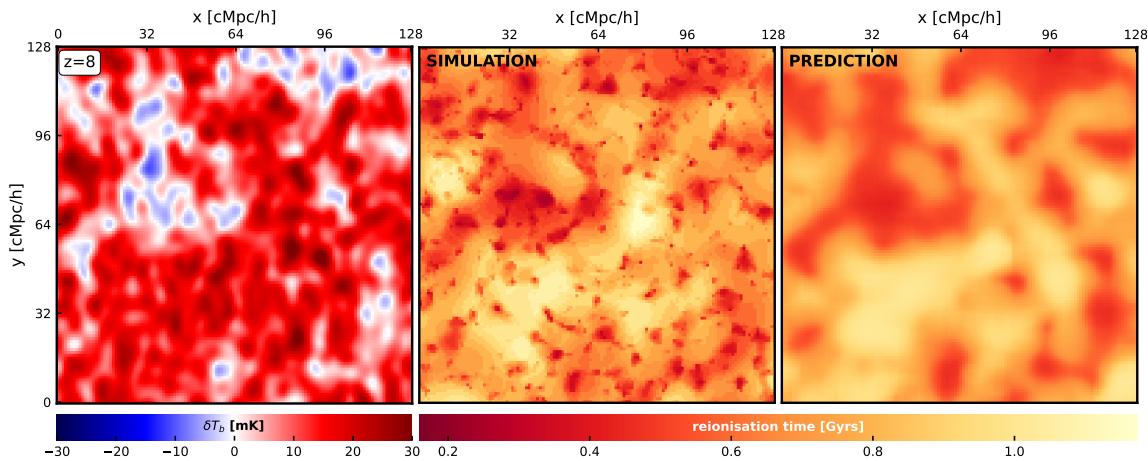


Figure 3.5: Convolutional neural network reconstruction of ionization fields from SKA-like noisy 21 cm maps at $z = 8$, showing input (left), ground truth (middle) and prediction (right) (Hiegel et al., 2023).

While this transition is rare and the resulting signal is faint due to foreground emissions being much brighter, the SKA will be largely able to observe it due to its unprecedented

sensitivity and the overwhelming abundance of the neutral hydrogen at those epochs.

In the context of the EoR, the 21cm signal probes the neutral hydrogen distribution which allows us to observe the contrast between ionized and neutral regions (notably the gaps between the ionizing bubbles) as reionization progresses.

3.3 Numerical methods

3.3.1 Lagrangian & Eulerian descriptions

In cosmological simulations, the Eulerian and Lagrangian descriptions represent two different ways of tracking the motion of mass elements (dark matter, gas, etc.) over time. In both of these methods, one needs to solve for gravity and hydrodynamics ; typically using Newton's and Poisson's equation for gravity and the Euler's fluid equations for hydrodynamics to accurately model the interactions and evolution of cosmic structures.

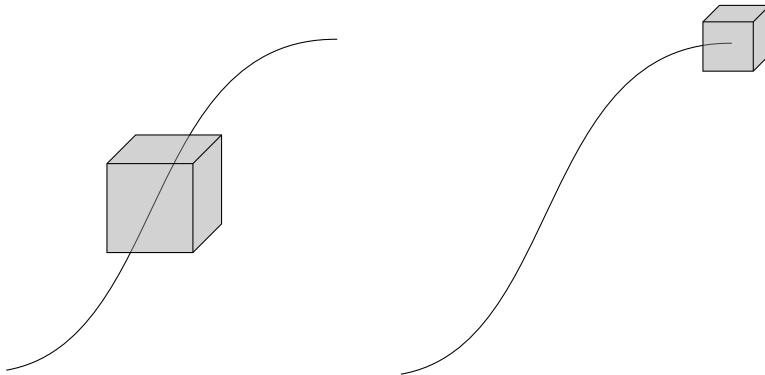


Figure 3.6: Eulerian (left) & Lagrangian (right) descriptions. The cube represents a simulation cell while the black line represents the path of an element of the simulation.

In Lagrangian descriptions, the simulation follows individual fluid elements or particles as they move through space and time. Each particle carry information like density, velocity, and pressure with them. There is no fixed grid. The method focuses on areas where matter is dense and solves the fluid and gravity equations by tracking how these particles move and interact with each other.

In Eulerian descriptions, space is divided in cells and the fluid's properties (density, velocity, pressure...) are tracked in this fixed grid (see Figure 3.6). The simulation computes how these properties change over time at each cell by solving the fluid equations and gravity on this static framework.

3.3.2 DYABLO

DYABLO ([Aubert & Durocher \(2021\)](#), [Delorme et al. \(2022\)](#)) is a new high-performance hydrodynamics simulation code developed at CEA since 2019 for modeling astrophysical fluids on Adaptive Mesh Refinement (AMR) grids. AMR ([Berger & Oliger \(1984\)](#), [Berger & Colella \(1989\)](#)) is a method used in Eulerian simulations to make calculations more

3.3. NUMERICAL METHODS

efficient. Instead of using the same size grid cells everywhere, AMR makes the grid thinner in important areas like regions with high density or sharp changes. This means the simulation can focus on details where they are significant without using too much computing power on less important regions.

The Strasbourg team contributes to the development of DYABLO by adding cosmological modules. The code is designed for exascale supercomputers (10^{18} floating point operations per second) thanks to its use of massively parallel hardware and hardware-agnostic framework. It uses the Kokkos library (Trott et al., 2022) which allows it to not only run on any existing GPU but also the future ones that are not yet designed or released. It also uses MPI for parallelism which makes it highly compatible with heterogeneous architectures including both CPU and GPU supercomputers.

The code solves equations of fluid dynamics, gravity and radiation which enables accurate simulations of a wide range of astrophysical phenomena. Its main goal is to simulate key processes in cosmology, notably to model the formation of the first cosmic structures and how their radiation reionized the intergalactic medium which provides insights into the early stages of structure formation in the Universe.

The gravitational field is computed by solving the Poisson equation :

$$\nabla^2 \Phi = 4\pi G\rho, \quad (3.5)$$

where Φ is the gravitational potential, G is the gravitational constant and ρ is the mass density.

The dynamics of the fluid are governed by the Euler equations in conservative form :

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0, \quad (3.6)$$

$$\frac{\partial(\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \otimes \vec{v}) = -\nabla p + \rho \vec{g}, \quad (3.7)$$

$$\frac{\partial(\rho E)}{\partial t} + \nabla \cdot (\rho E \vec{v}) = -\nabla \cdot (p \vec{v}) + \rho \vec{g} \cdot \vec{v}, \quad (3.8)$$

where ρ is the matter density which includes the contribution of gas (described in cells, Eulerian) and dark matter (described with particles, Lagrangian), \vec{v} is the velocity, P is the pressure, E is the total energy density (internal + kinetic) and Φ is the gravitational potential.

Radiation transport and cosmological expansion are included through additional modules. The next section focuses on the radiative transfer implementation in DYABLO.

3.3.3 Radiative transfer

In cosmological simulations, modeling the propagation of photons is critical for accurately understanding processes like the reionization. *Radiative transfer* is the name given to the description of how radiation is emitted, absorbed and scattered by matter when traveling in the domain.

DYABLO uses the *M1 radiative transfer scheme* ([Aubert & Teyssier, 2008](#)) which is a moment-based method which solves the radiative transfer equation :

$$\frac{1}{c} \frac{\partial I_\nu}{\partial t} + \vec{n} \cdot \nabla I_\nu = -\kappa_\nu I_\nu + \eta_\nu, \quad (3.9)$$

with I_ν the specific intensity, \vec{n} the direction of propagation, κ_ν the absorption coefficient, η_ν the emission coefficient and c the speed of light.

Instead of tracking individual photon rays, M1 studies the 3D fields of radiation energy $E_\nu(x, y, z)$ and radiation flux $\vec{F}_\nu(x, y, z)$. Two coupled equations involving those terms can be obtained by taking the first two moments of Equation 3.9 :

$$\frac{\partial E_\nu}{\partial t} + \nabla \vec{F}_\nu = -\kappa_\nu c E_\nu + S_\nu. \quad (3.10)$$

$$\frac{\partial \vec{F}_\nu}{\partial t} + c^2 \nabla \vec{P}_\nu = -\kappa_\nu c \vec{F}_\nu. \quad (3.11)$$

The first terms in the right hand sides of those equations are absorption terms while S_ν is the source term which is expressed as $S_\nu = \Delta N + \Delta N_{rec}$. Both of those term are photon number density, the first one represents the radiation coming from light sources such as stars or quasars while the second represents the radiation produced by the recombination of ionized hydrogen (HII). *The goal of this internship will be to have a physically motivated and numerically correct $\Delta N(x, y, z)$ so that it can be implemented in DYABLO.*

In Equation 3.11, \vec{P}_ν represents the pressure tensor (second order moment of the intensity). To close the system, an expression for \vec{P}_ν must be given. [Aubert & Teyssier](#) use the local closure $\vec{P}_\nu = \vec{D} E$ based on the Eddington tensor \vec{D} , which is constructed from E and \vec{F}_ν . This leads to a set of equations for E_ν and \vec{F}_ν which are solved using [Godunov \(1959\)](#) methods which follows :

$$\frac{\partial \vec{U}}{\partial t} + \frac{\partial \vec{F}}{\partial x} = \vec{0} \implies \frac{\vec{U}_i^{p+1} - \vec{U}_i^p}{\Delta t} + \frac{\vec{F}_{i+1/2}^p - \vec{F}_{i-1/2}^p}{\Delta x} = \vec{0}. \quad (3.12)$$

The p index represents the current time step while the i index represents the spatial index. In our case, we have $\vec{U}(x, y, z) = (E_\nu(x, y, z), \vec{F}_\nu(x, y, z))^T$ and $\vec{F}(x, y, z) = (\vec{F}_\nu(x, y, z), c^2 \vec{P}_\nu(x, y, z))^T$. To solve Equation 3.12, one needs to find the intercell value for $\vec{F}_{i \pm 1/2}^p$ which is called Lax-Friedrich numerical flux using the Lax-Friedrichs scheme ([Lax \(1954\)](#), [González et al. \(2007\)](#)) as :

$$\vec{F}_{i+1/2}^p = \frac{\vec{F}_i^p + \vec{F}_{i+1}^p}{2} - \frac{c}{2}(\vec{U}_{i+1}^p - \vec{U}_i^p). \quad (3.13)$$

One might notice that Equation 3.12 has a null right hand side which is not the case of Equation 3.10 & 3.11. This is because we apply the Godunov method on the homogeneous form of the radiative transfer system that focuses only on the transport part of the equation while Equation 3.10 & 3.11 also include the thermochemistry, namely the absorption and source terms. To handle both transport and thermochemistry, the scheme is split

3.4. INTERNSHIP OUTLINE

into two steps : a transport step which solves the homogeneous system (Equation 3.12) and a source step which applies absorption and source terms from Equation 3.10 and 3.11.

When photons sources such as stars are formed in DYABLO, the M1 scheme allows photons to propagate by updating the radiation field while taking into account photoionization, heating and radiation pressure. The strength of the M1 method lies in its computational efficiency and its local nature which enables it to scale well to large simulations.

3.4 Internship outline

Throughout the following chapter, we will derive and use the necessary tools to show that DYABLO's structure formation is consistent with theoretical frameworks.

Dark matter halos are the building blocks in which baryonic structures like galaxies and clusters form. They come from the gravitational collapse from the very first overdensities in the early Universe. Those small overdensities will grow under gravity until a point where they will self-collapse and form a dark matter halo.

The mass function describes the number density of dark matter halos as a function of their mass and is a key prediction of structure formation theory. In this work, we focus on two types of mass functions : the Halo Mass Function (HMF) and the Conditional Mass Function (CMF). The HMF gives the number density of halos per unit mass at a given redshift regardless of their environment. It provides a global statistical view of halo abundance in the Universe. In contrast, the CMF gives the number of halos within a specific, smaller region of the Universe conditioned by the size and density of the said region. We will first make sure that we can correctly code the theory of mass functions from scratch to be able to compare those results to a simulation output from DYABLO. After that, we will focus on the collapsed mass fraction which gives the proportion of mass that has collapsed into bound structures in a given region : $f_{coll} = M_{coll}/M_{region}$. Validating that DYABLO has the correct collapsed mass fraction will be the last step of this structure formation validation process. Figure 3.7 shows the steps that will be achieved to reach that result through the following chapter :

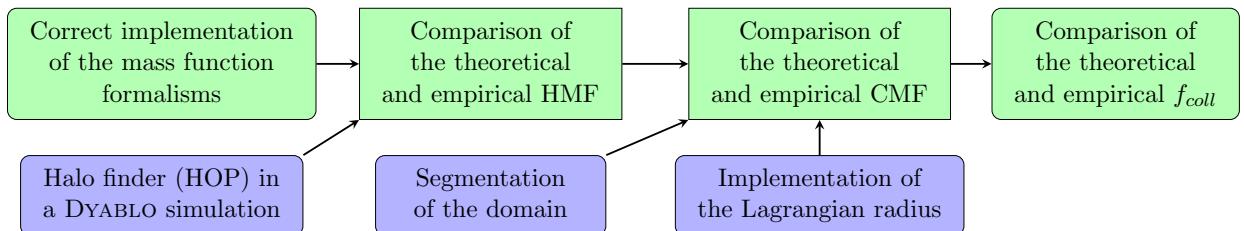


Figure 3.7: Validation roadmap diagram. Green nodes are validation steps while blue nodes are numerical techniques. All of the above will be described in their own section.

Our final goal will be to implement a physically motivated number density of photons ΔN emitted in each cell of the domain depending on the cell's properties. To do that,

we will use the previous result that gives the fraction of mass collapsed in a cell f_{coll} to compute df_* , the fraction of the gas particle converted into stellar mass in time step Δt of the simulation using (Meriot & Semelin, 2024) :

$$df_* = (f_{coll} - f_*) \frac{\Delta t}{\tau_{SF}}, \quad (3.14)$$

with f_* the fraction of gas particle already in stars, Δt the simulation time step and τ_{SF} the characteristic timescale for star formation set to 70 Gyr.

From that, one may find the number density of photons emitted

$$\Delta N = \rho_r \times df_* \times \dot{N} \times \Delta t, \quad (3.15)$$

with ρ_r the matter density in the region and \dot{N} the stellar emissivity constant.

Figure 3.8 shows the different possible steps to go from the information on a region which are the redshift z , the radius of the region R_0 and its intrinsic overdensity δ_{NL} to the computation of the variation in photon density ΔN , which is what we want to achieve. The most efficient path of the three will be explained in chapter 5.

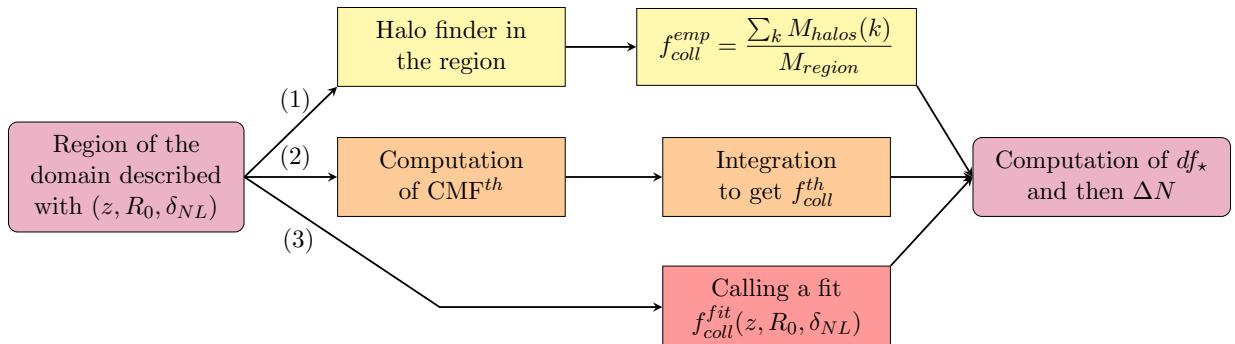


Figure 3.8: Development roadmap diagram

Chapter **4**

Dark matter halos mass functions

Contents

4.1	Simulation	19
4.1.1	Cosmological and numerical parameters	19
4.1.2	Detecting dark matter halos	20
4.2	Density field and power spectrum	21
4.3	Mass functions	23
4.3.1	Halo Mass Function	23
4.3.2	Conditional Mass Function	25
4.3.3	Implementation of the theoretical HMF & CMF	26
4.3.4	Validation of the abundance of halos	27
4.4	Collapsed mass fraction	31

4.1 Simulation

4.1.1 Cosmological and numerical parameters

In order to validate a module from a simulation, one would need a theory in which they can trust and outputs from a simulation, in our case, DYABLO.

Table 4.1 shows the numerical parameters that were used for the simulation. The cosmological parameters are the one from the Λ CDM shown in Table 3.1. They were used to produce initial conditions at large redshift. These initial conditions were given to me by D.Aubert and produced by GRAFIC ([Bertschinger, 2011](#)).

4.1. SIMULATION

L	N_{cells}	N_{part}	Δt
$h^{-1}\text{Mpc}$			Myr
164	256^3	256^3	[15, 60]

Table 4.1: Numerical parameters used in the simulation

Also, since the number of particle is known and given the cosmological parameters (see Table 3.1) we have $\bar{\rho} = 3\Omega_m H_0^2/8\pi G = M_{tot}/V_{tot}$, we may deduce the mass of one particle of the simulation, $M_{part} = M_{tot}/N_{part} = \bar{\rho} V_{tot}/N_{part} = 3.44 \times 10^{10} M_\odot$, which is roughly the weight of a mid-size galaxy. As explained previously, the smallest galaxies that were able to participate to reionization had a mass of roughly $10^8 M_\odot$, meaning that, with this particle mass, we are missing the smallest sources which motivates further the internship.

DYABLO was instructed to return data (position of the particles, matter density, ionization...) regularly at different redshifts to study of the evolution of the simulated Universe.

The following sections will only describe

the simulation at $z = 0$. The other redshifts will be studied after validating the first results at $z = 0$. The final position of the particles at $z = 0$ can be seen in Figure 4.1. The area of over/underdensity is already apparent, regions with higher density are expected to host more halos than their underdense counterparts.

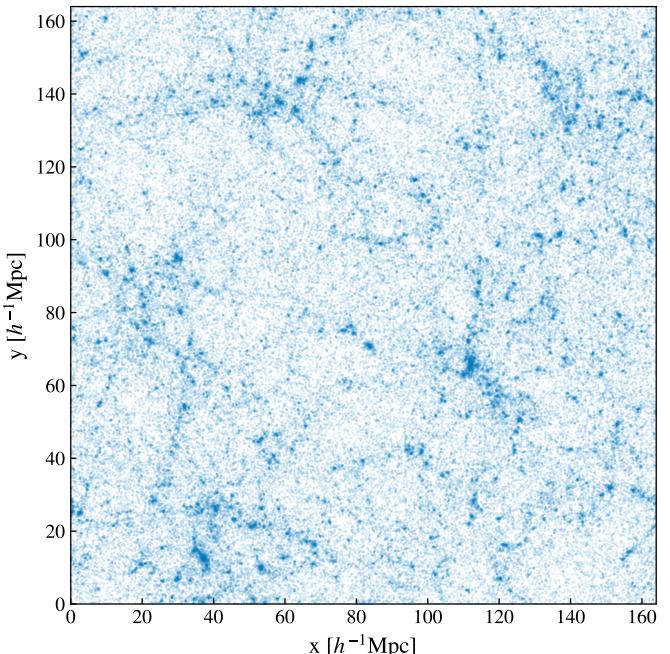


Figure 4.1: Position of one hundredth of the particles at the end of the simulation ($z = 0$)

4.1.2 Detecting dark matter halos

In simulations, dark matter halos are simply a group of particles that are gravitationally bounded together. The first step for studying them is to detect them in the outputs of the simulation. Multiple codes exists to be able to perform this task, the one used in our case is *Hierarchical Overdensity Peak (HOP)* developed by [Eisenstein & Hut \(1998\)](#). It works by looking for areas in the domain where the number of particles is the highest. Each particle is assigned a density value estimated using a smoothing kernel based on the distances to its nearest neighbors. Then, starting from a dense enough particle, the algorithm *hops* to the neighbor with the highest density and keeps hopping until it has found the highest density particle among its neighbors. All the particles that lead to this highest density particle are combined to form a halo and its mass is the sum of the mass of all the particles that are forming it.

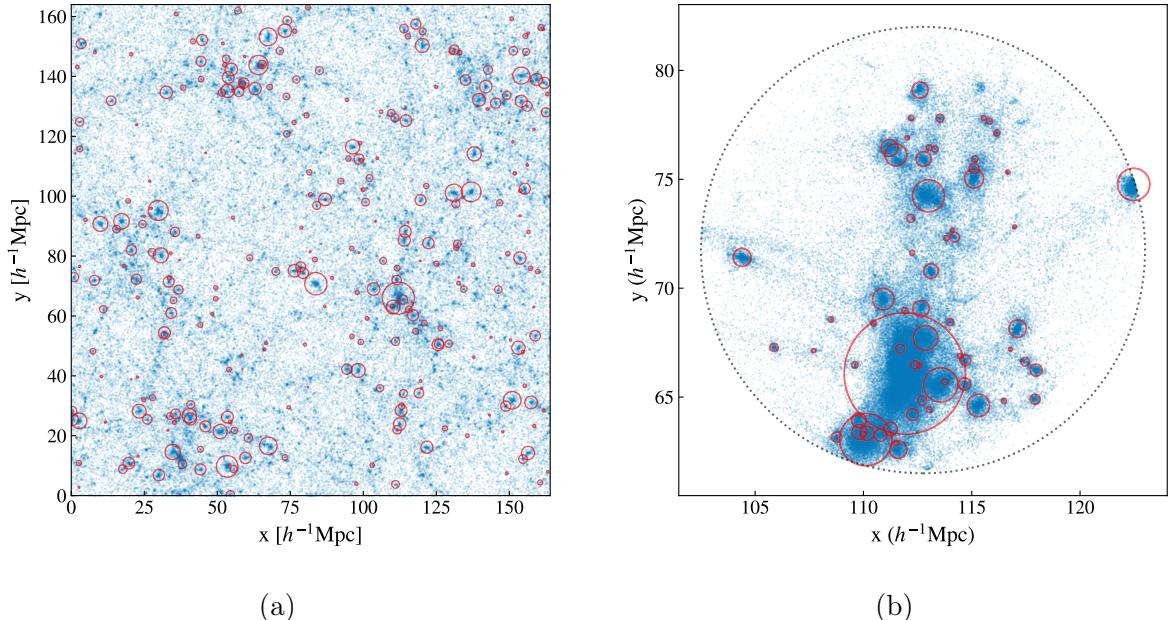


Figure 4.2: Halos (red) found among the particles (blue) forming them at $z = 0$, displayed in (a) The whole simulation output and showing one fifteenth halos (b) A dense ($\delta_{NL} = 8.59$) spherical region of radius $10.25 h^{-1}\text{Mpc}$.

In our case, 7084 halos were found with masses ranging in $[1.21 \times 10^{12}, 2.36 \times 10^{15}] M_{\odot}$ which corresponds to halos composed of 35 up to 68 607 particles.

Figure 4.2 shows the distribution of halos in the simulation with Figure 4.2a showing only a fifteenth of the halos for visual clarity. Moreover, in both figures, the circles representing the halos were scaled up according to their respective mass. We can see that, as expected, halos are much more clustered in regions of higher density. This is consistent with the large-scale structure of the Universe where dark matter halos form the cosmic web.

Figure 4.2b shows the 65 halos found in the densest region of radius $10.25 h^{-1}\text{Mpc}$ which has a non-linear overdensity δ_{NL} of 8.59. We can clearly see that lower mass sub-halos are inscribed in the more massive ones which indicates a merging process or hierarchical buildup. This structure reflects well the bottom-up nature of structure formation where smaller halos form first and later merge to make more massive systems.

4.2 Density field and power spectrum

The *matter power spectrum* shows how matter is spread out at different scales in the Universe. It helps understand how small under or overdensities in matter grew over time to make structures. Comparing it with observations (Figure 4.3) lets us test our models of how the Universe formed and evolves through time.

To rigorously define it, let's first define the *overdensity* in a region :

$$\delta(\vec{x}) = \frac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}}, \quad (4.1)$$

with $\rho(\vec{x})$ the mass density in the region and $\bar{\rho}$ the average density of the Universe. The value of overdensity quickly gives information on whether a region is overdense ($\delta > 0$) or underdense ($\delta < 0$). Moreover, the minimum value of the overdensity is -1 since a void region with no mass would have a density equal to zero ($\delta = -\bar{\rho}/\bar{\rho} = -1$).

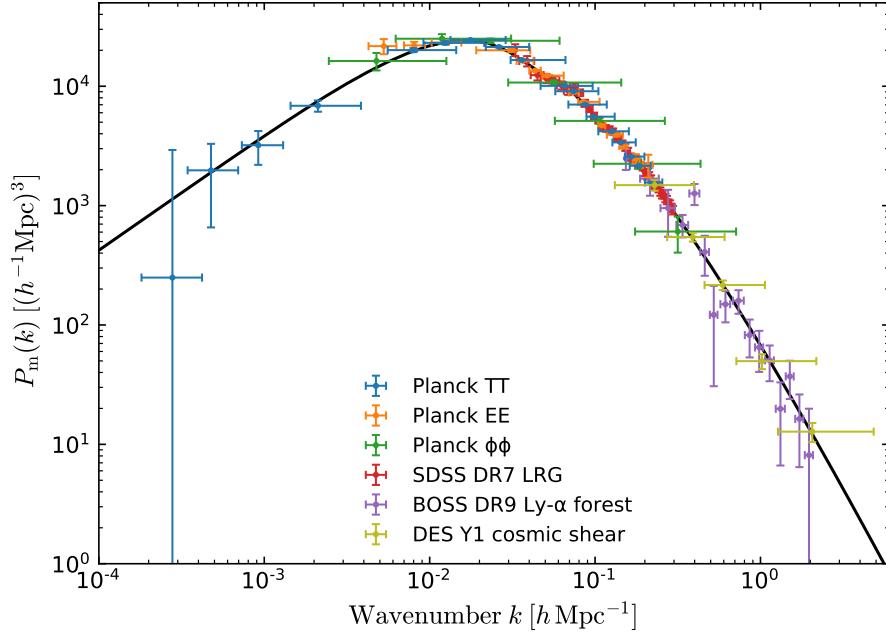


Figure 4.3: Matter power spectrum at $z = 0$ as a function of the wavenumber k inferred from multiple cosmological probes from [Planck Picture Gallery - ESA](#)

To study the distribution of density fluctuations across different scales, a Fourier transform (\mathcal{F}) is applied to move to a Fourier space¹ :

$$\delta(\vec{x}) = \sum_{\vec{k}} \delta(\vec{k}) e^{i \vec{k} \cdot \vec{x}} \quad \text{with} \quad \delta(\vec{k}) = \mathcal{F}(\delta(\vec{x})). \quad (4.2)$$

$\delta(\vec{k})$ gives the amplitude of density fluctuations as a function of wavenumber $k = 2\pi/\lambda$ where λ is the physical wavelength of a fluctuation mode. A large k represents small scales (short wavelengths) while a small k represents large scales (long wavelengths). The power spectrum is then defined as :

$$P(k) = \langle |\delta(\vec{k})|^2 \rangle. \quad (4.3)$$

This power spectrum describes how fluctuations at a given scale contribute to the Universe density.

It is a useful way to describe density fields and especially during the early stages of gravitational collapse known as Jeans instability when the overdensity is small ($\delta \ll 1$). In that case we can use a method called *perturbation theory* to study how these small fluctuations grow. Each wave (or Fourier mode) grows independently and while their

¹In practice, one uses the Fast Fourier Transform (FFT).

strength increases over time, the overall shape of the power spectrum stays the same.

The most important quantity to understand if a structure has collapsed is the overdensity (Equation 4.1). In linear perturbation theory ($\delta \ll 1$) the linear overdensity δ_L gives an estimate of how much a region has deviated from the background density. This approximation is not valid when the density grows and non-linear (NL) effects become significant.

The relation between δ_L and δ_{NL} has been derived by [Mo & White \(1996\)](#) :

$$\delta_L(\delta_{NL}) = \frac{\delta_c}{1.68647} \left(1.68647 - \frac{1.35}{(1 + \delta_{NL})^{2/3}} - \frac{1.12431}{(1 + \delta_{NL})^{1/2}} + \frac{0.78785}{(1 + \delta_{NL})^{0.58661}} \right). \quad (4.4)$$

δ_c is named critical overdensity and given the value of 1.68647 by [Sheth & Tormen \(2002\)](#). Every theory used here is based on the linear overdensity δ_L while measurements in simulations and observations give δ_{NL} , the formula allows a simple link between those two.

4.3 Mass functions

4.3.1 Halo Mass Function

One way to characterize those halos is the Halo Mass Function (HMF)² which gives the number density of halo per unit volume per unit mass at a given redshift.

[Press & Schechter \(1974\)](#) decided to predict what the HMF is without the need of expensive N-body simulations. First, they found that most structures collapse when the linear overdensity δ_L goes above δ_c (Figure 4.4).

By considering an arbitrary point in the Universe and placing a spherical region centered on this point and gradually increasing its radius, the average overdensity $\delta(R)$ within the sphere can be computed. Also, as the radius increases, the volume grows and $\delta(R)$ typically decreases to zero because it averages over a larger mass distribution : $\delta(R \rightarrow \infty) = 0$. For small enough radius, local density peaks may make it so $\delta(R_c) \geq \delta_c$ indicating that the region contains sufficient matter to collapse onto itself gravitationally.

Then, Press & Schechter computed what fraction of mass lies above δ_c by assuming that all this fraction of the mass has collapsed and formed halos.

Since they supposed matter density distribution in the Universe follows a Gaussian law, the probability of a region of size R to cross the threshold is given by :

$$P(\delta > \delta_c | R) = \int_{\delta_c}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma(R)} \exp\left(-\frac{\delta^2}{2\sigma(R)^2}\right) d\delta \quad (4.5)$$

with $\sigma(R)$ the variance of the density field.

²also sometimes called Unconditional Mass Function (UMF) in the literature.

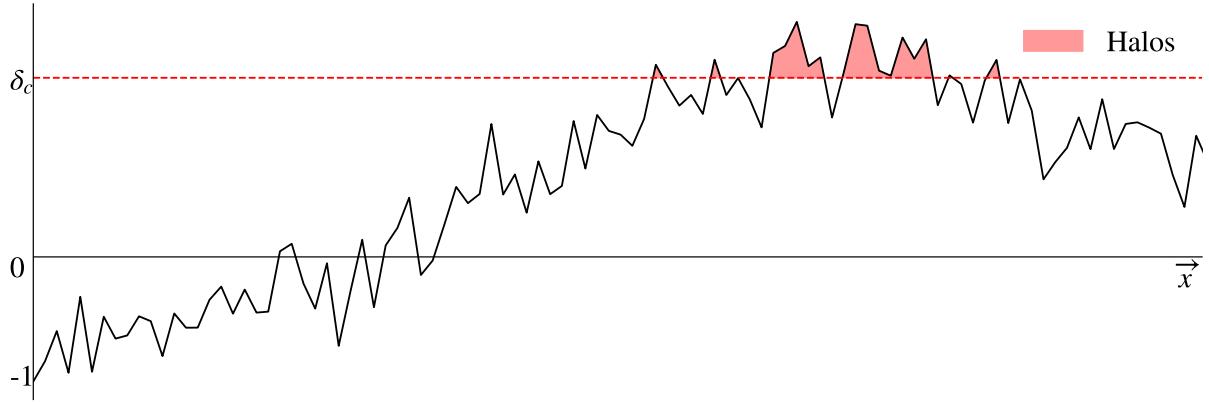


Figure 4.4: Overdensity $\delta(\vec{x})$ as a function of the radius \vec{x} of a spherical region centered at a fixed point in space. Regions where $\delta(R) > \delta_c$ (shaded red) are identified as containing collapsed halos.

The amplitude of density fluctuations³ $\sigma(R)$ depends on the scale R and is equal to :

$$\sigma^2(R, z) = \frac{1}{(2\pi)^3} \int P_m(k, z) W^2(k) d^3k, \quad (4.6)$$

with $k = 2\pi/\lambda$ and W is the Fourier transform of a window function (Mo, van den Bosch & White, 2010) given by, for example (Rubiño-Martín et al. (2008), Tramonte et al. (2017)) $W(x) = 3(\sin(x) - x \cos(x))/x^3$ which is used to smooth the field at a given scale.

For the rest of the report, unless stated otherwise, there always will be a direct link between a mass and a radius through $M = 4\pi\bar{\rho}R^3/3$, each mass will be bijectively associated to a radius, defining a sphere in which is contained the mass assuming the average density of the Universe in that sphere.

From Equation 4.5, Press & Schechter (hereafter PS) derived the following HMF :

$$\frac{dn_{PS}}{dM}(M, z) = \sqrt{\frac{2}{\pi}} \left| \frac{d\sigma}{dM} \right| \frac{\bar{\rho}}{M} \frac{\delta_c}{\sigma(M, z)^2} \exp \left(-\frac{\delta_c^2}{2\sigma(M, z)^2} \right). \quad (4.7)$$

dn/dM is the number of halo per unit mass per unit volume expressed in $Mpc^{-3}M_\odot^{-1}$. This equation aims to provide a statistical description of the population of dark matter halos in the Universe.

While reasonably accurate to predict the evolution and shape of the mass function of bound structures, the PS formalism misses some massive halos and overpredicts small ones. Moreover, this formalism assumes a spherical collapse of the structures which is a very simplistic model of gravitational collapse.

Sheth, Mo & Tormen (2001) decided to specify the latter model by not only assuming that the collapse is ellipsoidal but also that structures do not only form when the density crosses the fixed threshold δ_c but rather a moving threshold depending on the variance of the density field σ , meaning that smaller structures need to reach a higher density to

³The σ_8 seen in Table 3.1 is exactly $\sigma_8 = \sigma(R = 8 h^{-1}\text{Mpc}, z = 0)$.

collapse. They derived their own HMF :

$$\frac{dn_{ST}}{dM}(M, z) = A \sqrt{\frac{2a}{\pi}} \left| \frac{d\sigma}{dM} \right| \frac{\bar{\rho}}{M} \left[1 + \left(\frac{\sigma(M, z)^2}{a\delta_c^2} \right)^p \right] \frac{\delta_c}{\sigma(M, z)^2} \exp \left(-\frac{a\delta_c^2}{2\sigma(M, z)^2} \right), \quad (4.8)$$

with $a = 0.707$, $A = 0.3222$ and $p = 0.3$.

This formula is known to provide a better match to the empirical data.

4.3.2 Conditional Mass Function

As said previously the HMF gives a prediction for the abundance of dark matter halos on average in the Universe ; however when building a simulation, one may want to know this halo mass function not for the whole Universe but for a more localized region, typically inside a cell of the simulation to create stars inside for example. This is called the Conditional Mass Function (CMF). It is an essential ingredient to determine the population of unresolved sources since it gives a statistical prediction of halo abundances within specific environments which allows for more accurate modeling of small scale structures.

To obtain the CMF in a region characterized by a radius R_0 and a non-linear overdensity ⁴ δ_{NL} , the following transformation is applied to the HMF (Bond et al., 1991) :

$$\delta_c \rightarrow \delta_c - \delta_L; \quad (4.9a)$$

$$\sigma(R, z)^2 \rightarrow \sigma(R, z)^2 - \sigma(R_0, z)^2, \quad (4.9b)$$

with $\delta_L = \delta_L(\delta_{NL})$ by equation (4.4). It yields the Extended Press-Schechter (hereafter EPS) CMF :

$$\frac{dn_{EPS}}{dM}(M, z | R_0, \delta_{NL}) = \sqrt{\frac{2}{\pi}} \left| \frac{d\sigma}{dM} \right| \frac{\bar{\rho}}{M} \frac{\sigma(\delta_c - \delta_L)}{(\sigma^2 - \sigma_0^2)^{3/2}} \exp \left(-\frac{(\delta_c - \delta_L)^2}{2(\sigma^2 - \sigma_0^2)} \right), \quad (4.10)$$

where $\sigma = \sigma(M, z)$ and $\sigma_0 = \sigma(R_0, z)$.

The same can be done for the ST formalism, according to Sheth & Tormen (2002) :

$$\frac{dn_{cST}}{dM}(M, z | R_0, \delta_{NL}) = \sqrt{\frac{2}{\pi}} \left| \frac{d\sigma}{dM} \right| \frac{\bar{\rho}}{M} \frac{|T(\sigma^2 | \sigma_0^2)| \sigma}{(\sigma^2 - \sigma_0^2)^{3/2}} \exp \left(-\frac{(B(\sigma^2, z) - \delta_L)^2}{2(\sigma^2 - \sigma_0^2)} \right), \quad (4.11)$$

with $a = 0.707$ and where T is defined as :

$$T(\sigma^2 | \sigma_0^2) = \sum_{n=0}^5 \frac{(\sigma_0^2 - \sigma^2)^n}{n!} \frac{\partial^n (B(\sigma^2, z) - \delta_L)}{\partial (\sigma^2)^n}, \quad (4.12)$$

which is the first few terms of B , the moving barrier shape associated with ellipsoidal collapse defined as :

$$B(\sigma^2, z) = \sqrt{a} \delta_c (1 + \beta(a \frac{\delta_c^2}{\sigma^2})^{-\alpha}). \quad (4.13)$$

⁴Obtained by measuring the density ρ inside the region of radius R_0 and injected in Equation 4.1.

And it is quite easy to get analytically the n-th derivative of $B(\sigma^2, z)$ according to σ^2 :

$$\frac{\partial^n B}{\partial y^n} = \beta(a\delta_c^2)^{-\alpha+1/2} \frac{\alpha!}{(\alpha - n)!} y^{\alpha-n} \text{ for } n \geq 1 \text{ and } y = \sigma^2. \quad (4.14)$$

and $\alpha = 0.615$ and $\beta = 0.485$.

4.3.3 Implementation of the theoretical HMF & CMF

We need to assert that we can code the mass functions accurately. To do that we will reproduce the mass function plots from an article from the literature, for example, [Meriot & Semelin \(2024\)](#). We implemented every equation stated above to code the mass functions from scratch using Python and the *CAMB* (Code for Anisotropies in the Microwave Background) library ([Lewis et al., 2000](#)) which helped to get the power spectrum from given cosmological parameters.

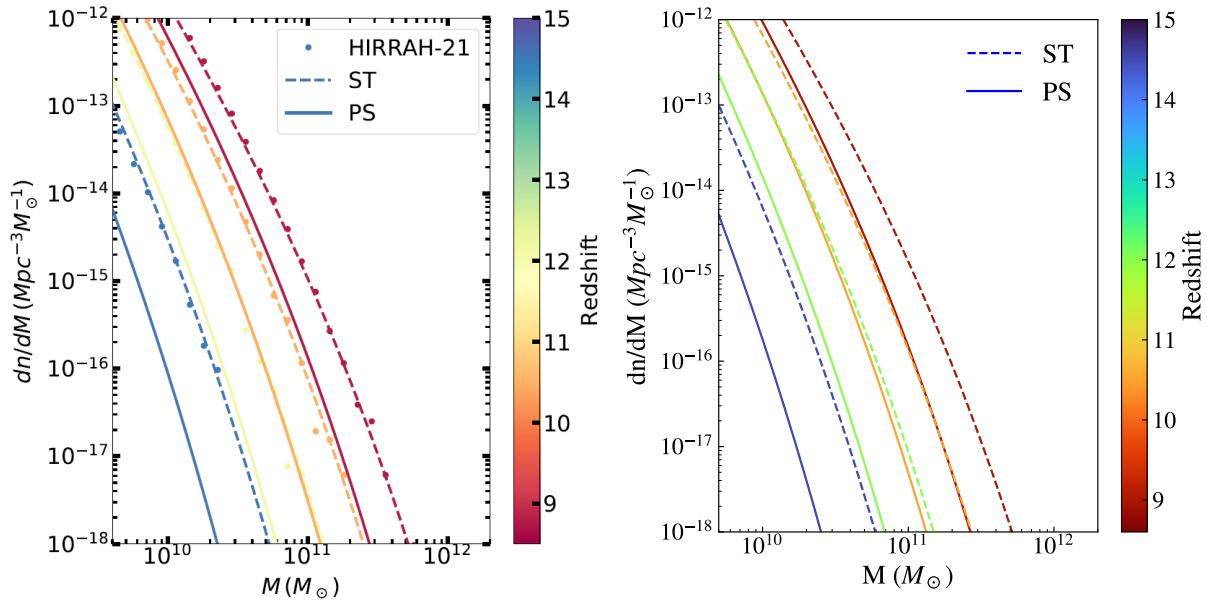


Figure 4.5: Left : Fig. 1 from [Meriot & Semelin \(2024\)](#). Right : My reproduction

Figure 4.5 shows the HMF curve for a Universe at different redshifts. On the general shape of the HMF, there are much more halos of small mass than higher mass because smaller halos form first and later merge to build larger halos ([Lacey & Cole, 1993](#)). Moreover, we can see that, a higher redshift (younger Universe) means a lower halo number density, this shows the progressive formation of cosmic structures over time. In the early Universe, density fluctuations had less time to grow under gravity which translates in a lower abundance of collapsed halos, especially for higher mass halos ([Davis et al., 1985](#)).

We also validated those results with the library *HMFcalc* ([Murray et al., 2013](#)) that computes halo mass functions in a single Python command. Although *HMFcalc* is useful for quickly computing the HMF, it does not compute the CMF. This is why coding ev-

erything ourselves is better since it gives us control over the full process and lets us apply the method to different cases without relying on a black-box tool.

Therefore we also coded the CMF from the equations stated in the previous section. We represent the CMF using the same parameters as Meriot & Semelin (2024) (left panel of Figure 4.6) ; the redshift is fixed at $z = 9.48$ in spherical regions of radius 5 Mpc of different overdensities :

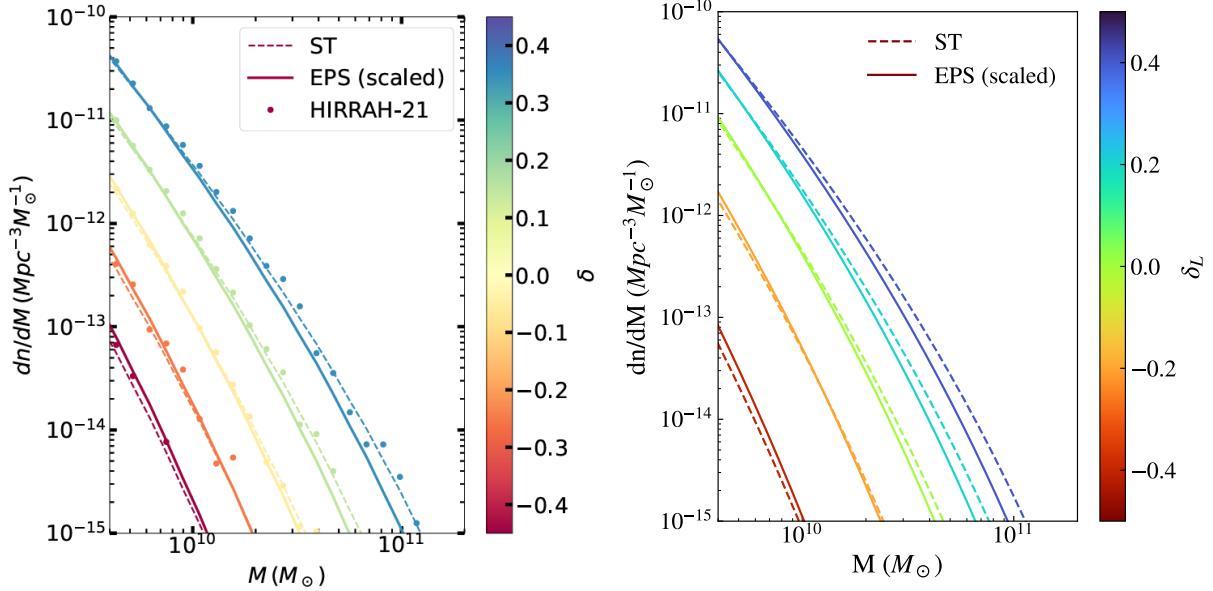


Figure 4.6: Left : Fig. 2 from Meriot & Semelin (2024). Right : My reproduction

Once more, there are much more low-mass halos than higher mass halos. Additionally, and as anticipated, the more a region is dense, the easiest it is for structure to gravitationally collapse onto themselves.

The previous figures (4.5 & 4.6) show that we are able to accurately reproduce the theory which will be useful when using it to compare to simulation outputs. For the rest of the report, we will only use the ST formalism when plotting theoretical curves because it provides a better fit to the empirical data.

4.3.4 Validation of the abundance of halos

4.3.4.1 Halo Mass Function

Since the different dark matter halos and their respective mass are available for the output of the simulation, one can simply plot the HMF of the output by doing the histogram of the mass of the halos and normalizing by the volume and the size of the bins which gives the number of halos per unit volume per unit mass.

Figure 4.7 shows both the theoretical HMF and the one obtained from the simulation and using the method described in subsection 4.3.3. Expectedly, it has the same general

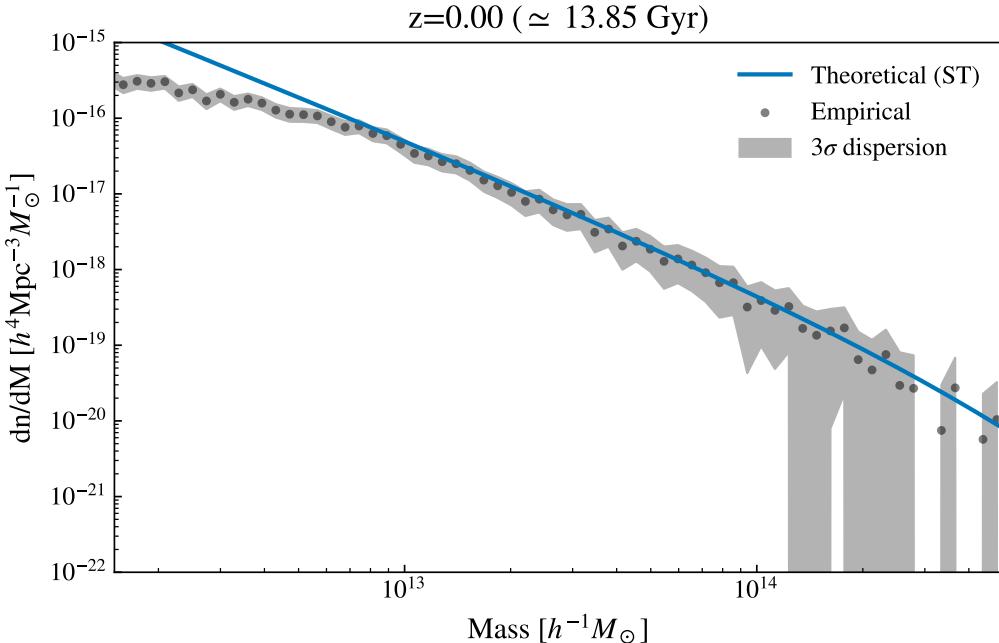


Figure 4.7: HMF at $z = 0$ from the ST formalism (blue curve) and empirical from the simulation (black dots) with their 3σ dispersion represented as the grey shaded area

shape as the previous mass function graph (Figure 4.5) with higher number density of lower mass halos than higher mass ones. We can first see that the simulation follows well the theory for halos that respect $M_h \gtrsim 7 \times 10^{12} h^{-1} M_\odot$ but not below. This is due to the limited resolution of the simulation. The simulation lacks the particle mass and force resolution needed to reliably resolve the formation and evolution of smaller halos. Consequently, the empirical halo mass function (HMF) deviates from the theoretical expectation in the low-mass regime, while agreement remains good at higher masses where resolution effects are less significant. This is important that it works well at the medium and high mass end because, in a later part of the report, we will be *extrapolating* the theoretical HMF at the low mass end in the simulation to *expect* halos without having to resolve them with high resolution.

While the HMF provides a global statistical view of halo abundance, studying the CMF helps understanding environmental dependence and needs smaller and more localized regions of the simulation domain.

4.3.4.2 Conditional Mass Function

4.3.4.2.a Segmenting the domain

In order to study the CMF, the domain of the simulation needs to be segmented in multiple regions so that the results may be statistically significant. If the domain is divided in not enough regions, the regions will most likely have a density close to the Universe's average one and the CMF will be similar to the HMF which does not provide much information. If the domain is divided in too many regions, there will not be enough

halos in the said regions to be able to properly understand their properties thanks to the CMF. It was chosen to study $8^3 = 512$ cubic regions in the domain with a 30.60 Mpc of width.

We will then consider the circle inscribed in those cubes since the ST and PS formalisms are considering spherical regions as well. The radius of the spheres is then $R_0 = 30.60/2 = 15.30$ Mpc.

It's important to note that when using spheres, we only get $V_{sphere}/V_{cube} = (4\pi R^3/3)/(2R)^3 = \pi/6 \simeq 0.52$ of the volume that a cube would have. While only about 52% of the domain volume is actually sampled when using inscribed spheres instead of full cubes, this is acceptable under the cosmological principle which ensures that the statistical properties of the Universe are homogeneous on large scales. Therefore even a partial sampling of the full volume provides representative statistics for the CMF especially given the large number of sampled regions.

4.3.4.2.b Lagrangian radius

One may understand from the CMF equations ((4.10) & (4.11)) that the *formalisms require for the following equation to always be true :*

$$\sigma(M, z) > \sigma_0(M, z). \quad (4.15)$$

This requirement ensures the result of the CMF remains real. Breaking this inequality would result in imaginary values for physical quantities like the halo number density, which is does not make sense in this context.

Studying what may cause this problem will help identify limitations of the model and may point where the CMF formalism requires modification or should not be applied.

A spherical region of radius R_0 in the Universe can define two masses :

$$M_r = \frac{4}{3}\pi\rho_r R_0^3 \quad \text{and} \quad M_0 = \frac{4}{3}\pi\bar{\rho}R_0^3 \quad (4.16)$$

where M_r is the true mass contained within R_0 and M_0 is the mass that would have the region if the said region is characterized by having the Universe's average density $\bar{\rho}$.

The available mass range to study in the region of radius R_0 is then $[0, M_0[$ because going above M_0 would break the condition (4.15) since, for all redshift, $\sigma(M)$ is a decreasing function of M : the larger the region, the smaller the variance.

However, when segmenting the domain of a simulation in regions, some may have a larger density than the Universe's one ($\delta_{NL} > 0$). Consequently, the equivalent radius R of the total mass in the region $M_r = 4\pi\bar{\rho}R^3/3$ would be larger than R_0 .

This leads to a serious problem : *overdense regions cannot be fully studied* because the mass range available for study is limited to $[0, M_0[$ while the actual mass content of the region exceeds this upper bound.

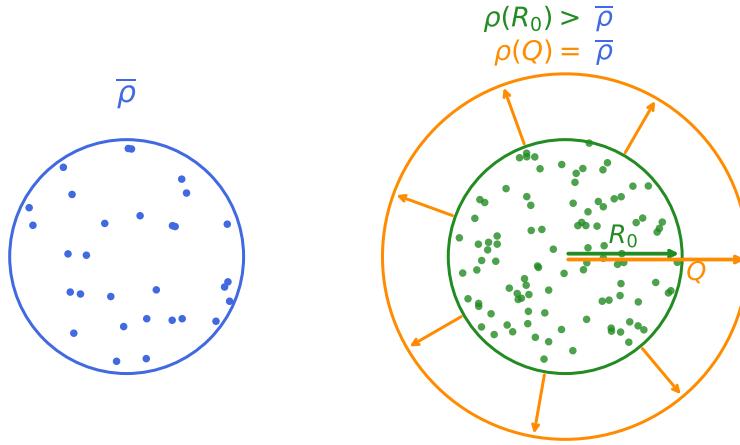


Figure 4.8: Schematic representation of the concept of Lagrangian radius

To overcome this problem, Rubiño-Martín et al. (2008) or Tramonte et al. (2017) decided to name Q , the Lagrangian radius (contrarily to the Eulerian radius R_0) that defines a new region of mass M_r with the density of the region at $\bar{\rho}$.

$$Q = \left(\frac{3M_r}{4\pi\bar{\rho}} \right)^{\frac{1}{3}}. \quad (4.17)$$

Therefore, with the definition of R_0 in Equation 4.16 and by remembering with Equation (4.1) that $\rho_r = \bar{\rho} (1 + \delta_{NL})$, one can simply find that :

$$Q = R_0 \left(\frac{\rho_r}{\bar{\rho}} \right)^{\frac{1}{3}} = R_0 (1 + \delta_{NL})^{\frac{1}{3}}. \quad (4.18)$$

By replacing the Eulerian radius R_0 by the Lagrangian radius Q for the spherical region of interest, it means that we are *virtually increasing the radius so that larger regions are studied* (Figure 4.8) when having $\delta_{NL} > 0$ which allows to study the exact range of mass possible for halos in the region of interest $[0, M_r]$ without breaking the condition (4.15).

4.3.4.2.c Results

After validating the implementation of the theoretical models using the empirical halo mass function (HMF) derived from the output of a DYABLO simulation at $z = 0$, we segment the domain into subregions, apply the the Lagrangian radius and develop a method to determine the minimal mass required to identify a halo. All of those steps allow us to study subregions based on their overdensity which characterizes the local environment through the CMF.

Figure 4.9 shows the CMF for some of the available ranges from the output. Each solid line corresponds to a subregion respecting the condition on δ_{NL} while the black dashed line corresponds to the average CMF of all the subregions. The black dots correspond to the simulation CMF with their respective error bars. Bottom right plot is the normalized

histogram of the subregions depending on their overdensity. On that plot, the different vertical dashed lines corresponds to the limits of the overdensity classes.

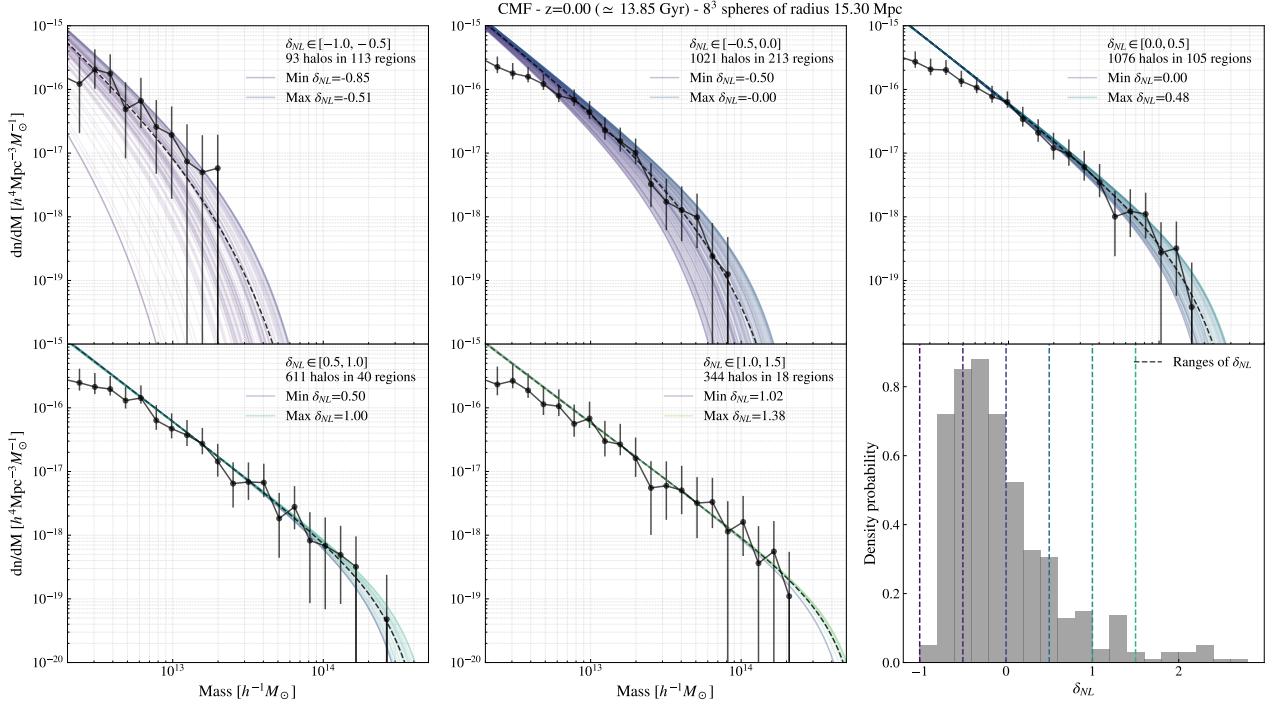


Figure 4.9: CMF for different classes of δ_{NL} at $z = 0$.

One caveat that might lead to misinterpretation of the curves is that, since the problem is highly non-linear, we have $\langle \frac{dn}{dM}(M|\delta_{NL}) \rangle \neq \frac{dn}{dM}(M|\langle \delta_{NL} \rangle)$ where $\langle \rangle$ represents the average over all the subregions ; the average CMF of the subregions is not representative of the CMF of the average subregions respecting the condition on δ_{NL} . Figure 4.9 plots $\langle \frac{dn}{dM}(M|\delta_{NL}) \rangle$.

As shown in the figure, the empirical data closely follows the average CMF of the regions, particularly at medium and high masses, as expected. However, at the low-mass end, the data deviates from the average CMF, which was also anticipated. This shows that not only DYABLO correctly forms dark matter halos but also that we are now able to make predictions of their distribution across different environments which is quantified by the region overdensity.

4.4 Collapsed mass fraction

To go further, one may want to look at the collapsed mass fraction which is a key parameter when doing cosmological simulations because it gives the fraction of mass that has currently or previously collapsed in a given region.

When considering a region of radius R that contains a mass M_r and of volume $V_r = 4\pi R^3/3$, one can integrate the CMF to get the theoretical collapsed mass M_{coll} inside the region that is heavier than a certain parameter M_{min} . Empirically, in a simulation, one

4.4. COLLAPSED MASS FRACTION

only need to store the mass of halos and sum them up when necessary :

$$\begin{cases} M_{coll}^{th} = V_r \int_{M_{min}}^{M_r} \frac{dn_c}{dM}(M, z, R_0, \delta_{NL}) M dM; \\ M_{coll}^{emp} = \sum_{k \in H_r} M_h(k), \end{cases} \quad (4.19)$$

with H_r the halos found in the region and M_h their respective mass.

The collapsed mass fraction can then be found by dividing M_{coll} by the total mass in the region :

$$f_{coll} = \frac{M_{coll}}{M_r}. \quad (4.20)$$

Moreover, an analytical formula has been found for the EPS formalism :

$$f_{coll}^{EPS} = \text{erfc} \left(\frac{\delta_c - \delta_L}{\sqrt{2(\sigma(M_{min})^2 - \sigma_0^2)}} \right), \quad (4.21)$$

where erfc is the complementary of the error function. The cST formalism does not hold such analytical formula.

Since we have access to the halos, we may also compare the collapsed fraction in every region empirically and theoretically. If DYABLO does structure formation correctly, we should obtain the results along the $x = y$ line meaning $f_{coll}^{th} = f_{coll}^{emp}$.

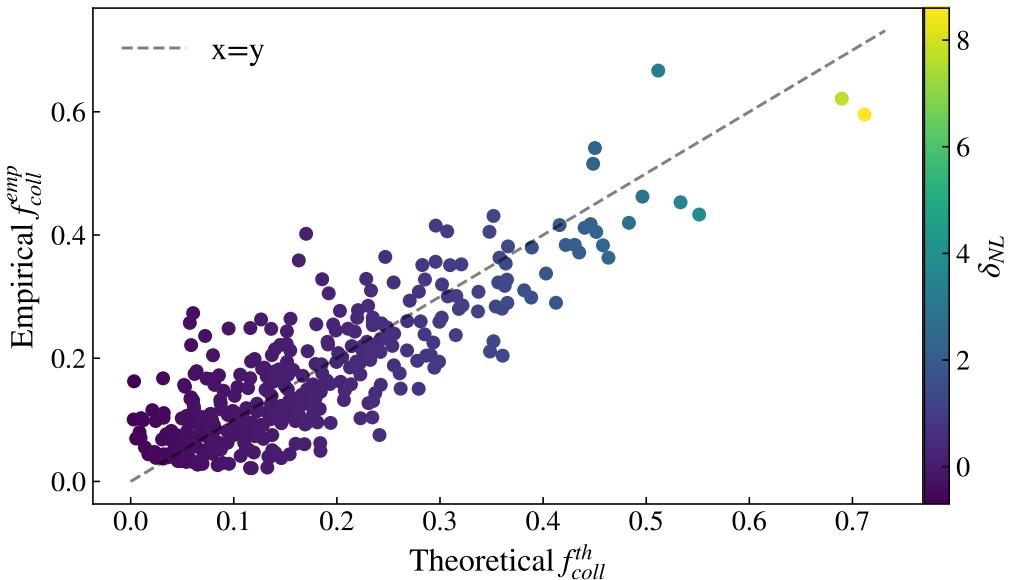


Figure 4.10: Comparing the empirical and theoretical collapsed mass fraction at $z = 0$

While the markers are not exactly on the $x = y$ line, we can still follow the trend to follow this line, they are no persistent under or overestimation by the simulation.

The simulation we used to get those results goes only to $z = 2.48$ ($\simeq 2.64$ Gyr old Universe). Since, DYABLO's aim is to properly study the EoR, the redshift of study should be $z \geq 6$. This is something we should test that has not been done yet. The issue with

the current simulation is the fact that it cannot create non linear structures at $z \geq 6$ due to low resolution. To overcome this problem, we could run a high-z simulation with the same number of particles in a small volume, however, according to [Iliev et al. \(2014\)](#), small volume simulations may not capture the large scale ionization structures necessary for an accurate representation of the EoR. An another option would be to do an extremely costly simulation with a large number of particles. A large simulation both spatially and temporally would allow halos to form and be resolved to be able to fully complete the validation of structure formation in DYABLO ; we plan to do that eventually.

For now, we continue to work with our successful results for $z \in [0, 2.48]$.

Figures 4.9 & 4.10 show that DYABLO behaves correctly at $z = 0$. The rest of the results, proving that DYABLO behaves correctly at all the redshifts available for our simulation in the Appendix C.

Chapter **5**

Improving DYABLO's ionization module

Contents

5.1	Computing the fraction of collapsed mass	35
5.2	Implementation of the photon source term	37

5.1 Computing the fraction of collapsed mass

Now that the structure formation in DYABLO has been validated, we may use theoretical frameworks instead of purely empirical ones to use results from the structure formation module. By using f_{coll}^{th} instead of f_{coll}^{emp} , we are using the theoretical CMF rather than the empirical one because *it accounts for low mass halos*. However, to get f_{coll}^{th} , one needs to compute an integral to find M_{coll}^{th} which takes ~ 1 second to compute. This is *too long* for a computation that is needed in every simulation cell at every time step.

Therefore, we will find a fit to f_{coll} which will give an accurate result almost instantaneously. Finding a fit in three dimensions (z, R_0, δ_{NL}) is quite challenging, therefore, we will set¹ R_0 to 1 Mpc which is restrictive but corresponds to the simulation's resolution. The Python library *PySr* ([Cranmer, 2023](#)) is a high-performance symbolic regression tool that uses machine learning to efficiently find fitting functions.

It is important to remember that functions also have their own complexities, for example, an hyperbolic sinus is much more computationally expensive than basics additions or subtraction. By finding a middle-ground between the best possible fit and computationally cheap functions, *PySr* finds the following fit (shown in Figure 5.1) :

$$f_{coll}^{fit}(\delta_{NL}, z | R_0 = 1 \text{ Mpc}) = \tanh\left(0.609 \phi(\delta_{NL}, z)\right), \quad (5.1a)$$

¹since this is work in progress, we aim to find the three dimensional fit at a later stage.

$$\text{with } \phi(\delta_{NL}, z) = \frac{z + 1.77}{\delta_{NL} + 1.25} + \frac{0.430}{\cosh(z)}. \quad (5.1b)$$

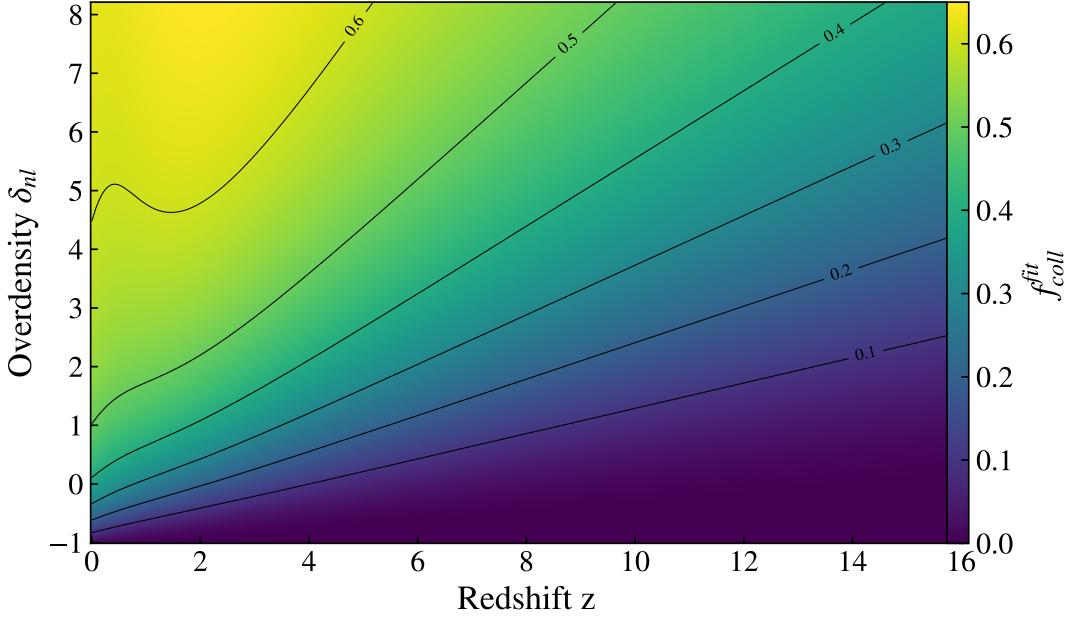


Figure 5.1: Plot of the f_{coll} fit found by *PySr*. R_0 was fixed at 1 Mpc.

The result of the fit is coherent in a sense that higher density region have a higher collapsed fraction. Additionally, structures at high-z do not have time to collapse significantly, which results in lower values of f_{coll} which is also correctly captured by the fit.

From Equation 3.14, one can compute the fraction of gas that is converted into stellar mass at each time step. Thus, one can know the Star Formation Rate (SFR) at any redshift between each time step of the simulation. After implementing Equation 3.14 in DYABLO and by retrieving the aforementioned value, we may plot the SFR evolution as a function of redshift (Figure 5.2) and compare it to the literature (Bouwens et al., 2015). The offset of the curve is tuned through the value of $\tau_{SF} = 70$ Gyr ($\in [10, 100]$ Gyr according to Meriot & Semelin (2024)). Moreover, it is important to note that our SFR computation is quite crude because our time steps are rather large (cf. Table 4.1) which means that we are averaging over relatively long periods of time which can smooth out short term fluctuations

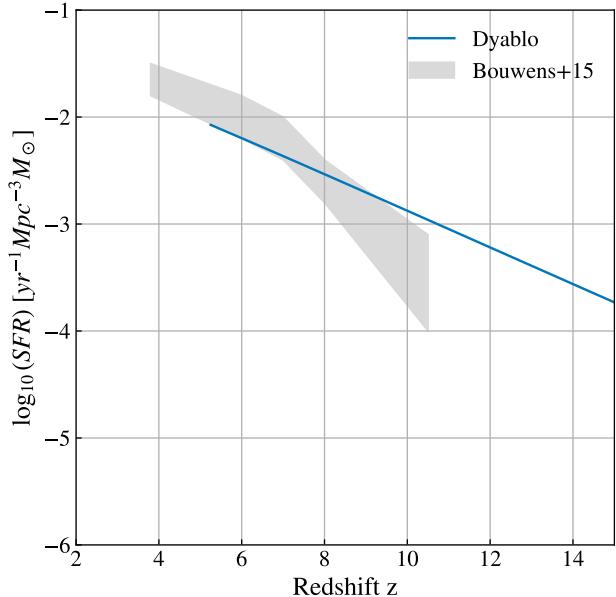


Figure 5.2: Star Formation Rate (SFR) as a function of redshift.

of the SFR. Therefore, only the overall trend of decreasing SFR with higher redshift is reproduced and roughly matches the observations from [Bouwens et al. \(2015\)](#) which is sufficient for our model.

5.2 Implementation of the photon source term

It is assumed that only newly formed stars are producing ionizing photons, therefore, at each time step only the stellar mass derived from df_* (Equation 3.14) contributes to ionization. We only consider those since the production of ionizing photons is dominated by massive stars that emit almost all of their ionizing radiation within a few million years of their formation ([Leitherer et al., 1999](#)) meaning :

$$\Delta N = \rho_r \times df_* \times \dot{N} \times \Delta t. \quad (5.2)$$

In that expression, $\rho_r \times df_*$ defines the new density of newly formed stars. \dot{N} is a constant representing the stellar ionizing emissivity expressed photons/s/ M_\odot . The value of \dot{N} may be set by the user and we will first need to tune it so it gives valid reionization scenarios. All of this yields ΔN which is the new density of photons created in the cell.

This photon density variation is implemented within the M1 scheme for radiative transfer inside the DYABLO code (subsection 3.3.2).

After implementing the method for photons production and tuning \dot{N} for a correct reionization scenario ($\dot{N} = 1.5 \times 10^{48}$ photons/s/ M_\odot), we may launch a simulation using the same parameters as Table 3.1 & 4.1. We may then plot the ionizing fraction as a function of redshift (Figure 5.3).

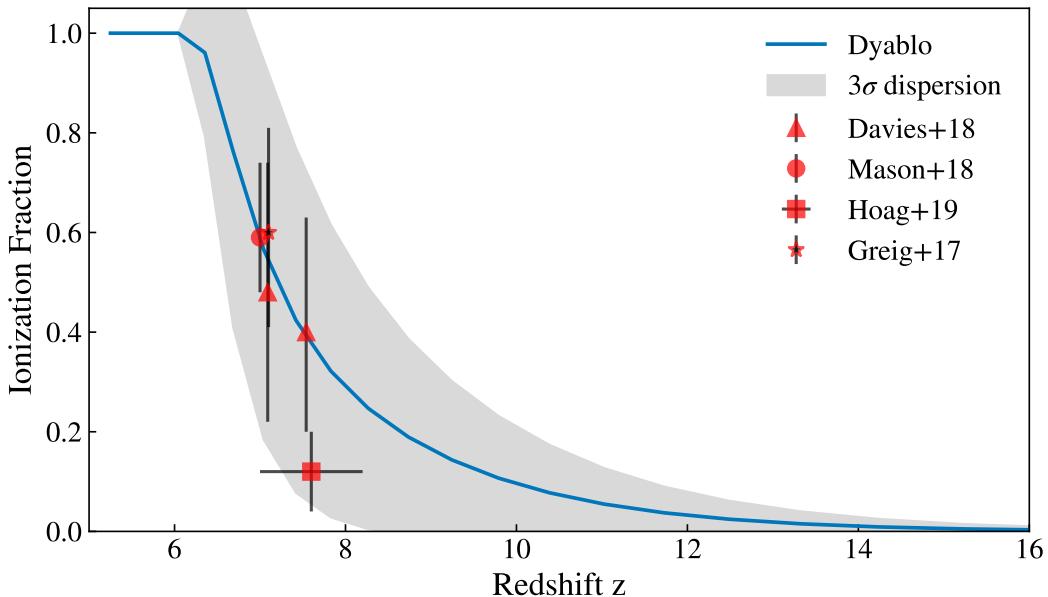


Figure 5.3: Ionizing fraction as a function of redshift

This graph shows that using the fit for f_{coll} and tuning well \dot{N} not only produces a reionization scenario consistent with observational constraints and in agreement with the literature (Davies et al. (2018), Mason et al. (2018)), Hoag et al. (2019) & Greig et al. (2017)) but also leads in a model that is now physically motivated.

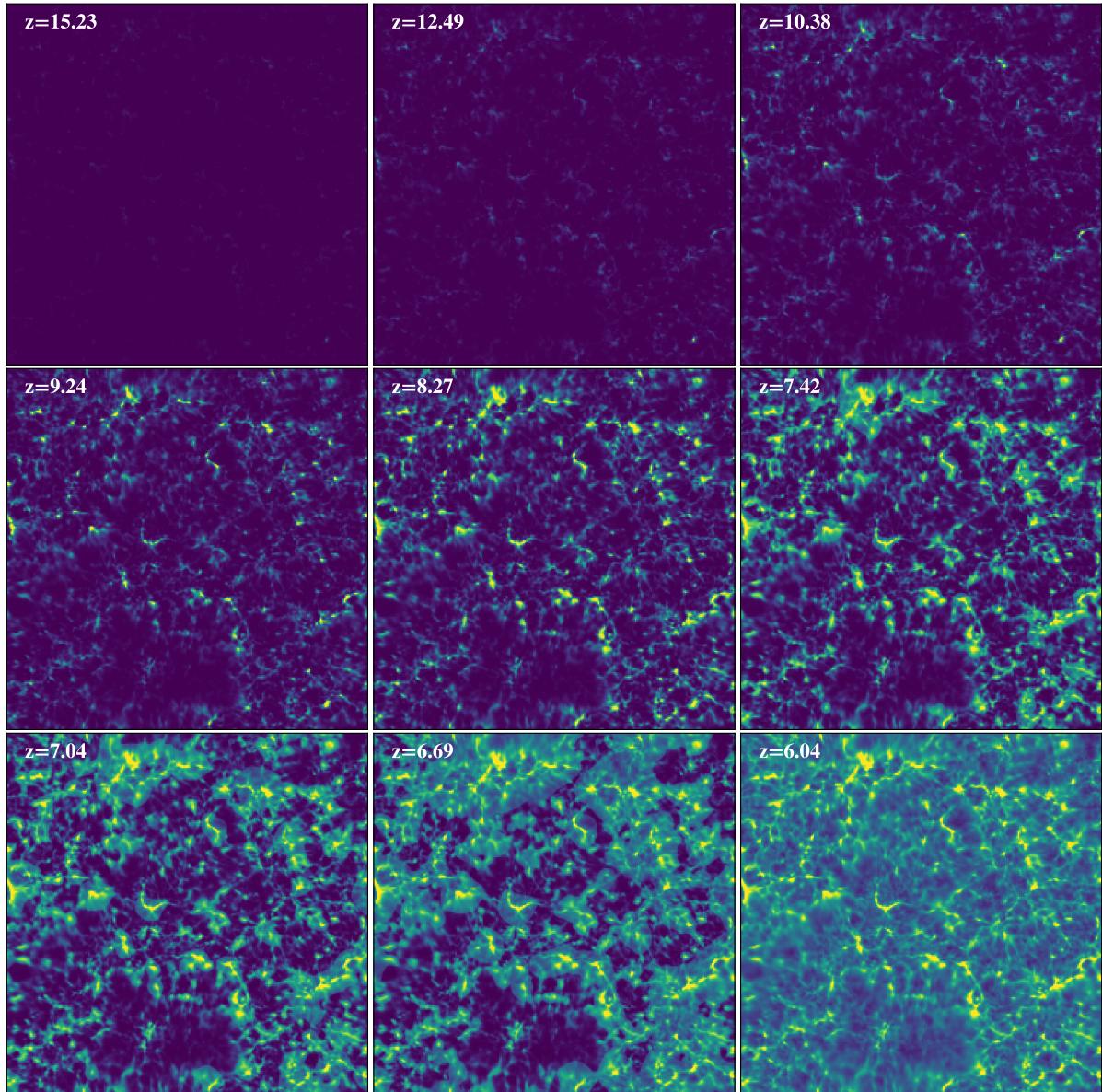


Figure 5.4: Snapshots of the evolution density of ionized gas as a function of redshift

The value of $\dot{N} = 1.5 \times 10^{48}$ photons/s/ M_\odot for the stellar emissivity constant has been manually tuned to get the reionization scenario that follows well the literature however Ocvirk et al. (2020) set that value to $\dot{N} = 4.32 \times 10^{46}$ photons/s/ M_\odot which is approximately two orders of magnitude lower than what we tuned. We believe it comes from our numerical approximation that stars only emit ionizing photons during the numerical time step that they are born. However, this time step can vary and may not be consistently representative of the emission period during which massive stars produce ionizing radia-

tion.

We can also plot the density of ionized gas (Equation 3.4) at an arbitrary slice of the domain (Figure 5.4), defined as the product of the density and the ionization fraction. This provides a much more intuitive and visual representation of the reionization process. On this simulation output, the EoR starts at $z \sim 15$ with small regions that start ionizing. Those are ionization bubbles that continue to grow as time passes. One can also notice that the ionization bubbles start developing where the density is the highest which is expected from the underlying physics of reionization : high density regions are where the first luminous sources form and produce ionizing photons that ionize their surroundings.

One limit for now is the fit that has been set only for regions of size 1 Mpc which is quite restrictive.

Chapter **6**

Conclusion

To conclude, the goals of this internship which are to validate and develop physics modules for reionization simulation code DYABLO has been successfully met.

Indeed, numerical simulations have become an essential tool in modern astrophysics but need to be rigorously coded and validated in order to give physically accurate results. One of those simulation code is DYABLO which aims at simulating the Universe's history. It is within this framework that this internship takes place.

When coding cosmological simulations, one wants to make sure all the physical processes associated to the formation of structure operate correctly. Dark matter halos, which are the building blocks in which baryonic structures like galaxies and clusters form, play a major role in this process. By comparing the mass functions which give the number density of dark matter halos per mass interval at a given redshift of theoretical formalisms to the output of DYABLO, we found that DYABLO correctly models them at any redshift.

DYABLO, and most generally, every numerical simulation code, cannot solve all physical scales due to finite computational resources. This means that simulations are constrained by limited resolution which requires the use of subgrid models to approximate the effects of unresolved small scale physics, in our case, star formation. We used the previous results to be able to confidently use theoretical frameworks for subgrid structure formation and star formation. Implementing those in the M1 radiative transfer scheme in DYABLO allowed to have more physically accurate reionization scenarios.

While the validation of the subgrid structure formation module is completely finished, there is still some work to do for the development of the ionization module. Indeed, the fit we use to estimate the fraction of collapsed mass in a cell assumes the cell to be exactly of 1 Mpc of width which is a problem for two reasons. First, a user of DYABLO may choose a different cell size meaning the current fit lacks flexibility. Second, even if a 1 Mpc cell size is desired, the use of AMR inherently introduces variations in cell sizes which make the cell width assumption incompatible with simulations that use AMR.

This internship was an overall success and lead to satisfying results.

Bibliography

- Aubert D., Durocher A., 2021, Ginea and Dyablo, in Siebert A., et al., eds, SF2A-2021: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. pp 473–475
- Aubert D., Teyssier R., 2008, A radiative transfer scheme for cosmological reionization based on a local Eddington tensor, *MNRAS*, 387, 295
- Berger M. J., Colella P., 1989, Local Adaptive Mesh Refinement for Shock Hydrodynamics, *Journal of Computational Physics*, 82, 64
- Berger M. J., Oliger J., 1984, Adaptive Mesh Refinement for Hyperbolic Partial Differential Equations, *Journal of Computational Physics*, 53, 484
- Bertschinger E., 2011, GRAFIC-2: Multiscale Gaussian Random Fields for Cosmological Simulations, <https://ascl.net/1106.008>
- Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations, *ApJ*, 379, 440
- Bouwens R. J., et al., 2015, UV Luminosity Functions at Redshifts $z \sim 4$ to $z \sim 10$: 10,000 Galaxies from HST Legacy Fields, *ApJ*, 803, 34
- Cranmer M., 2023, Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl, *arXiv e-prints*, p. arXiv:2305.01582
- Davies F. B., et al., 2018, Quantitative Constraints on the Reionization History from the IGM Damping Wing Signature in Two Quasars at $z > 7$, *ApJ*, 864, 142
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, The evolution of large-scale structure in a universe dominated by cold dark matter, *ApJ*, 292, 371
- Delorme M., Durocher A., Brun A. S., Strugarek A., 2022, Novel numerical methods for solar convection: The Dyablo Whole-Sun adaptative mesh refinement code, in Richard J., et al., eds, SF2A-2022: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. pp 209–213

BIBLIOGRAPHY

- Eisenstein D. J., Hut P., 1998, HOP: A New Group-Finding Algorithm for N-Body Simulations, *ApJ*, **498**, 137
- Friedmann A., 1922, Über die Krümmung des Raumes, *Zeitschrift fur Physik*, **10**, 377
- Friedmann A., 1924, Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes, *Zeitschrift fur Physik*, **21**, 326
- Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, The Growth of H II Regions During Reionization, *ApJ*, **613**, 1
- Godunov S. K., 1959, A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics, *Mat. Sb. (N.S.)*, **47**(89), 271
- González M., Audit E., Huynh P., 2007, HERACLES: a three-dimensional radiation hydrodynamics code, *A&A*, **464**, 429
- Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, Are we witnessing the epoch of reionisation at $z = 7.1$ from the spectrum of J1120+0641?, *MNRAS*, **466**, 4239
- Hcéres 2023, Technical report, Rapport d'évaluation de l'unité UMR 7550 - Observatoire astronomique de Strasbourg (ObAS). Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres), https://www.hceres.fr/sites/default/files/media/publications/rapports_evaluations/pdf/C2023-EV-0673021V-DER-PUR230023325-RF.pdf
- Hiegel J., et al., 2023, Reionisation time field reconstruction from 21 cm signal maps, *A&A*, **679**, A125
- Hoag A., et al., 2019, Constraining the Neutral Fraction of Hydrogen in the IGM at Redshift 7.5, *ApJ*, **878**, 12
- Hubble E., 1929, A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae, *Proceedings of the National Academy of Science*, **15**, 168
- Iliev I. T., Mellemma G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, Simulating cosmic reionization: how large a volume is large enough?, *MNRAS*, **439**, 725
- Lacey C., Cole S., 1993, Merger rates in hierarchical models of galaxy formation, *MNRAS*, **262**, 627
- Lax P. D., 1954, Weak solutions of nonlinear hyperbolic equations and their numerical computation, *Communications on Pure and Applied Mathematics*, **7**, 159
- Leitherer C., et al., 1999, Starburst99: Synthesis Models for Galaxies with Active Star Formation, *ApJS*, **123**, 3
- Lemaître G., 1927, Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques, *Annales de la Société Scientifique de Bruxelles*, **47**, 49

- Lewis A., Challinor A., Lasenby A., 2000, Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models, *ApJ*, **538**, 473
- Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, The Universe Is Reionizing at $z \sim 7$: Bayesian Inference of the IGM Neutral Fraction Using Ly α Emission from Galaxies, *ApJ*, **856**, 2
- Merlot R., Semelin B., 2024, The LORELI database: 21 cm signal inference with 3D radiative hydrodynamics simulations, *A&A*, **683**, A24
- Mo H. J., White S. D. M., 1996, An analytic model for the spatial clustering of dark matter haloes, *MNRAS*, **282**, 347
- Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution. Cambridge University Press, doi:10.1017/CBO9780511807244
- Murray S. G., Power C., Robotham A. S. G., 2013, HMFcalc: An online tool for calculating dark matter halo mass functions, *Astronomy and Computing*, **3**, 23
- Ocvirk P., et al., 2020, Cosmic Dawn II (CoDa II): a new radiation-hydrodynamics simulation of the self-consistent coupling of galaxy formation and reionization, *MNRAS*, **496**, 4087
- Planck Collaboration et al., 2020, Planck 2018 results. VI. Cosmological parameters, *A&A*, **641**, A6
- Planck Picture Gallery - ESA 2018, Planck Picture Gallery, <https://www.cosmos.esa.int/web/planck/picture-gallery>
- Press W. H., Schechter P., 1974, Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation, *ApJ*, **187**, 425
- Robertson H. P., 1935, Kinematics and World-Structure, *ApJ*, **82**, 284
- Rubiño-Martín J. A., Betancort-Rijo J., Patiri S. G., 2008, A prescription for the conditional mass function of dark matter haloes, *MNRAS*, **386**, 2181
- Sheth R. K., Tormen G., 2002, An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier, *MNRAS*, **329**, 61
- Sheth R. K., Mo H. J., Tormen G., 2001, Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes, *MNRAS*, **323**, 1
- Spergel D. N., et al., 2007, Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology, *ApJS*, **170**, 377
- Tramonte D., Rubiño-Martín J. A., Betancort-Rijo J., Dalla Vecchia C., 2017, Testing the conditional mass function of dark matter haloes against numerical N-body simulations, *MNRAS*, **467**, 3424

BIBLIOGRAPHY

Trott C. R., et al., 2022, Kokkos 3: Programming model extensions for the exascale era, *IEEE Transactions on Parallel and Distributed Systems*, 33, 805

Walker A. G., 1937, On Milne's Theory of World-Structure, *Proceedings of the London Mathematical Society*, 42, 90

Appendix

Appendix A - Metrics

A metric is a mathematical tool that defines how distances and time intervals are measured in spacetime. The whole field of special relativity was built on the observation that the speed of light is always equal independently on which inertial frame the experiment takes place.

When moving from an inertial frame to another one according to special relativity, the following metric (Minkowski metric) is always true :

$$ds^2 = c^2 dt^2 - dx_1^2 - dx_2^2 - dx_3^2, \quad (6.1)$$

with c the speed of light, x_i the Cartesian coordinates and ds is the spacetime interval between two events.

This expression and the field of special relativity is valid at small scales but fails when trying to describe large-scale dynamics of the Universe like the expansion. When building a cosmological framework, the Minkowski metric needs to be replaced by a metric that takes into account the possibility of a dynamic & curved spacetime. General relativity provides this framework.

[Friedmann \(1922, 1924\)](#), then [Lemaître \(1927\)](#), [Robertson \(1935\)](#) and finally [Walker \(1937\)](#) derived the most general form of the spacetime metric compatible with the Cosmological Principle, named Friedmann-Lemaître-Robertson-Walker (FLRW) metric :

$$ds^2 = c^2 dt^2 - a(t)^2 \left(\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin(\theta)^2 d\phi^2 \right), \quad (6.2)$$

where (r, θ, ϕ) are the comoving spherical coordinates and k is the spatial curvature constant and may have take three different values that depends on the Universe shape :

- $k = 0$: The Universe is flat. No curvature.
- $k = 1$: The Universe is closed. Like the surface of a sphere, parallel lines eventually converge and the Universe is finite in size.
- $k = -1$: The Universe is open. Like a saddle, parallel lines diverge and the Universe is infinite.

Appendix B - Friedmann's derivation of the density parameters

From the cosmological principle and FLRW metric applied to the Einstein's equation, one can derive the Friedmann equations which describe how the Universe expands or contracts depending on its content and its geometry :

$$\frac{\dot{a}^2 + kc^2}{a^2} = \frac{8\pi G\rho + \Lambda c^2}{3}; \quad (6.3a)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + \frac{3p}{c^2} \right) + \frac{\Lambda c^2}{3}. \quad (6.3b)$$

where G is Newton's gravitational constant, Λ the cosmological constant and ρ is total energy density of the Universe (with matter, radiation...).

The first Friedmann equation (6.3a) explains how the expansion rate is determined by the contents and curvature of the Universe while the second equation (6.3b) explains how the Universe is accelerating or decelerating depending on the balance between attractive (matter and radiation) and repulsive (dark energy) components.

Finding constraints on the latter components will efficiently parametrize the cosmological model. By reordering equation (6.3a) :

$$-kc^2 = \dot{a}^2 \left(1 - \frac{8\pi G\rho}{3H^2} - \frac{\Lambda c^2}{3H^2} \right) \quad (6.4)$$

$$= \dot{a}^2 \left(1 - \frac{8\pi G\rho}{3H^2} (\rho_{matter} + \rho_{rad} + \rho_\Lambda) \right) \quad (6.5)$$

By defining the *critical density* as $\rho_c(z) = 3H^2(z)/8\pi G$, the dimensionless *density parameters* that quantify the relative contributions of the energy components to the total energy density of the Universe can be defined as :

$$\begin{cases} \Omega_m(z) = \frac{\rho_{matter}(z)}{\rho_c(z)} & \text{(matter)} \\ \Omega_r(z) = \frac{\rho_{rad}(z)}{\rho_c(z)} & \text{(radiation)} \\ \Omega_\Lambda(z) = \frac{\Lambda c^2}{3H^2(z)} & \text{(dark energy)} \\ \Omega_k(z) = -\frac{kc^2}{a^2 H^2(z)} & \text{(curvature)} \end{cases} \quad (6.6)$$

The first Friedmann equation (6.3a) can then be written using these density parameters:

$$\Omega_m(z) + \Omega_r(z) + \Omega_\Lambda(z) + \Omega_k(z) = 1. \quad (6.7)$$

This relation is valid for any value of the spatial curvature k and it describes how the total energy in the Universe is divided between its components.

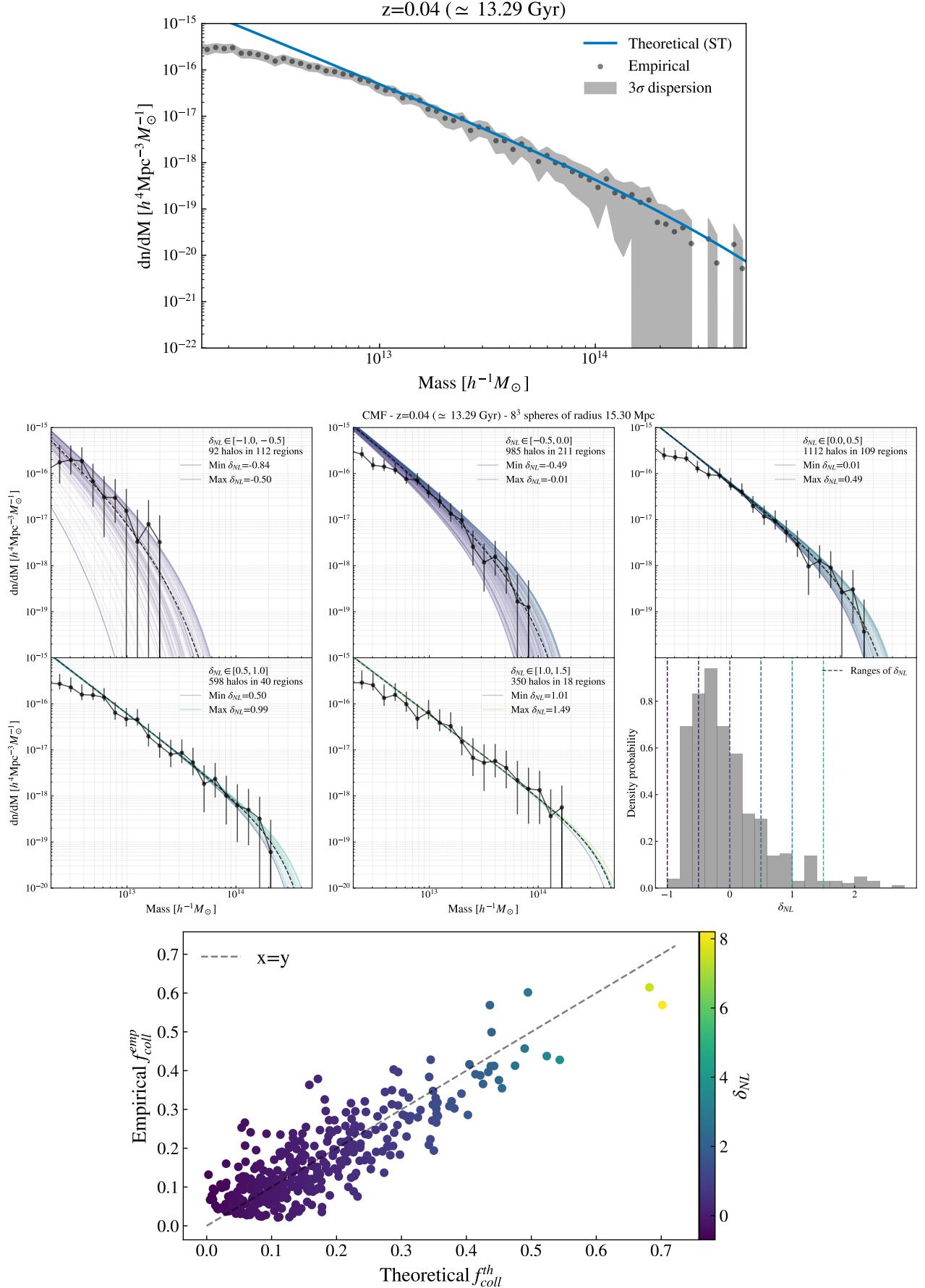
Appendix C - Results on DYABLO's structure formation at any redshift

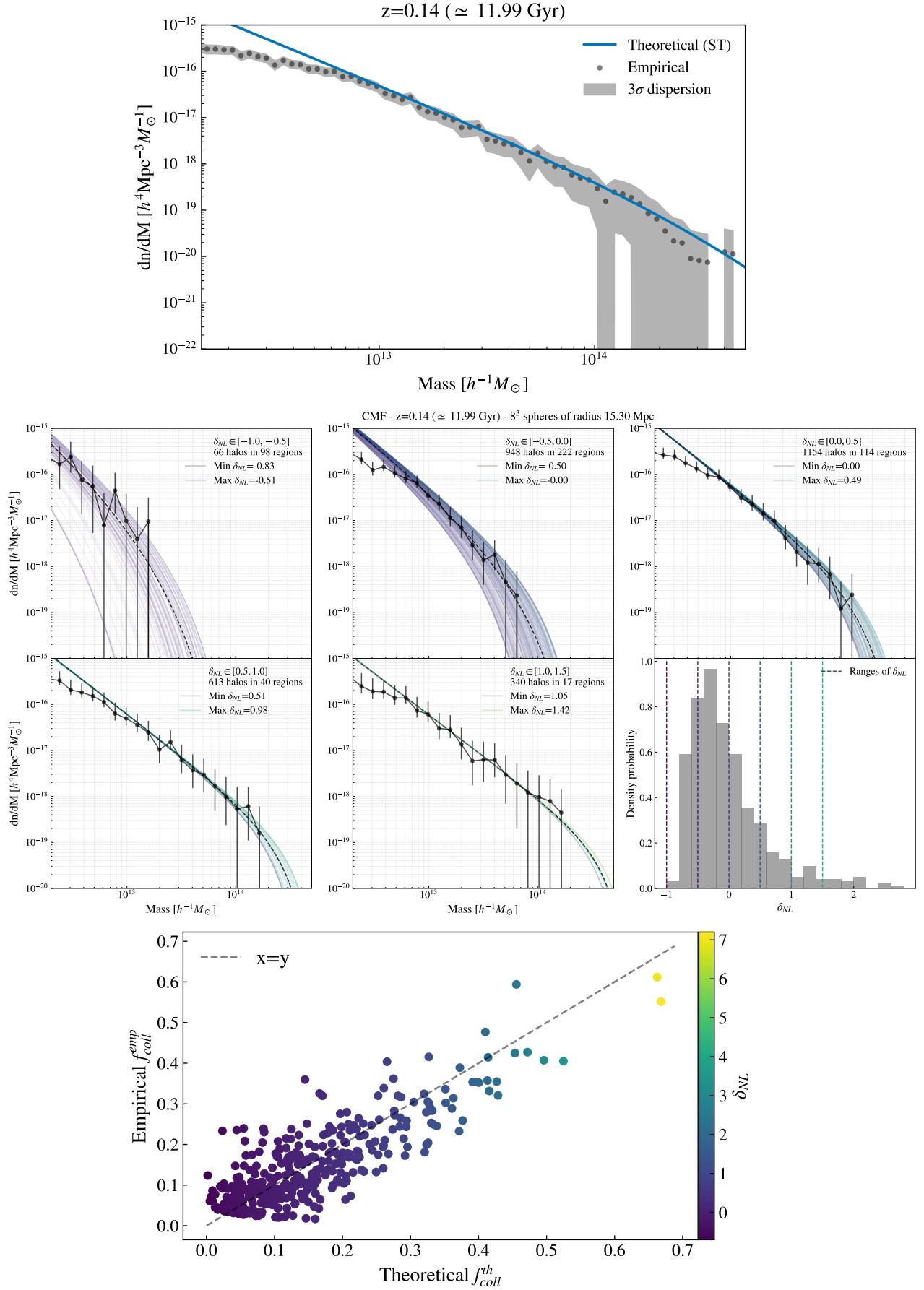
[Click to go back to the results about structure formation in DYABLO.](#)

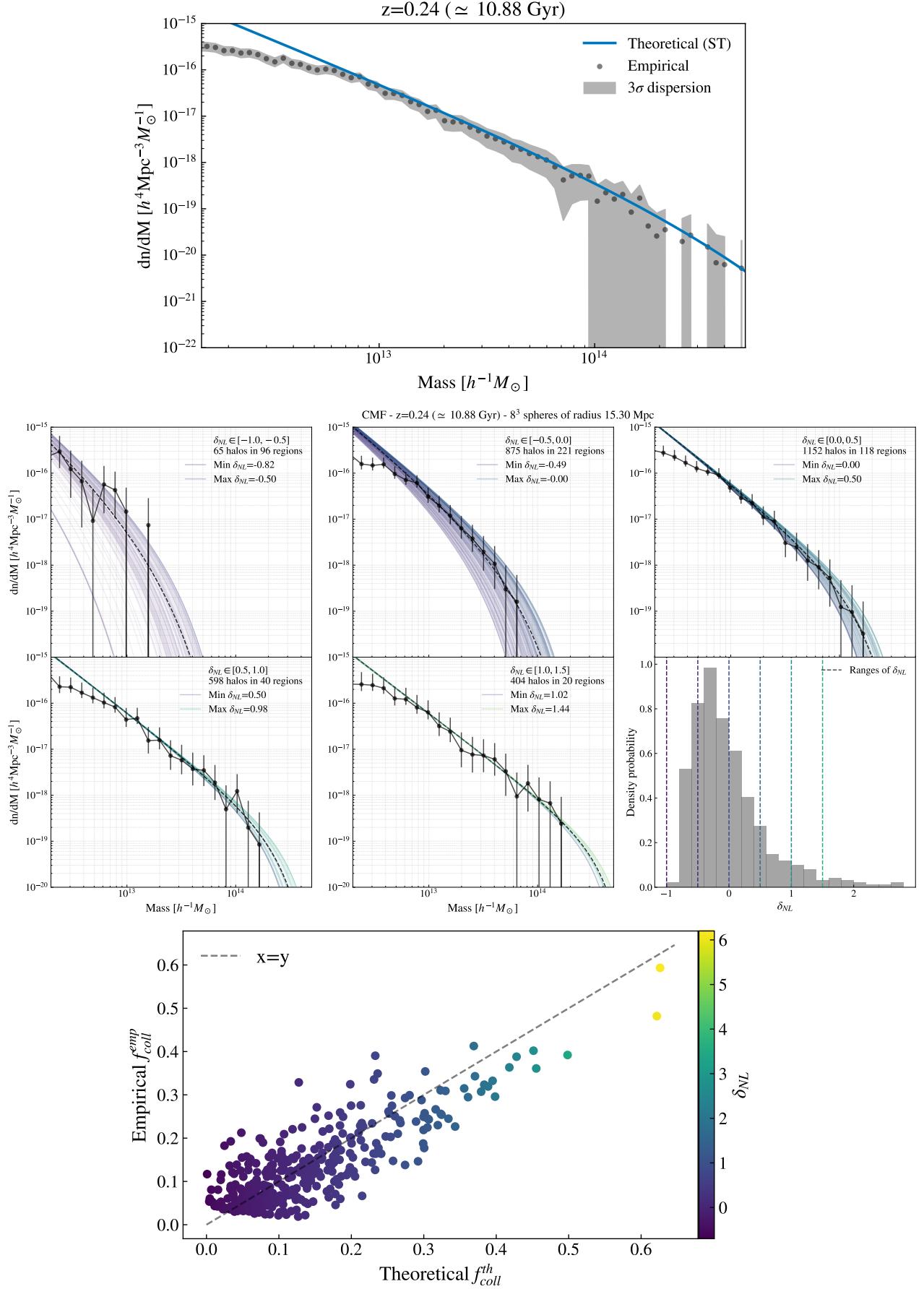
Each of the following page shows at a given redshift :

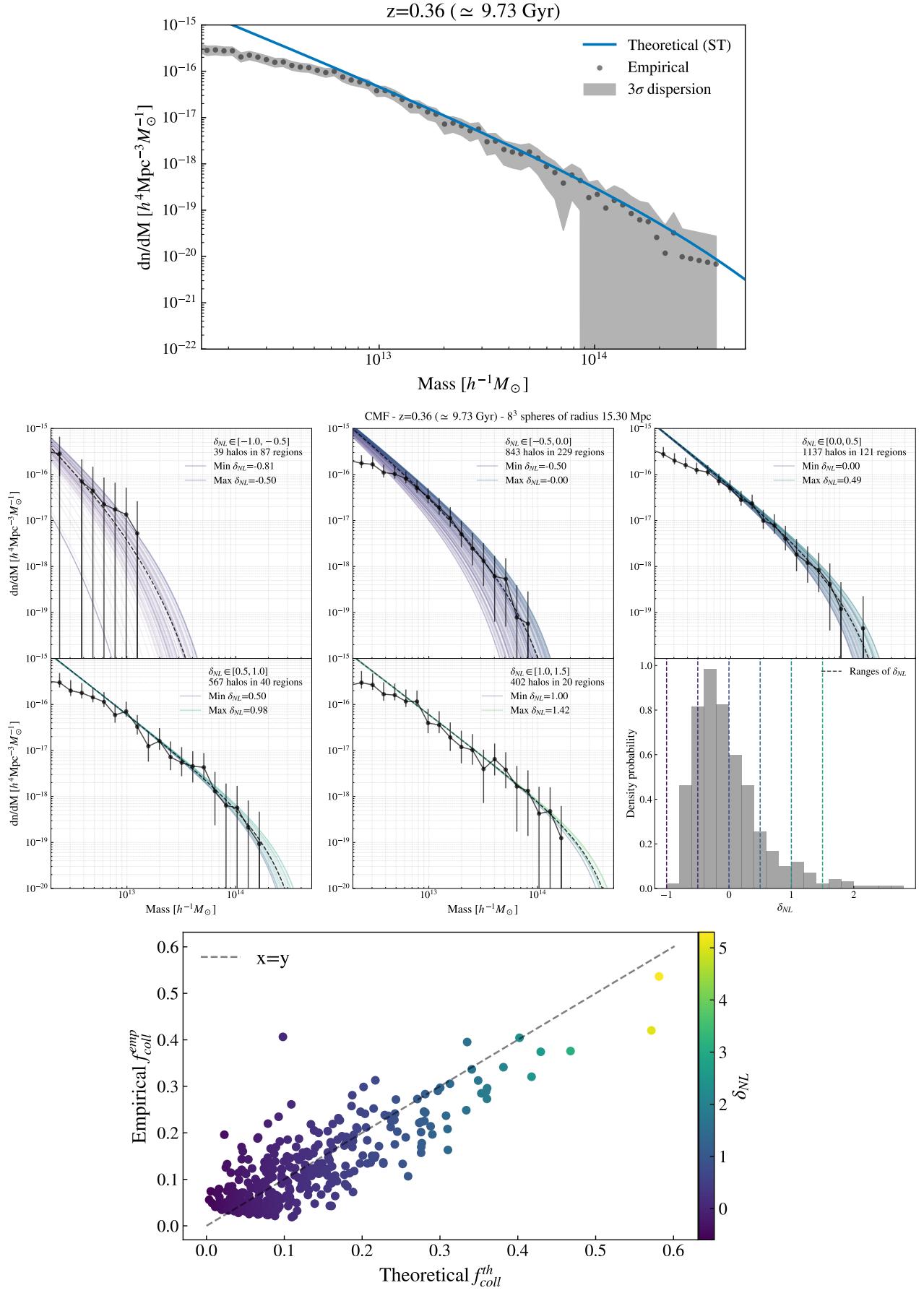
- Top plot : the Halo Mass function ;
- Middle plot : the Conditional Mass Function depending on the environment ;
- Bottom plot : the comparison between the empirical et theoretical fraction of collapsed mass.

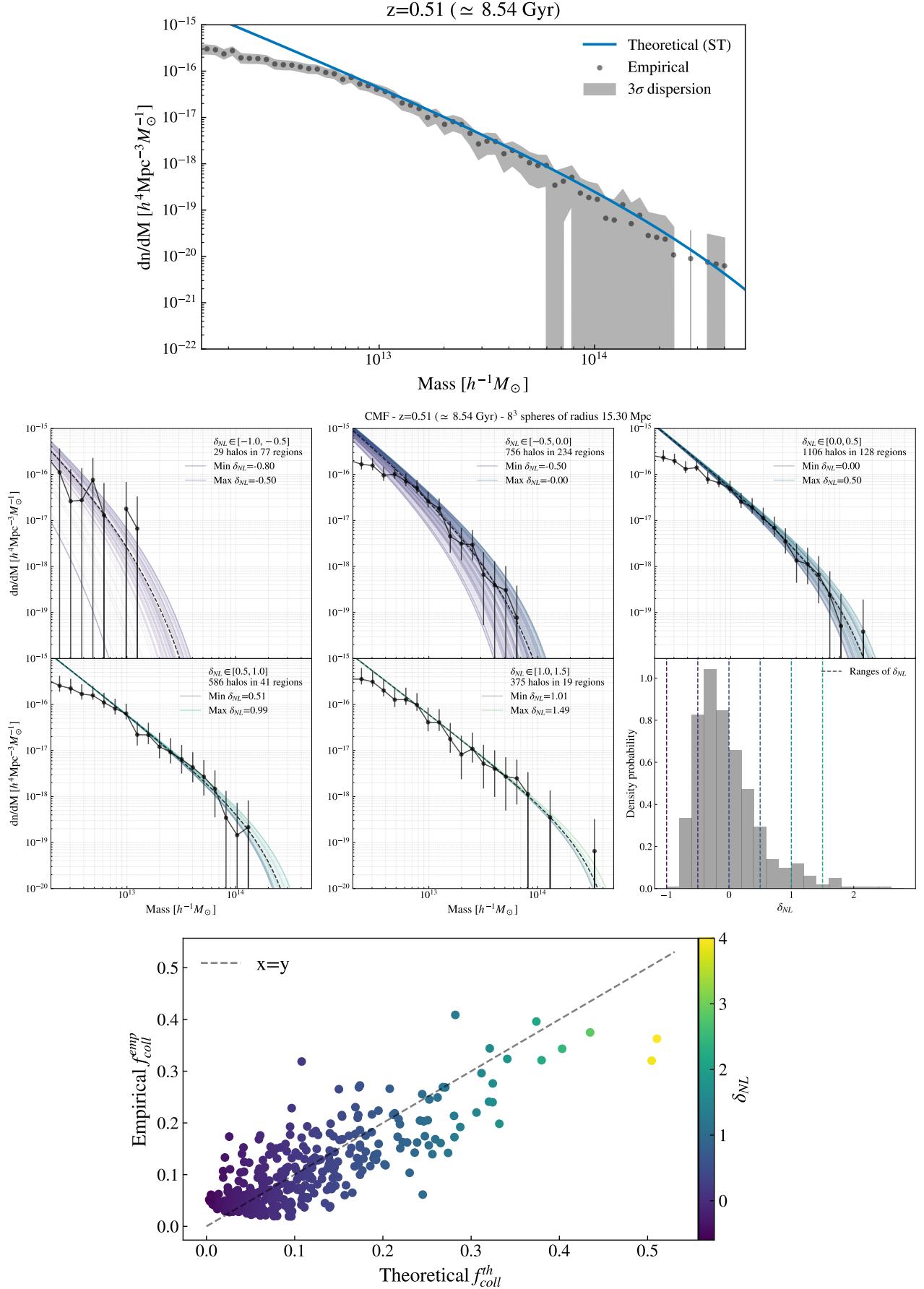
At high-z, structures do not have time to collapse significantly, which leads to inaccurate results with high variance.

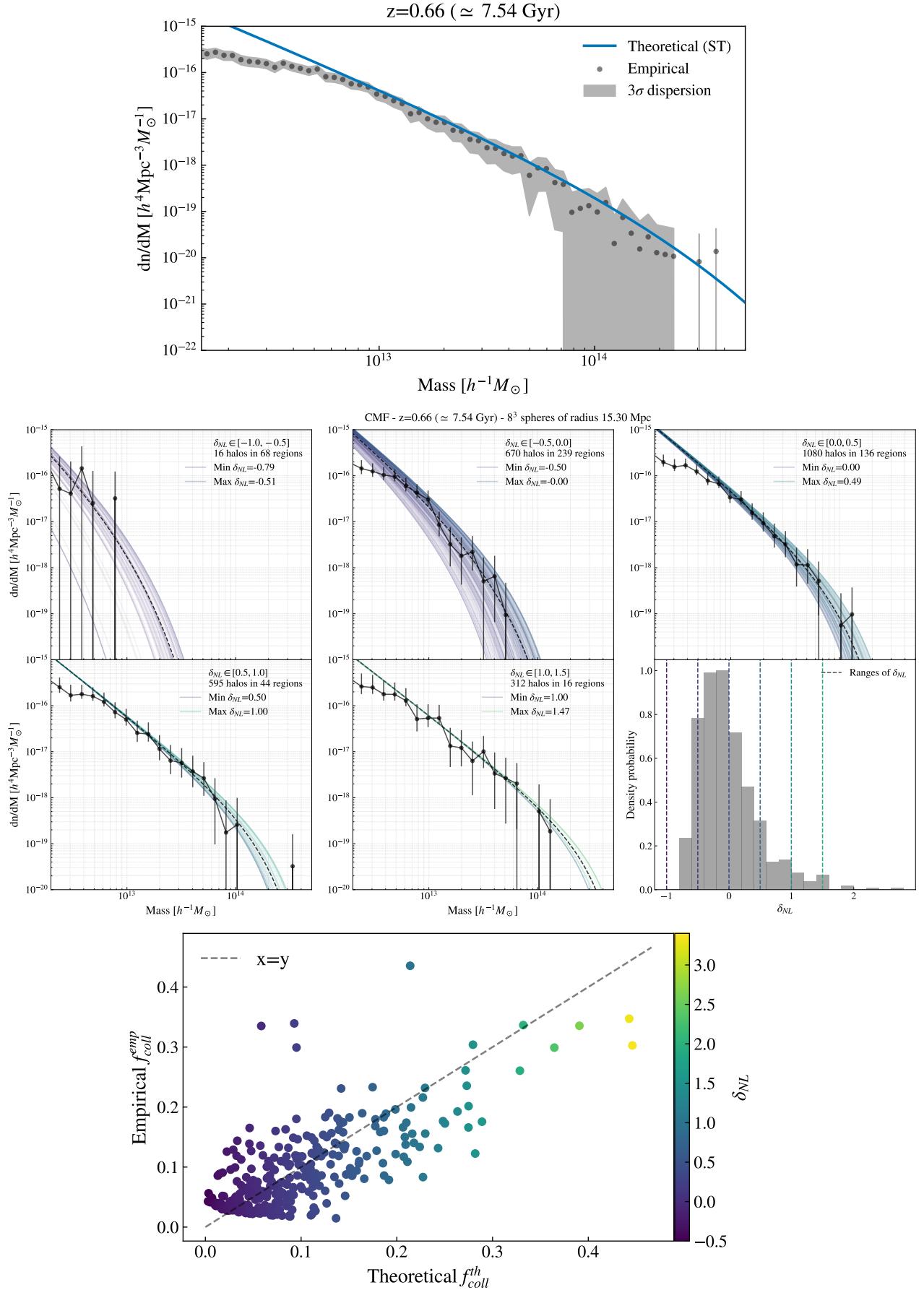
Figure 6.1: Results at redshift $z \approx 0.04$ (≈ 13.29 Gyr)

Figure 6.2: Results at redshift $z \approx 0.14 (\simeq 11.99 \text{ Gyr})$


 Figure 6.3: Results at redshift $z \approx 0.24 (\simeq 10.88 \text{ Gyr})$

Figure 6.4: Results at redshift $z \approx 0.36 (\simeq 9.73 \text{ Gyr})$

Figure 6.5: Results at redshift $z \approx 0.51$ (≈ 8.54 Gyr)

Figure 6.6: Results at redshift $z \approx 0.66 (\approx 7.54 \text{ Gyr})$

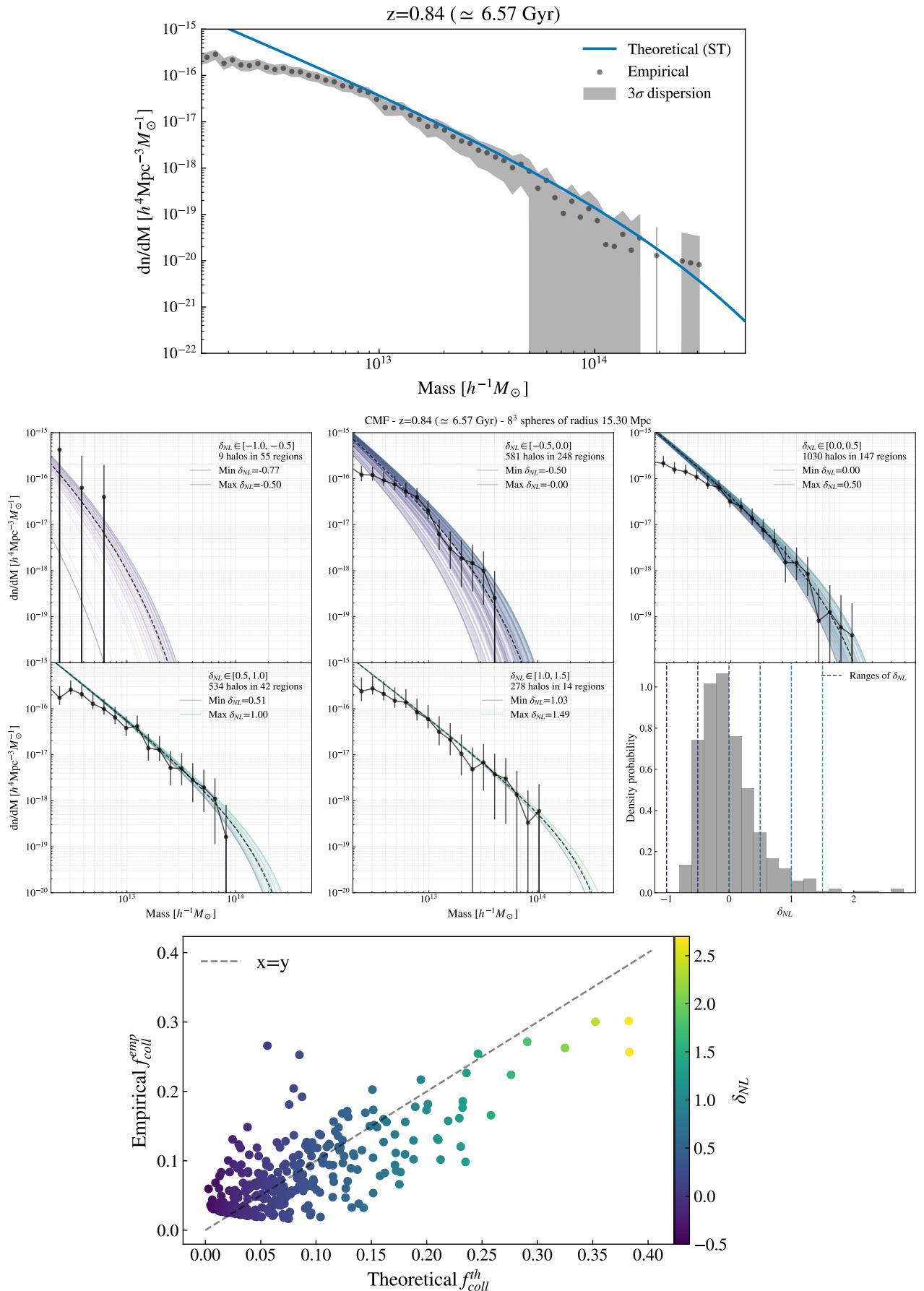
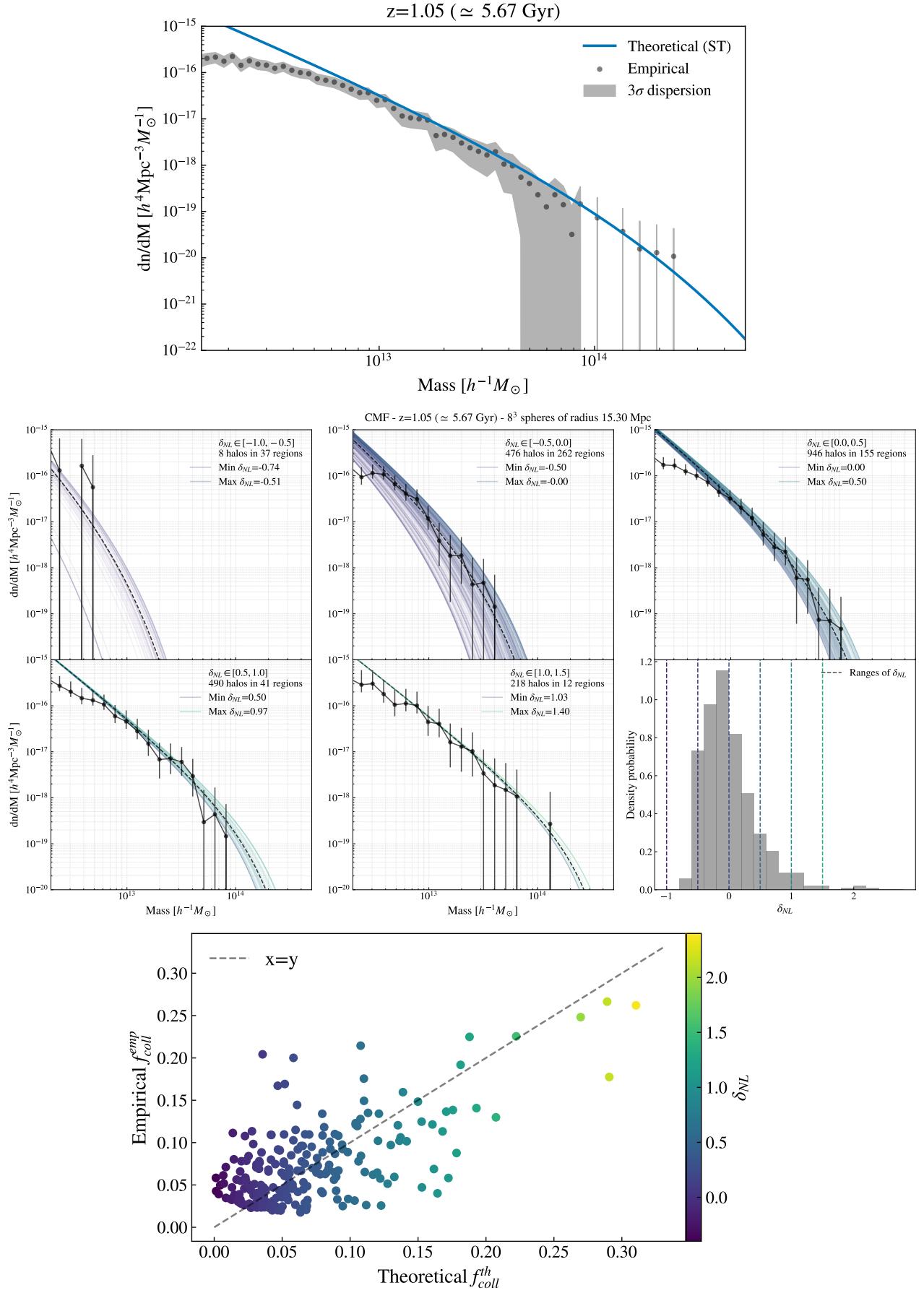
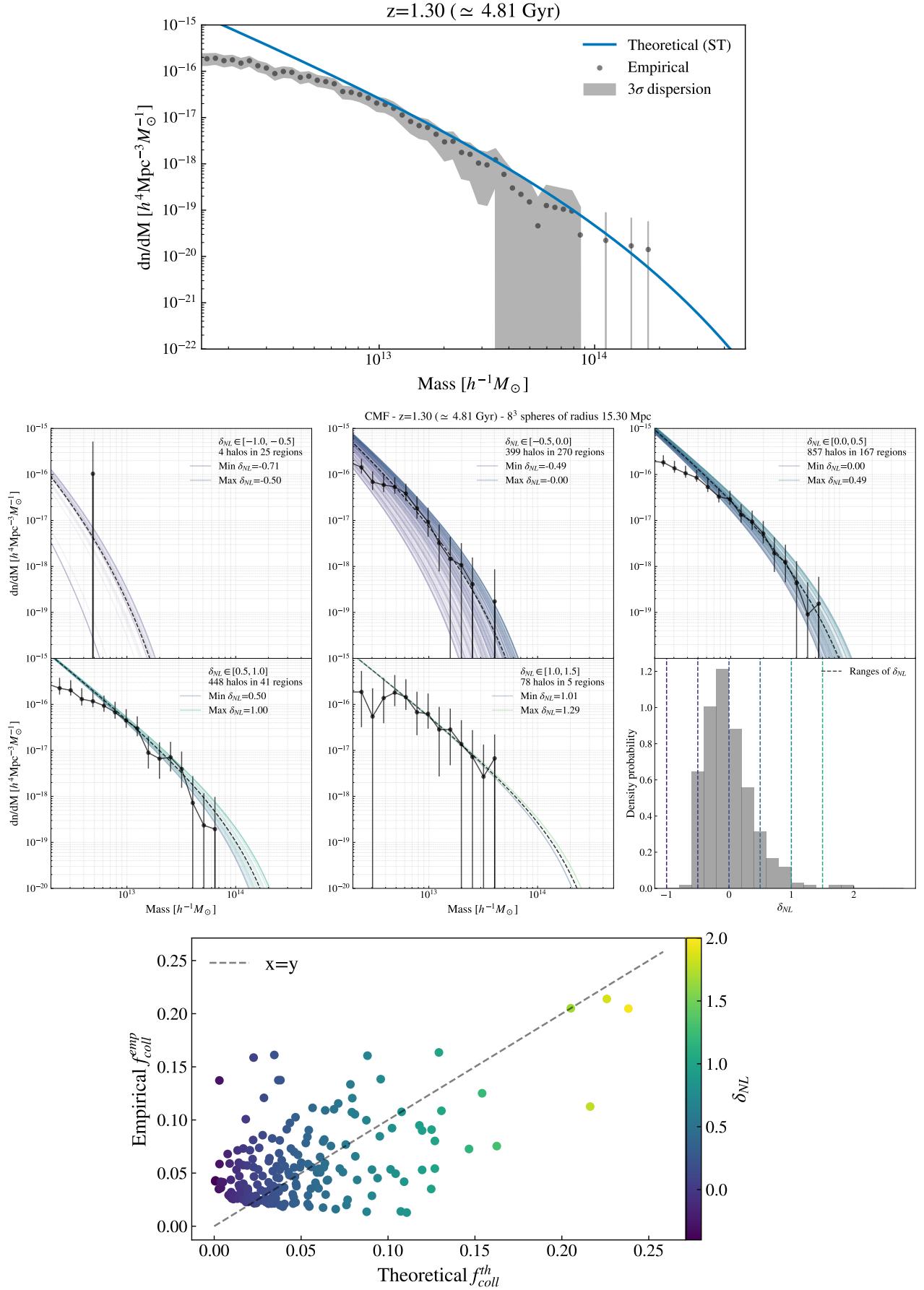
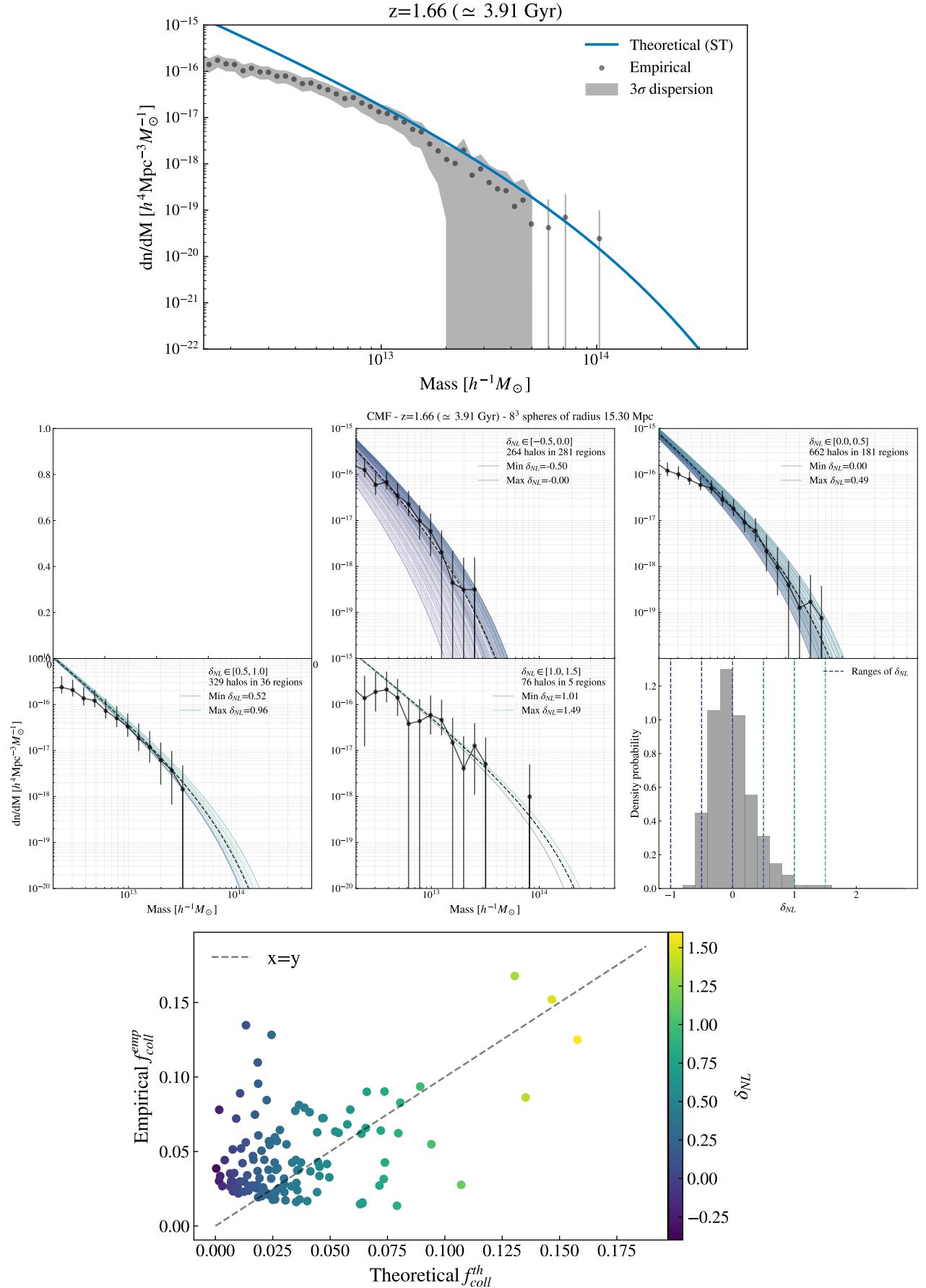
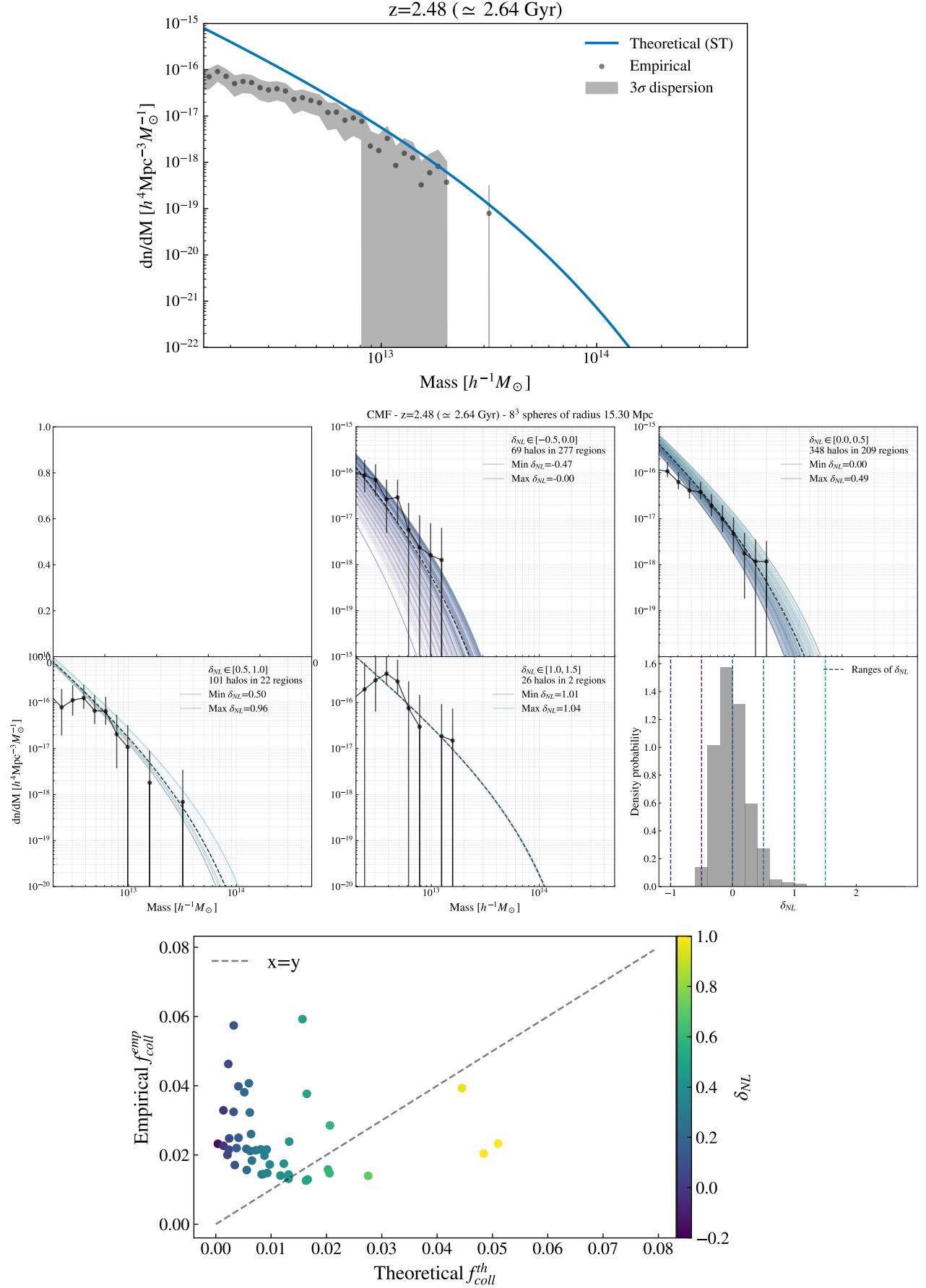


Figure 6.7: Results at redshift $z \simeq 0.84$ ($\simeq 6.57$ Gyr)

Figure 6.8: Results at redshift $z \approx 1.05 (\approx 5.67 \text{ Gyr})$


 Figure 6.9: Results at redshift $z \approx 1.30 (\simeq 4.81 \text{ Gyr})$

Figure 6.10: Results at redshift $z \approx 1.66 (\simeq 3.91 \text{ Gyr})$


 Figure 6.11: Results at redshift $z \approx 2.48 (\simeq 2.64 \text{ Gyr})$

Résumé & Abstract

Résumé

Développement et Validation de Modules de Physique pour le Nouveau Code de Simulation de Réionisation DYABLO

L'astrophysique moderne repose de plus en plus sur des simulations numériques pour tester des modèles théoriques et visualiser des processus complexes sans avoir recours à des expérimentations coûteuses ou temporellement longues. Dans ce contexte, le code de simulation DYABLO joue un rôle majeur. Il s'agit d'un code hydrodynamique haute performance qui modélise les fluides astrophysiques et étant conçu pour étudier l'Epoque de Réionisation en utilisant des superordinateurs à l'aide de grilles à raffinement de maillage adaptatif.

Une modélisation précise de la formation des structures et de la réionisation est cruciale pour comprendre l'évolution de l'Univers jeune. Ce travail se concentre sur ces deux aspects via DYABLO. Le premier est la validation du module de formation des structures qui a été codé mais pas encore validé. Pour évaluer sa validité, les fonctions de masse de formalismes théoriques connus de formation des structures (Press-Schechter & Sheth-Tormen) seront comparées aux résultats de simulations de DYABLO. Une seconde partie de ce travail vise à améliorer le module de réionisation en utilisant les résultats de la première partie. En effet, la validation de la formation des structures permet d'utiliser en toute confiance des cadres théoriques pour modéliser des phénomènes non résolus comme la formation des étoiles qui émettent des photons et ionisent l'environnement environnant, conduisant à une estimation plus précise du nombre de photons produits.

Ce projet a été un succès : la validation a confirmé la cohérence du module de formation des structures avec les attentes théoriques et nous avons pu utiliser ces résultats comme base pour améliorer les scénarios de réionisation produits par DYABLO.

Mots-clés : réionisation, calcul haute performance, formation des structures, simulations numériques

Abstract

Development and Validation of Physics Modules for the New Reionization Simulation Code DYABLO

Modern astrophysics increasingly relies on numerical simulations to test theoretical models and visualize complex processes without the need for expensive or time-consuming experimentation. In this context, the simulation code DYABLO plays a major role. It is a high-performance hydrodynamics code that models astrophysical fluids designed to study the Epoch of Reionization in supercomputing environments using adaptive mesh refinement grids.

Accurate modeling of structure formation and reionization is crucial to understand the early evolution of the Universe. This work has a focus on both of those aspects through DYABLO. The first is the validation of the structure formation module which has been coded but not yet validated. To assess its validity, the mass functions of known structure formation theoretical formalisms (Press-Schechter & Sheth-Tormen) will be compared to outputs of DYABLO. A second part of this work aims at improving the reionization module using the results from the previous part. Indeed, the validation of the formation of structures allows to confidently use theoretical frameworks to model unresolved phenomena like the formation of stars which emit photons and ionize the surrounding environment leading to a more accurate estimate of the number of photons produced.

This project has been successful : the validation confirmed the consistency of the structure formation module with theoretical expectations and we were able to use those results as a basis to improve reionization scenarios produced by DYABLO.

Keywords : reionization, high performance computing, structure formation, numerical simulations