Scuola Internazionale Superiore di Studi Avanzati - Trieste

COSMIC NEUTRAL HYDROGEN AS TRACER OF THE LARGE SCALE STRUCTURE OF THE UNIVERSE

Author:
Isabella Paola Carucci

Supervisors:
Matteo Viel
Andrea Lapi

A thesis submitted in fulfilment of the requirements for the degree of Doctor Philosophiae

in

ASTROPHYSICS

October 2017

SISSA - Via Bonomea 265 - 34136 TRIESTE - ITALY

COSMIC NEUTRAL HYDROGEN AS TRACER OF THE LARGE SCALE STRUCTURE OF THE UNIVERSE

Author:
Isabella Paola Carucci

Supervisors:
Matteo Viel
Andrea Lapi

A thesis submitted in fulfilment of the requirements for the degree of Doctor Philosophiae

in

ASTROPHYSICS

Trieste, October 2017

Dedico questa tesi a Trieste:
dalle insondabili pedocere della Costa,
su su per il Carso, fin lo svettante Nanos

e a chi mi ha accompagnata in questo percorso,
in particolare alla Cumpa, ad Ana,
a Nicola, a Carolina e a Donato.

FOREWORD

This thesis summarises part of the research I carried out during my Ph.D. in collaboration with Matteo Viel, Francisco Villaescusa-Navarro, Pier Stefano Corasaniti and Andrea Lapi.

A significant part of the work presented has already appeared in

i) Carucci et al., 2015 [37],

ii) Carucci, Villaescusa-Navarro, and Viel, 2017 [35],

iii) Carucci, Corasaniti, and Viel, 2017 [34].

The numerical resources employed are acknowledged in the above publications.

My research has also lead to other published works that are not included in this thesis: Cardoso et al., 2013 [31], Cardoso et al., 2013 [32] and Carucci et al., 2014 [36].

Trieste, October 2017

SUMMARY

The nature of the most abundant components of the Universe, dark energy and dark matter, is still to be uncovered. The study of the matter distribution in the Universe and of its clustering properties may reveal key insight on the origin of these phenomena. However, matter is not directly observable, but can be mapped through tracers.

In this study we consider neutral hydrogen, which pervades space from the time of recombination up to present day. It has a characteristic line emission at around 21cm that being redshifted gives a measure of comic distance and makes it possible to reconstruct the three-dimensional density field over a wide range of redshift and scales. In the upcoming years a new generation of experiments will map the distribution of neutral hydrogen from unresolved galaxies up to the high-redshift Universe, by employing a technique called intensity mapping.

In this thesis we assess the potential of the 21cm intensity maps for constraining cosmological models. We analyse competitive and realistic dark energy and dark matter models with state-of-the-art numerical simulations and show how different cosmologies produce distinctive and detectable effects in the 21cm signal. We compute radio telescope forecasts showing how these models will be distinguishable in an unprecedented way.

The drawback of 21cm observations is the intrinsic weakness of the signal compared to the expected large foregrounds. Motivated by this, we study the cross-correlation signal between 21cm intensity maps and the transmitted Lyman-α forest flux, representing a powerful way to extract more information, by isolating instrumental and astrophysical systematics.

Thesis outline

This thesis is organised as follows. In Chapter 1 we sketch the overall picture for motivating this study and putting it into context, while in Chapter 2 there is a simple illustration of some statistical tools we encounter throughout the thesis.

In Chapter 3 we present and compare the modelling methods we employ for distributing neutral hydrogen in our cosmological simulations, together with discussing the current available observational data we can rely on.

Chapter 4 and Chapter 5 address the potentiality that 21cm intensity mapping surveys would have for constraining cosmology and unveiling the nature of dark energy and dark matter. We start in

Chapter 4, where we present hydrodynamic simulations of warm dark matter (WDM) cosmologies, study the effect of WDM on matter and eventually on the 21cm signal at redshifts $z = 3 - 5$. We quantify this effect and check its detectability with forthcoming radio observations. In Chapter 5 we extend this analysis to other dark matter models and dynamical dark energy models, at redshifts $z = 1 - 2$. Both in Chapter 4 and Chapter 5 the goal is to consider alternative but realistic cosmological models and show how 21cm observations can remarkably tell them apart.

Chapter 6 is devoted to illustrate how we model the Lyman-α forest in simulations and it is preparatory for next Chapter 7, where we investigate the cross-correlation signal between 21cm and Lyα and show what kind of advantages such study could give for inferring the parameters and shrinking their uncertainties, for identifying systematics in the 21cm maps and for testing how much we can rely on linear theory for the modelling of the observed fields.

Finally, we summarise the main results and future perspectives in Chapter 8.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Tiger got to hunt, bird got to fly;
Man got to sit and wonder, "Why, why why?"
Tiger got to sleep, bird got to land;
Man got to tell himself he understand.

— Kurt Vonnegut [164]

In this thesis I illustrate my contribution to the study of the large scale structure of the Universe. I focus on the information we can gain by observing the radiation emitted by neutral hydrogen using a technique called intensity mapping. What do we mean by large scale structure? And why detecting neutral hydrogen is a good idea for describing it? This introduction is meant to answer these questions and to give a general overview before entering into the details of my work.

1.1 THE LARGE SCALE STRUCTURE OF THE UNIVERSE

The Universe is remarkably structured: the arrangement of clusters, galaxies and clouds reveals us walls, voids and filaments, all contributing to what we call the large scale structure of the Universe. This field has been one of the most dynamic of fundamental science since *i)* a great wealth of observational data has been and will be produced, and *ii)* a theoretical framework has emerged, the ΛCDM cosmological model, for understanding these observations.

Indeed, ΛCDM has been shown to be very successful in explaining a wide variety of observables: the anisotropies in the cosmic microwave background (CMB), the clustering of galaxies at low redshift, the abundance of galaxy clusters, the baryonic acoustic oscillations (BAO) observed both in the spatial distribution of galaxies and in the Lyman-α forest. However, this success is linked to the existence of 2 dark components.

As shown in Figure 1.1, while ordinary matter[1] contributes just 4.9% of the Universe mass-energy inventory, we have no understanding of the precise nature of the major ingredients: dark energy (68.3%) and dark matter (26.8%). In next sections, we motivate the need for the existence of these dark components, and we summarise what we know so far about them.

1 Also called *baryonic* matter, i.e. the fields of Standard Model: all that makes up stars, galaxies but also the Earth and us.

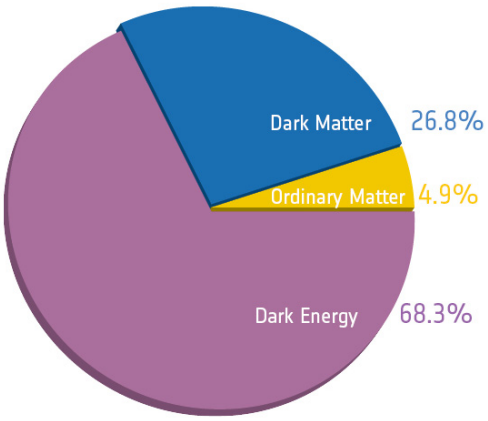


Figure 1.1: Planck high precision CMB map [120] has allowed to extract the most refined values yet of the Universe components (assuming the ΛCDM cosmological model).

1.1.1 *Dark energy*

Since the studies on supernovae IA of almost 20 years ago [119, 128], we know that the Universe is not only expanding (that was an idea that Hubble and more entertained already in the 1920s) but its expansion is accelerating. Whatever is the thing or the mechanism that pushes the Universe to ever faster expansion, we call it dark energy (DE).

The simplest DE scenario (the one adopted by the ΛCDM standard model) consists of a positive cosmological constant Λ in Einstein's equation of general relativity: a constant energy density filling space homogeneously [33]. What is the physical nature of Λ? Could the DE phenomenon be of complete different origin? At present, answering these questions is one of the most striking problem in physics. Moreover, by employing Λ we are left with some issue [167], mainly:

IN THE CONTEXT OF QUANTUM FIELD THEORY Λ can be naturally interpreted as the energy contribution of quantum vacuum fluctuations; however, the observed value for Λ is roughly 119 orders of magnitude lower than what Standard Models fields account for[2] .

A COINCIDENCE PROBLEM Λ is comparable to the present matter density (despite having extremely different dependence on the scale factor): its value must have been fixed very precisely at

2 A similar result can be obtained also by employing the supersymmetric theory.

early times in order to assure a structure formation period long enough to have distributed matter as we observe today[3].

A substantial variety of DE models beyond the cosmological constant have been invoked to solve either one or both of these problems (see [43] for a review). We find models in which a dynamical field relaxes to the value of Λ via some mechanism, as the ones we consider in Chapter 5. Another viable possibility is to discard the DE hypothesis and look at modifications of the gravity theory itself, as for instance scalar-tensor theories that effectively make the gravitational coupling G varying instead of being constant[4].

1.1.2 *Dark matter*

Dark matter (DM) interacts with ordinary matter just gravitationally, it does not interact via electromagnetism and thus neither emits nor reflects light: we have not been able to observe it directly. Evidences of its interactions via gravity come through different observational techniques, spanning a remarkably large range of scales:

AT THE HORIZON SCALE The detection of the power spectrum of the CMB shows temperature fluctuations $\Delta T/T \sim 10^{-5}$ [120], that are 2 orders of magnitude smaller than the value we would expect tracing back present day structures[5].

AT LARGE SCALES Spectroscopic redshift surveys measure the spatial distribution of galaxies and how this distribution evolves in time [150]. The outcomes are well fit with the hypothesis of cold DM and with the idea that structure formation proceeds hierarchically, by merging of smaller into larger objects.

AT CLUSTERS OF GALAXIES SCALE We invoke DM to sum up the mass distribution we measure through *i)* the motions and mass of individual galaxies [143, 174], *ii)* the distortion of the images of the background galaxies due to dense cluster cores (gravitational lensing, e.g. [44, 92]) and *iii)* the measurements of the hot intra-cluster gas temperature, assuming hydrostatic equilibrium [131].

AT GALAXIES SCALE Rotation curves of stars in galaxies behaves as a DM component surrounds the system [129].

3 This problem is solved if we assume the anthropic principle, i.e. observations of the Universe must be compatible with the conscious and sapient life that observes it.

4 This is the class of gravity theories that I consider in my works [31, 32], whose results I do not include in this thesis.

5 If we account for DM, that did not interact electromagnetically with the CMB photons, then structures started forming before CMB seeped out.

In brief, DM is the most abundant gravitating matter, thus drives the formation of all visible structures. By mapping it with the different techniques shown above, we find a loose network of filaments, growing over time, which intersect in massive structures at the locations of clusters of galaxies. This picture is consistent with predictions of gravitationally induced structure formations (from the Millennium run [146] of 2005, to which have followed a series of ever more sophisticated simulations), in which the initial, smooth distribution of DM collapses into filaments then into clusters, forming a gravitational scaffold into which gas can accumulate, and stars can be built.

The physical nature of DM is still to be uncovered. The opportunity is seized by particle physicists, who have been proposing a zoo of particles acting DM-like (see [60] for a review). However, no detection of DM particle candidates in Earth laboratories has yet been made, or ultimate annihilation or decaying signal being detected [26].

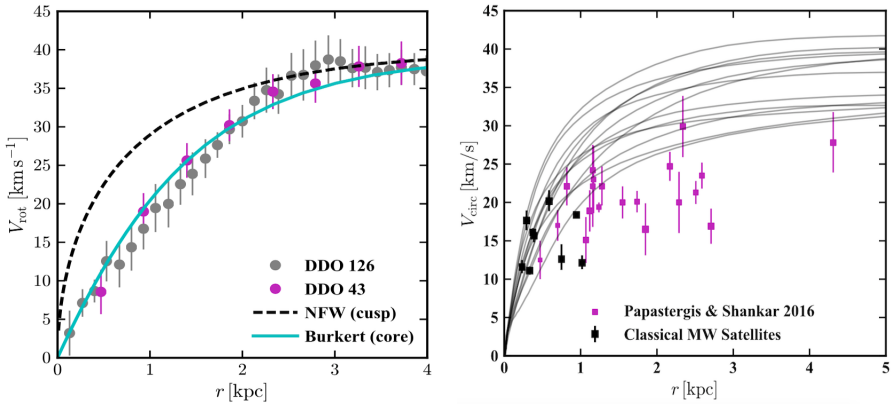
The standard model postulates DM to be cold (CDM): weakly interacting particles whose velocity dispersion in the early Universe was too small to erase structure on a galactic or sub-galactic scale. As just said, the theory has been very successful in explaining cosmic structure over a remarkable span of redshift [18, 52, 117], but in the last $\sim$ 25 years the CDM predictions have been challenged by an escalation of small scale observations [166]. We can summarise these discrepancies in 3 main issues (see [28] for a more complete list):

MISSING SATELLITES Because CDM preserves primordial fluctuations down to very small scales, simulated halos retain a large amount of substructure, predicting hundreds of sub-halos in contrast to the $\sim$ 10 satellites of the Milky Way [76, 101].

CUSP-CORE PROBLEM CDM collapse leads to denser halos (in normalisation) and more cuspy profiles (in inferred density profile slope) than what derived from rotation curves of DM dominated galaxies (e.g. [61, 100, 104]), see Figure 1.2a.

TOO-BIG-TO-FAIL The local Universe contains too few galaxies with central densities indicative of halos of virial mass $M_{vir} \simeq 10^{10} M_{\odot}$: halos of this mass are generally believed to be too massive to have failed to form stars, so the fact that they are missing is hard to understand [22, 114, 154], see Figure 1.2b.

Part of the solution of these conflicts lies in baryonic physics: for the missing number of satellites it could be that galaxy formation becomes increasingly inefficient as the halo mass drops, and for the cusp-core issue it has been claimed that baryonic processes (e.g. gas cooling, star formation, supernovae and active galactic nuclei feedbacks) alter the structure of dark matter halos (e.g. [107, 121, 127]). However, these findings rely on hydrodynamic simulations only and, more importantly, for very low mass galaxies and for field dwarf



(a) CDM expectation for a galaxy rotation curve of a $V_{max} \simeq 40$ km/s (dashed black line): it rises quickly, reflecting a central density profile that rises as a cusp with $\rho(r) \propto 1/r$. The data points show the rotation curves of 2 example galaxies of the same size [110], which are more slowly rising and better fit by a density profile with a constant density core ([29], cyan line).

(b) Data points show the circular velocities of Milky Way satellite galaxies (black) and field dwarfs galaxies (magenta). The grey curves are predictions for CDM halos from fully self consistent hydrodynamic simulations that span the same stellar mass range as the observed galaxies. Details of data and simulations in [28].

Figure 1.2: Plots taken from [28]. On the left for the cusp-core problem, on the right for the too-big-to-fail.

galaxies (i.e. those mostly affected by the too-big-to-fail problem) the impact of baryonic feedback is actually found to be negligible [28, 166].

Alternatively, these small scale issues could point towards the need of changing the cold DM hypothesis itself. The ideal candidates are models presenting dissipative effects that can suppress the amplitude of small density fluctuations, fully retaining the success of CDM at large scales and alleviating the small scale issues summarised above. In this thesis, we explore a series of viable models, starting in Chapter 4 from the vanilla possibility of making DM *warm*, i.e. with free-streaming velocities in the early Universe large enough to erase the small scale primordial fluctuations, and extending the analysis in Chapter 5 with models characterised by non-thermal distribution functions. The bottom line is that these scenarios are (and need to be) indistinguishable from CDM at middle and large scales, hence being hard to be tested. We will show that cosmic neutral hydrogen observations could play a fundamental role for discriminating these models.

Concluding this section, the big picture is the following: DE is responsible for the accelerated expansion of the Universe (at late times), while initial fluctuations in the DM density field grow via gravitational instabilities to form DM halos, and galaxies are believed to form within these density peaks once they are sufficiently massive

for the baryons to be able to cool and condense. DE and DM are fundamental for our understanding, yet we directly observe galaxies only.

1.1.3 *Why do we need tracers?*

Effects of both DE and DM are embedded in the spatial distribution of matter in the Universe, hence a way to constrain the cosmological model is measuring the statistical properties of the matter distribution and compare them against predictions of the theory. The problem resides in the fact that the DM cosmic web is not directly observable. Our knowledge depends on the spatial distribution of its tracers: galaxy clustering, abundance of galaxy clusters, Lyman-α forest and so on. These observables indeed trace matter, but each of them in a differently biased way, that it is crucial to study. Main subject of this thesis is cosmic neutral hydrogen observations and, especially in Chapter 7, we focus on delineating its biased relation to the underlying matter density field.

The primary statistical tool used to characterised the spatial properties of the tracers and, eventually, of the matter field is the power spectrum[6]. It describes the clustering of matter as function of scale[7].

In Figure 1.3 we see the comparison between the linear matter power spectrum as predicted by ΛCDM and some data points that refer to a variety of observables [8]: the agreement is remarkable. Different observables contribute to the overall knowledge of matter distribution, and eventually cosmology, by sampling the matter power spectrum at different scales. In next sections (and more broadly in the whole thesis) we show how intensity maps of the 21cm radiation by neutral hydrogen will contribute to this picture, potentially revolutionising the field of cosmological observations.

1.2 NEUTRAL ATOMIC HYDROGEN

Neutral atomic hydrogen (HI) is the protagonist of this thesis. We later see why it will be crucial for cosmology, here we start by asking how much HI is there in the Universe.

---

6 In this thesis we use it extensively, in Chapter 2 we formally define it and briefly describe how we calculate it.

7 On large scales, gravity competes with cosmic expansion, and structures grow according to linear theory. In this regime, the density contrast field is Gaussian, Fourier modes evolve independently, and the power spectrum is sufficient to completely describe the density field. On small scales, gravitational collapse is non-linear, and can be computed accurately using numerical simulations, as indeed we do in this thesis.

8 This figure is old, as it is of 2002 [149]. At present this curve could be more sampled and points would have smaller error bars. We choose this particular plot for clearness and simplicity sake.

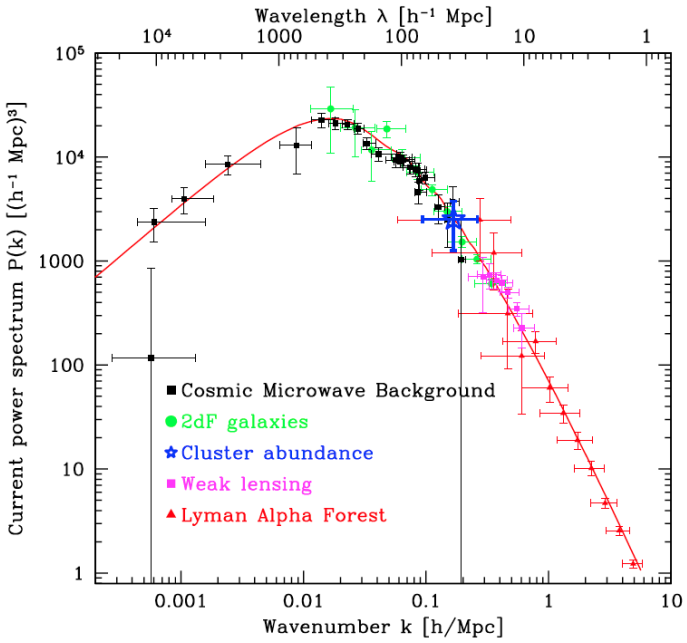


Figure 1.3: The linear matter power spectrum P(k) predicted by the concordance ΛCDM model (solid curve), i.e. not a fit to the observational measurements shown. Data catalogues and details in [149], where the plot is taken from.

In this introduction we already came across the CMB, the cosmic microwave photons that could seep away from the first hot and dense phase when the Universe starts cooling down and expanding, according to the Hot Big Bang theory. This epoch is called recombination: protons and electrons are able to recombine and to form neutral hydrogen atoms. Hence, after recombination the gas is left in a cold and neutral state. Tiny density fluctuations later start growing under the action of gravity, collapsing into bound objects inside which stars and galaxies eventually form [41]. These objects start emitting ultraviolet (UV) radiation that carves out ionized regions around them. After a sufficient number of ionizing sources have formed, the ionized fraction of the gas in the Universe rapidly increases until hydrogen becomes fully ionized. This period, during which the cosmic gas goes from neutral to ionized, is known as the Universe's Epoch of Reionization (EoR), that takes place in the redshift range $6 \lesssim z \lesssim 10$ [13]. Understanding how EoR is at play is one of the primary goals of present-day cosmology (a review in [172]). This thesis is not aimed at doing so as it focuses on the post-reionization Universe, however we do encounter EoR discussions throughout, mainly because EoR sets the total amount of neutral hydrogen left in the Universe.

Indeed, after EoR the bulk of surviving HI is thought to reside in in pockets of high matter density regions, where it is protected from the ionizing background since clouds which are predominantly

composed of HI (above column density of $N_{HI} > 2 \times 10^{20}$ atoms cm^{-2} [169]) must absorb, in an outer layer, the majority of incident photons capable of ionizing hydrogen ($h\nu > 13.6$ eV): this mechanism is called self-shielding. HI can be used as a tracer of the matter distribution of the Universe and is detectable either in absorption or in emission.

1.2.1 *Signal in emission: the 21cm radiation*

HI in its ground state consists of an electron bound to a proton. Both the electron and the proton have intrinsic magnetic dipole moments (spins), that can be either parallel or anti-parallel. The parallel state, populated by collisions, has a slightly higher energy, so that a transition to the anti–parallel state produces radiation at a wavelength of 21cm (a frequency of 1420 MHz). However, this *spin-flip* transition is highly forbidden with an extremely small transition rate of 2.9×10^{-15} s^{-1} that corresponds to around 10 million years for spontaneous emission, i.e. very unlikely to be detected in a laboratory on Earth. The history of 21cm observations is indeed uniquely astrophysical! Due to the vast amount of hydrogen in the Universe as well as the efficiency of collisions, the probability of observing the 21cm emitted photons greatly increases and it is indeed one of the most important astrophysical probes [58], providing detailed information on the gas distribution and its relation to star formation and gas kinematics.

Unfortunately, this knowledge is limited to the local Universe galaxies ($z \ll 1$): the weakness of the 21cm transition implies that extremely large telescopes would be needed to carry out such studies in high-z galaxies[9]. This is why gas-rich galaxies at high redshifts are most easily detected through their absorption signatures in the optical spectra of background quasar, as we will see below and more in detail in Section 3.3.

But another possibility of exploiting the 21cm signal consists of giving up the detection of individual galaxies and starting to profit from the large scale and high redshift potential of HI thanks to *intensity mapping*.

1.2.1.1 *The intensity mapping technique*

As just discussed, the HI emission from individual galaxies is too weak to be measured efficiently, thus a technique called intensity mapping (IM) has been proposed: measuring the total 21cm radiation over large angular scales without needing to resolve the individual sources, i.e. integrating all the radiation coming to single large pixels [11, 14, 15, 27, 39, 84, 95] . This results in a map of the large scale fluctuations in 21cm intensity, similar to a CMB map or better

9 For example, in [73] they calculate that detecting a single $z \sim 2.5$ galaxy with an optimistic HI mass $M_{HI} \sim 6.5 \times 10^9 M_\odot$ would require ~ 360 hours with the SKA telescope.

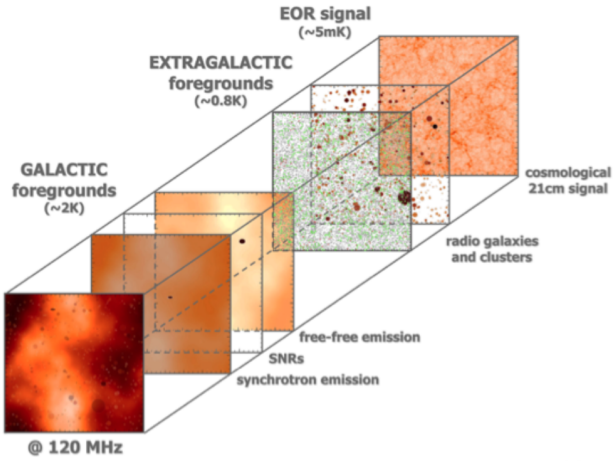


Figure 1.4: Image taken from [172] showing the various cosmological and galactic components that contribute to the measured signal at a given frequency. It refers to EoR signal, but same foregrounds are at play for the signal coming from the redshifts analysed in this thesis.

to *many* CMB maps: the signal depends on redshift too, hence the encoded information is 3D. These experiments will be faster than usual galaxy surveys, since there will be no need to resolve discrete objects. Still comparing with galaxy surveys: the collected data of 21cm IM experiments will look like a continuous field (maps of pixels), thus disclosing to more data analysis possibilities and allowing to exploit, for example, the well trained CMB-related analysis machinery; thanks to the narrow channel bandwidths of modern radio receivers, we will measure redshifts too, bypassing one of the most difficult aspects of performing galaxy surveys. On top of this glowing portfolio of advantages, it sums up the present and prospected instrumental effort for performing 21cm IM experiments: among present and future facilities we highlight the Canadian Hydrogen Intensity Mapping Experiment (CHIME)[10], the Baryon acoustic oscillations In Neutral Gas Observations (BINGO)[11], the Ooty Radio Telescope (ORT)[12], the Five-hundred-meter Aperture Spherical radio Telescope (FAST)[13], the Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX)[14] and ultimately the Square Kilometre Array (SKA)[15] and its pathfinders (see [27] for a more complete list and a discussion on the different instrumental configurations). As we will show in Section 3.4.2, in this thesis we focus on SKA phase-1, considering the MID and the LOW

10 http://chime.phas.ubc.ca/
11 http://www.jb.man.ac.uk/research/BINGO/
12 http://rac.ncra.tifr.res.in/
13 http://fast.bao.ac.cn/en/
14 http://www.acru.ukzn.ac.za/~hirax/
15 https://www.skatelescope.org/

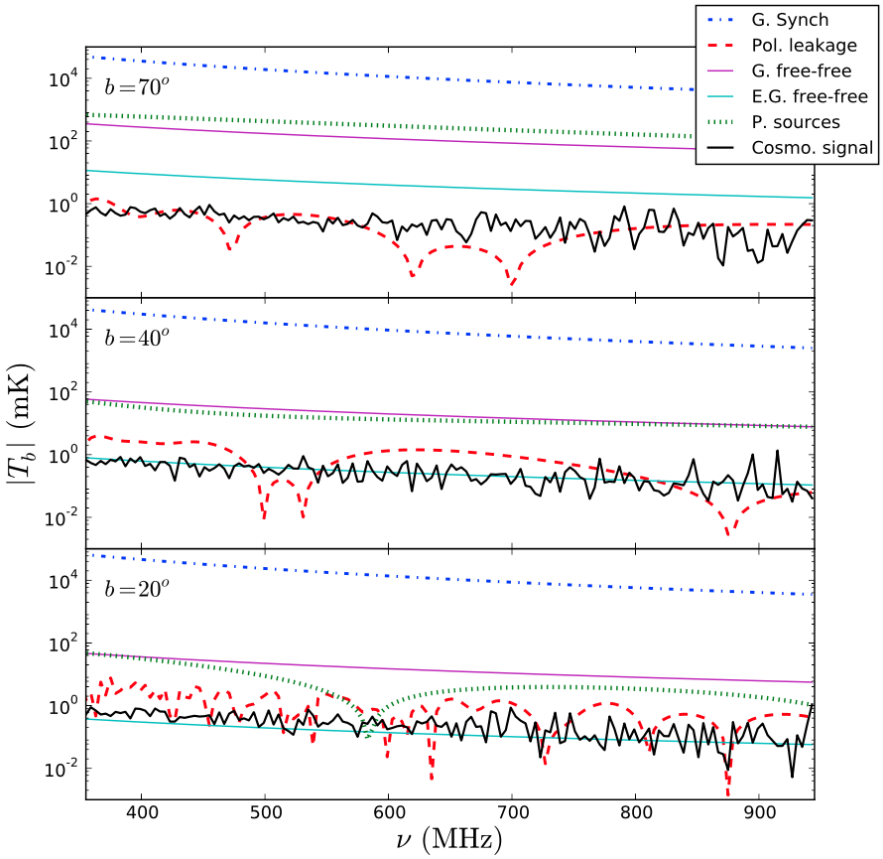


Figure 1.5: Plot taken from [5] showing the frequency dependence of the different foregrounds and the cosmological signal along lines of sight with different galactic latitudes (given in the top-right corner of each panel)..

arrays according to the wavelength (redshift) we study in the specific case [162].

Summarising, 21cm IM surveys will be characterised by:

i) large volumes sampled in comparatively small observing times;

ii) the advantage of the spectroscopic nature of the signal;

iii) a number of suitable instruments developing over the coming decade.

As follows, IM constitutes a promising method to efficiently observe the distribution of matter in the Universe on cosmological scales. Indeed, we already find in recent literature efforts to exploit this new cosmological probe. In [27] they assess how well a number of cosmological observables and parameters will be measured by a variety of upcoming IM experiments; in [30] they consider the access to ultra-large scales that these kind of surveys will have, allowing for detection of effects of primordial non-Gaussianity and placing very

stringent bounds on different models of inflation; tests of general relativity and modified gravity are explored in [69], while in [159] they assess the detectability of the baryonic acoustic oscillations (BAO).

However, it is a novel technique and as such has to face several issues, e.g. the calibration of the radio telescopes[16], the understanding the possible instrumental effects [27], and:

LACK OF A THEORETICAL FRAMEWORK As we will discuss in Chapter 3, there is no consensus on the how to model the signal, i.e. the distribution of HI in the post-reionization Universe ($1 < z < 5$), as observations are still missing. The value of $\Omega_{\rm HI}$ itself, the cosmological abundance of HI, is still quite uncertain, together with its redshift evolution (discussion in Section 3.3.1), but it is crucial for optimising the surveys, e.g. the lower $\Omega_{\rm HI}$, the harder is to detect the HI signal and the more aggressive the foreground cleaning needs to be.

FOREGROUND CONTAMINATION Contaminating signals are expected to be between 4 and 6 orders of magnitude higher than the cosmological HI signal, see Figure 1.4 and 1.5. We summarise the most severe ones:

i) Galactic synchrotron emission: the radiation caused by high-energy cosmic ray electrons accelerated by the galactic magnetic field.

ii) Galactic free-free emission: free electrons accelerated by the ions of the warm ionized medium.

iii) Extragalactic radio sources: bright radio galaxies (as active galactic nuclei) and star forming galaxies.

iv) Earth-wise contaminants: artificial radio frequency interference and atmospheric (ionosphere) noise.

Thus, the possibility of using 21cm IM as cosmological probe vitally depends on the availability of accurate foreground removal techniques (e.g. [6, 170]). Another way to tackle this issue is through cross-correlation studies, as we propose in Chapter 7, where we study the cross-correlation between the 21cm maps and the Lyman-α forest flux at redshift $z = 2.4$.

In this thesis we focus on the post-reionization Universe, however the redshifted 21cm emission line is a crucial observable also for probing the EoR (e.g. [65, 88]). To date, a number of experiments have sought to measure this high-redshift 21 cm emission, using LO-

16 Especially for single dish experiments, calibration errors constitute some of the most challenging foregrounds [27], e.g. drifts in the gain (due to instrumental temperature variations) and spillover (due to the poorly characterized beam).

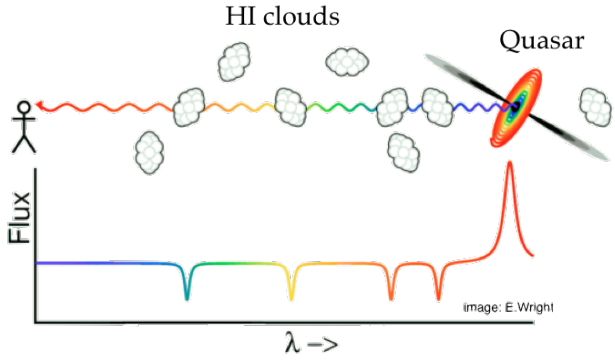


Figure 1.6: Diagram of the Lyman-α forest: as the light from the quasar travels towards the observer, it encounters blobs of neutral gas that gets ionized, leaving absorption features in the observed spectra.

FAR[17], the GMRT[18], the MWA[19] and PAPER[20]. The task is incredibly hard (see discussion in [172]) and no signal has been detected yet, but encouraging results come from the LOFAR collaboration who got first limits on the EoR 21cm power spectra, in the redshift range $z = 7.9 - 10.6$ [115]. Needless to say, their efforts will be greatly valuable for detecting the post-reionization signal too, because it shares the same foregrounds as the EoR signal.

1.2.2 *Signal in absorption: the Lyman-α forest*

We discussed above how challenging is to detect the 21cm emission line of gas-rich high-z galaxies. Instead, these objects are easily detected through their *absorption* signatures in the optical spectra of background quasars. The highest HI column density system found ($N_{HI} > 2 \times 10^{20}$ atoms cm^{-2}) are called damped Lyman-α absorber systems (DLAs) and are the largest repository of neutral gas at high redshifts. In Section 3.3 we better describe DLAs and we use their statistical properties to test our models of HI distribution.

DLAs are called *damped* systems because they are indeed identified by a depression in the quasars' continuous spectra. Actually, all quasars at high redshift exhibit also huge numbers of more narrow absorption lines, starting at the wavelength of the quasar's own Lyman-α emission line and extending blueward. These correspond to lower column density ($N_{HI} < 10^{14.5}$ atoms cm^{-2}) absorption systems that are believed to smoothly trace the cosmic web filaments and sheets [156]. The picture is the following: when light from the quasar travels

17 http://www.lofar.org
18 http://www.gmrt.ncra.tifr.res.in
19 http://www.mwatelescope.org
20 http://eor.berkeley.edu

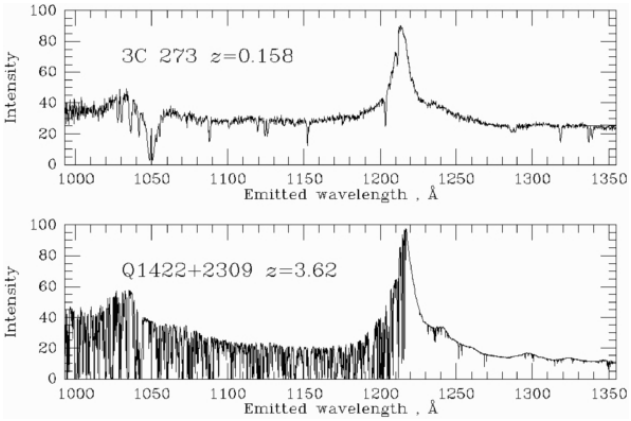


Figure 1.7: Spectra taken from the Quasars and Active Galactic Nuclei archive at http://pages.astronomy.ua.edu/keel/agn/. The high redshift spectra (bottom panel) displays so many absorption lines that they are said to be a *forest* of lines.

through the hydrogen clouds, each cloud imprints an absorption line of Lyman-α onto the continuum spectrum, as shown in Figure 1.6.

As the clouds are receding due to cosmic expansion, a series of redshifted lines called Lyman-α *forest* is formed. These features encode information about the distribution and density of cold gas along the line of sight to the quasars, thus probing an otherwise invisible component of cosmic gas.

This component evolves strongly with cosmic time, since we see dramatically more absorbers toward higher redshifts, as in the 2 example spectra shown in Figure 1.7, top panel quasar at $z = 0.16$, bottom panel $z = 3.6$. The availability of accurate Lyα forest spectra for a large number of quasars from the Baryon Oscillation Spectroscopic Survey (BOSS)[21] allows 3D statistics to be done [141], e.g. the clustering properties of the Lyα forest have been recently used to detect the BAO peak at $z = 2.34$ [56].

In Chapter 6 we describe the methods we use to model the Lyα forest in our simulations, and in Chapter 7 we study the cross-correlation signal of 21cm IM - Lyα forest flux.

21 http://www.sdss3.org/surveys/boss.php

2

THEORETICAL FRAMEWORK AND STATISTICAL TOOLS

In this chapter we present the statistical tools we employ for the analysis of the results of this thesis, together with a description of the model of gravitational instability needed to describe the distribution of matter in the Universe. We follow mainly Peebles, 1980 [116]. We do not aim at giving exhaustive derivations, our purpose is merely to clarify the quantities we later use and sketch how we calculate them.

2.1 THE POWER SPECTRUM

The matter[1] density fluctuations can be described in terms of the density contrast

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x})}{\bar{\rho}} - 1\,, \tag{2.1}$$

where $\rho(\mathbf{x})$ is the matter density at the comoving position $\mathbf{x}$ and $\bar{\rho}$ is the mean background density $\langle\rho(\mathbf{x})\rangle$. The density contrast can be written as a sum over Fourier modes

$$\delta(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\delta(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{2.2}$$

with the inverse Fourier transform

$$\delta(\mathbf{k}) = \int d^3x\delta(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}\,. \tag{2.3}$$

Assuming that the field is statistically homogeneous and isotropic, we have that the two point correlation function $\langle\delta(\mathbf{x_1})\delta(\mathbf{x_2})\rangle = \xi(\mathbf{x_1},\mathbf{x_2})$ is invariant under translations and rotations, which means that is only function of the separation distance $r = |\mathbf{x_1} - \mathbf{x_2}|$ between the two points where it is evaluated. The power spectrum P(k) is the Fourier transform of the two point correlation function, or alternatively

$$\langle\delta(\mathbf{k_1})\delta(\mathbf{k_2})\rangle = (2\pi)^3\delta_D(\mathbf{k_1}+\mathbf{k_2})P(k_1)\,, \tag{2.4}$$

with δ_D the Dirac delta.

In this thesis, the field considered is interpolated on a grid using a Cloud in Cell (CIC) algorithm [83], we then make use of the Fast Fourier Transforms[2] (FFTs) to calculate the power spectra.

1 The same mathematical framework applies to other fields too, as for the Lyman-α flux in Chapter 6, or for the 21cm signal considered many times in this thesis.

2 https://en.wikipedia.org/wiki/Fast_Fourier_transform

2.2 GRAVITATIONAL INSTABILITIES IN AN EXPANDING UNIVERSE: LINEAR THEORY

The growth of structure via gravitational collapse is a non-linear process which, even in an expanding background, quickly departs significantly from the solutions of the linearised equations. However, linear perturbation theory is an extremely useful tool when studying structure formation, since it contains vital information which can be used to understand the late-time distribution of the matter density field.

Assuming a dark matter dominated Universe, in the Newtonian limit[3], at sub-horizon scales, the linearised Poisson, continuity and Euler equations that describe the evolution of the overdensity field δ are given by

$$\nabla^2\phi = \frac{3}{2}\Omega_m\delta \tag{2.5}$$

$$\dot{\delta} + \frac{1}{a}\nabla\cdot\mathbf{v} = 0 \tag{2.6}$$

$$\dot{\mathbf{v}} + H\mathbf{v} = -\frac{\nabla\phi}{a}\,. \tag{2.7}$$

where the derivatives are taken with respect to the cosmic time t and where

- ϕ gravitational potential
- Ω_m matter density parameter
- a scale factor, describing the expansion of the Universe
- $\mathbf{v}$ velocity field.

and we define $H \equiv \dot{a}/a$ the expansion rate. By differentiating the continuity Equation 2.6 with respect to t and combining the result with the Poisson and Euler equations 2.5-2.7, we find a second order linear differential equation for the linear density contrast

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}\frac{\Omega_m}{a^2}\delta = 0\,. \tag{2.8}$$

In the above equation only time derivatives appear, hence solutions can be factorised into a spatial and a time dependent solutions, associated with a growing $D_+(t)$ and a decaying mode $D_-(t)$:

$$\delta(\mathbf{x}, t) = D_+(t)A(\mathbf{x}) + D_-(t)B(\mathbf{x})\,, \tag{2.9}$$

where $A(\mathbf{x})$ and $B(\mathbf{x})$ are two arbitrary functions describing the initial density field. In the case of matter dominated Universe, which holds over the range $3100 \gtrsim z \gtrsim 0.5$ in which most structures form, $H = 2/(3t)$, hence

$$D_+(t) = t^{2/3} \quad \text{and} \quad D_-(t) = t^{-1}\,. \tag{2.10}$$

3 Assuming weak gravitation ($\delta \ll 1$) and velocities to be non-relativistic ($v \ll c$).

After some time the growing solution will dominate, i.e. structures form. Thus, the decaying mode is generally neglected and the solution gets normalised by the growing mode today ($a_0 = 1$), so we define the growth factor as

$$D(a) \equiv \frac{D_+(a)}{D_+(a_0)} \,. \tag{2.11}$$

2.2.1 *The transfer function*

Let us continue following the evolution of the matter fluctuations. We consider collisionless dark matter particles, their velocities can cause the overdensities to disperse and this process makes fluctuations erase on scales $\lesssim \sigma t$, with σ the velocity dispersion of the particles at time t. As time passes, σ decreases and the damping becomes smaller. In standard ΛCDM model no suppression is predicted, i.e. the dark matter particle must have mass $\gtrsim 1$ keV [18]. Such particles become non-relativistic at $z \gtrsim 5 \times 10^6$ and their random velocities at decoupling are $\lesssim 50$ km/s: dark matter particles satisfying these constraints are called cold dark matter (CDM), that we already illustrated in Section 1.1.2.

Precise calculations analogous to this one can be performed also for baryons and other dark matter models, as in Chapter 4-5. The results are accounted by the so called transfer function:

$$T(k,z) \equiv \frac{\delta(k,z)}{\delta(k,z=\infty)} \frac{\delta(0,z=\infty)}{\delta(0,z)} \,. \tag{2.12}$$

where $\delta(0)$ is the overdensity on large scales, which obeys the Equation 2.8. Thus $T(k,z)$ represents the factor by which the linear fluctuations with wavenumber k are enhanced or suppressed relative to large scale fluctuations[4]. In order to derive $T(k,z)$, in this thesis we make use of the code CAMB [81] together with fitting formulas later presented.

2.2.2 *The matter power spectrum*

If linear theory were valid up to present times, the current matter power spectrum $P(k)$ would be proportional to the product of the primordial $P_{ini}(k)$ and of the square of the transfer function and of the growth factor:

$$P(k,z) \propto D^2(z) T^2(k) P_{ini}(k) \,. \tag{2.13}$$

However, as discussed in Section 1.1.3, we do not observe directly matter, but observables that trace matter in a biased way. The idea that

4 In the case of CDM, the growth is self-similar, hence $T(k,z) = T(k)$

baryons condense and form galaxies in dense environment has been generalised by Sheth and Tormen, 1999 [137] in the peak-background split framework. The statistical relation between the DM and luminous matter distribution is described by a bias b_A, with A standing for galaxy or other observables, such that at linear order it holds

$$P_A(k) = b_A^2 P(k)\,. \tag{2.14}$$

At small scales the linear bias relation breaks down mainly because of non-linear structure formation and acquires non-local features as scale dependency, as we check in Chapter 7.

2.3 THE NON-LINEAR REGIME

In this thesis, we follow the highly non-linear dynamical regime in which DM particles enter and get bound into stable objects (halos), making use of numerical simulations, as analytical tools break down at describing the matter clustering.

A semi-analytical model that has proven useful to address non-linear clustering is the halo model (see [46] for a review). In this thesis we do not employ it, but we refer to it. The assumption made is that at late times all the matter distribution is contained only in virialised DM halos. Then, the total matter power spectrum ca be split into 2 terms

$$P(k) = P_{1h}(k) + P_{2h}(k)\,, \tag{2.15}$$

where P_{1h} is the 1-halo term describing correlations on small scales between mass particles within the same halo and P_{2h} is the 2-halo term arising from large scale correlation of matter in different halos.

3

ON THE MODELLING OF THE HI SPATIAL DISTRIBUTION

The spatial distribution and the clustering properties of neutral hydrogen in the Universe are unknown. The closest we got so far has been measuring:

i) the clustering of HI selected galaxies at $z \sim 0$ from the ALFALFA survey (Martin et al., 2012 [91]), where they measure a HI bias of $b_{HI} \simeq 0.8$,

ii) the product $\Omega_{HI} \times b_{HI} = 0.62 \times 10^{-3}$ i.e. the HI cosmic abundance times its linear bias, from intensity mapping (IM) observations at $z \simeq 0.8$ performed with the Green Bank Telescope by Switzer et al., 2013 [147] (in Section 3.3.1 we see why 21cm IM surveys are sensitive to this product and how much uncertain Ω_{HI} and b_{HI} are on their own),

iii) the bias of the Damped Lyman-α systems (DLAs) at $z \sim 2.3$ by the BOSS collaboration [62]: $b_{DLAs} = 2.17$ (in next paragraph we say more on DLAs).

There is overall consensus that after reionization ($z < 6-5$) almost all HI belongs to galaxies (and consequently to dark matter halos) as the gas elsewhere has been ionized by the background light. Last argument is endorsed by Lyman-α forest data, that enable us to measure *how few* neutral gas is left in the intergalactic medium (IGM) [106, 171]. Briefly, we do not know which halos bear which amount of neutral hydrogen at a given redshift, but we know the greatest part of HI is in halos.

Coming back to the forest data, main tracers of HI are indeed the above cited DLAs: objects with HI column density $N_{HI} > 10^{20}$ atoms cm^{-2} displaying a 21cm line in absorption in the spectrum of distant quasars [63]. DLAs are believed to host most ($\sim$ %90) of the cosmic neutral gas up to $z \simeq 5$ [62] and this is why in the literature it is very common to interchange equivalently their bias b_{DLAs} with the HI bias b_{HI}, but in [38] they claim the 2 biases are significantly different, b_{DLAs} being a number-of-object weighted quantity and b_{HI} a mass weighted one.

Indeed, putting on the same foot all the measurements listed above, it is hard to reconcile them in a unique HI spatial distribution prescription (e.g. none of [9, 10, 51, 74, 113, 163] agrees to each another), winding up with the recent work by [112] where it is argued that there must be a crucial change over time in HI hosting systems properties.

In this thesis we make use of different methods to distribute HI in the large scale structure. They are different in spirit and actual realization and are supported by different observational evidences. Hence, their joint use empowers our conclusions as we are fully exploring the possible and viable ways at our disposal to distribute cold gas in simulations.

3.1 HALO BASED MODELS

Since the ionizing UV light background prevents the formation of large neutral gas clouds unless they are self-shielded, it is believed that in the post-reionization Universe, most HI resides in the halos and only a negligible fraction is diffuse[1]. This allows us to make use of *halo based* models whose main ingredient is the function $M_{HI}(M, z)$ that represents the average HI mass that a dark matter halo of mass M hosts at redshift z.

In the literature we can find quite a few versions of the relation $M_{HI}(M, z)$: it has been calibrated on hydrodynamic simulations [51, 74, 163], inspired from observations [9] or parametrised and fitted on data [10, 38, 113]. Some of them are summarised in Figure 3.1a. There is no agreement on the modelling of $M_{HI}(M, z)$: the prescriptions vary in shape, amplitude and slope. Summarising, the models that are built to fit the high DLAs bias measured at high redshift favour more massive halos in Figure 3.1a as compared to the others that are aimed at matching the 21cm data coming from lower redshift.

In this thesis we make use of 2 halo based HI mass assignments. The first one, dubbed *halo based model A*, was originally presented in in Bagla, Khandai, and Datta, 2010 [9]. The second, *halo based model B*, is described in Villaescusa-Navarro et al., 2016 [163].

We acknowledge that this modelling of the $M_{HI}(M, z)$ function implicitly neglects the intrinsic scatter expected in it[2]. However, the data so far available is not enough to constrain the HI - halo mass relation (but see the very recent work of [67] that goes in this direction). Reassuringly, such halo based models have been already extensively investigated at similar redshifts against other methods such as post-processing pseudo radiative-transfer calculations (as the one described later in Section 3.2), and the results have been proven to be consistent among them: e.g. [37, 161, 162].

1 This point has been checked and validated in simulations too [161].

2 Moreover, if the scatter correlates with environment, cross-correlations might be seriously affected by this approximation (as the study in Chapter 7).

MODEL A Following Bagla, Khandai, and Datta, 2010 [9], we model the function $M_{HI}(M)$ as:

$$M_{HI}(M) = \begin{cases} f\frac{M}{1+(M/M_{max})} & \text{if } M_{min} \leqslant M \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

The lower mass cut-off M_{min} takes into account that a minimum hydrogen density (clustered in a minimum halo potential well) is needed to have self shielding and prevent the gas to be fully ionised. The fall-off above M_{max} is introduced to mimic observations: galaxies in dense environments as galaxy clusters are HI-poor. Finally, according to [9], M_{max} and M_{min} corresponds to halos with circular velocities v_{circ} of 200 and 30 km s^{-1} respectively.

MODEL B Following the results of Villaescusa-Navarro et al., 2016 [163], where they run high-resolution and zoom-in hydrodynamic simulations of individual clusters of galaxies, we model the $M_{HI}(M, z)$ as

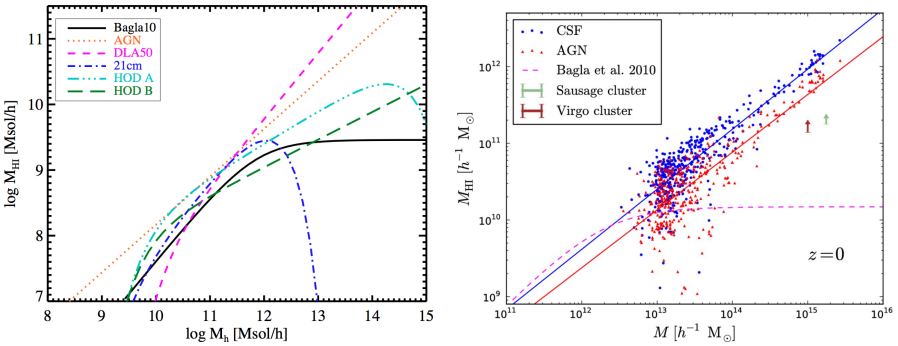
$$M_{HI}(M) = \begin{cases} f\,M^{\alpha} & \text{if } M_{min} \leqslant M \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

where we set $\alpha = 3/4$. This power law index regulates how fast is HI mass accreted onto halos and both hydrodynamic simulation [163] and semi-analytic models [74] have a preference for $\alpha \leqslant 1$ since processes such as tidal interactions, ram-pressure stripping and mergers, which tend to remove HI from the galaxies, appear to be more efficient than those that stimulate the cooling of hot gas.

As for model A, M_{min} accounts for the fact that there is a HI density threshold below which the gas becomes fully ionised and unable to effectively self-shield from UV radiation. According to [163], the mass parameter M_{min} corresponds to a dark matter halo[3] with circular velocity $v_{circ} = 25$ km s^{-1}. Notice that also on the value of M_{min} there is no agreement between the different models, still due to a tension between the 21cm low redshift galaxies data and DLAs: the first leads to host DM halos of mass $10^9 - 10^{11} M_\odot$ (in agreement with simulations e.g. [122, 125]) whereas the latter to more massive halos of at least $10^{11.5} M_\odot$. In this thesis, we discuss case by case the choice of M_{min} we make and sometimes we are forced to use numerical resolution arguments.

Figure 3.1b is from [163] where our model B is originally presented; they show with points the corresponding mass in HI versus the halo mass of the simulated clusters of galaxies (blue and red colours correspond to different feedback prescriptions in their simulations). Solid

3 Let us highlight that in [163] they infer the $M_{HI}(M, z)$ relation (solid lines in Figure 3.1b) extrapolating their fit on scales below the mass resolution of their simulations.



(a) Figure from [118]. Different M_{HI} available in the literature at $z = 1$. Our model A corresponds to the black solid line, model B to the dotted orange line.

(b) Figure from [163]. Mass of HI in simulated galaxy clusters (points). The red and blue lines are our model B. Model A is the pink dashed line. The arrows represent a lower limit on the HI mass hosted by real clusters.

Figure 3.1: HI mass functions M_{HI} as function of the host halo mass M.

lines are the relation $M_{HI}(M, z)$ of our model B, that is compare to model A in dashed magenta: their discrepancy shows up for masses of $10^{13} h^{-1} M_\odot$ and higher, here at $z = 0$. For most of the box sizes of simulations analysed in this thesis, combined to the higher redshift considered, this high mass range is scarcely probed, hence the choice between model A or B is less influential compared for example to the choice of minimum mass threshold M_{min} and the normalisation factor f (see Section 5.2.1 for a discussion on the M_{min} and volume effects). Anyway, at the beginning of my Ph.D. we used model A of Equation 3.1 then stop and start employing model B of Equation 3.2 because the first one was found to poorly fit the bias of the DLAs [161] and to be in disagreement with the bounds coming from clusters of galaxies (shown with arrows in Figure 3.1b), whereas in [38] they show how model B is capable to reproduce both the abundances and the clustering of the DLAs. In other words, model B is generally truer than A but our conclusions do not change much if employing model B instead of A.

In both models, the halo mass threshold M_{min} (and M_{max} too for model A) is chosen in correspondence of the dark matter halo circular velocity, thus we make use of the relation

$$M(v_{circ}, z) = 10^{10} M_\odot \left(\frac{v_{circ}}{60\,\mathrm{km\,s^{-1}}} \right)^3 \left(\frac{1+z}{4} \right)^{-1.5} , \tag{3.3}$$

derived from virialisation arguments [9].

Either using model A or B, we are left with the calibration parameter f to be tuned. We set it such that the total amount of HI in

the halo catalogue considered reproduces the observed cosmic abundance. The cosmic HI density is given by

$$\Omega_{HI}(z) = \frac{1}{\rho_c} \int_0^\infty \frac{dn}{dM}(M,z)\, M_{HI}(M,z) dM, \tag{3.4}$$

where ρ_c is the present critical density and dn/dM the halo mass function. We can write the following relation:

$$f = \frac{\Omega_{HI} L^3 \rho_c}{\sum_{i=0}^{N_{halo}} M_{HI}(M_i)}, \tag{3.5}$$

where the function M_{HI} is either Equation 3.1 or Equation 3.2 depending on the model chosen, L is the box size of the simulation and the index i runs over all the dark matter halos of the simulation considered, of which N_{halo} is their total number.

The value of Ω_{HI} is poorly constrained by observations, as we will see in Section 3.3.1. We conservatively set $\Omega_{HI} = 10^{-3}$, consistently with observational results at $3 \lesssim z \lesssim 5$ presented in [105, 124]. Assuringly, the conclusions we derive on the HI clustering are not affected by this choice as Ω_{HI} only controls the amplitude of the 21cm power spectrum (but it is crucial on the observational point of view, e.g. see discussion on the Ω_{HI} - b_{HI} degeneracy in Section 7.3.2.1).

By setting the normalisation f as in Equation 3.5 for each snapshot and simulation, we guarantee to have the same reference value of Ω_{HI} independently of redshift and cosmology. This holds for the halo based methods, whereas it is not the case for the particle based method, as we see in next section.

Since these models consist in assigning HI in a deterministic manner to dark matter halos of a given mass, we can compute the linear HI bias by solving the integrals

$$b_{HI}(z) = \frac{\int_0^\infty b(M,z) n(M,z) M_{HI}(M,z) dM}{\int_0^\infty n(M,z) M_{HI}(M,z) dM}, \tag{3.6}$$

where $b(M,z)$ and $n(M,z)$ are the bias and mass function of halos, as in [137] and [136].

3.2 PARTICLE BASED MODEL

In this method we assign neutral hydrogen to all gas particles in the simulation according to their physical properties. Compared to the halo based method, we do not rely on any definition of dark matter halo or on any assumption on the amount of HI outside halos and therefore we can also predict the amount of HI in different places, as filaments and cosmic voids.

We proceed as depicted in [51]: for every single gas particle in the simulation we compute the neutral hydrogen fraction in photo-ionization equilibrium with the Ultra Violet (UV) background and

correct that fraction to account for both HI self-shielding and formation of molecular hydrogen. Hydrogen of star forming particles is split in totally neutral for the cold phase, and totally ionized for the hot phase, i.e. we consider than the HI/H fraction of star forming particles in photo-ionization equilibrium is equal to its multi-phase cold gas fraction. The strength of the UV background, at a particular redshift, is corrected by demanding that the mean flux of the Lyman-α forest reproduces the observations [12], with the approach described in [161]. For every gas particle we then compute the radial column density profile $N_{HI}(r)$ making use of the SPH spline kernel $W(R)$[4]:

$$N_{HI}(r) = \frac{0.76\, m}{m_H} \left(\frac{HI}{H}\right) \int_r^{r_{SPH}} W(r')\, dr', \tag{3.7}$$

with

m/m_H	gas particle mass in units of hydrogen atom mass
HI/H	HI fraction in photo-ionization equilibrium
r_{SPH}	SPH smoothing length.

If a radius r exists such that $N_{HI}(r) = 10^{17.2}$ cm^{-2}, then the sphere from $R = 0$ to $R = r$ is considered to be self shielded against the external UV radiation and its HI/H fraction is set to 0.9. For the surrounding spherical shell ($R > r$) it still holds the neutral fraction HI/H coming from the photo-ionization equilibrium assumption.

The last step consists in correcting the HI/H fraction computed in the previous step to account for the formation of molecular hydrogen. Using the density and internal energy, we compute the pressure of gas particles P and use it to correct the HI/H fraction computed previously using the observed relation between the surface densities of HI and H_2 with the pressure of the disk galaxy [16]:

$$R_{mol} = \frac{\Sigma_{H_2}}{\Sigma_{HI}} = \left(\frac{P/k_B}{3.5 \times 10^4\ \mathrm{cm^{-3}K}}\right)^{0.92} \tag{3.8}$$

with k_B the Boltzmann constant. H_2 is assigned only to star forming particles. We are implicitly assuming that the relation of Equation 3.8, which arises from observations at low redshift, holds also at high redshift which may not be the case (see for instance [51, 79]).

3.3 THE HI COLUMN DENSITY DISTRIBUTION FUNCTION

Up to present days our understanding of the morphology of the hydrogen content in the high redshift Universe comes mainly from quasar absorption spectra, where the presence of the Lyman-α transition lines is a conclusive evidence of intervening gas between us and the sources. From these observations we can also infer that most of

4 In the hydrodynamic SPH simulations the spatial distribution of gas is discretised into a finite number of gas particles with a given kernel and radius.

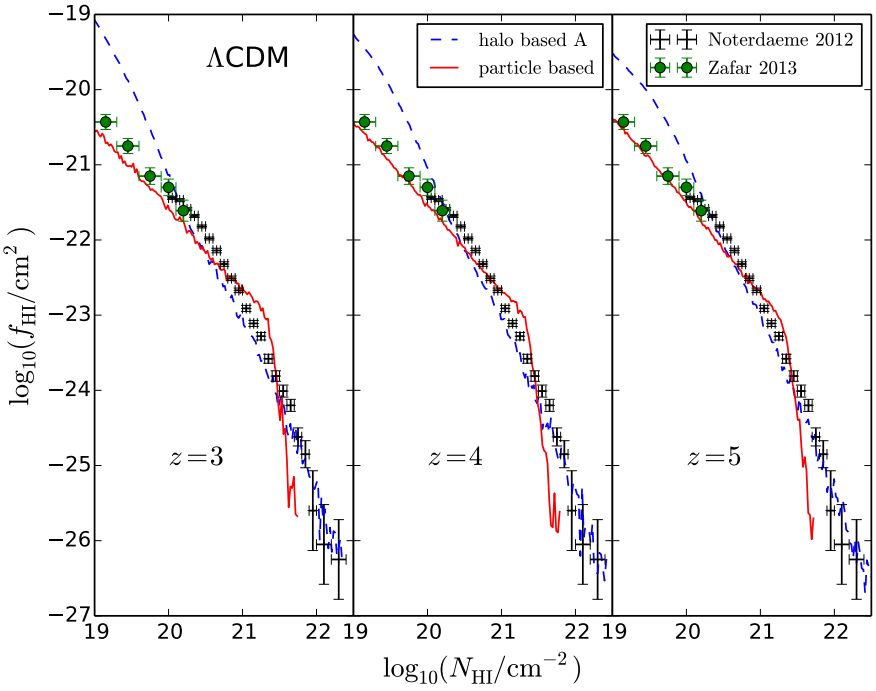


Figure 3.2: HI column density distribution function $f_{HI}(N_{HI})$ obtained by assigning HI to gas particles employing: halo based model A (dashed blues lines) and particle based (solid red). Observational data is plotted with black crosses for [106] and with green circles for [171]. We show results at redshifts $z = 3$ (left), $z = 4$ (middle) and $z = 5$ (right). The simulation used is the ΛCDM - $30\,h^{-1}$Mpc described in Chapter 4.

the hydrogen in the intergalactic medium is ionized, and the neutral hydrogen is mainly contained in the so called damped Lyman-α absorber systems (DLAs) with HI column densities N_{HI} above 2×10^{20} atoms cm^{-2} and secondarily in Lyman limit systems (LLS), at lower HI column densities. In order to test the 2 ways we distribute HI in our simulations[5], we compare the abundance of HI absorbers in our simulations with observational data from DLAs [106] and LLS [171]. We assume that the observed HI column density distribution function does not vary in the redshift range considered.

We test our 2 HI mass assignment models (halo and particle based) in the ΛCDM hydrodynamic simulation described in Section 4.1.1: 512^3 CDM and 512^3 baryon particles followed in a box of $30\,h^{-1}$Mpc comoving size.

We compute the column density distribution function $f_{HI}(N_{HI})$ for both HI assignment methods. We project the HI particles position on a plane and pick random lines of sight perpendicular to the plane. For each line the HI column density is computed using the SPH radius

5 We never make use of simulations with halo mass resolution lower than M_{min} whenever applying the halo based model.

and HI mass of all gas particles. Details of the computation can be found in [161].

Results are shown in Figure 3.2 for both the halo based model A and particle based method for redshift $z = 3, 4$ and 5 from left to right, together with the observational measurements. The HI column density distribution obtained by distributing the neutral hydrogen with the two different methods is in overall good agreement with the data.

For the halo based model A (dotted lines in Figure 3.2) the agreement works better for column densities higher than 10^{21} cm^{-2} while the abundance of absorbers with lower column densities is overestimated by this method. Our results are redshift dependent, as the overestimation is weaker at higher redshifts, which is in turn due to the lack of those low mass halos not yet formed, that are where low column density absorbers reside.

Again, since we look at redshift $z \geqslant 3$, results would not change considerably if the model B is employed. Actually, the mismatch with data is present for low column density objects, i.e. low mass halos, which are covered in the same way by both model A and B.

By simulating the HI distribution using the particle based method (solid red lines in Figure 3.2), the abundance of absorbers with large column densities is underestimated at all redshifts while the agreement is better than that of the halos based method for low HI column density absorbers: the particle based method goes beyond the strict HI-in-halo paradigm and can more genuinely account for low dense HI environments that are sharply excluded or included by the uncertain M_{min} parameter of Equation 3.1. Ultimately, the mismatch in the abundance of LLS between our results and observations does not weaken the conclusions of this work, since their contribution to the total HI distribution is marginal compared to the DLAs, which are instead well reproduced in the halo based method.

Overall, the 2 different HI assignment methods are somehow complementary: one overestimates what the other underestimates, leaving the data points wrapped by the two approaches. Exploiting both the 2 methods makes our conclusions stronger as we are somehow exploring the range of possibilities for spatially distribute neutral hydrogen.

We stress that even if the halo based and the particle based methods are greatly different both in spirit and in the actual details of the implementation, as discussed in earlier sections, yet, looking at the $f_{HI}(N_{HI})$ results, they reproduce to some extent observations if combined appropriately.

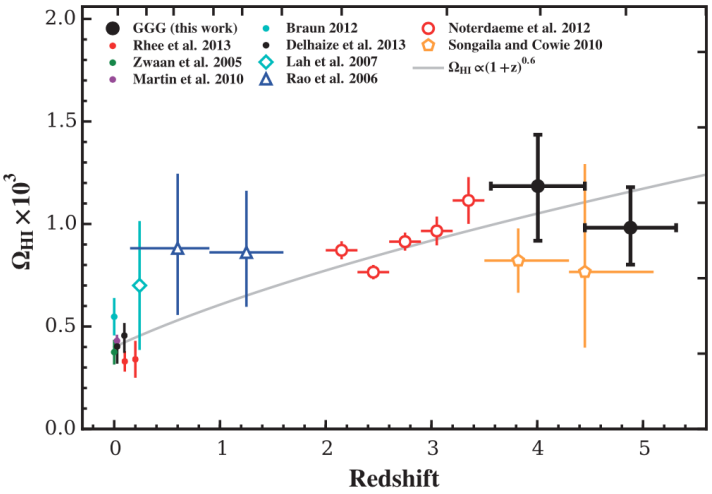


Figure 3.3: Measurements of HI at different redshifts. Details in [48], where the plot is taken from.

3.3.1 *About the cosmic abundance of HI:* Ω_{HI}

Measurements of the column density distribution function allow to constrain the cosmic HI abundance Ω_{HI} in the redshift range $1.5 < z < 5$, as shown in Figure 3.3, and at lower redshifts from the HI mass function from surveys such as HIPASS [173] and ALFALFA [23]. Even within large error bars, it is seems Ω_{HI} is very slowly evolving with redshift [48].

In this thesis we assume a fiducial value of $\Omega_{HI} = 10^{-3}$, independent of redshift (except when employing the particle based method in Chapter 4). In 21cm IM the value of Ω_{HI} is entangle with the bias, which is instead completely unknown. In Chapter 7 we discuss more in details the $\Omega_{HI} - b_{HI}$ degeneracy.

When employing the particle based method in Chapter 4, we let our radiative transfer algorithm set the Ω_{HI} value in our simulations, hence in that case we do deal with redshift evolving HI cosmic abundance. However, we need to stress that, as for observations, also theoretical predictions fail to converge to a common value or a common z-evolution of Ω_{HI} (e.g. see Figure 14 in [48]).

3.4 THE 21CM SIGNAL

The intensity of the 21cm signal is controlled by the spin temperature T_{spin} of HI, which is defined by

$$\frac{n_1}{n_0} = 3e^{(-T_\star/T_{spin})}, \tag{3.9}$$

where n_1 and n_0 are the number densities of electrons in the excited and ground states for the spin-flip transition, $T_\star = 0.068$K is the temperature of the 21cm radiation. The ratio in Equation 3.9 sets

the intensity of the emission of a HI cloud, however we need to take into account the background radiation being transmitted through the cloud, together with the absorption and emission balance within the cloud, i.e. we need to solve a radiative transfer problem (for a detailed description see [130]).

Briefly, the HI cloud shines in 21cm with an intensity effectively described by a so called *brightness temperature* T_b and radio telescopes detect this temperature against the CMB background, i.e. a differential brightness temperature $\delta T_b \equiv T_b - T_{CMB}$.

Hence, when quantifying the 21cm observations in this thesis, we refer to the 21cm power spectrum, defined as

$$P_{21cm}(k) = \langle \delta T_b(\mathbf{k})(\delta T_b)^*(\mathbf{k}) \rangle, \tag{3.10}$$

where δT_b is given by

$$\delta T_b(\nu) = \overline{\delta T_b}(z) \left[\frac{\rho_{HI}(\mathbf{s})}{\bar{\rho}_{HI}} \right], \tag{3.11}$$

with $\rho_{HI}(\mathbf{s})$ being the neutral hydrogen density in the redshift-space position $\mathbf{s}$ and (following [65])

$$\overline{\delta T_b}(z) = 23.88 \; \bar{x}_{HI} \left(\frac{\Omega_b h^2}{0.02} \right) \sqrt{\frac{0.15}{\Omega_m h^2} \frac{(1+z)}{10}} \; \text{mK}, \tag{3.12}$$

where $\bar{x}_{HI} = \bar{\rho}_{HI}/\bar{\rho}_H$ is the average neutral hydrogen fraction.

The 21cm power spectrum P_{21cm} is proportional to the HI power spectrum in redshift space, hence to calculate it we practically start from the spatial distribution of neutral hydrogen in real-space previously described, we then displace the HI particle positions to obtain their coordinates in redshift-space by doing

$$\mathbf{s} = \mathbf{x} + \frac{1+z}{H(z)} \mathbf{v}_{los}(\mathbf{x}), \tag{3.13}$$

with z being the redshift of observation, $\mathbf{v}_{los}$ the line of sight component of the peculiar velocity and $H(z)$ the Hubble parameter.

Notice that the amplitude of P_{21cm} depends on Ω_{HI}^2, hence the HI cosmic abundance that does not influence the HI power spectrum is actually important in the determining the power spectrum of the observed signal. We have seen in the previous section how the value of Ω_{HI} is rather uncertain, however throughout this thesis we are mostly interested in *ratios* of the power spectra of signals, comparing non-standard cosmologies with ΛCDM (Chapter 4-5), therefore most of our conclusions are not affected by this issue (assuming Ω_{HI} would not change among cosmologies).

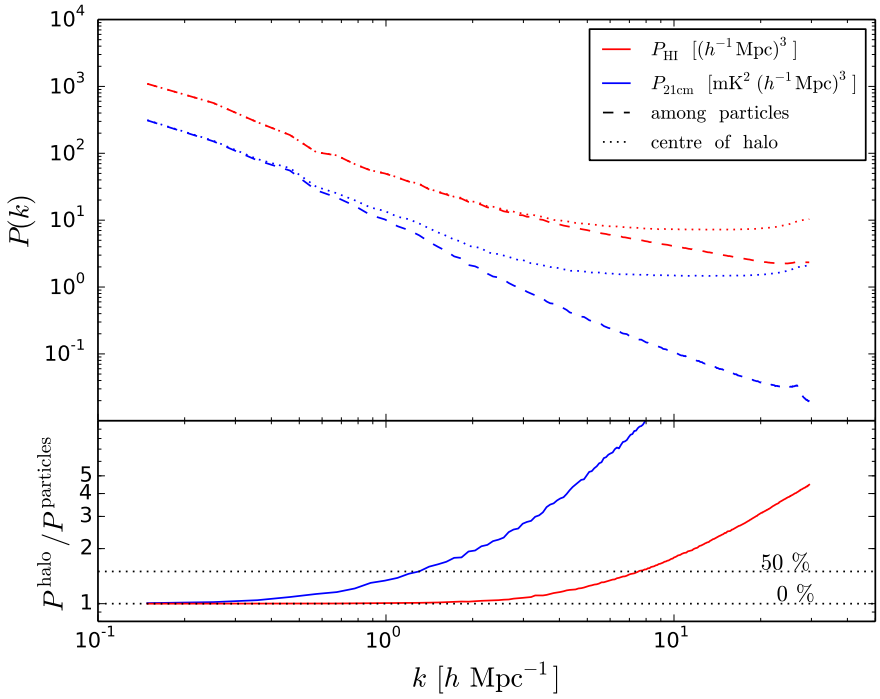


Figure 3.4: *Upper panel:* The HI power spectrum in real space $P_{HI}(k)$ (in red) and the 21cm $P_{21cm}(k)$ (in blue), distributing the HI at the centre of each halo (dotted lines) or spreading it among all its gas particles(dashed lines). *Lower panel:* Ratio between the two $P(k)$ computed with the HI at the centre of halos over the one with HI evenly distributed.

3.4.1 *Dependence on the HI halo density profile*

Concerning the halo based models described in Section 3.1, while the clustering of HI on large scales is fully determined by the function $M_{HI}(M,z)$, on smaller scales it depends on the way the HI is distributed within halos: the HI density profile $\rho_{HI}(r|M,z)$. For example, in the hydrodynamic simulations of Chapter 4, we split M_{HI} of the halo among the gas particles that belong to it, effectively building an unique $\rho_{HI}(r|M,z)$ per halo, dictated by the hydrodynamics of the simulation. But in principle the halo based method can be used in dark matter only simulations, that can be bigger and more resolved at lower computational expenses than hydrodynamic ones, hence must be exploited.

The simplest way is to place all M_{HI} of an halo in its centre of mass. How much does this affect the small scale clustering of HI? At which scale does this impact the HI power spectrum and eventually the 21cm signal?

We practically address these questions by using an hydrodynamic simulation that we name 60-512, run using the TreePM+SPH code GADGET-III [144], with cosmological parameters in agreement with recent Planck data [120]:

$$\Omega_m = 0.3175 \qquad \Omega_b = 0.049 \qquad \Omega_\Lambda = 0.6825$$
$$h = 0.6711 \qquad n_s = 0.9624 \qquad \sigma_8 = 0.834\,.$$

The 60-512 simulation follows the evolution of 512^3 cold dark matter particles and 512^3 baryon particles within a periodic box of comoving size of $60\,h^{-1}$ Mpc from $z = 99$ down to $z = 3$. Star formation is modelled using the effective multi-phase model of Springel & Hernquist [145]. The code also simulates radiative cooling by hydrogen and helium and heating by an uniform UV background. 60-512 has mass resolutions of $m_{CDM} = 1.2 \times 10^8\,h^{-1}M_\odot$ and $m_{baryon} = 2.2 \times 10^7\,h^{-1}M_\odot$ and softening length with value $l_{soft} = 2.9\,h^{-1}$kpc.

After having determined the amount of HI per halo M_{HI} following the halo based model B of Equation 3.2, we model the HI density profile in 2 ways.

1. AMONG HALO PARTICLES We distribute the HI evenly among all gas particles belonging to the host halo, i.e. we set $\rho_{HI}(r|M,z) = \rho_g(r|M,z)$, where $\rho_g(r|M,z)$ represents the density profile of gas within a dark matter halo of mass M at redshift z.

2. AT THE CENTRE OF HALO We place all M_{HI} in the centre of mass of each halo, i.e.

$$\rho_{HI}(r|M,z) = M_{HI}(M,z)\delta(\mathbf{r})\,,$$

where $\delta(\mathbf{x})$ is the Dirac delta. Thus, we are collapsing the HI density distribution into a single point located in the halo centre.

In Figure 3.4 we show the HI (in red) and 21cm power spectra (in blue) that we obtain using the two different HI density profiles, with the ratio of the profile 2 over profile 1 in the bottom panel. As expected, the HI density profile only affects the 1-halo term and therefore the difference among the two $\rho_{HI}(r|M,z)$ only shows up on relatively small scales ($k \gtrsim 2.5\ h\text{Mpc}^{-1}$). On the other hand, dispersion velocities within halos propagate into large scales through redshift-space distortions and thus worsen the power spectrum convergence in the 21cm case. Indeed, we find that the 21cm power spectra start deviating from each other on scales $k \lesssim 0.03\ h\text{Mpc}^{-1}$, and differ by a $\sim 50\%$ already at $k \sim 1.5\ h\text{Mpc}^{-1}$.

Concluding, the approximation of placing HI in the halo centres is suitable just when interested at quite large scales. As we do in Chapter 5-7, where we are interested on the amplitude and shape of the power spectra on large scales, where the different models for the HI density profile produce identical results.

3.4.2 *SKA and the modelling of its instrumental noise*

The Square Kilometre Array (SKA) is a large radio telescope project aimed to be built in Australia, New Zealand, and South Africa. It

would have a total collecting area of approximately one square kilometre. It will have 3 different instruments working at different redshifts bands and it will be built in 2 phases. In this thesis we consider the phase 1 (2018-2023) representing about 10% of the capability of the whole telescope. We make use of the specifications described in the SKA website[6] In particular, the instruments we take into considerations are:

SKA1-MID To be built in South Africa. It will cover a frequency range from 350 MHz to 14 GHz and will have a total of 250 antennae of 15 m diameter each (this is also includes 60 antennae from the MeerKAT telescope[7]).

SKA1-LOW To be built in Australia. It will cover a frequency range $50-300$ MHz and will have a total of 911 dishes of 35 m diameter each. The SKA1-low has a total collecting area of 0.88 km^2, much bigger than that of SKA1-MID (of 0.044 km^2).

In this thesis, we forecast errors for the above instruments in interferometer mode[8], i.e. the 21cm emission is measured by a set of stations (antennae or dishes) and the observations of pairs of stations separated by a given baseline are cross-correlated: the intensity map is then reconstructed by combining pairs with different baselines.

For forecasting errors, we model the system noise only[9], assuming that astrophysical foregrounds, radio frequency interference and others nuisances have been already removed from the observed data. Follow [161], we can write

$$\sigma^2[P_{21cm}(k,\theta)] = [P_{21cm}(k,\theta)+P_N(k,\theta)]^2 \tag{3.14}$$

$$= \left[P_{21cm}(k,\theta)+\frac{T_{sys}^2}{2Bt_0}\frac{D^2\Delta D}{n(k_\perp)}\left(\frac{\lambda^2}{A_e}\right)^2\right]^2 \tag{3.15}$$

where

6 https://www.skatelescope.org/

7 http://www.ska.ac.za/science-engineering/meerkat/

8 Another option is to use single-dish measurements, which possess less angular resolution but can sample much lager areas of the sky [27]. However, in this thesis we never study large enough scales to need the single-dish configuration.

9 We notice that we are neglecting the contribution to the error from shot-noise. In [38] it was shown that this term is subdominant for the scales considered in this thesis.

T_{sys}	system temperature of the telescope
D	comoving distance to the redshift of observation
ΔD	radial width of observation
λ	wavelength of observation
	(i.e. corresponding 21cm line at the redshift of detection)
B	bandwidth of the measurement
	(chosen accordingly to redshift of observation)
t_0	total observation time
$n(k_\perp)$	number density of the interferometer baselines
	sensitive to the transverse mode $k_\perp$
A_e	effective collective area of a station.

T_{sys} is the sum of the temperature of the sky at this redshift T_{sky} and of the temperature of the receiver T_{rcvr}, which in turn are defined as

$$T_{sky} \simeq 60(300\text{MHz}/\nu_{HI}(z))^{2.55}$$

with $\nu_{HI}(z) = 1420/(1+z)$ MHz, and

$$T_{rcvr} = 0.1T_{sky} + T_{inst}$$

with T_{inst} the instrument temperature, which is $T_{inst} = 40$K for SKA1-LOW and $T_{inst} = 28$K for SKA1-MID.

The array geometry of the telescope enters through the factor $n(k_\perp)$, that depends on the spatial distribution of the stations that we calculate using the baseline density distributions in appendix A of [162]. The effective collective area of SKA1-MID is $A_e = 140$ m^2. For SKA1-MID can it be expressed as

$$A_e(\nu) = A_{e,crit} \times \begin{cases} (\nu_{crit}/\nu)^2 & \nu > \nu_{crit} \\ 1 & \nu \geqslant \nu_{crit}\,. \end{cases} \tag{3.16}$$

with the critical $\nu_{crit} = 110$ MHz and $A_{e,crit} = 925$ m^2.

In this work, we mainly work with the error on the monopole of the 21cm signal power spectrum. To calculate the noise error for the spherically averaged power spectrum, we need to average over θ for a fixed k, but the error varies consistently with θ. The optimal choice is to adopt an inverse-variance weighting scheme [161]:

$$P_N(k) = \left[\sum_\theta \frac{1}{P_N^2(k,\theta)}\right]^{-1/2}. \tag{3.17}$$

Finally, what is left to be tuned is the total observation time t_0.

4

DARK MATTER: HOW MUCH WARM?

We already introduced in Section 1.1.2 how warm dark matter (WDM) has been proposed to explain the discrepancies on small scales between observations and standard CDM predictions. Indeed, while CDM has power all the way to very small scales, the higher streaming velocities of WDM at high redshift prevent it from initially collapsing into small halos with shallow gravitational wells [19, 26].

The WDM mass has already been constrained by a number of different observations. The tightest limits come from the Lyman-α forest[1] ($m_{WDM} > 1.4 - 5.3$ keV, [72, 158]), but we also have constraints coming from: phase-space considerations [21, 70], Milky Way satellite abundance [87], gravitational lensing [98], the earliest epoch of star formation [111] and the luminosity function of high-z galaxies [47, 80, 97]. In these works the advocated values of m_{WDM} are different and sometimes in disagreement, e.g. for solving the missing-satellite problem in [87] they assert $m_{WDM} \gtrsim 2$ keV is needed, while in [134] a mass of $m_{WDM} < 2$ keV is claimed for solving the too-big-to-fail problem.

In brief, the picture is complex, also because we need to keep in mind that baryonic processes are at play and there is lack of consensus on how to model those. The interest is on WDM with mass of the order of few keV, and in such warm cosmologies structure formation is nearly indistinguishable from standard CDM. Here lies the importance and novelty of our study.

We investigate the impact of WDM in terms of 21cm intensity mapping (IM) and find how the suppression of power present in the linear and non-linear matter power spectra of WDM, results in an increase of power in terms of neutral hydrogen (HI) and 21cm power spectra: the differences of these scenarios in 21cm IM maps are significant and makes them eventually distinguishable.

We perform the study at $z = 3 - 5$, i.e. after reionization (EoR) has taken place. These redshifts are relatively new for probing WDM models in the sense that they partially overlap only with Lyman-α forest data, while are lower compared to early structure formation regimes investigated by [54, 89, 139]. Moreover, this redshift range is at the same time far from the complex astrophysics of reionization and also possibly poorly affected by galactic feedbacks, making our modelling more reliable.

1 Although they depend on the assumed intergalactic medium temperature-density relation and on the data set used.

Our analysis is limited to WDM, however the transfer function induced (i.e. the way the matter power spectrum is modified) can be traced back to other viable non-standard DM models[2], most notably the thermally produced sterile neutrino [1]. Hence, the effects studied here can be extended to other models, for example as we do in Chapter 5.

This is how we proceed: we perform hydrodynamic simulations and distribute HI a-posteriori with the methods described in Chapter 3, we quantify the WDM effect and address significance for the telescope SKA1-LOW by including a realistic noise modelling.

4.1 SIMULATIONS AND MATTER DISTRIBUTION

4.1.1 *Hydrodynamic simulations*

Our simulation suite comprises a set of 5 high-resolution hydrodynamic N-body simulations, one for each of the 5 cosmological model under analysis: 1 reference model with standard cold DM and 4 models with warm DM, each of them with a different thermal relic mass of the DM particles:

$$m_{WDM} = \quad 1 \text{ keV}, \quad 2 \text{ keV}, \quad 3 \text{ keV} \quad \text{and} \quad 4 \text{ keV}.$$

All simulations are run using the TreePM+SPH code GADGET-III [144] with which we follow the evolution of 512^3 CDM/WDM and 512^3 baryon (gas+stars) particles within simulation boxes of comoving sizes equal to $30\,h^{-1}$Mpc. The mass resolution of our simulations is

$$m_{DM} = 1.50 \times 10^7\ h^{-1} M_\odot \quad \text{mass of the dark matter particle}$$
$$m_b \quad = 2.74 \times 10^6\ h^{-1} M_\odot \quad \text{mass of baryon particle (gas+stars).}$$

The gravitational softening is set to 1/40 of the mean inter-particle linear spacing, i.e. $1.5\ h^{-1}$kpc. The values of the cosmological parameters are the same in all the models and in agreement with the latest results of the Planck satellite [120]:

$$\Omega_m = 0.3175 \qquad \Omega_b = 0.049 \qquad \Omega_\Lambda = 0.6825$$
$$h = 0.6711 \qquad n_s = 0.9624 \qquad \sigma_8 = 0.834\ .$$

We model star formation using the effective multi-phase model of Springel & Hernquist [145]. The code also simulates radiative cooling by hydrogen and helium, and heating by a uniform Ultra Violet (UV) background. We modify both the cooling routine and the UV background to obtain a desired thermal history, corresponding to the reference model of [158].

We generate initial conditions at $z = 99$ using the Zel'dovich approximation, using same seed for all cosmologies. For the model with

2 Recently in [103] they propose a new general way of fitting the transfer function for including more DM models.

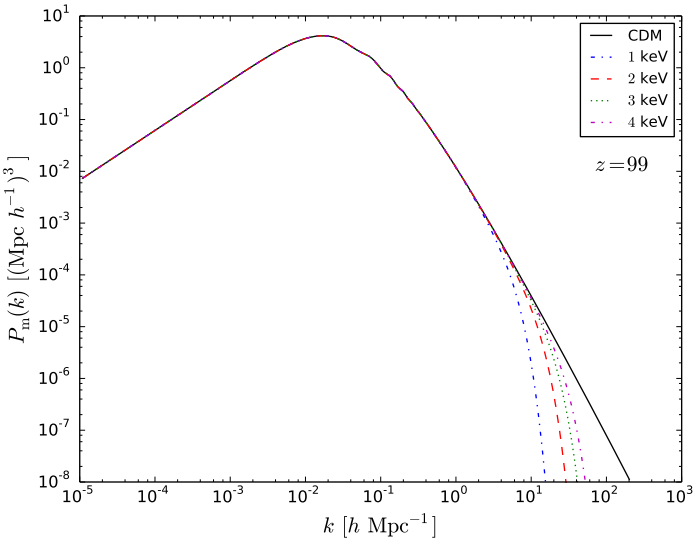


Figure 4.1: The initial condition $P_m(k)$ of all 5 simulations. The CDM (solid black line) one is calculated with CAMB, that has been multiplied by the transfer function Equation 4.2 for getting the WDM.

CDM the matter power spectrum and the transfer functions are computed using the publicly available Einstein-Boltzmann solver code CAMB [81], whereas for the models with WDM the power spectra, $\tilde{P}_a(k)$, are calculated as [157]

$$\tilde{P}_a^{\Lambda WDM}(k) = T_{lin}^2(k) P_a^{\Lambda CDM}(k) \tag{4.1}$$

where a stands for either baryonic and the non-baryonic matter components and with

$$T_{lin}(k) = \left(1 + (\alpha k)^{2\nu}\right)^{-5/\nu} \tag{4.2}$$

where $\nu = 1.12$ and

$$\alpha(m_{WDM}) = 0.049 \left(\frac{1\ \text{keV}}{m_{WDM}}\right)^{1.11} \left(\frac{\Omega_{WDM}}{0.25}\right)^{0.11} \left(\frac{h}{0.7}\right)^{1.22} h^{-1}\text{Mpc}\,. \tag{4.3}$$

In Figure 4.1 we show the initial conditions for all 5 simulations. Indeed, the WDM models are completely equivalent to CDM at large scales. The 1 keV WDM $P_m(k)$ starts diverging at $k \simeq 3h\,\text{Mpc}^{-1}$, at bigger k's for higher mass m_{WDM}. The $P_m(k)$ cut-off shown in Figure 4.1 is that discussed in Section 1.1.2, responsible for making these WDM models alleviate the small scale issues CDM is challenged by.

Furthermore, for the cosmological models with WDM we add thermal velocities, on top of the peculiar velocities, to the non-baryonic particles. For a given WDM particle, the modulus of the thermal ve-

locity vector is drawn randomly from a Fermi-Dirac distribution with a mean equal to [19]

$$\bar{V}_{WDM}(z) = 0.012(1+z)\left(\frac{\Omega_{WDM}}{0.3}\right)^{1/3}\left(\frac{h}{0.65}\right)^{2/3}\left(\frac{keV}{m_{WDM}}\right)^{4/3} \text{ km/s} \tag{4.4}$$

and the direction of the velocity vector is taken randomly[3].

We analyse snapshots at $z = 3$, 4 and 5 and identify the dark matter halos using the Friends-of-Friends (FoF) [52] algorithm with linking parameter length set to $b = 0.2$.

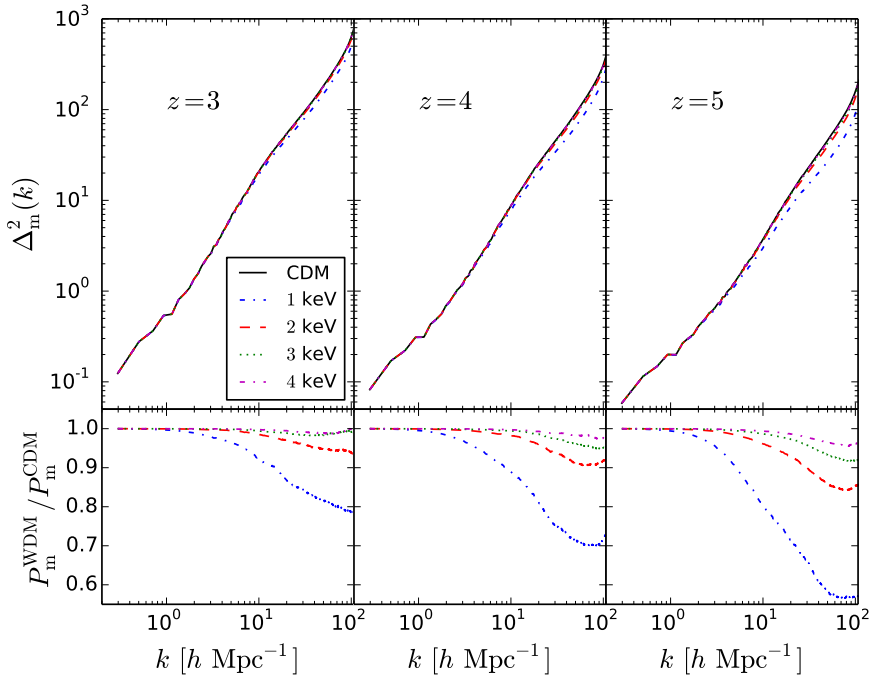
4.1.2 *Impact of WDM on the matter distribution*

We now examine the impact of WDM on the spatial distribution of matter. In particular, we focus our attention on the matter power spectrum and the halo mass function.

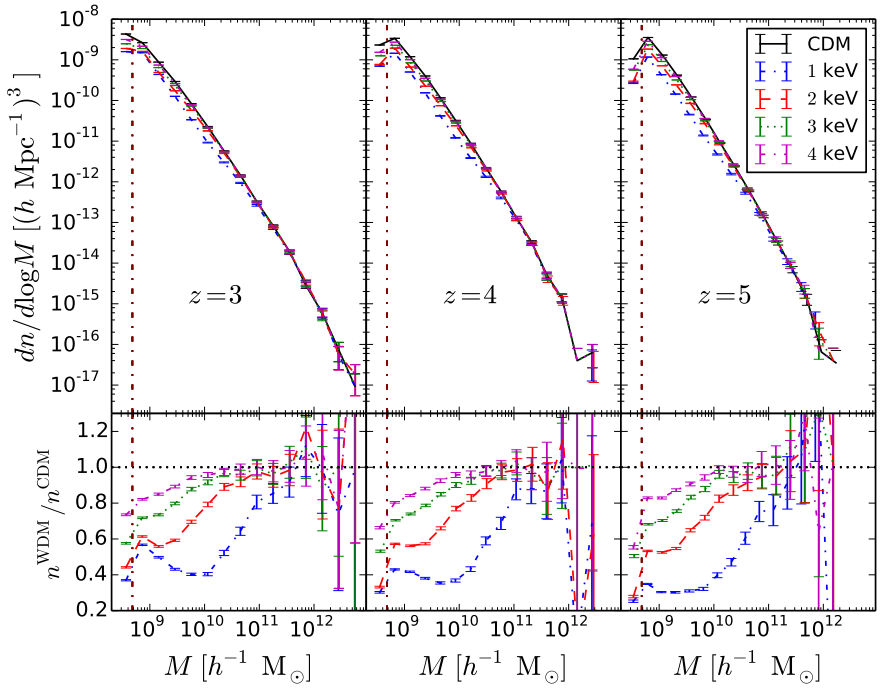
In the top panels of Figure 4.2a we show the total matter (baryons and DM) dimensionless power spectrum $\Delta^2_m(k) = k^3 P_m(k)/(2\pi^2)$ for each of our 5 different cosmologies at the redshifts analysed: $z = 3$, 4 and 5. The bottom panels represent instead the differences between warm models and the cold one. Since WDM can not substantially cluster on scales smaller than its free-streaming length, the matter power spectrum in WDM cosmologies presents a relatively sharp cut-off on small scales; as expected, this suppression is larger at higher redshift and for smaller masses m_{WDM}, i.e. warmer models. The clustering on scales larger than the free-streaming length is not affected by the thermal velocities of the WDM particles, thus the amplitude of the matter power spectrum is the same for all the models, i.e. at large scales cold and warm DM models are indistinguishable, as discussed in Section 1.1.2 and at the beginning of this chapter.

These cut-off at the matter power spectra level directly translate in different abundances of halos of different mass. Indeed, let us have a look at the halo mass function: we plot them in Figure 4.2b for all 5 scenarios. In the top panels we plot the halo mass function at $z = 3$ (left), $z = 4$ (middle) and $z = 5$ (right). In the bottom panel, we plot the difference between each warm model and the reference CDM. As it can be seen in the figure, WDM induces a suppression on the abundance of low mass dark matter halos, with respect to the abundance of halos in the CDM model, which increases with redshift

3 But see also [7], where they point out how, since simulation particles are a coarse-grained representation of the phase-space distribution, this additional kick is equivalent to inducing a local coherent motion of a large ensemble of microscopic particles. However, the $\bar{V}_{WDM}$ of Equation 4.4 is several order of magnitudes smaller than that arising during the non-linear gravitational collapse, so it has a negligible effect on the halo mass function, that we are most interested in. At the same time in [132] they claim that adding the extra velocity kick is crucial for preventing *spurious fragmentation*, a numerical issue linked to WDM simulation that we discuss in next paragraph.



(a) Dimensionless power spectrum of total matter (DM + baryons), $\Delta^2_{\rm m}(k) = k^3 P_{\rm m}(k)/(2\pi^2)$.



(b) Halo mass function. The dash-dotted vertical line highlights the mass of a FoF group with 32 DM particles. Error bars represent the uncertainty in the halo mass function assuming that the number of halos follows a Poissonian distribution.

Figure 4.2: Matter power spectrum (a) and halo mass function (b) at redshifts $z = 3$ (left), $z = 4$ (middle) and $z = 5$ (right). Results for the CDM model is shown with a solid black line whereas the colored lines represent the results for the WDM models: 1 keV mass (dash-dotted blue), 2 keV (dashed red), 3 keV (dotted green) and 4 keV (dash-dotted magenta). The bottom panels show the ratios between the WDM models and the reference CDM model.

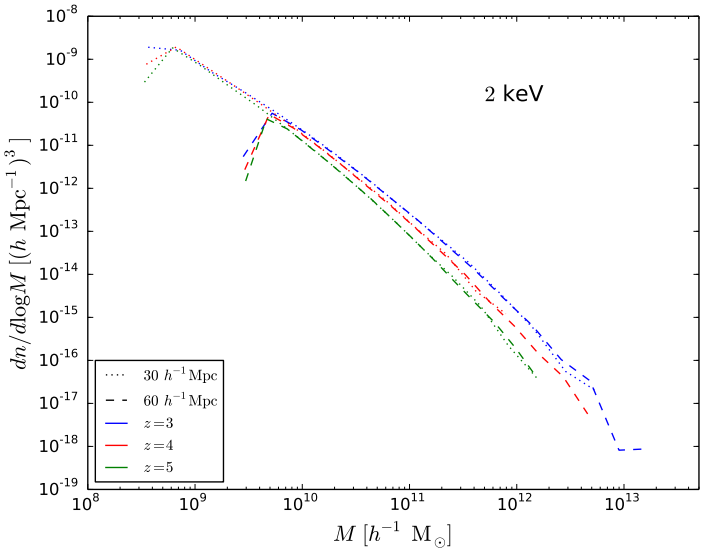


Figure 4.3: Halo mass function of the 2 keV WDM cosmology simulation at redshifts $z = 3$ (blue), $z = 4$ (red) and $z = 5$ (green) for the FoF catalogues derived from the $30\,h^{-1}$Mpc box size simulation mainly used in this work (dotted) and a $60\,h^{-1}$Mpc (dashed).

and decreases with the mass of the WDM particles. As we shall see later, this effect will drive the changes we observe in the 21cm power spectrum between the models with cold and warm DM.

Error bars plotted in Figure 4.2b are calculated assuming the number density of halos follow a Poissonian distribution. Thus, this is why the relatively large errors in the halo mass functions are associated with the most massive halos, i.e. with halos of masses $M \gtrsim 10^{11} h^{-1} M_\odot$: we find few of them in our simulation boxes. However, despite their limited number, they heavily influence the results of this analysis, since they are expected to host a considerable amount of the HI we are willing to model.

To reduce the uncertainty on this leg of the halo mass function, we check its convergence with what obtained in a double size simulation ($60\ h^{-1}$Mpc) for all cosmologies. For example, in Figure 4.3 we show the 2 keV WDM cosmology case, for all redshifts. The $30\ h^{-1}$Mpc halo mass function for massive halos overlaps with the $60\ h^{-1}$Mpc counterpart, assuring us that we are dealing with a correct amount of massive halos in our analysis.

4.1.2.1 *A closer look to the halo mass function*

In deriving the mass function of warm DM simulations, a numerical artefact known as *spurious fragmentation* could produce an increase of halos at a low mass scale, see for example [66, 133, 165]. We believe that our simulations, except the model with 1 keV as we will later discuss, are not affected by such effect, both because we artificially

m_{WDM} [keV]	z	α	M_0 [10^9 $h^{-1}M_\odot$]
1	3	-1.72	0.887
	4	-1.92	0.960
	5	-2.06	0.737
2	3	0.571	3.98
	4	0.389	3.80
	5	0.286	4.81
3	3	0.830	1.86
	4	0.726	1.28
	5	0.427	1.74
4	3	0.958	1.44
	4	1.04	0.890
	5	0.522	1.05
∞	3	1.26	1.04
	4	2.77	0.491
	5	1.69	0.455

Table 4.1: Parameter values for the mass function fit of equation Equation 4.5.

boost late time thermal velocities at the beginning of our simulations (and this is shown to alleviate this issue in [132]) and because we do not reach enough mass resolution for our WDM models, which have higher thermal masses compared to those in the literature. This can be directly seen from the behaviour of the mass function curves in Figure 4.2b, that do not show any characteristics steep power law upturn for low halo masses. However, we see an upturn in the ratio n^{WDM}/n^{CDM} in the bottom panels of Figure 4.2b for the 1 keV case (and slightly for the 2 keV), as it raises weakly until 10^9 $h^{-1}M_\odot$ and this behaviour could be explained by spurious fragmentation. It is also true that the state of the art for numerical fragmentation is focused on redshift $z = 0$ halos, therefore a direct comparison of our results with other works can not be easily done.

Briefly, these simulations are not warm and resolute enough to be seriously affected by this numerical problem. Later in Chapter 5 we employ simulations where we do have to deal with numerical fragmentation and we will enter in more details.

Still looking at the ratio n^{WDM}/n^{CDM} in the bottom panels of Figure 4.2b, a turnaround or a change in slope can be noticed towards the very lower mass halos. We think this behaviour is due to the proximity to the resolution limit of our simulations (vertical dotted line in

Figure 4.2b is a 32 DM particle halo mass limit), and therefore the abundance of those low mass halos is slightly affected by resolution[4].

We fit the WDM mass functions obtained from the simulations by parametrizing their deviation from the mass function predicted by the Sheth & Tormen [138] formula[5]

$$\frac{dn^{sim}_{WDM}}{d\log M}(M) = \frac{dn^{ST}_{WDM}}{d\log M}(M)\left[1 - \alpha e^{-M/M_0}\right], \tag{4.5}$$

where α and M_0 are free parameters. The best fit values for the different WDM models and redshifts are shown in the Table 4.1. We find that the mass function of the 1 keV WDM cosmology is the only one that is under predicted by the Sheth & Tormen model. We exclude the possibility of fragmentation, because n^{sim}_{WDM} lacks the associated steeper raise for low halo masses (as already discussed above), although this effect could point to a transitional and physical regime before fragmentation takes place (see figure B1 in [132]).

We also find that the mass function of the CDM model is over predicted by the Sheth & Tormen formula at these redshifts, although only in the low mass end; indeed, in Table 4.1 we show also the best fit values for $m_{DM} = \infty$, i. e. cold DM.

Overall, for WDM masses between 3 and 4 keV we observe a suppression in power up to $\sim$ 10% on very small scales and a reduction in the number of 10^9 $h^{-1}M_\odot$ halos of the order of 20-40% compared to the CDM case.

4.2 IMPACT OF WDM ON THE HI DISTRIBUTION

We distribute HI in the simulations following 2 methods: the halo based model A of Section 3.1 and the particle based of Section 3.2. As already discussed in Chapter 3, in the halo based model A, we set lower and upper cut-off in halo mass that host HI and this could seem too coercive: variations in these parameters could lead to an important difference in the HI bias. We overcome this by adopting also a second HI distribution method that overlooks any DM halo issue: the particle based method.

The magnitude of M_{min} is of the order of 10^9 $h^{-1}M_\odot$, so to distribute HI consistently, our simulations need to resolve dark matter halos up to this mass halos or lower. We achieve this by setting the box size of our simulations to 30 h^{-1}Mpc, among the highest values possible considering the number of particles we are dealing with: 2×512^3.

4 With the exception of the 1 keV model, that displays this change in slope a decade before the resolution mass range. Again, our warmest model might experience spurious fragmentation.

5 This is achieved by computing the Sheth & Tormen mass function using the linear power spectrum of the WDM model as in Equation 4.1.

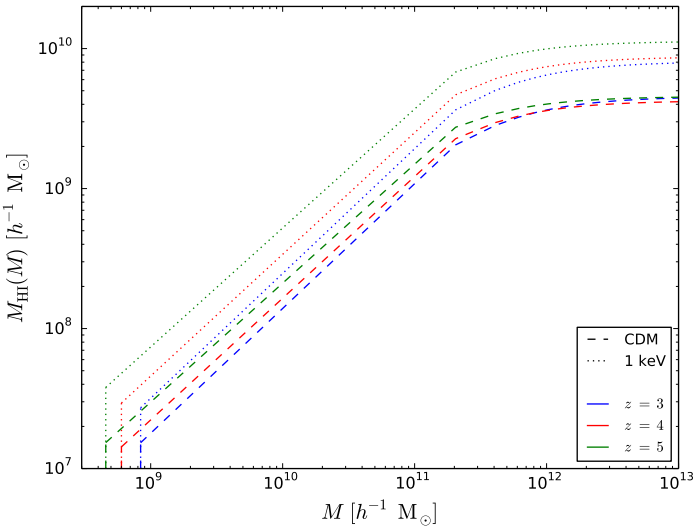


Figure 4.4: The function $M_{HI}(M)$ relating the HI mass and total mass of a dark matter halo of mass M, employed in the halo based model A as described by the equation Equation 3.1, for the CDM (dashed lines) and 1 keV WDM (dotted lines) cases, for redshifts $z = 3, 4, 5$ (blue, red and green). The values of the parameter f are 0.0183, 0.0239, 0.0337 for the CDM and 0.0325, 0.0491, 0.0835 for the 1 keV, redshift 3 to 5.

We show the function $M_{HI}(M)$ of Equation 3.1 that the halo based method prescribes for the models with CDM and 1 keV WDM in Figure 4.4. The $M_{HI}(M)$ function steeply increases up to M_{max}, then displays a plateau. The values of M_{min} and M_{max} do not change among different cosmologies, but the normalisation does, through the parameter f of Equation 3.1 where we set $\Omega_{HI} = 10^{-3}$: M_{HI} gets higher at fixed M_{halo} for warmer DM scenarios.

We do not have any guarantee that the halo based model works also for the warm cosmologies, i.e. we cannot distinguish the effect of WDM from the physics behind the HI abundance in different environments. However, the additional use of the particle based method partially takes care of this issue.

An overall picture of the spatial distribution of matter and of neutral hydrogen is given in the snapshots of Figure 4.5, where we confront the CDM scenario (left column) with the 1 keV WDM (right) at redshift $z = 3$. In the top panels we colour code the total matter density contrast: CDM displays much more clustered field than WDM. The spatial distribution of HI also differs in the two cosmologies (middle and bottom panels). By using the halo based method, HI is present just in halos (middle panels), while by employing the particle based (bottom panels), the HI distribution is smoother since HI is assigned to single gas particle in the simulation, i.e. the yellow dots in the middle panels are replaced by a more web-like structure in the bottom. Still the cold - warm clustering differences are noticeable by eye for both HI mass assignments.

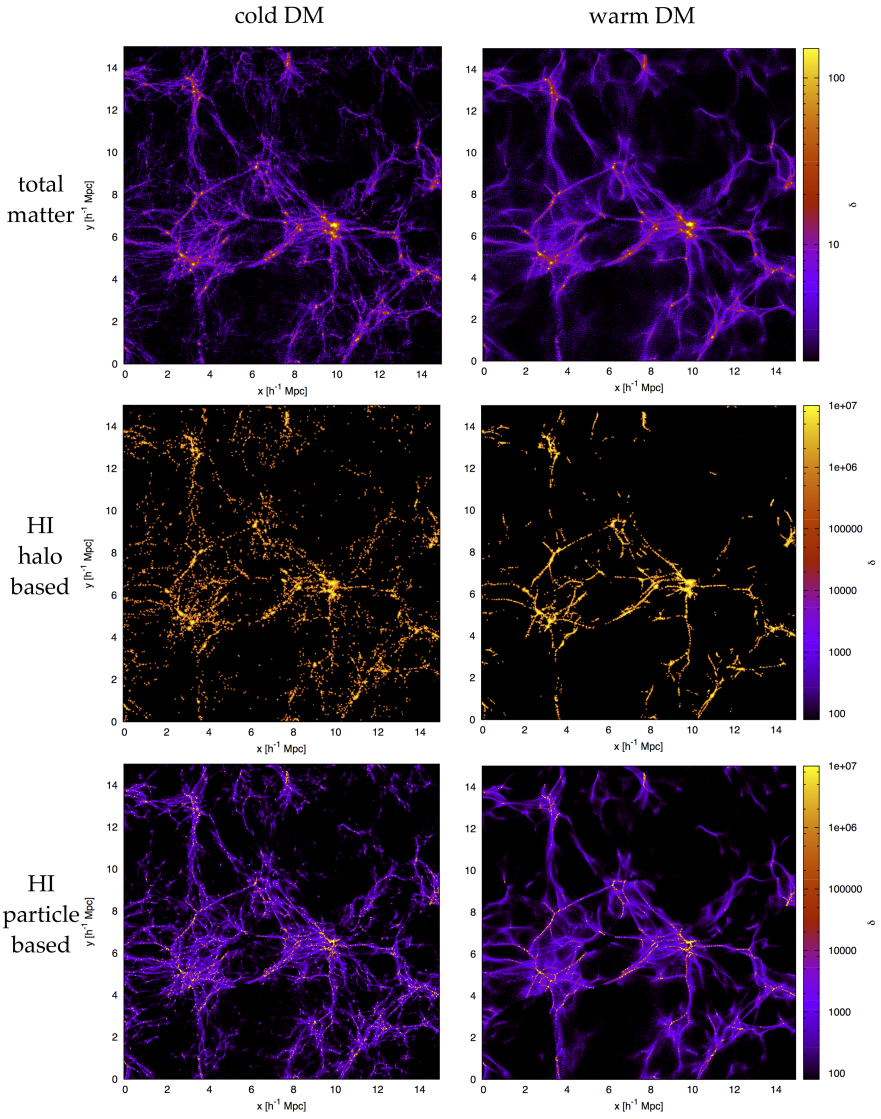


Figure 4.5: Spatial distribution at redshift $z = 3$ of the density contrast of total matter (DM + baryons) and of the HI placed according to the halo and particle based models, upper, middle and low row respectively, for the CDM (left column) and 1 keV WDM (right column) scenarios. We zoom in a $(15\,h^{-1}\mathrm{Mpc})^2$ region, taking a slice of $2\,h^{-1}\mathrm{Mpc}$ width.

We perform a more quantitative analysis of the HI distribution. In next Section 4.2.1 we start by comparing the HI column density distribution function that we obtain for each cosmological model with observations, then we investigate the differences in terms of HI power spectrum and of the HI bias $b^2_{HI}(k) = P_{HI}(k)/P_m(k)$.

4.2.1 *The HI clustering and power spectrum*

We now investigate the impact of WDM on the spatial distribution of neutral hydrogen. First of all we compare our scenarios with the available data. We compute the HI column density function in all 5 simulations and compare them with available observations, as done in Section 3.3. Results are shown in Figure 4.6 for both the halo based and particle based method, together with the observational measurements. For clarity, we only show the results for 2 cosmological models: the models with CDM and 1 keV WDM, since these are the extreme scenarios that together span the range of halo mass functions probed by all other models.

As already discussed in Section 3.3, the HI column density distribution obtained by distributing the neutral hydrogen with the 2 different methods considered is in overall good agreement with the data. Moreover, they are somehow complementary: the halo based overestimates what the particle based model underestimates, leaving the data points wrapped by the two approaches.

In Figure 4.6 we show in blue the CDM prediction and in red the WDM one. Since both scenarios are equivalent in terms of the abundances of the massive halos, they also agree on the higher HI column density leg of the distribution function, interestingly also for the particle based model that does not rely directly on the halos information of the cosmology. Instead, the low HI column density objects exhibit differences in number for the different cosmologies, for both methods. In general and not surprisingly, the warm DM scenario gives less of these objects (since they lack of small mass halos), but it would be pretentious to discriminate the DM models by looking at this statistic since the HI assignment methods themselves do not converge (see discussions in Chapter 3).

We continue analysing the HI field. The HI power spectra $P_{HI}(k)$ of all 5 scenarios (cold and warm DM) is plotted in Figure 4.7a for HI distributed following the halo based method A and in Figure 4.7b for the particle based one. The 5 $P_{HI}(k)$ get closer to each other, at all redshifts, on large scales, although they never exactly converge as happen for the total matter power spectra $P_m(k)$ showed in figure 4.2a.

In the bottom panels of Figure 4.7a and Figure 4.7b, we plot the ratio between the HI power spectra of the models with WDM to the one of the model with CDM. As expected, the lighter the WDM mass,

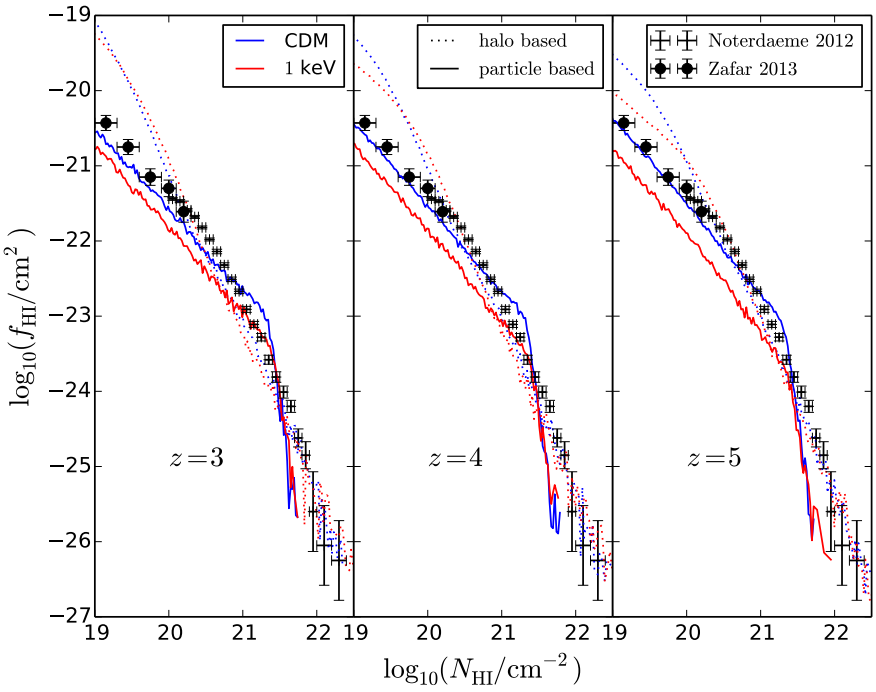
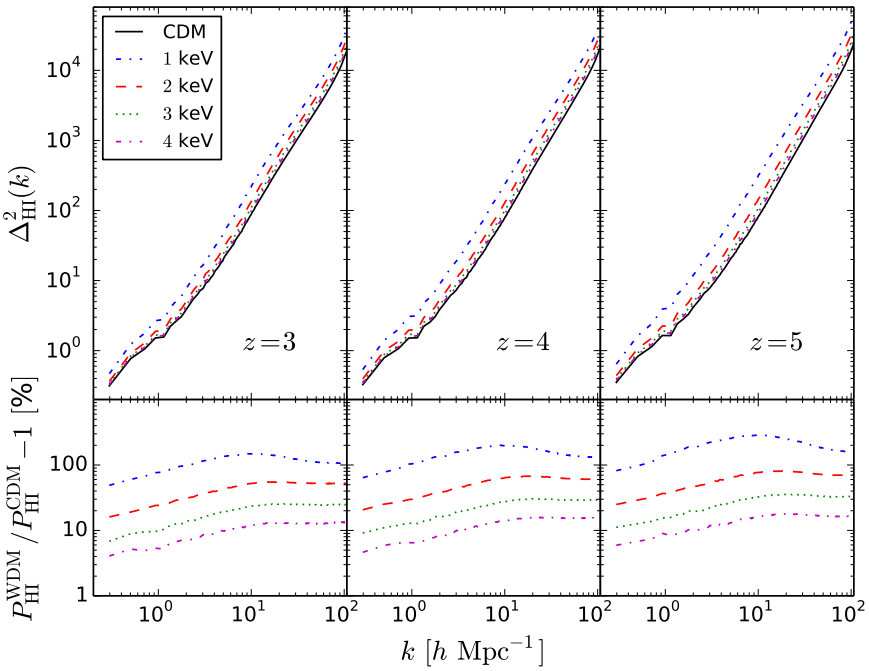


Figure 4.6: HI column density distribution function $f_{HI}(N_{HI})$ obtained by assigning HI to gas particles employing 2 different methods: halo based A (dotted lines) and particle based (solid lines). Observational data are plotted in black for [106] (crosses) and for [171] (circles). We show results at redshifts $z = 3$ (left), $z = 4$ (middle) and $z = 5$ (right). In each panel we plot the results for the CDM simulation in blue whereas results for the 1 keV WDM cosmology are plotted in red.

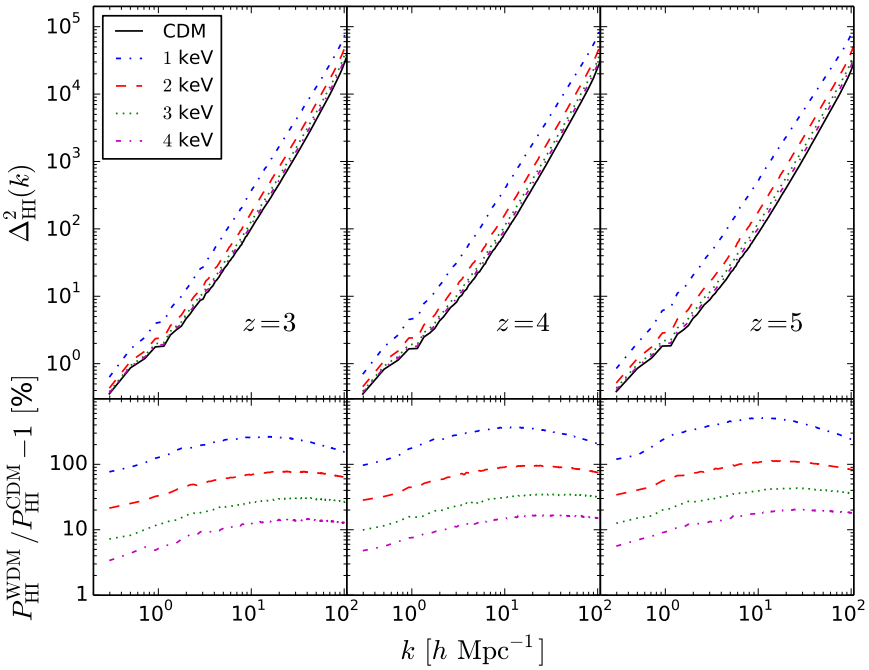
the bigger is the discrepancy with the CDM case, with an increase of power in the HI clustering. We find that whereas the HI power spectrum for the 1 keV model has an amplitude roughly a factor 2 higher than the one from the CDM model, almost independently on the model used to distribute the HI, differences between the 4 keV and CDM models are smaller than $\sim 10\%$.

With the halo based prescription, the difference in HI power spectrum is at least of order $\sim 5\%$ for the 4 keV WDM cosmology at redshift $z = 3$ and goes up to over order $\sim 100\%$ for the 1 keV WDM at $z = 5$. By employing the particle based, the differences are even higher, reaching more than 300% in the most extreme case of 1 keV at $z = 5$.

The reason why the amplitude of the HI power spectrum is higher in the WDM models with respect to the CDM is manifest when HI is distributed according to the halo based method: we need to accommodate the same amount of HI in all simulations and, since there is a deficit of low mass halos in the warm cosmologies (in comparison to the cold one), we need to put more HI into the remaining halos, i.e.



(a) HI distributed according to the halo based method A.



(b) HI distributed according to the particle based method.

Figure 4.7: Dimensionless HI power spectrum, $\Delta^2_{\rm HI}(k) = k^3 P_{\rm HI}(k)/(2\pi^2)$, at redshifts $z = 3$ (left), $z = 4$ (middle) and $z = 5$ (right) obtained by assigning HI to gas particles according either to the halo based method (a) or to the particle based method (b). In each panel we plot the results of the CDM simulation with a continuous black line and for the WDM ones: 1 keV mass (dash dotted blue), 2 keV (dashed red), 3 keV (dotted green) and 4 keV (dash dotted magenta). The relative difference, in percentage, of the WDM models with respect to the CDM scenario is shown in the bottom panels. We set the k range up to the Nyquist frequency ($k \simeq 107\,h\,{\rm Mpc}^{-1}$).

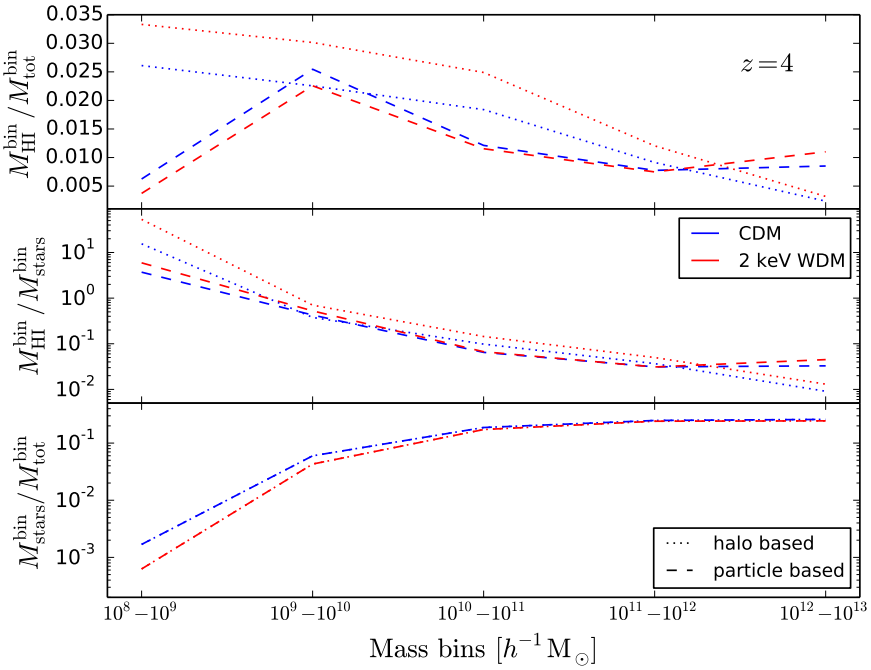


Figure 4.8: Ratios of the total HI mass over the matter M_{HI}/M_{tot} (upper panel), over the stellar mass M_{HI}/M_{stars} (middle panel) and the total stellar mass over the total mattel M_{stars}/M_{tot} (lower panel) calculated as function of the halo mass bins within which the ratio is performed, for the CDM (blue lines) and 2 keV WDM (red lines) cosmologies with the halo based (dotted lines) and particle based (dashed lines) HI assignment methods at redshift $z = 4$.

the more massive ones. Thus, the HI gets more clustered in the more massive halos (that are also those to have a larger bias).

When the HI is modelled using the particle based method, the reason of larger clustering power of the neutral hydrogen in the WDM cosmologies is not that obvious, so it is assuring that the results with both methods point toward the same effect.

To further investigate this issue we perform the following test. We select all DM halos in a given mass interval and compute the sum of the HI mass, stellar mass and total mass within those halos for the models with CDM and 2 keV WDM. In Figure 4.8 we plot the ratio between the total HI mass and the total mass contained in the halos in the selected mass range (top panel), the ratio between the total HI and stellar mass (middle panel) and the ratio between the total stellar and total mass (bottom panel).

Looking at the ratio $M_{HI}^{bin}/M_{tot}^{bin}$ in the top panel of Figure 4.8, we see the cut-off of the HI in the high mass halos predicted by the halo based model A, whereas the particle based results do not show such decline. From that panel it is also clear that halos of the same mass have more HI in the models with WDM with respect to the CDM model, when the HI is modelled using the halo based method.

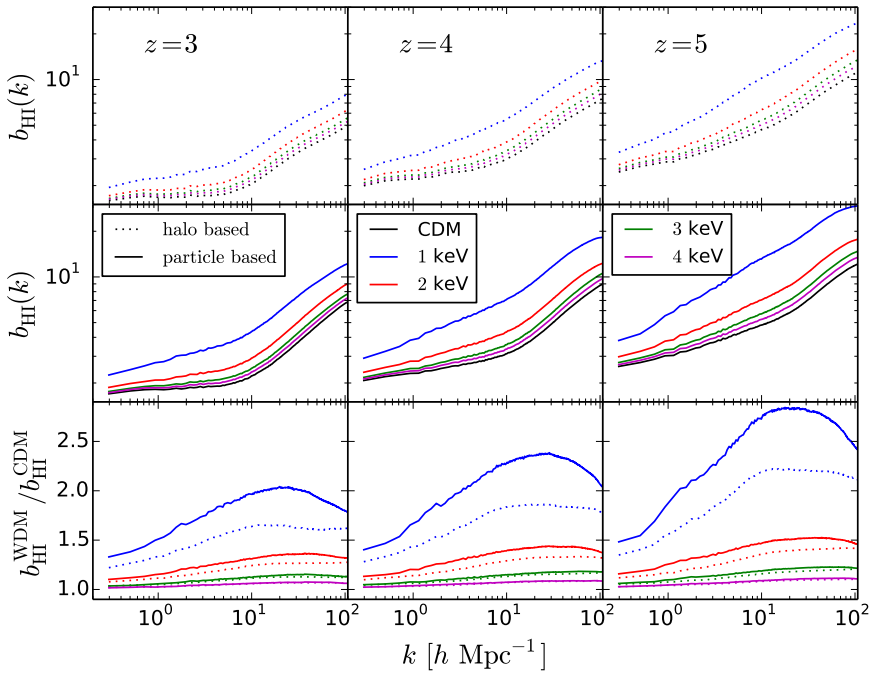


Figure 4.9: HI bias $b_{\rm HI}(k)$ at $z = 3$, 4 and 5 from left to right when the HI distribution is modelled using the halo based (top row) and the particle based (middle row) methods. In each panel we plot the results for the CDM model with black lines whereas: 1 keV WDM model (blue), 2 keV (red), 3 keV (green) and 4 keV (magenta). The ratio between the bias of models with WDM over that of CDM is displayed in the bottom panels.

$M_{\rm HI}^{\rm bin}/M_{\rm tot}^{\rm bin}$ is more similar in the two different cosmologies when the HI is modelled using the particle based method. These results point out that the amount of HI per total mass in a given dark matter halo is almost the same between halos in CDM and WDM cosmologies. Thus, the larger amplitude of the HI power spectrum in the models with WDM arises because the bias of the dark matter halos is higher in the WDM models.

The ratio $M_{\rm HI}^{\rm bin}/M_{\rm stars}^{\rm bin}$ in the middle panel of Figure 4.8 is higher for WDM and comparable among the halo and particle based methods for the central mass range bins. The ratio of the stellar mass over total mass $M_{\rm stars}^{\rm bin}/M_{\rm tot}^{\rm bin}$ in the lower panel of course does not change with the different HI assignment methods, and for less massive halos it is slightly lower for WDM: it could be sign of less active star formation in the WDM halos (but see for example [54, 89]). The latter result also points towards the necessity of studying the Epoch of Reionization (EoR) in WDM scenarios, as we discuss in the thesis outlook in Section 8.2.

To check how the total matter clustering properties are reflected by those of neutral hydrogen, we analyse in more detail the different bias that HI exhibit in cold and warm DM cosmologies.

After having computed the total matter and the neutral hydrogen power spectra, $P_m(k)$ and $P_{HI}(k)$, we can make an estimate the HI bias, $b_{HI}(k)$, defined as:

$$b_{HI}^2(k) = \frac{P_{HI}(k)}{P_m(k)} \,. \tag{4.6}$$

The box size of our simulations ($30\,h^{-1}$Mpc) is not big enough to study the *true* scale-independent linear bias predicted by linear theory (as we do in Chapter 7). Still we compute the quantity b_{HI} defined in Equation 4.6 to compare the HI clustering with the matter field of each of the cosmology considered.

In Figure 4.9 we plot the HI bias at redshifts $z = 3$ (left), $z = 4$ (middle) and $z = 5$ (right) when the HI distribution is modelled using the halo based A (top row) and the particle based (middle row) methods. The spatial distribution of neutral hydrogen is more biased in the WDM models: the bias increases with decreasing WDM mass, at all redshift and for both methods. The particle based produces a slightly higher HI bias in comparison to the results obtained by employing the halo based model. This can be more easily seen in the bottom panels of Figure 4.9, where we plot the relative difference in the bias between the models with WDM and CDM, for both methods.

Briefly, neutral hydrogen is more strongly clustered in warm than in cold DM cosmologies, and indeed its bias is higher in the first case.

The detectability of a WDM signature through the 21cm signal is mainly possible thanks to the boost of the HI bias shown above. We show in next section that this boost can easily be higher than the instrumental noise of, for example, SKA1-LOW, making possible to tell apart these warm models from CDM.

4.2.2 *The 21cm power spectrum and SKA1-LOW forecasts*

We compute the 21cm power spectrum as in Equation 3.10 for the 5 cosmological models analysed. For the halo based method $\Omega_{HI}^{ref} = 10^{-3}$ is fixed in all scenarios by construction, whereas for the particle based, we have different values of Ω_{HI}^{sim} for each simulation, as previously pointed out.

Since we want to investigate differences between models, we can assume that the value of Ω_{HI} is fixed by independent observations (such as the abundance of DLAs and LLS), than a further normalization is needed when computing the 21cm power spectrum, to force all models to have the same value of Ω_{HI}[6]:

$$(\Delta_{21cm}^2(k))^{norm} = \Delta_{21cm}^2(k) \left(\frac{\Omega_{HI}^{ref}}{\Omega_{HI}^{sim}} \right)^2 , \tag{4.7}$$

6 Notice that by doing this we are assuming that the clustering properties of the neutral hydrogen do not change, i.e. $P_{HI}(k)$ does not change.

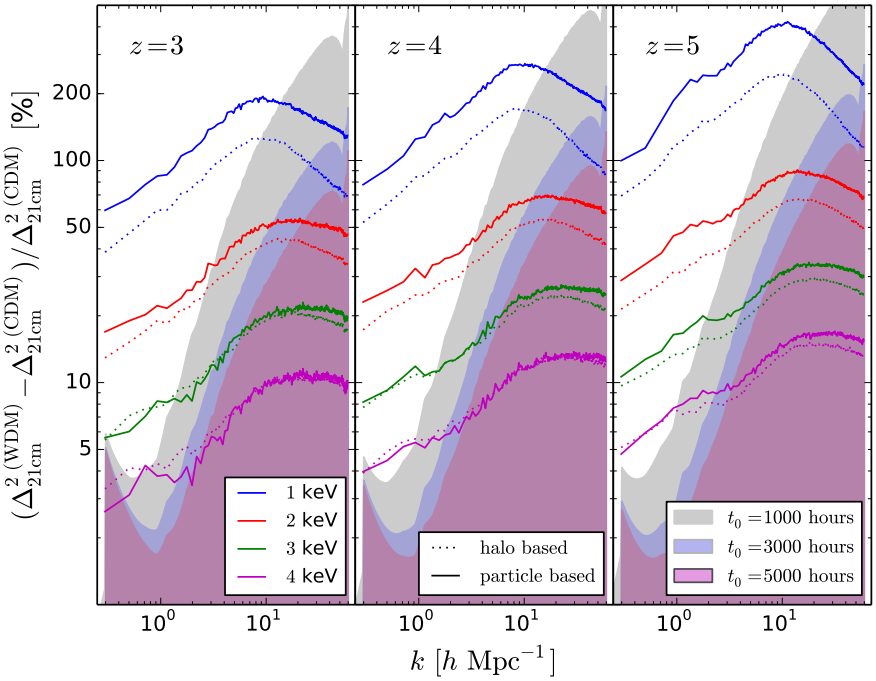


Figure 4.10: Relative difference between the 21cm power spectrum of the models with WDM and CDM when the HI distribution is modeled using the halo based method A (dotted lines) and the particle based method (solid lines). Results are shown at $z = 3$, 4 and 5 from left to right. The error on the 21cm power spectrum of the model with CDM, normalized to the amplitude of the 21cm power spectrum, $\sigma\left[\Delta^{2\,(\mathrm{CDM})}_{21\mathrm{cm}}\right]/\Delta^{2\,(\mathrm{CDM})}_{21\mathrm{cm}}$, is shown with a shaded region for three different observation times: $t_0 =$ 1000 (grey), $t_0 = 3000$ (blue) and $t_0 = 5000$ hours (fuchsia). For clarity, we show the error on $\Delta^{2\,(\mathrm{CDM})}_{21\mathrm{cm}}$ from one HI assignment method only because both are very similar and practically overlap for this plot size.

where $\Omega^{\mathrm{sim}}_{\mathrm{HI}}$ is the value of Ω_{HI} directly obtained by employing the particle based method, shown in Figure 4.11.

In Figure 4.10 we show the relative difference in the 21cm power spectrum between the WDM models compared to the cold DM one. Dotted lines represents the results when the halo based model is employed whereas solid lines for the particle based model.

Not surprisingly, following the HI power spectrum, the amplitude of the 21cm power spectrum is higher in the models with warm rather than cold DM, as consequence of having HI stronger clustered in those cosmologies. Thus, the amplitude of the signal is higher in the WDM scenarios, making it more easily detectable.

To quantify the detectability of the 21cm line in intensity mapping for the warm DM cosmology, we compute the errors with which the SKA1-LOW radio telescope will measure the signal power spectrum, $\sigma^2[P_{21\mathrm{cm}}(k,\theta)]$, as described in Section 3.4.2. We model the system noise only, assuming that astrophysical foregrounds, radio frequency

interference and others nuisances have been already removed from the observed data.

The shaded areas in Figure 4.10 show the quantity $\sigma\left[\Delta^{2\,(\mathrm{CDM})}_{21\mathrm{cm}}\right]/\Delta^{2\,(\mathrm{CDM})}_{21\mathrm{cm}}$, where $\sigma\left[\Delta^{2\,(\mathrm{CDM})}_{21\mathrm{cm}}\right]$ represents the 1σ error on the 21cm dimensionless power spectrum of the model with CDM, computed using Equation 3.15. Results are shown for observations times equal to $t_0 = 1000$ hours (grey), $t_0 = 3000$ hours (blue) and $t_0 = 5000$ hours (fuchsia).

With a reasonable observational time of $t_0 = 1000$ hours [123] the WDM models with 1, 2, and 3 keV can be fully distinguished from CDM. On the other hand, the model with 4 keV WDM is consistent with CDM at the $\sim 1\sigma$ confidence level at $z = 5$. With longer observing times, e.g. 3000 hours, even the model with 4 keV can be distinguished from the model with CDM at more than 2σ at $z = 5$. We note that on large scales, the error budget is dominated by cosmic variance, thus, the error magnitude barely changes by increasing the observation time. Extending the observation time to $t_0 = 5000$ hours, also the 4 keV case is above the error at $\sim 2.5\sigma$ confidence level for 21cm emission coming from redshift $z = 3$. These results are valid for both the HI assignment methods considered.

4.3 THE Ω_{HI} - WDM MASS DEGENERACY

We need to acknowledge that other effects could mimic the same increase in 21cm signal, as for example a higher value of the HI density parameter Ω_{HI}, whose uncertainty is double transferred to the detectable power spectrum. Measurements of this parameter still display much uncertainties and do not seem to precisely converge to a common value (see discussion in Section 3.3.1). We summarise the different values of Ω_{HI} that we obtained by modelling the neutral hydrogen distribution using the 2 different methods, as a function of redshift in Figure 4.11, together with values obtained from observations reported in [105, 124].

Interestingly, Figure 4.11 points out a globally lower values of Ω_{HI} for warmer cosmologies, that for example agrees with [54], where they show how the suppression of small scale structure leads to a notably delayed and subsequently more rapid stellar assembly, i.e. a faster exhaustion of the hydrogen reservoir in galaxies. Again, this is effect is crucial to be investigated in EoR, as we will point out in Section 8.2.

Let us discuss more quantitatively the role played by the uncertainty of the Ω_{HI} value in measuring the 21cm power spectrum and how we can disentangle it from the uncertainty of the *warmness* of DM.

As displayed in Figure 4.10, the warmness of DM increases the amplitude of the 21cm power spectrum, but also affects its shape. A higher or lower value of Ω_{HI} causes an overall increase or decrease

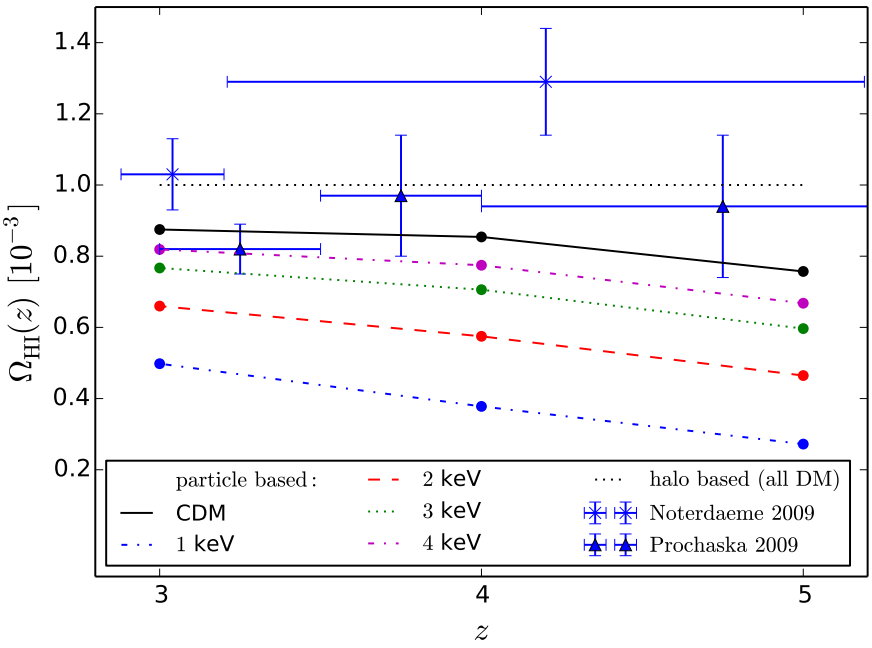


Figure 4.11: Value of the parameter $\Omega_{HI}(z)$, as a function of redshift, obtained by distributing HI according to the halo based method A (dotted line, all cosmologies) and to the particle based method (coloured lines). Observational measurements are displayed with error bars in blue, crosses for [105] and triangles for [124].

of the amplitude of the 21cm power spectrum, independently of the scale. This scale dependence difference could help breaking the Ω_{HI} - WDM mass degeneracy. We explicitly check this by performing a Fisher matrix calculation on the computed 21cm power spectra.

The Fisher matrix analysis[7] quantifies the amount of information that the 21cm power spectrum as observable carries about the two parameters Ω_{HI} and m_{WDM}. Practically, starting from our prior (the CDM case with $\Omega_{HI} = 10^{-3}$), we calculate how much the 21cm power spectrum varies by varying Ω_{HI} and m_{WDM}. Indeed, the inverse of the Fisher matrix is the covariance matrix, i.e. the Gaussian uncertainties linked to the parameters (Ω_{HI}, γ):

$$F^{-1} = C = \begin{bmatrix} \sigma_\Omega^2 & \sigma_{\Omega\gamma} \\ \sigma_{\gamma\Omega} & \sigma_\gamma^2 \end{bmatrix}, \tag{4.8}$$

7 A quick and clear reference for this kind of analysis is in [45].

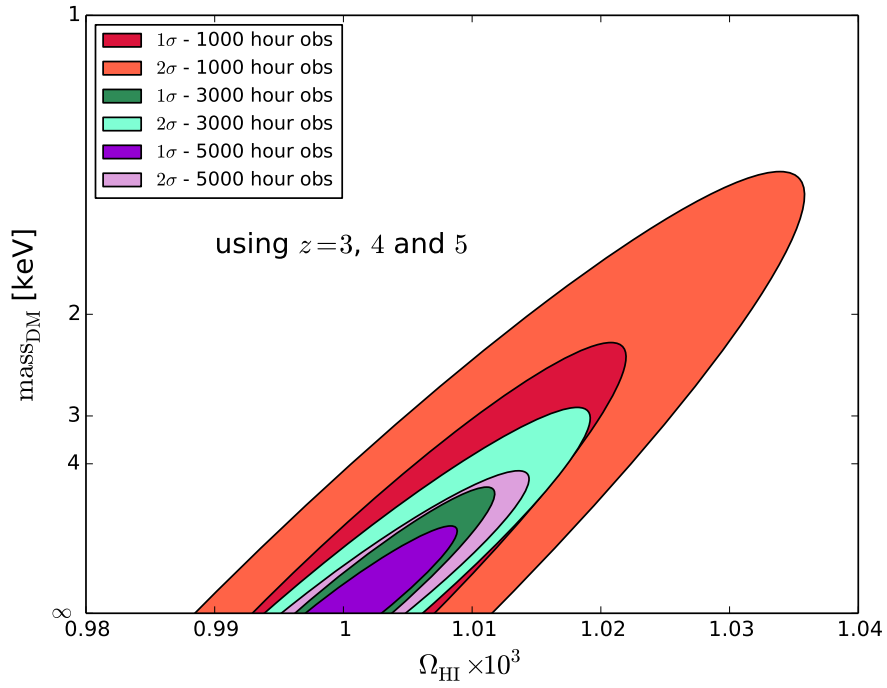


Figure 4.12: 1σ and 2σ contours (dark and light areas) of the values of Ω_{HI} and m_{WDM} determined using the 21cm power spectrum measured by SKA1-LOW with three different observation times: 1000, 3000 and 5000 hours (red, green and violet). The Fisher matrix analysis is performed using information coming from redshifts $z = 3$, 4 and 5.

where we have redefined the variable $\gamma = 1/m_{WDM}$, so to have $\gamma = 0$ for the CDM case. We follow the framework in [160] and write the elements (A, B) of F for a single redshift bin as

$$\begin{aligned} F_{AB} &= \frac{1}{2}\int \frac{d^3k}{(2\pi)^3}\frac{V_p h y s}{\sigma^2[P_{21cm}(k)]}\frac{\partial P_{21cm}(k)}{\partial\theta_A}\frac{\partial P_{21cm}(k)}{\partial\theta_B} \\ &= \sum_i N(k_i)\frac{1}{\sigma^2[P_{21cm}(k_i)]}\frac{\partial P_{21cm}(k_i)}{\partial\theta_A}\frac{\partial P_{21cm}(k_i)}{\partial\theta_B}, \end{aligned} \quad (4.9)$$

with θ either Ω_{HI} or the γ parameter, $\sigma^2[P_{21cm}(k)]$ the instrumental noise of SKA calculated in the previous section and $N(k)$ the number of independent modes k. We use the information coming from all redshifts at our disposal, $z = 3$, 4 and 5, i.e. we sum the 3 Fisher matrices altogether before inverting it [8]. The difference within a change in Ω_{HI} can be derived analytically, since it acts as a multiplying factor on the amplitude of the 21cm power spectrum. Instead, the 21cm power spectrum variation with a different m_{WDM} is calculated numerically using the results shown in Figure 4.10.

The analysis results are summarised in Figure 4.12 (the larger the area spanned in the parameter space, the larger the uncertainty),

8 This can be done with the assumption that the 3 measurements at $z = 3$, 4 and 5 are independent.

where we plot the 1σ and 2σ contours referring to the measurements of the 21cm power spectrum of 3 different observational scenarios (different observing time) of the SKA1-LOW. Already with a 1000 hours of observations, a lower limit on the mass of the WDM can be set to $m_{WDM} \geqslant 2$ keV.

By the time radio telescopes as SKA will start collecting data referring to these high redshifts, our knowledge of Ω_{HI} will be much tighter constrained, for example by the abundance of HI absorbers coming from ongoing Lyman-α forest observations.

5

HOW TO DISTINGUISH INDISTINGUISHABLE COSMOLOGIES

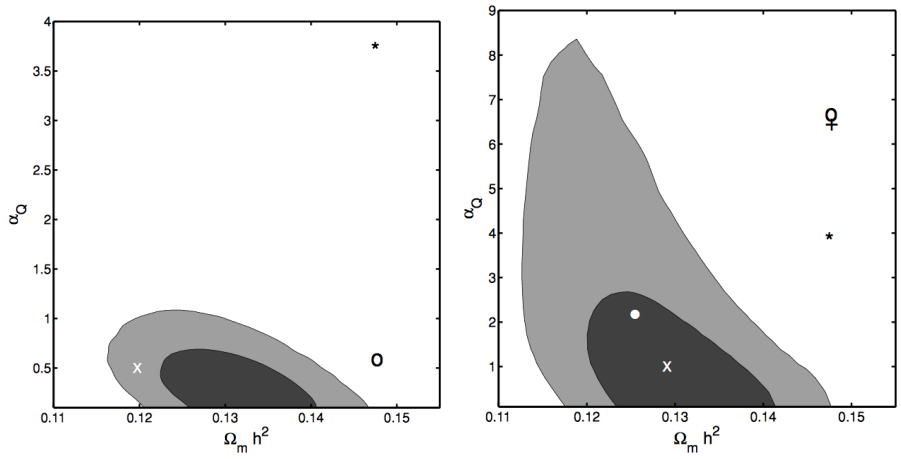
In the spirit of the work started in Chapter 4, in this chapter we describe the imprint on the 21cm signal in intensity mapping (IM) power spectrum left by non-standard cosmological models, otherwise statistically indistinguishable from ΛCDM and likewise successful at explaining currently available cosmological observations. We consider 2 models of dynamical dark energy (DE) and 2 models of non-cold dark matter (DM), and we compare them with standard ΛCDM. For the DM models, the same mechanism studied in Chapter 4 is at play producing an increase in the 21cm signal with respect to the cold DM case. Interestingly, and of totally different origin, also in the non-Λ DE scenarios we find an increase in signal amplitude, due to the differences in the halo mass function as these models experience a slightly different matter clustering. The 2 classes of cosmologies (non-standard DE and DM) leave a distinct signature on the 21cm spectra across a wide range of scales, thus allowing to disentangle their individual effects.

We proceed by considering halo catalogues from N-body simulations of dynamical DE models and DM scenarios. We limit our analysis to redshifts $z = 1$ and $z = 2$, which are common to all simulations. After having modelled HI using the prescription described in Chapter 3, we calculate the 21cm power spectra, study the imprints of the different cosmologies and we quantify these effects considering the radio telescope SKA.

5.1 COSMOLOGICAL MODELS AND N-BODY SIMULATIONS

5.1.1 *Dynamical dark energy*

We already discussed in Section 1.1.1 how observational probes suggest that the Universe is now experiencing an accelerated expansion and in the standard ΛCDM model such phenomenon is linked to a positive cosmological constant Λ, whose physical nature is unclear. Generally, whatever is the mechanism that can explain the accelerated expansion we call it dark energy (DE) and DE models beyond Λ have been proposed.



(a) Ratra-Peebles potential model. (b) Supergravity potential model.

Figure 5.1: Figures from Alimi et al., 2010 [4]. 1 and 2 σ confidence regions in the $\Omega_m h^2 - \alpha_Q$ plane (α_Q is the quintessence self-interaction potential parameter), from the combined analysis of the UNION SN Ia Hubble diagram and WMAP-5 CMB data. On the left panel the RP model case, on the right SUGRA. A white cross mark (X) indicates the choice of realistic model parameters done in [4], while the other marks refer to model parameter values assumed in the literature (see [4] for details). The ΛCDM model corresponds to the limiting value $\alpha_Q \to 0$, which is indeed within the 1σ confidence region.

Here we focus on dynamical DE scenarios[1] for which there exists a light minimally coupled scalar field, often referred to as *quintessence*, whose late time dynamics is responsible for driving the cosmological expansion: as the scalar field rolls down in its self-interaction potential, the kinetic energy becomes small compared to the potential energy and the field pressure becomes negative enough to drive the acceleration (see [155] for a review).

We use a set of large volume N-body simulations from the Dark Energy Universe Simulations (DEUS) database[2] of three flat DE models:

i) ΛCDM-W5: a standard cosmological model with cosmological constant Λ,

ii) RPCDM-W5: a quintessence model with dynamical equation of state as given by the scalar field evolution in a Ratra-Peebles [126] self-interacting potential,

iii) SUCDM-W5: a quintessence model with supergravity [25] self-interacting potential.

1 They are said *dynamical* because effectively DE is characterised by an equation of state $w(a) = w_0 + (1-a)w_a$ which varies with the scale factor a (see Table 5.1), as opposed to Λ for which is constant $w = -1$.

2 http://www.deus-consortium.org/deus-data/

Model	Ω_m	σ_8	w_0	w_a
ΛCDM-W5	0.26	0.80	−1	0
RPCDM-W5	0.23	0.66	−0.87	0.08
SUCDM-W5	0.25	0.73	-0.94	0.19

Table 5.1: Cosmological model parameters of realistic DE models calibrated against WMAP-5 and SN Ia observations. The other cosmological parameters are set to $\Omega_b = 0.04$, $h = 0.72$, $n_s = 0.96$. Notice that w_0 and w_a the DE equation of state parameters of the Linder-Chevalier-Polarski parametrization [40, 82] best-fitting the time evolution of the quintessence-field equation of state.

The cosmological parameters of all above models have been calibrated in [4] in order to reproduce the cosmic microwave background power spectra from WMAP-5 observations [77] and the luminosity distances from SN-Ia measurements [78], as shown in Figure 5.1. Quintessence models tend to fit the data by requiring lower values of Ω_m, as marked by the orientation of the degeneracy line in Figure 5.1: this is because these models are characterised by a less accelerated cosmic expansion, therefore, under the flatness requirement, they require a large amount of DE (and lower matter density) to fit the data. This less pronounced acceleration will effect the halo mass function we are interested in, as we see later.

In Table 5.1 we show the tuned parameters, the other cosmological parameters are set to $\Omega_b = 0.04$, $h = 0.72$, $n_s = 0.96$ for all models considered. As we are interested in modelling the HI cloud distributions using N-body halo catalogues, we use data from the DEUS simulations with the largest available mass resolution. These have box size of comoving 162 Mpc h^{-1} and contain 1024^3 particles (corresponding to mass particle resolution of $m_p = 2.5 \cdot 10^8\ M_\odot\ h^{-1}$). We analyse snapshots at $z = 1$ and 2.3.

After having calibrated these models and having obtained *realistic* cosmologies, i.e. statistically indistinguishable from ΛCDM, in [4] they analyse their linear and non-linear power spectra, showing how the scalar field dynamics impact the cosmic expansion and the evolution of matter perturbations. These dynamical DE models are characterised by more clustering power on large linear scales due to DE perturbations while on small scales the growth factor is suppressed (see discussion on the Ω_m value above): the net results is a different distribution of power between small and large scales with respect to the ΛCDM scenario.

Summarising, since matter density fluctuations grow less efficiently for these dynamical DE models than in ΛCDM, these alternative scenarios are characterised by an overall lower level of clustering. In particular, they exhibit matter power spectra in the range $0.1 \lesssim k \lesssim$

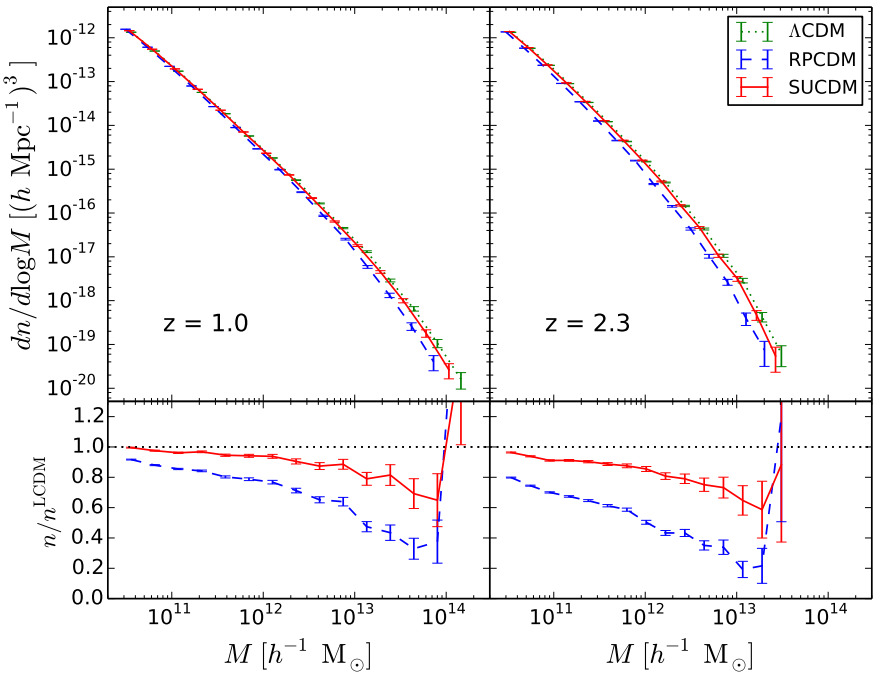


Figure 5.2: Halo mass functions of the 3 cosmological models and relative difference with respect to the ΛCDM prediction (bottom panels) at $z = 1$ and $z = 2.3$ (left and right respectively).

$1\,\mathrm{Mpc}\,h^{-1}$ and $z \lesssim 2$ that are lower than the ΛCDM prediction with deviations as large as $20 - 40\%$ [4]. We thus expect differences in the halo abundances that should cause different 21cm signals.

The halos in each simulation box are identified using the FoF halo finder [52] with linking length parameter $b = 0.2$. To limit numerical systematic errors we only retain halos with more than 100 particles corresponding to a minimum halo mass of the catalogues of $M_{\mathrm{min}}^{162} = 2.5 \cdot 10^{10}\ \mathrm{M}_\odot\ h^{-1}$.

We compute the halo mass functions, that we show in Figure 5.2. Realistic quintessence cosmologies yield a DM density that is less structure than that of the standard ΛCDM, leading to a suppressed abundance of the most massive halos. This is in antithesis to the warm DM cosmologies seen in Chapter 4, for which there are the low mass halos to be missing.

5.1.2 *Non-standard dark matter*

The ΛCDM model is constituted by a perfectly collisionless *cold* dark matter component (CDM), i.e. with negligible thermal velocities on all scales at high redshift. However, tensions exist at small scales between CDM predictions and various observational probes and, as discussed in Section 1.1.2, there are diverse theoretical DM models that can alleviate these tensions while preserving the success of CDM on large

scales. They do so by suppressing gravitational clustering at small scales, i.e. DM can not significantly cluster below a specific length, this way they feature a matter power spectrum that at middle/large scales overlaps with that of CDM and at a specific small scale exhibits a cut-off.

In Chapter 4, we deal with thermal warm DM, whose free-streaming length is responsible for the suppression of the amplitude of matter density fluctuations. We extend the work of Chapter 4 and consider other 2 families of DM models composed by:

i) ULADM: ultra-light axion DM, in the literature sometime called *fuzzy* DM. They affect structure formation due to their large de Broglie wavelength, i.e. their natural oscillating axion field introduces a damping scale in the matter power spectrum ([90] for a review);

ii) LFDM: DM particles emerged from a scalar field undergoing a phase transition near matter-radiation equality (late-forming DM). This delayed structure formation alters the small scale clustering [3, 50].

In brief, these models are characterised by a suppression of the amplitude of matter density fluctuations at small scales below a characteristic length that for ULADM depends on the axion particle mass m_a, for LFDM models depends on the phase transition redshift z_t, as for warm DM studied in Chapter 4 depends on the thermal relic mass m_{WDM}. The matter power spectrum produced by the ULADM and LFDM models exhibits a distinctive damped oscillatory tail at the small scales, instead of a smooth cut-off as for the vanilla WDM models (see Figure 5.3).

We consider N-body simulations presented in Corasaniti et al., 2017 [47]. They are run with RAMSES [152] with box size of comoving 27.5 Mpc h^{-1}, that follow 1024^3 particles. They consist of 3 ULADM models[3] with axion mass of:

$$\begin{array}{rccc} & \text{ULADM-1} & \text{ULADM-2} & \text{ULADM-3} \\ m_a[\text{eV}] = & 1.56\times10^{-22} & 4.16\times10^{-22} & 1.54\times10^{-21} \end{array}$$

and of 3 late-forming DM models with transition redshift:

$$\begin{array}{rccc} & \text{LFDM-1} & \text{LFDM-2} & \text{LFDM-3} \\ z_t = & 5\times10^{5} & 8\times10^{5} & 15\times10^{15}. \end{array}$$

3 Note that the ULADM models investigated are in disagreement with the recent constraints obtained at $z > 3$ from the Lyman-α forest which result in a lower limit at the 2σ level of $\sim 2\times10^{-21}$ eV [71] for a conservative analysis. However, given the intrinsic complementarity of the two observables (21cm IM power spectrum probes the HI in mass while the forest probes the HI in volume, as we see in Chapter 7) it is important to explore these models to confirm or disproof the limits obtained with the forest (that are obtained at redshifts higher than those we consider here).

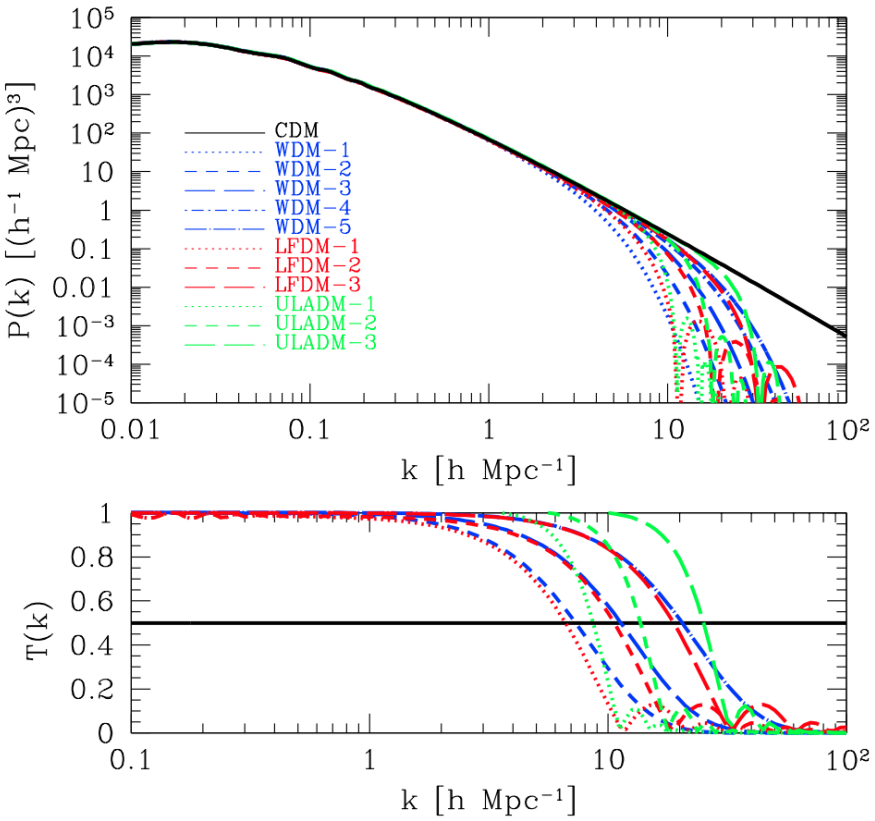


Figure 5.3: Figure from Corasaniti et al., 2017 [47]. Linear matter power spectra (top panel) and transfer functions (bottom) at $z = 0$ for CDM (black solid line), ULADM models (green lines) and LFDM (red lines), corresponding to the models studied in this chapter. (The blue lines refer to cosmologies we do not take into account.)

The cosmological model parameters have been set to those of a reference ΛCDM simulation (ΛCDM-S) of the same box size and with equal number of particles, setting cosmological parameters to:

$$\Omega_m = 0.3 \qquad \Omega_b = 0.046 \qquad h = 0.7$$
$$n_s = 0.99 \qquad \sigma_8 = 0.8\ .$$

Thus all 7 simulations are characterised by a particle mass resolution of $m_p = 1.61 \cdot 10^6\ M_\odot\ h^{-1}$.

In Figure 5.3, taken from [47], the black line is the linear matter power spectrum of the reference ΛCDM-S model compared to the ULADM models (green) and LFDM models (red), in blue are WDM models that do not correspond to those studied in Chapter 4[4].

These non-cold DM models are not indistinguishable from standard ΛCDM in the same sense as the dynamical DE models studied in the previous section: as for the WDM models of Chapter 4, they are

4 The WDM models displayed in Figure 5.3 have thermal masses between 0.7 and 2.4 keV, hence lower (warmer models) than those studied in Chapter 4 and already excluded by Lyman-α studies [72, 158].

identical to ΛCDM at middle-large scales, and differ in sub-structures number and small scale clustering.

We analyse snapshots at $z = 1$ and 2 and identify halos using the FoF halo finder [52], although we need to remove the spurious artificial halos from our catalogues.

5.1.2.1 *Removing the fake halos from the catalogues*

Indeed, since we are dealing with cosmological models with suppressed spectra at small scales, the sampling of the Poisson noise between the cut-off scale of the power spectrum and the Nyquist frequency of the simulations leads to the formation of spurious numerical halos, which cause an unphysical upturn of the halo mass function at low masses [66, 133, 165]. In the literature this phenomenon is often called *spurious fragmentation*. We already discuss it in Section 4.1.2.1, although in the warm DM simulations employed in Chapter 4, we find spurious fragmentation negligible since in that case:

i) the matter power spectrum cut-off is at smaller k than those of the ULADM and LFDM considered here[5],

ii) the redshifts considered ($z = 3 - 5$) are higher and the number of spurious halos increases as the simulation evolve from earlier to later times [2].

Hence, for the simulations considered here we need to take care of spurious fragmentation, since it could bias the way we populate the halos with HI. In the literature there are a few methods for removing these unphysical halos: crudely applying a mass cut in the halo catalogue [165], extrapolating the contribute of the artificial halos from the characteristic power law of the upturn of the mass function [132], going back to initial condition and identify false halos as groups of particle associated to Lagrangian patches [85] (relying on the idea that genuine proto-halos are spheroidal). We decide to use the most recent method described in Agarwal and Corasaniti, 2015 [2], which is shown to work independently of simulation resolution and of cosmological model. In [2] they find that spurious halos have systematically larger spin values, are highly elliptical or prolate and significantly deviate from virial equilibrium. Practically, their method consists in retaining halos with at least 300 particles and with deviation from the virial state in the range:

$$0 < \frac{2K}{|E|} < 1.5\,, \tag{5.1}$$

where K is the total kinetic energy of the halo and $|E|$ its gravitational potential energy. To illustrate the effect of spurious halos in

5 As rule of thumb, ULADM-1, ULADM-2, ULADM-3 and LFDM-1, LFDM-2, LFDM-3 exhibit their power spectrum cut-off at roughly the same k of warm DM scenarios with thermal masses of 1, 1.5 and 2.4 keV respectively (see Figure 2 of [47]).

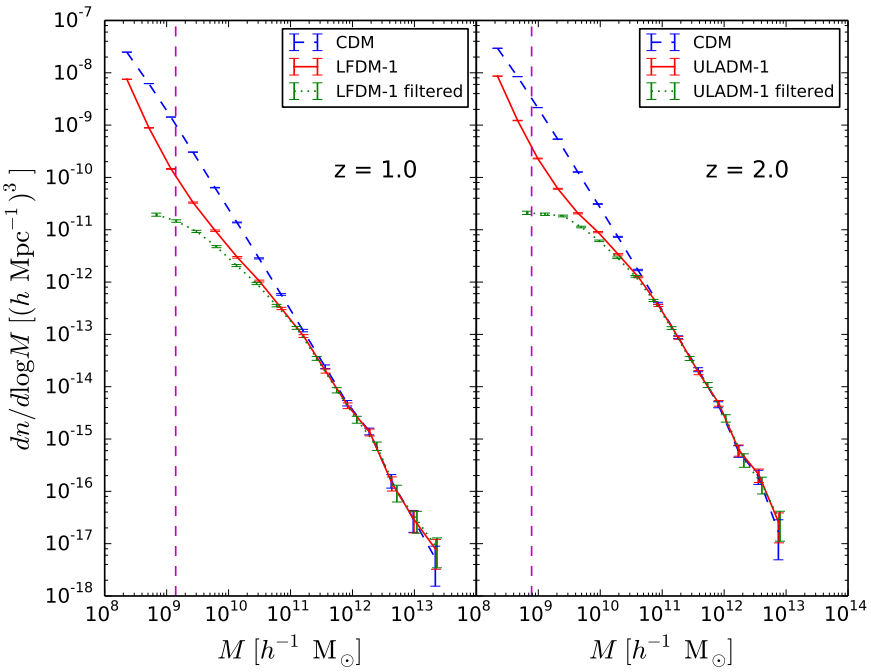


Figure 5.4: Halo mass functions for the ΛCDM-S simulation in dashed blue line. On the left at $z = 1$ and also for the LFDM-1 simulation using its bare FoF catalogue (solid red) and having filtered out the spurious halos (dotted green); on the right at $z = 2$ for the UALDM-1 simulation with same colour coding. Error bars correspond to the Poisson errors in each mass bin.

our simulations and the method used to remove them, we plot in Figure 5.4 the halo mass functions at $z = 1$ for LFDM-1 and at $z = 2$ for ULADM-1 (left and right panels respectively). In each panel, the blue dashed line represent the ΛCDM-S mass function, while the non-standard DM model is in red before the spurious halo filtering and in green dotted after. For both non-standard models it is evident how the mass function converges to that of CDM at the high mass end: on large scales our models are identical to ΛCDM (as already discussed in Chapter 4). At low masses we see the characteristic upturn which indicates the presence of spurious halos. After halo selection, the upturn disappears and we recover the expected low mass halo abundance suppression [6].

Finally, in Figure 5.5, we plot the ratio of the halo mass function of non-standard DM models relative to the reference ΛCDM simulation at $z = 1$ (top panel) and 2 (bottom panel), after having removed spurious halos. As expected these models exhibit suppressed halo abundances at low masses compared to the ΛCDM case, with the mass scale cut-off depending on the specificities of the underlying

6 For consistency, we apply the spurious halos selection method to all cosmologies. Practically it has no effect on the halo mass function of ΛCDM-S: it removes just few halos, the least relaxed (see discussion in [2] and its Figure 7).

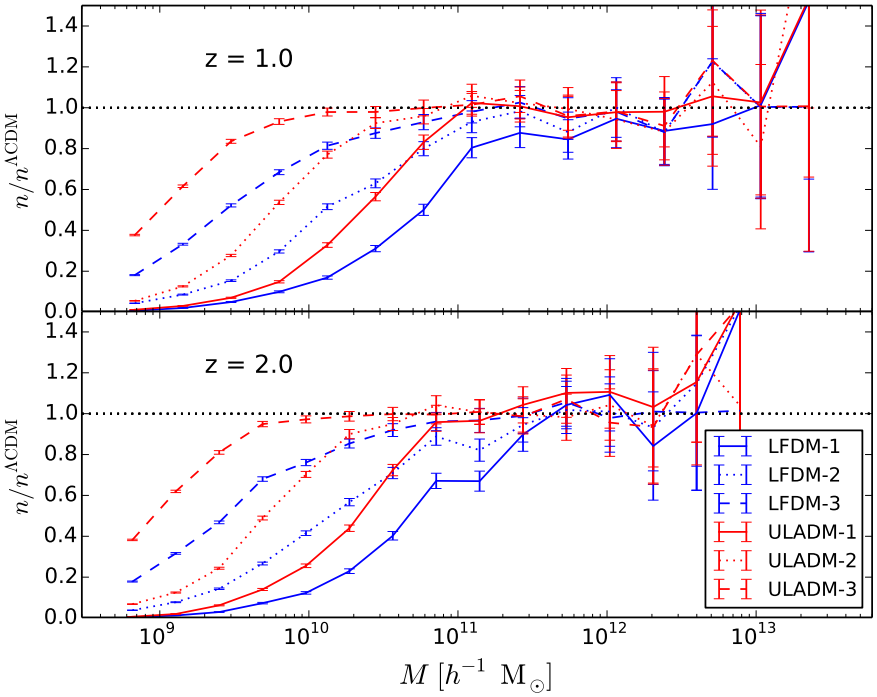


Figure 5.5: Ratios of the halo mass function of the non-standard DM models relative to the reference ΛCDM-S scenario at $z = 1$ (top panel) and 2 (bottom panel). Error bars are the Poissonian errors.

DM model. The larger the ULADM particles mass or equivalently the higher the phase transition redshift of LFDM models the lower, the mass scale cut-off and the smaller the deviation from ΛCDM in the simulated mass range.

By comparing the trends in Figure 5.5 and in the bottom panel of Figure 5.2, we can notice that non-standard DE and DM models exhibit different mass dependent deviations from ΛCDM. As we will see, this results in having different differences for the HI clustering signal compared to ΛCDM.

5.2 IMPACT OF NON-STANDARD MODELS ON THE HI DISTRIBUTION

Both for the dynamical DE models described in Section 5.1.1 and for the non-standard DM models of Section 5.1.2, we distribute HI following the halo base method B described in Section 3.1, placing all HI at the centre of each hosting halo (in Section 3.4.1 we illustrate why this approximation does not bias our results) and setting $\Omega_{\mathrm{HI}} = 10^{-3}$ (i.e. setting f in Equation 3.2), a value in agreement with observational data as discussed in Section 3.3.1. Subsequently, we proceed at calculating the 21cm signal power spectrum for each scenario, following

Section 3.4, checking how the different matter clustering properties of these cosmologies are reflected into the 21cm signal.

The halo base method B parameters M_{min} and α are also responsible for the clustering amplitude of the signal. We believe that their values should be fixed for all cosmologies, ΛCDM and alternative ones. M_{min} indicates the minimum halo mass (gravitational potential well) for hydrogen to self-shield and stay neutral against cosmic UV background, while α is the trade-off between the cooling processes and the mechanical stripping of cold gas from halos, responsible for the amount of HI inside halos. I.e. these two parameters are of astrophysical origin and do not depend on the background cosmology. However, if cosmology changes e.g. the halo density profiles, we could speculate that the HI physics could change and the parameters M_{min} and α vary. The issue is that there is not much data available at our disposal (as discussed in Chapter 3) to match observations for all cosmologies individually. Hence, the conservative choice we make is to fix astrophysics (through the values of M_{min} and α) and check what changes between cosmologies, focusing on the ratios between their different 21cm $P(k)$.

Our goal is to study the imprints of DE and DM models on the HI (and eventually 21cm) power spectrum. Since the HI distribution is modelled upon the results of N-body simulations, we first evaluate the impact of numerical effects due to the finite volume and the mass resolution of the halo catalogues.

5.2.1 *Minimum halo mass and volume effects*

The HI distribution model we follow depends on the specification of a minimum halo mass containing the neutral gas cloud (M_{min}). Although largely uncertain (as discussed in Section 3.1), a conservative guess is to set $M_{min} \sim 10^9\ M_\odot$ (simply making use of Equation 3.3), however such a mass scale is resolved only in the case of non-standard DM halo catalogues from simulations with box size of 27.5 Mpc h^{-1} and 1024^3 DM particles. On the other hand, in the dynamical DE model simulations, the minimum halo mass is about 10 times larger[7]. Thus, we now evaluate the impact of M_{min} on the HI power spectrum of our scenarios.

In Figure 5.6 we plot the HI power spectrum at $z = 1$ for the ΛCDM-W5 model simulation (green dotted line) and from the ΛCDM-S catalogue having set $M_{min} = M_{min}^{27.5}$ (red dashed line) and $M_{min} = M_{min}^{162}$ (blue solid line) in Equation 3.2, with $M_{min}^{27.5}$ and M_{min}^{162} being the smallest halo mass resolved in the simulations of 27.5 Mpc h^{-1} and 162 Mpc h^{-1} respectively. The power spectrum from the larger simula-

7 Notice that this M_{min} value is anyway compatible with what found in [38], where it is tuned by requiring $M_{HI}(M)$ to reproduce the observed bias of the Damped Lyman α systems at $z = 2.3$.

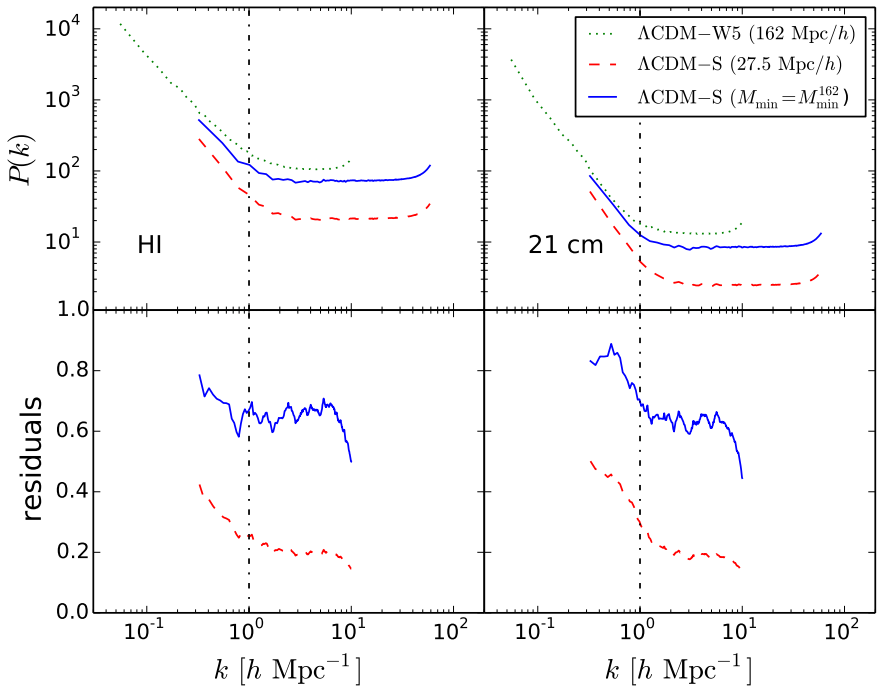


Figure 5.6: Power spectrum of HI (left panels) and of the 21cm signal (right panels) at $z = 1$ for the different ΛCDM simulations: ΛCDM-W5 in green dotted line and the ΛCDM-S catalogue having set $M_{\rm min} = M_{\rm min}^{27.5}$ (red dashed) and $M_{\rm min} = M_{\rm min}^{162}$ (blue solid). In the bottom panels the ratios of the ΛCDM-S power spectra over the ΛCDM-W5 one.

tion box covers a wider range of low-k modes than the smaller box, the latter however extends to larger k modes as the corresponding simulation has higher spatial and mass resolution. It is important to notice that in both cases the spectra exhibit an unphysical flattening of power at $k \gtrsim 1$ Mpc^{-1} h, which is due to the lack of modelling the HI distribution within the halos (see discussion in Section 3.4.1). We can see that the HI power spectrum from the 27.5 Mpc h^{-1} box size halo catalogue increases in amplitude when increasing the value of $M_{\rm min}$. This is because the total amount of HI is fixed and by increasing $M_{\rm min}$ we assign more HI mass to more massive halos which are more clustered then low mass ones, thus leading to a larger amplitude of the HI power spectrum. Notice also that as we set $M_{\rm min} = M_{\rm min}^{162}$, the HI power spectrum of the 27.5 Mpc h^{-1} box size halo catalogue lies closer to that of the ΛCDM-W5 model with difference of $\sim 20-30\%$, however this cannot be attributed uniquely to a volume effect since the two ΛCDM simulations have slightly different values of the cosmological parameters.

Since we do not have a larger volume simulation of the ΛCDM-S model, we address the effect due to the discrepancy of cosmological parameters using halo catalogues of the LFDM-3 model for which

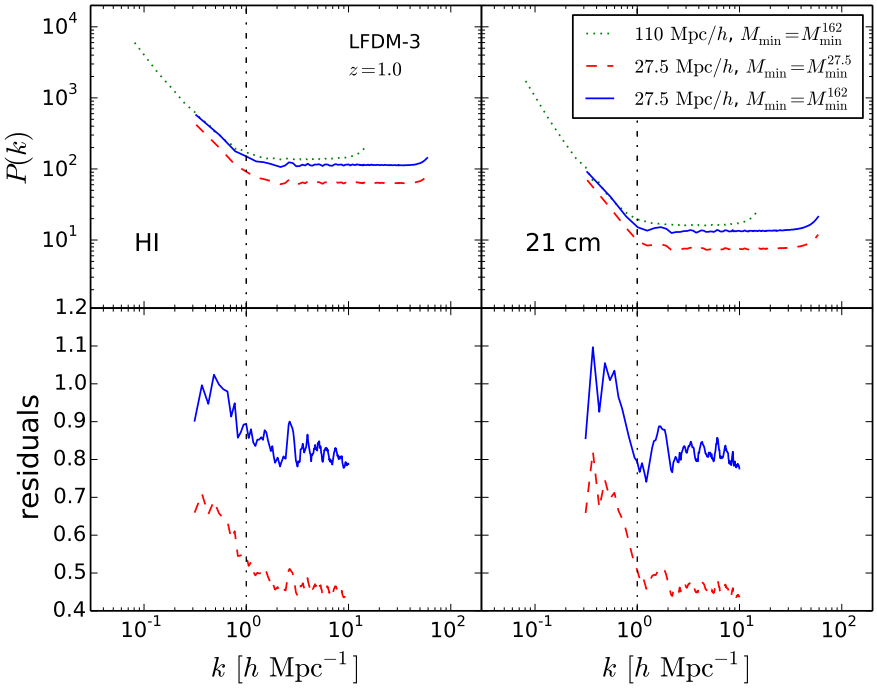


Figure 5.7: Power spectrum of HI (left panels) and of the 21cm signal (right panels) at $z = 1$ of the LFDM-3 scenario for different simulation box size and value of $M_{\rm min}$ as in Equation 3.2. In the bottom panels their ratios.

we have an additional simulation with box size of 110 Mpc h^{-1} and 2048^3 particles, i.e. 64 times bigger in volume than the simulation we are analysing (and 8 times smaller DM mass resolution). The corresponding HI power spectra are shown in Figure 5.7. As in the previous case (Figure 5.6), we notice a flattening of the power spectrum for $k \gtrsim 1\,h$ Mpc^{-1} due to the lack of HI modelling inside each halo. Moreover, we can see that increasing $M_{\rm min}$ from $M_{\rm min}^{27.5}$ to $M_{\rm min}^{162}$ leads to a larger amplitude of the HI spectrum by an amount similar to that found in ΛCDM-S case and that the HI spectrum from the 27.5 Mpc h^{-1} catalogue with $M_{\rm min} = M_{\rm min}^{162}$ (blue solid line) differs by less than $\lesssim 10\%$ from that of the 110 Mpc h^{-1} box size simulation with $M_{\rm min}$ set to the same value (green dotted line). The latter results are independent of cosmology this time (we make use of the LFDM-3 model only).

Briefly, Figure 5.6 and Figure 5.7 show how numerical effects modify the HI power spectrum in the same way across different models. This allows us to compare the relative difference of the HI spectra among models as they are largely insensitive to volume and mass resolution effects.

As we do not model the HI gas distribution within halos, hereafter we only consider the HI spectra for $k \lesssim 1\,h$ Mpc^{-1}.

5.2.2 *Cosmological parameters dependence of HI spectra*

We now focus on the signatures of the non-standard cosmological models on the HI spectra relative to the standard ΛCDM scenario.

In Figure 5.8 we plot the HI spectra for the DE (top figure) and DM (bottom figure) models at $z = 1$ and 2.3. In the bottom panels we show the relative differences with respect to the reference ΛCDM model.

In the case of non-standard DE models (Figure 5.8a) the differences among the HI spectra at large scales can be understood in terms of the *evolution* of the halo mass function, since the growth rate is altered in these scenarios: the growth of the power spectrum tends to slow down once structures become more and more virialised [4].

At $z = 2$ the RPCDM and SUCDM models have spectra with amplitude larger than in the ΛCDM case. This is consistent with the fact that at this redshift the abundance of halos in the non-standard DE models is suppressed compared to the reference ΛCDM (see Figure 5.2) consequently, since we distribute in the halo catalogues the same HI total mass to all models, the HI clouds are more clustered in RPCDM and SUCDM than in the ΛCDM, because of the smaller number of halos at our disposal. At $z = 1$, the halo abundance in the non-standard DE model is still suppressed compared to the ΛCDM, but higher than at $z = 2.3$, hence the amplitude of the HI spectra decreases.

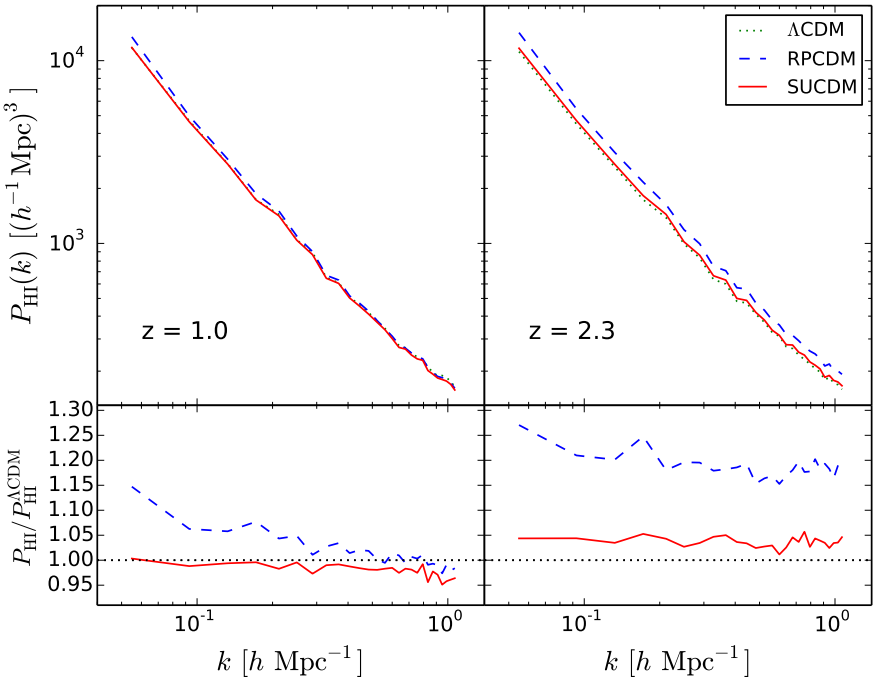
We notice that at small scales the SUCDM has less suppressed spectrum than the RPCDM, again in agreement with the growth of structures differences studied in [4], where they claim the effect in SUCDM is larger than RPCDM compared to ΛCDM.

For the non-standard DM models shown in Figure 5.8b, the linear growth rate is identical to that of the reference ΛCDM model, thus the relative differences in the HI spectra at $z = 1$ and 2 are entirely due to the suppressed abundances at low halo masses (see Figure 5.5), i.e. the effect illustrated in Chapter 4 for WDM applies here too.

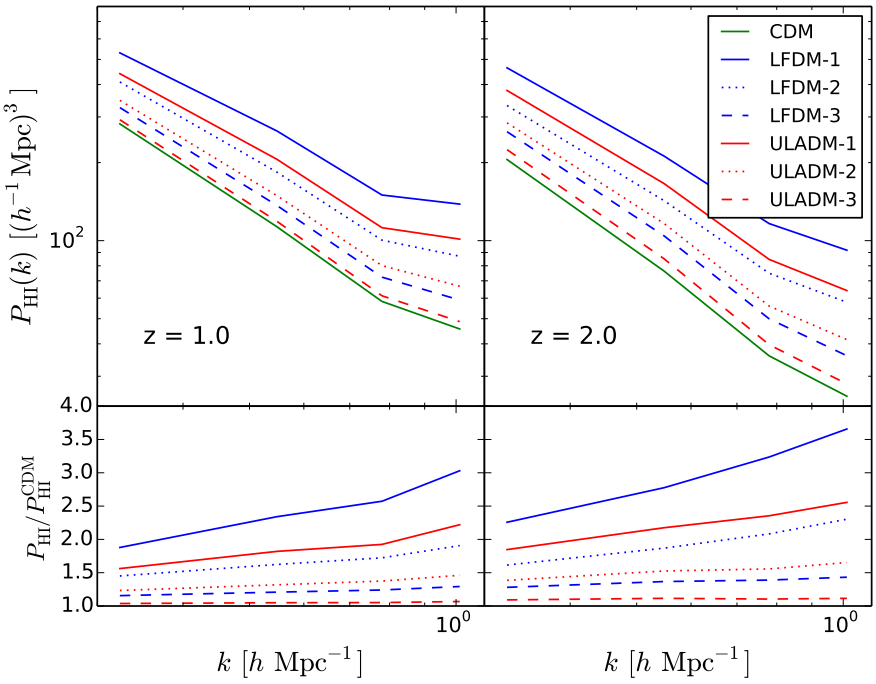
5.3 21CM INTENSITY MAPPING POWER SPECTRA AND SKA1-MID FORECASTS

We now focus on the 21cm IM power spectrum. Differently from the HI spectra, the 21cm encodes additional cosmological information since it traces in redshift-space the location of HI clouds whose peculiar velocities alter the clustering signal (as illustrated in Section 3.4).

In Figure 5.9a we plot the relative difference of the 21cm power spectrum between the non-standard DE models and the reference ΛCDM-W5 at $z = 1$ (left panel) and $z = 2.3$ (right panel), likewise in Figure 5.9b we plot the spectra in the case of the non-standard DM models relative to the reference ΛCDM-S. It is worth noticing

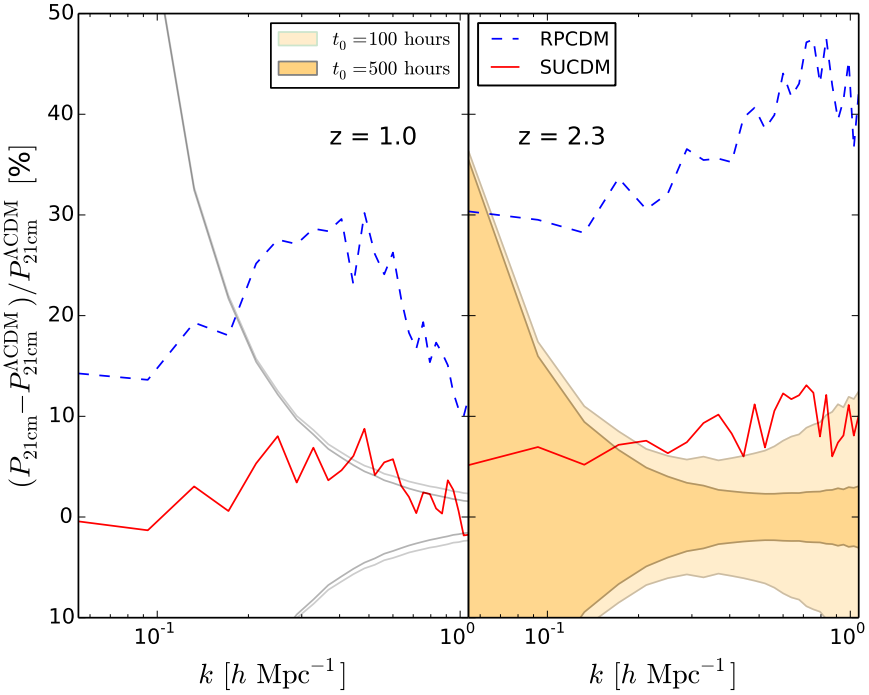


(a) Non-standard DE models.

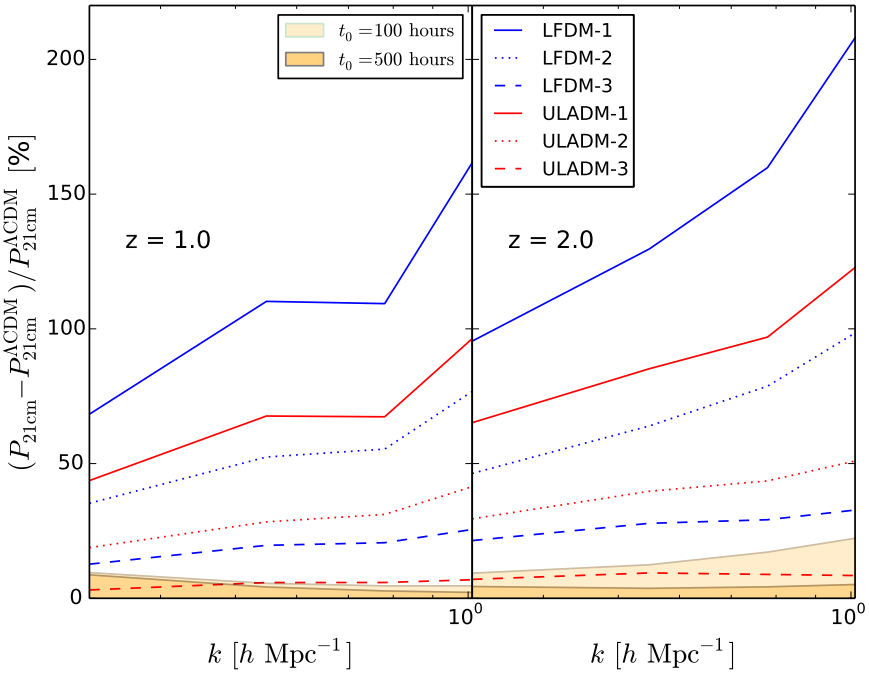


(b) Non-standard DM models.

Figure 5.8: HI power spectra and relative difference with respect to the ΛCDM prediction (bottom panels) at $z = 1$ and $z = 2.3\,(2)$ (left and right respectively).



(a) Non-standard DE models relative to the reference ΛCDM-W5 model.



(b) Non-standard DM models relative to the reference ΛCDM-S model.

Figure 5.9: 21cm intensity map power spectrum of the non-standard cosmological models relative to ΛCDM predictions at $z = 1$ (left panels) and 2.3 (2) (right panels). The shaded area represents the expected errors from SKA1-MID measurements for the reference ΛCDM model, $\sigma[P_{21cm}^{\Lambda CDM}(k)]/P_{21cm}^{\Lambda CDM}(k)$, assuming $t_0 = 100$ (light shaded area) and $t_0 = 500$ (dark shaded area) observing hours.

that the 21cm spectra of the DE models differ from the ΛCDM case not only by an amplitude factor but also on the scale dependence of the signal, in particular we may notice a change of the slope of the spectra at small scales. In the case of the non-standard DM models the differences with respect to the standard cosmological scenario show a very different trend, with differences increasing at small scales. This suggests that in principle the imprints of DE and DM models can be distinguished from one another through 21cm intensity mapping measurements.

To be more quantitative, we estimate the instrumental error on 21cm power spectrum interferometric measurements expected from the SKA1-MID radio telescope for the reference ΛCDM models. Neglecting systematic effects such as foreground subtraction, we can cast the error on the 21cm detection as shown in Section 3.4.2. More specifically, we employ Equation 3.15 and tune the total observation time: $t_0 = 100$ and 500 hours for the following analysis.

We show the above calculated 1σ errors on 21cm power spectrum measurements in Figure 5.9a and Figure 5.9b as shaded regions for $t_0 = 100$ (light shaded area) and $t_0 = 500$ (dark shaded area) observing hours respectively. We stress again that we assume astrophysical and atmospheric foreground contaminations and radio interferences have already been removed from data.

We can see that the RPCDM model can be distinguished from the ΛCDM at high-statistical significance. Even the SUCDM model, characterised by a cosmic expansion and a linear growth rate similar to that of the ΛCDM, can be potentially distinguished at more that 1σ at $z = 2.3$ in the range of scales corresponding to $0.02 \lesssim k\,[\mathrm{Mpc}^{-1}\,h] \lesssim 1$.

Similarly, the non-standard DM models considered here should be detectable or ruled out with future SKA observations. We notice that, compared to the WDM models in Chapter 4, the differences with respect to CDM here are much higher, due to the fact that here we have considered models with a cut-off in the power spectrum of different shape and, especially, present at smaller k [8].

8 In [47] they find that, having equal half modes (i.e. wavenumber k at which the transfer function of a given DM model is half that of the correspoinding CDM one) the ULADM is the scenario with the greatest number of small halos, followed by WDM and then by LFDM.

6

ON THE MODELLING OF THE LYMAN-α FOREST

As already discussed in Section 1.2.2, neutral hydrogen (HI) in the intergalactic medium (IGM) can be detected thanks to observations of the Lyα forest, a characteristic pattern in the absorption spectra of quasars. It represents a powerful way to trace highly ionized low density gas clouds [96], being able to trace density fluctuations in the high redshift range $z \sim 2 - 4$.

The Lyα forest is set by the interplay of gravitational collapse, expansion of the Universe, and reionization processes due to the build-up of a background of UV photons emitted by active galactic nuclei (AGN) and star forming galaxies. There is no analytic solution for the small scale evolution of the matter and baryon density fluctuations over time. In order to precisely describe the behaviour of the IGM, it is therefore necessary to treat the problem numerically.

In this chapter we illustrate how we model the Lyα forest on high-resolution hydrodynamic simulations to later perform in Chapter 7 the cross-correlation study with the 21cm intensity mapping[1] . We focus on 2 observationally most relevant statistics of the transmitted flux: the line-of-sight power spectrum, and the 3D power spectrum.

6.1 EXPLOITING THE SHERWOOD SIMULATIONS

We rely on high-resolution hydrodynamic simulations that belong to the Sherwood suite of Bolton et al., 2017 [20]: they are tailored to be in very good agreement with the observational data relative to the low density intergalactic medium. We consider 2 of these simulations, named:

80-2048:	follows 2×2048^3 (CDM + baryons) particles within a box of comoving size of $80\,h^{-1}$Mpc ,
160-1024:	follows 2×1024^3 (CDM + baryons) particles within a box of comoving size of $160\ h^{-1}$ Mpc.

We make use of 2 simulations for performing convergence tests; next in Chapter 7 we use 160-1024 only, which has the bigger volume, for the cross-correlation study at large scales. The simulations are run with periodic boundary conditions using the TreePM+SPH code GADGET-III, an updated and extended version of GADGET-II [144].

1 We remark that we have also modelled the Lyα on N-body only simulations exploiting the fluctuating Gunn-Peterson approximation (FGPA) we later introduced in Section 6.2.1.

The values of the cosmological parameters are in agreement with recent Planck data [120]:

$$\Omega_m = 0.308 \qquad \Omega_b = 0.0482 \qquad \Omega_\Lambda = 0.692$$
$$h = 0.678 \qquad n_s = 0.961 \qquad \sigma_8 = 0.829\ .$$

Both 80-2048 and 160-1024 simulations evolve from $z = 99$ to $z = 2.4$, the latter being the snapshot considered in this analysis. In Table 6.1 we show the mass resolution and softening length for the 2 different runs.

Name	Box size $[h^{-1}\mathrm{Mpc}]$	m_{CDM} $[h^{-1}M_\odot]$	m_{baryon} $[h^{-1}M_\odot]$	l_{soft} $[h^{-1}\mathrm{kpc}]$
80-2048	80	4.3×10^6	7.8×10^5	1.56
160-1024	160	2.8×10^8	5.1×10^7	6.25

Table 6.1: Summary of the resolution parameters of the 2 different simulations.

MAIN PHYSICAL INGREDIENTS OF THE HYDRODYNAMICS

The hydrodynamics is presented in detail in [20]. Here we report the fundamentals. Star formation processes are not properly followed in the simulations: gas particles with temperature $T < 10^5$ K and an overdensity $\delta_\rho > 1000$ are converted to collisionless particles, resulting in a significant increase in computation speed at the expense of removing cold, dense gas from the model. The simulations do not include metal line cooling. The photoionisation and photo-heating of the hydrogen and helium gas is calculated using the spatially uniform ionising background model presented in Haardt and Madau, 2012 [68], where HI reionization happens at $z \sim 12$. Moreover, the gas is assumed to be optically thin and in ionisation equilibrium. We stress that 80-2048 and 160-1024 share the same thermal history, since for both of them the cooling routine and the UV background have been modified following the reference model of [158].

We extract mock Lyα absorption spectra skewers from the simulations. Each skewer represents a line of sight (l.o.s.) to a quasar. We have 3 different catalogues, containing 2500, 1600 and 900 spectra. They are built from regular grids of 50^2, 40^2 and 30^2 points along the x, y and z directions, respectively (next in Section 6.2.2 we show how regularising the spectra position on a grid does not bias our results). Each spectrum contains 2048 pixels, evenly distributed along the simulation box length. In Table 6.2 a summary of the absorption spectra catalogues.

The gas density, weighted temperature, the neutral fraction and gas peculiar velocities are extracted following the SPH interpolation

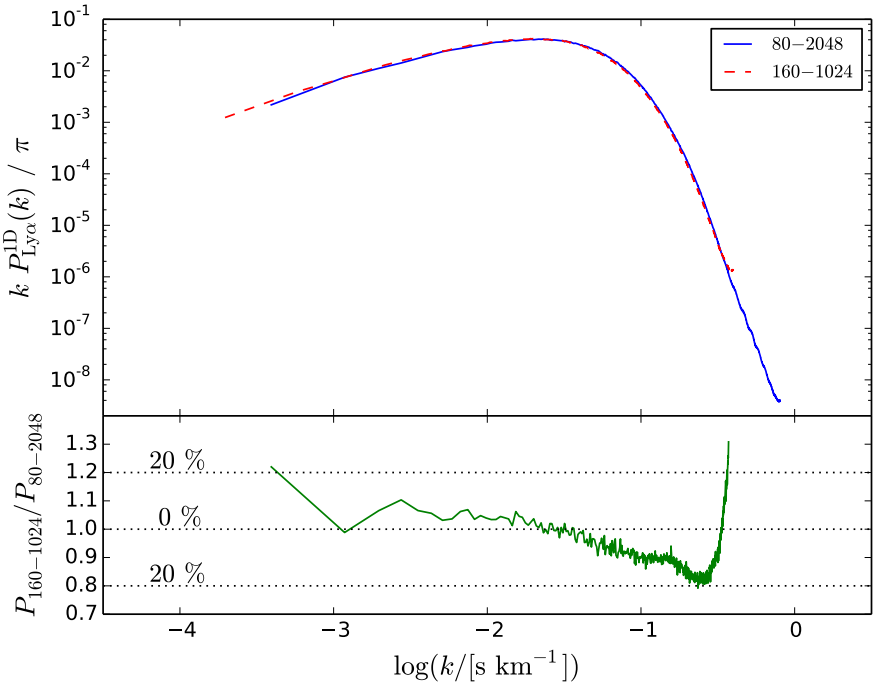


Figure 6.1: *Upper panel:* The dimensionless 1D Lyα flux power spectrum for the 80-2048 (blue solid) and the 160-1024 (red dashed) simulation. *Lower panel:* Ratio of the 2 power spectra; we highlight the 20 % difference region.

scheme described in [153]; for the Lyα optical depth τ along each l.o.s. we make use of the Voigt profile approximation, as in [151]. Once τ is determined in every pixel, we define the absorption flux as

$$F = e^{-\tau}\,. \tag{6.1}$$

The flux F is the field we analyse. For computing its power spectrum $P_{Ly\alpha}(k)$ we compute the flux contrast as

$$\delta_F(\mathbf{x}) = \frac{F(\mathbf{x}) - \langle F\rangle}{\langle F\rangle} \tag{6.2}$$

where $\langle F\rangle$ is the flux mean; later in Section 6.2.1 we discuss some caveats related to $\langle F\rangle$.

6.1.1 *The Lyα forest auto-power spectrum*

We first consider the one-dimensional Lyα power spectrum $P^{1D}_{Ly\alpha}(k)$. For each spectrum we compute the flux contrast δ_F in every pixel along the l.o.s., hence the flux power spectrum. In Figure 6.1 we show in the upper panel the 2 dimensionless power spectra

$$k\,P^{1D}_{Ly\alpha}(k)/\pi$$

Catalogue	Number of spectra	Direction of spectra	Pixels per spectrum
1	50^2	x	2048
2	40^2	y	2048
3	30^2	z	2048

Table 6.2: Summary of the Lyα absorption spectra catalogues, for each simulation.

with solid blue line for the 80-2048 simulation and in dashed red for 160-1024. The shown power spectra have been computed by averaging the measured power spectrum from each individual skewer from all catalogues (5000 spectra in total: 2500 along x plus 1600 along y plus 900 along z). We find that our results converge against resolution at a 20% level.

Next we compute the 3D Lyα power spectrum $P_{Ly\alpha}(k)$, shown in Figure 6.2. This is achieved by considering the 3D δ_F, interpolated on a grid that covers the whole simulation box. There is good overlap among the six power spectra (giving that these catalogues have not been normalised, see later in Section 6.2.1), especially on large scales, with the 80-2048 spectra slightly flattening for $k \lesssim 0.2 h\,\mathrm{Mpc}^{-1}$. At smaller scales, the 160-1024 spectra display an increase of power compared to the 80-2048 spectra because of the lack of small scale information. The same effect is visible also among the catalogues belonging to the same simulations: as the number of spectra per catalogue decreases (i.e. employing a smaller number of skewers in the box) we see an increase of power due to a sub-sampling of the field.

6.2 ON SOME STANDARD PRACTICES IN GENERATING MOCK LYα FOREST SPECTRA

In this section we discuss some of the issues we have encountered when computing the 3D power spectrum of the Lyα forest. We make use of the 60-512 hydrodynamic simulation presented in Section 3.4.1, which is smaller and numerically easier to handle than the Sherwood's. We work at redshift $z = 3$. We extract mock Lyα absorption spectra skewers as described in Section 6.1. All absorption spectra used in this section contain 256 pixels each and are taken along the x direction of the 60-512 simulation box.

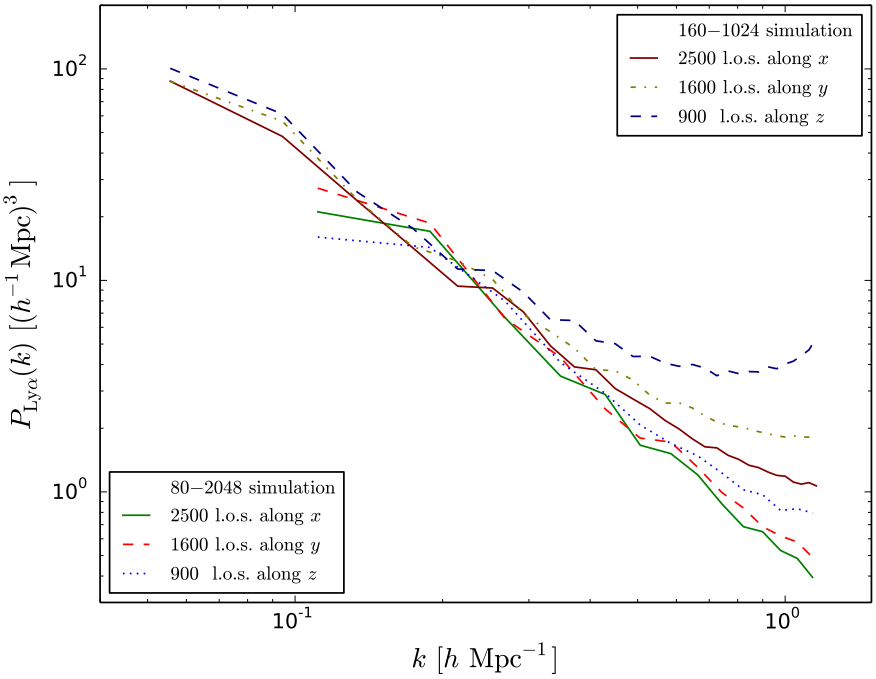


Figure 6.2: The 3D Lyα flux power spectrum computed for the 3 different catalogues (with spectra along different directions) for both simulations.

6.2.1 *Normalising the spectra with* τ_{eff}

The amplitude of the 3D Lyα power spectrum $P_{\mathrm{Ly}\alpha}(k)$ depends on the actual observed cosmological mean transmitted flux $\langle F\rangle_{\mathrm{obs}}$, measured in e.g. [12, 59, 75]. When the absorption spectra are recovered artificially by piercing the simulation box with skewers, the mean value $\langle F\rangle_{\mathrm{cat}}$ of the catalogue of spectra varies, due primarily to the number of skewers that are drawn, as already seen earlier in Figure 6.2, i.e. the more we sample the box, the more flux we get overall. Of course in a survey we have a finite amount of lines of sight ($nlos$). The question is: how should the mean flux be computed to get sensible results for the 3D Lyα power spectrum?

A possibility is to take a real survey l.o.s. density, for example we know that BOSS roughly detects 15 quasars spectra per deg^2 [141], but this would make a very low number of skewers in a typical hydrodynamical simulation box and let arise other computational problems (see next Section 6.2.2).

It is also possible to take the $nlos$ necessary to have $\langle F\rangle_{\mathrm{cat}} = \langle F\rangle_{\mathrm{obs}}$, i.e. to match the simulation flux mean with the measured one, so it would be naively expected that the simulation $P_{\mathrm{Ly}\alpha}(k)$ converges with the measured one. But this is not completely true as we now discuss.

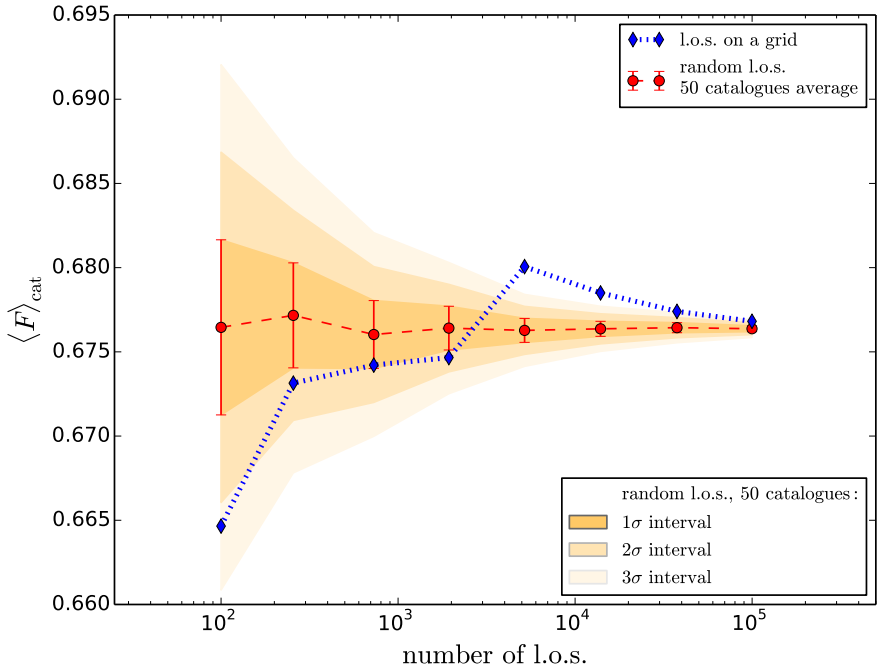


Figure 6.3: Mean flux $\langle F \rangle$ of the Lyα absorption spectra catalogues, plotted versus the nlos in each catalogue. 8 catalogues have the l.o.s. placed on a regular grid (blue diamonds). Each red point represents the average of the mean flux from 50 different catalogues with same number of random l.o.s.. The error bars display the 1σ variation from the 50 catalogues, while the shaded orange areas show the 1σ, 2σ and 3σ deviations.

In Figure 6.3 we show $\langle F \rangle_{cat}$ coming from catalogues with the following nlos: 10^2, 16^2, 27^2, 44^2, 72^2, 118^2, 194^2 and 316^2. The skewers have been placed either on a $\sqrt{nlos} \times \sqrt{nlos}$ regular grid (blue diamonds, each point corresponds to one catalogue) or randomly (red dot, each point correspond to the average among 50 different catalogues). In case of random l.o.s., we also plot the 1σ error bar on the average $\langle F \rangle_{cat}$ and show with shaded area the 2σ and 3σ intervals too. The scatter is big for the catalogues with smaller nlos, nevertheless their average $\langle F \rangle_{cat}$ agrees with what we obtain with a huge amount of l.o.s.. We thus conclude that the catalogue mean flux is not very sensitive to the value of the surface density with which the Lyα field is sampled through skewers, i.e. the quantity $\langle F \rangle_{cat}$ is a property of the specific realization itself, marginally dependent on the choice of nlos.

Throughout the literature (e.g. [12, 20, 86, 153, 168]), the general practice consists in artificially change the amplitude of the UV background strength to obtain the desired mean transmitted flux:

$$\langle e^{-B\tau_i} \rangle = \langle e^{-\tau_{eff}} \rangle = \langle F \rangle_{obs} \,, \tag{6.3}$$

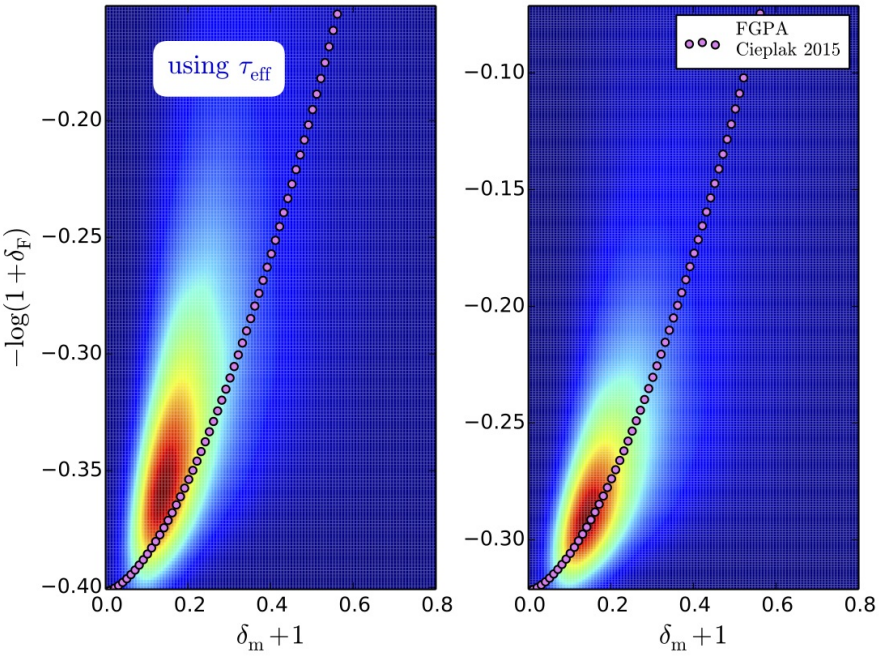


Figure 6.4: Density of points (red colour for high density values, blue for low) for the scatter of $(\delta_m + 1)$ - $(-\log(1 + \delta_F))$ calculated for the 60-512 simulation pierced by 100 random l.o.s. with 256 pixels each. The magenta dotted line is the FGPA relation of Equation 6.4 with the parameters in [42]. *Left panel:* The spectra have been normalised by using an effective optical depth $\tau_{\rm eff}$. *Right panel:* The spectra have not been normalised.

where τ_i is the optical depth of a single pixel, B is an ad-hoc chosen parameter representing the variation in the amplitude of the UV background and $\tau_{\rm eff}$ is called effective optical depth (e.g. [12]).

By construction, employing $\tau_{\rm eff}$ makes $\langle F\rangle_{\rm cat} = \langle F\rangle_{\rm obs}$, but the shape and amplitude of the Lyα power spectrum is affected by this change, particularly on small scales and not only: by linearly shifting the pixel optical depth, we are non-linearly changing its flux.

To clarify this point, we make use of the fluctuating Gunn-Peterson approximation (FGPA, [49]) that relates the matter and transmitted flux fields through:

$$F = e^{-A(1+\delta_m)^\alpha}, \tag{6.4}$$

where δ_m is the matter density contrast and for the other parameters we adopt the values analytically computed in [42]: $A = 0.3((1+z)/(1+2.4))^{4.5}$ and $\alpha = 1.6$. Although neglecting some small scales physics, the FGPA has been first derived analytically and has had great success in explaining statistical observed properties of the Lyα forest. We check whether FGPA holds for our mock spectra. We generate 2 catalogues composed by the same 100 random l.o.s. (i.e. BOSS resolution in the 60^2 $(h^{-1}\mathrm{Mpc})^2$ simulation area) with 256 pixels each, but one of them is normalised to $\langle F\rangle_{\rm obs}$ using $\tau_{\rm eff}$ of Equation 6.3. We

interpolate the flux F of each pixel and the matter particles onto a 256^3 grid covering the whole simulation volume, obtaining 3D density maps: of flux and of matter. We compare these 2 density maps in Figure 6.4: we colour code the density of points in the F − matter scatter plot and we plot with magenta dots the FGPA relation[2] of Equation 6.4. The left panel refers to the normalised catalogue, the right panel to the non-normalised. We find that the flux normalisation shifts the whole flux field (see the different y scale on the 2 panels), but also spoils the flux − matter relation predicted by the FGPA: the right panel (non-normalised spectra) has a much better agreement with FGPA that gets lost after the $\tau_{\rm eff}$ -normalisation (left panel).

To have convergence between a mock Lyα catalogue $P_{Ly\alpha}(k)$ and observations or to compare different simulations Lyα power spectra, instead of employing the effective $\tau_{\rm eff}$, a wiser choice would be to shift whole spectra a posteriori, rather than degrade the small scale information of the flux.

None of the Lyα spectra catalogues of this thesis has been normalised using $\tau_{\rm eff}$, at the expenses of having a small (yet understood) large scale discrepancies in $P_{Ly\alpha}(k)$ (as in Figure 6.2); still, the statistical properties of the Lyα forest flux are remarkably recovered (see later in Section 7.3.2 the bias and redshift space distortion parameter estimation).

6.2.2 *Placing lines of sight on a regular grid*

The question on the proper number of l.o.s. to use and on the way to place them in the simulation box is related to another central issue. To compute $P_{Ly\alpha}(k)$ we need a good coverage of skewers in the box because we discretise the box into cells, make 3D flux density maps and perform discrete Fourier transforms: having cells not crossed by a line of sight would be a concern because no flux information is available, hence it will be not possible to assign a field value in a rigorous way.

Among possible solutions we can think of, there are:

i) computing the 3D Lyα power spectrum masking out the regions not sampled by the mock spectra,

ii) working in configuration-space (i.e. compute correlation functions), as in [56, 140].

Unfortunately, these solutions are computationally expensive and, in the case of i), technically hard to implement[3].

2 We stress again that we take the parameters of Equation 6.4 from the theoretical work by [42]: nothing has been calibrated a posteriori, we blindly compare our results with [42].

3 The needed mask would be the whole simulation box without the *straws* where we place the lines of sight, and we would need to Fourier transform it too.

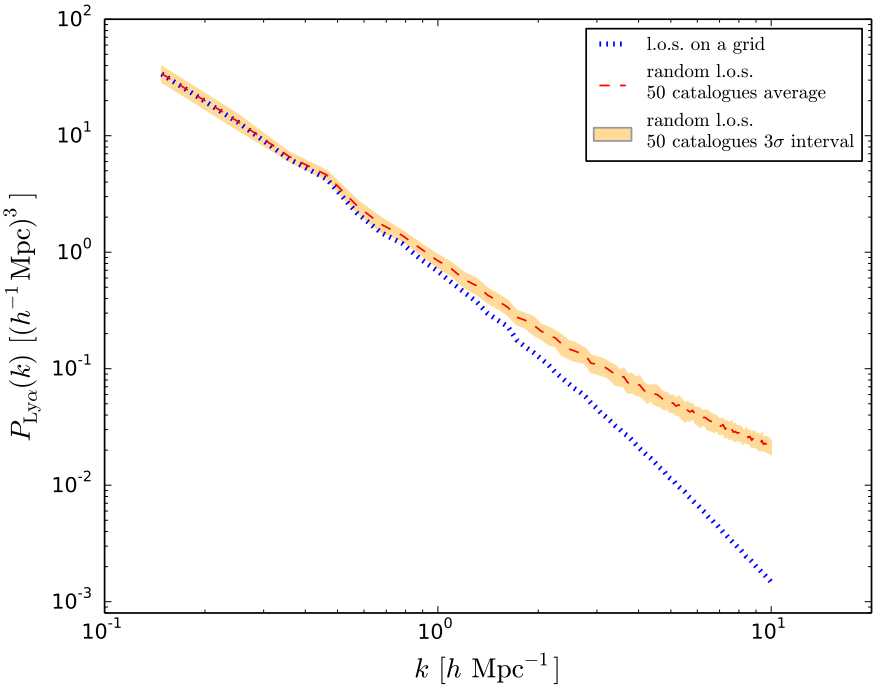


Figure 6.5: 3D Lyα flux power spectra $P_{Ly\alpha}(k)$ at $z = 3$ obtained by placing the skewers into a regular grid (dotted blue) or distributing them randomly (dashed red). For the regular grid we employ 256^2 spectra while for the randomly distributed skewers we create 50 different catalogues, each of them containing 256^2 spectra, and we show their average and 3σ variation.

The standard practice is to place the skewers into a regular grid, this way the whole box is sampled uniformly, although real quasars are not on a grid. We check what is the effect of regularising the spectra position. In Figure 6.5 we plot with a red dashed line the average of 50 Lyα power spectra, each corresponding to a catalogue with $256^2 = 65536$ l.o.s. randomly placed, and with a dotted blue line the $P_{Ly\alpha}(k)$ of a catalogue of the same number of l.o.s. but placed on a 256×256 regular grid in the $y - z$ plane. The 2 power spectra converge at large scales, the sub-sampling effect discussed above becomes relevant at scales as small as $k \gtrsim 1$ hMpc^{-1}. Thus, we are allowed to employ this approximation for the cross-correlation study next in Chapter 7, where we look at scales as large as $k \lesssim 0.2$ hMpc^{-1}.

We conclude that our method to compute the 3D Lyα power spectrum is robust, on large scales, against the different ways of sampling the Lyα field. Given the good agreement on both the 1D and 3D Lyα power spectra among the 2 simulations (on large scales for the 3D power spectrum and on all scales in the 1D case) we conclude that the spatial distribution of gas in the Lyα forest is converged against resolution in the simulation 160-1048, that we employ next in Chapter 7 for large scale cross-correlation studies.

7

HELPING OUT THE 21CM INTENSITY MAPPING: CROSS-CORRELATION WITH THE LYMAN-α FOREST FLUX

In Section 1.2.1.1 we illustrated how major drawbacks of the 21cm intensity mapping (IM) technique are systematics and especially foregrounds, expected to be order of magnitudes higher than the cosmological signal. This motivates the study presented in this chapter, where we cross-correlate the signals of 21cm IM and Lyα forest. Both probes originate from cosmic neutral hydrogen (HI): 21cm IM probing HI in emission, the Lyα forest in absorption. Hence, their cross-correlation can provide a coherent and comprehensive picture of the HI content of our Universe in the post-reionization era, probing both its mass content and volume distribution.

In this chapter we perform the analysis in the fully non-linear regime using state-of-the-art hydrodynamic simulations. We compute the auto-power spectra of both fields together with their cross-power spectrum at $z = 2.4$, finding them completely anti-correlated. We derive the bias parameters and test how well and up to which scale linear theory is able to describe the 21cm field[1]. We proceed at determining the detectability of the cross-correlation signal with SKA1-MID and a BOSS-like survey, and check how much cross-correlation is able to shrink the uncertainties on values and degeneracies of the bias parameters of the 21cm field.

7.1 WHY CROSS-CORRELATING

Cross-correlation analysis between independent probes are a useful tool to infer further information about the Universe while providing extra safety checks on known (and unknown) systematics affecting the data sets. Indeed, if we have 2 independent observables $\{A, B\}$ their estimate can be split into a signal and a noise term as $\hat{A} = S^A + N^A$ (same applies for B). Then, the 2-point correlator is given by

$$\begin{aligned} \langle \hat{A}\hat{B} \rangle &= \langle (S^A + N^A)(S^B + N^B) \rangle \\ &= \langle S^A S^B \rangle + \cancelto{0}{\langle N^A S^B \rangle} + \cancelto{0}{\langle S^A N^B \rangle} + \cancelto{0}{\langle N^A N^B \rangle} \end{aligned} \tag{7.1}$$

where we used the fact that for uncorrelated noise the signal-noise and noise-noise cross terms vanish: is this true for the probes we want to study?

1 As discussed in Section 1.2.1.1, a correct theoretical template for 21cm IM is still missing, but it is crucial for best benefiting from future observations.

We have already discussed in Section 1.2.1.1 systematics and foregrounds affecting the 21cm IM, those influencing Lyα observations are conveniently of totally different kind [75, 93], as for example i) proper modelling and subtraction of the quasar continuum flux, ii) incorporating the fluctuations of the ionizing source, iii) uncertainties in the intergalactic medium temperature-density relation and iv) contamination of the spectra by metal lines. Moreover for 21cm observations we use radio telescopes, for the Lyα forest spectrographes: instrumental issues are of different kind too.

Hence, mixing these 2 different probes allows to extract signals hidden in noisy data, providing us a clean and unbiased measurement of the cross-correlation signal.

The redshift choice of $z = 2.4$ is convenient for our purposes since it belongs to the z-range for which the highest number of quasars is observed (hence most Lyα forest data available, for example see [141]) and for which 21cm observations are planned to be carried out (see the instrument list in Section 1.2.1.1).

7.2 HALO CATALOGUES AND 21CM MODELLING

We rely on 2 high-resolution hydrodynamic simulations that belong to the Sherwood simulation suite [20], already presented in Chapter 6: 80-2048 and 160-1024. They are labelled this way since 80-2048 follows the evolution of 2×2048^3 cold dark matter plus baryon particles within a periodic box of linear comoving size of $80\,h^{-1}$Mpc, 160-1024 follows the evolution of 2×1024^3 cold dark matter plus baryon particles within a periodic box of linear comoving size of 160 h^{-1} Mpc. They evolve from $z = 99$ to $z = 2.4$, the latter being the snapshot considered in this analysis.

The main physical ingredients of the hydrodynamics are presented in Chapter 6. Here we stress that these simulations are tailored to give converged properties for low density intergalactic medium statistics as probed by the Lyα forest: on this side they are the state-of-the-art simulations. On the other hand, the implemented schematic star formation makes the amount of cold gas contained in halos unreliable[2]. Thus, we conservatively avoid using the particle based scheme

2 In the hydrodynamic SPH simulations the spatial distribution of gas is discretised into a finite number of gas particles with a given kernel and radius. The phase in which the gas is (ionized, neutral, forming molecules) can be found through radiative-transfer calculations. While this is the most robust way to model the spatial distribution of HI, it is also the most computationally expensive. Unfortunately, a radiative-transfer calculation will not output the correct HI distribution in our simulations. The reason for this is because the hydrodynamic simulations have been run using the so-called "quick-Lyα" flag: the code follows the full hydrodynamic evolution of the gas until this reaches a given density and temperature threshold; at that stage the code will transform the gas particle into a collisionless star. This technique allows to speed up calculations by avoiding modelling the gas in the inter-stellar

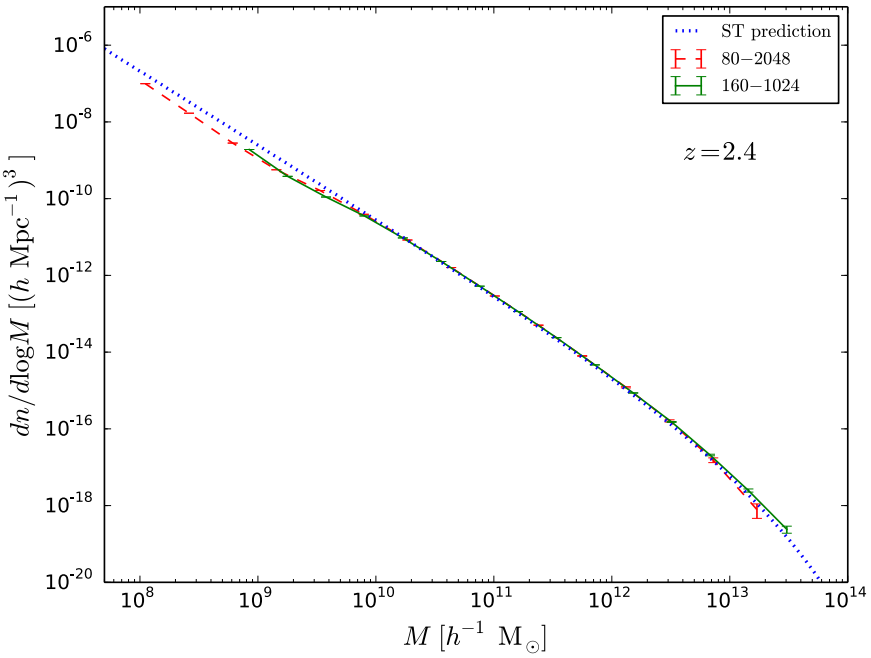


Figure 7.1: Halo mass functions of the 80-2048 and 160-1024 simulations (red dashed and green solid lines), compared with the Sheth & Tormen prediction [138] (blue dotted line). Uncertainties are computed assuming that the number of halos follows a Poissonian distribution.

for distributing HI, and we adopt to the halo based one (for the HI distributing methods see Chapter 3).

We identify halos using the Friend-of-Friends (FoF) [52] algorithm with a linking parameter length of $b = 0.2$. As already discussed in Section 3.1, the halo information is crucial for this analysis since our 21cm maps are derived from it, this is also why we adopt 2 simulations: the 160-1024 simulation (whose volume is 8 times bigger than that of 80-2048) is the one we use for studying the shape and amplitude of the cross-signal on linear and mildly non-linear scales ($k < 0.2\ h\mathrm{Mpc}^{-1}$) in next Section 7.3, but we make use of the 80-2048 simulation (with higher mass and spatial resolution) to check the convergence of the halo mass function and, eventually, of the 21cm signal[3].

In Figure 7.1 we plot the halo mass function for the 80-2048 and 160-1024 simulations in red and green respectively, while the dotted blue line shows the Sheth & Tormen prediction [138] for the corresponding redshift $z = 2.4$. We find that the halo abundances from the simulations agree well with each other in the common mass range and also

medium (ISM), where most of the HI is located. This kind of simulations produce too many stars that make the gas reservoir unreliable.

3 Analogously, convergence tests with the 2 simulations are performed for the Lyα field in Chapter 6.

with the ST prediction. The higher mass resolution of 80-2048 allows us to sample the low mass end of the halo mass function while the larger box size of 160-1024 is better suited to explore the high mass end. Thus, the combination of both simulations enables us to sample the halo mass function on a wide range of masses as shown in Figure 7.1, which is important for our modelling of the 21cm signal.

For distributing HI in both simulations, we make use of the halo based model B of Equation 3.2 described in Section 3.1. M_{min} corresponds to halos with circular velocity of $v_{circ} = 25$ km s^{-1} [163] and we set $\Omega_{HI} = 10^{-3}$. We have model the HI density profile as:

$$\rho_{HI}(r|M,z) = M_{HI}(M,z)\delta(\mathbf{r}) \tag{7.2}$$

where $\delta(\mathbf{x})$ is the Dirac delta. Thus, we are collapsing the HI density distribution into a single point located in the halo centre. This is of course a very crude approximation, but given the fact that in our simulations the spatial distribution of gas is unreliable, this is the most simple choice that does not involve ad-hoc assumptions. We show in Section 3.4.1 that the approximation of Equation 7.2 for the HI density profile does not bias our conclusions, since our analysis focuses at studying the cross-correlation signal at large scale[4].

Since we employ a halo based method, a N-body simulation would have been enough for our purposes, with the advantage to be less computational expensive to run and to reach larger scales more easily. On the other hand, we remind that a precise modelling of the Lyα forest requires high-resolution hydrodynamic simulations as we describe in detail in Chapter 6. We decide to use state-of-the-art hydrodynamic simulations that accurately model the properties of the Lyα forest at the expenses of neglecting the contribution of HI outside halos to the global 21cm signal (which was found negligible in [37, 161]) and avoiding modelling the intrinsic scatter in the $M_{HI}(M,z)$ function, that represents the average HI mass that a dark matter halo of mass M hosts at redshift z. However, we notice that we also tested a method to model the Lyα signal in pure dark matter simulations (see the fluctuating Gunn-Peterson approximation introduced in Section 6.2.1).

Following Section 3.4, we compute the 21cm signal. We do a direct comparison between the 2 simulation 21cm power spectra in Figure 7.2, upper panel, with the residuals $P_{21cm}^{160-1024}(k)$ over $P_{21cm}^{80-2048}(k)$ in green in the lower panel. We find that the shape of both power spectra agree on all scales, with a $\sim$ 20% offset in amplitude among the two: HI is more clustered in 160-1024. The larger simulation box of 160-1024 is able to capture the large-scale modes (which are not present in 80-2048) thus enhancing the HI clustering on lower scales and producing a higher signal amplitude.

4 The clustering of HI on large-scales is fully determined by the function $M_{HI}(M,z)$, while on smaller scales it depends on the way the HI is distributed within halos.

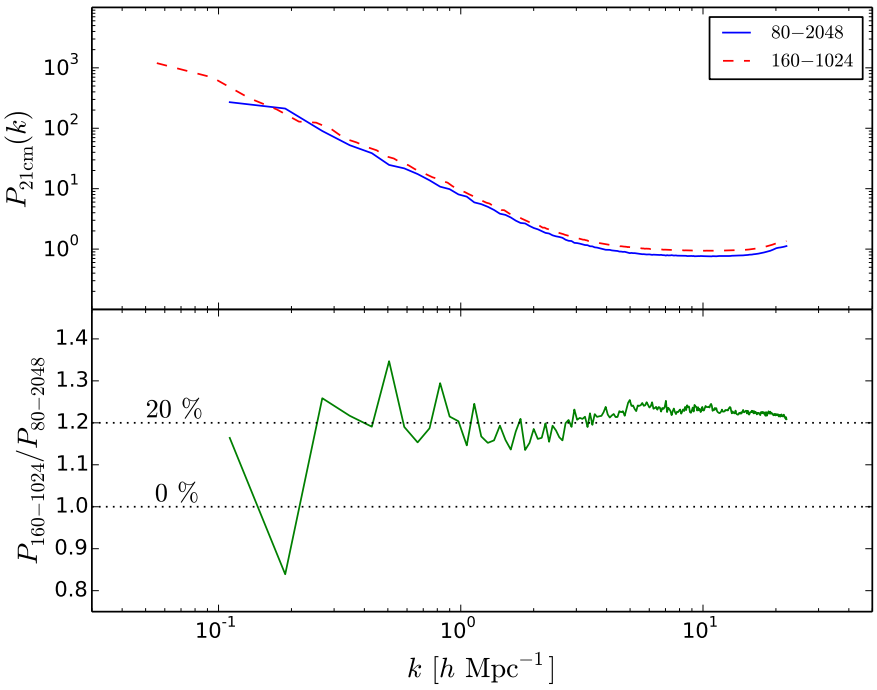


Figure 7.2: The upper panel displays the 21cm power spectrum from the 80-2048 (solid blue) and 160-1048 (dashed red) simulations at $z = 2.4$ while the lower panel shows their ratio.

These volume effects in the 21cm power spectrum are discussed also in Section 5.2.1, where we find same order of discrepancies. It affects the overall amplitude of the signal, for which we need anyway to take into account the effect of the poorly constrained Ω_{HI} (see discussion in Section 3.3.1).

7.3 THE 21CM - LYα CROSS CORRELATION

As anticipated above, in the analysis of this section we employ the 160-1024 simulation since we are interested in investigating the shape and amplitude of the cross-power spectrum on as large as possible scales.

Having modelled both fields (in Chapter 6 we illustrate and test how we model the Lyα flux field on 160-1024), in Figure 7.3a we depict the HI density contrast in redshift-space (upper snapshot) and the Lyα flux contrast (bottom snapshot), at $z = 2.4$. It is evident how HI tracks the halo positions in the simulation box, thus revealing the web-like structure of the matter in the Universe. We notice that since we are not modelling the distribution of HI inside each halo, the presence of fingers-of-God in Figure 7.3a is suppressed.

Our HI modelling is fully deterministic. By using our HI mass function prescription of Equation 3.2 with the halo mass function and halo

bias from the Sheth & Tormen prediction [136, 138], we can compute Equation 3.6 for the value of the HI bias $b_{HI}(z = 2.4) = 1.45$. The latter b_{HI} might seem low compared for example to the bias of DLAs (see discussions in Chapter 3). Anyway, the purpose of this study is to be able to retrieve the input HI model values when fitting the results of the simulations within the linear theory framework. We also emphasize that a higher value of the HI bias will turn out to make the HI bias more scale-dependent [161], and therefore shifting to large scales the onset of non-linearities, a situation we want to avoid given the relatively small volume of our simulations.

We compute the cross-correlation of the 21cm and Lyα forest signals. In Figure 7.4 we show with a dotted green line the 21cm power spectrum $P_{21cm}(k)$, in dashed blue the 3D Lyα flux power spectrum $P_{Ly\alpha}(k)$ and in solid magenta the absolute value of the 21cm-Lyα cross-power spectrum $P_X(k)$.

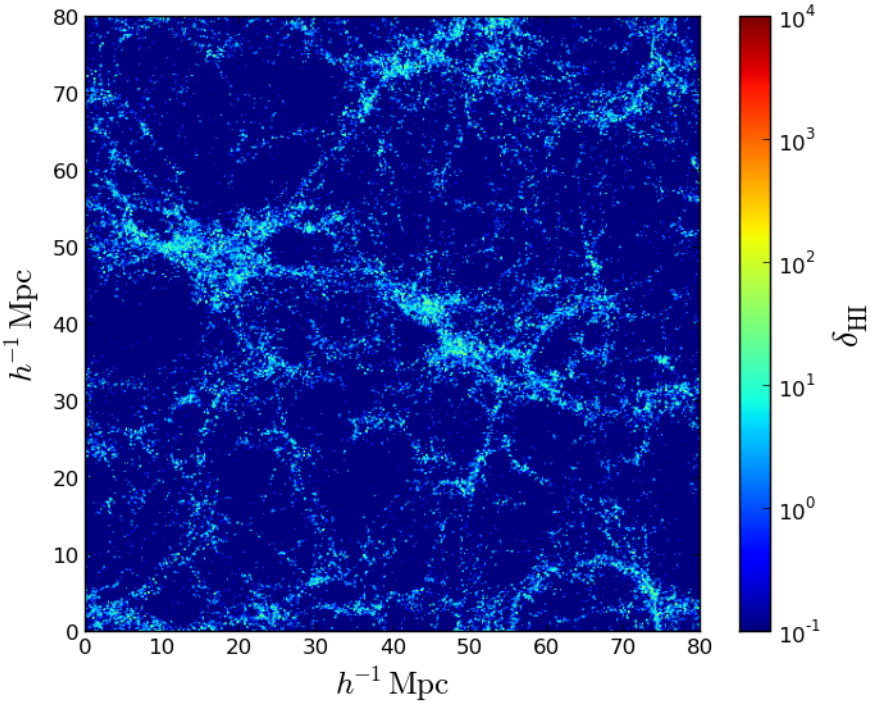
$P_X(k)$ is plotted in absolute value because it is negative, i.e. the fields are anti-correlated. Qualitatively, this results agrees with the picture of having the HI responsible for the 21cm radiation in dense environments as galaxies, whereas the Lyα forest arises mainly from low density, highly ionized, gas clouds in the intergalactic medium mostly residing in the filaments of the cosmic web. We thus expect underdensities in the Lyα flux in places where 21cm overdensities are located (e.g. halos): this is exactly what being anti-correlated means. By visually inspecting the two fields in Figure 7.3, we indeed notice how HI-poor regions (Figure 7.3a, dark blue) correspond to high Lyα forest transmitted flux regions (Figure 7.3b, yellow and red). We notice that this result agrees with recent observations by Mukae et al., 2017 [102], where they analyse HI-rich galaxies at $z \sim 2-3$ together with Lyα forest spectra finding anti-correlation between the two fields.

In the bottom panel of Figure 7.4 we plot the cross-correlation coefficient, defined as

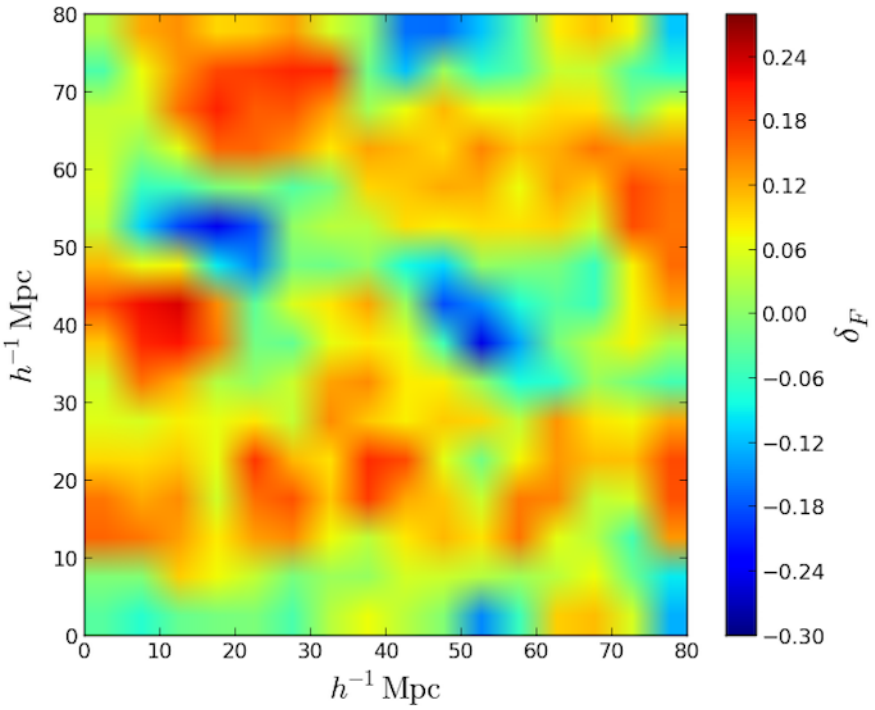
$$r(k) = \frac{P_X(k)}{\sqrt{P_{21cm}(k) P_{Ly\alpha}(k)}} \,. \tag{7.3}$$

As expected from the arguments given above, we find the sign of the cross-correlation to be negative on all the scales, indicating that the fields are anti-correlated. On the largest scales probed by our simulation the value of the cross-correlation is close to -1, while for scales $k \geqslant 0.2\ h\mathrm{Mpc}^{-1}$ the value of the cross-correlation increases. We can naively associate this scale to non-linearities, as linear theory predicts a scale-independent cross-correlation coefficient. Thus, since we are interested in extracting cosmological information from linear scales, we limit our analysis to modes with wavenumbers $k < 0.2\ h\mathrm{Mpc}^{-1}$.

In a nutshell: we are looking at signals coming from different regions in the sky. By considering together 21cm radiation and Lyα flux, we are looking at HI both in emission and in absorption, thus



(a) Density contrast of HI in redshift-space.



(b) Density contrast of the Lyα forest flux.

Figure 7.3: Spatial distribution of the 2 fields at $z = 2.4$. We show the whole 80-2048 simulation box, taking a slice of $10\,h^{-1}$ Mpc width.

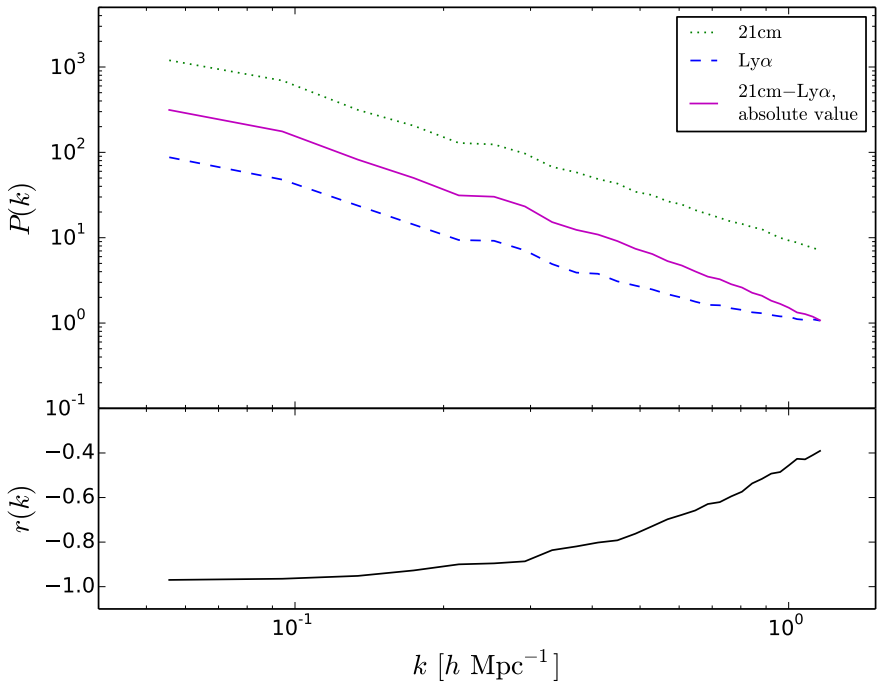


Figure 7.4: The upper panel displays the power spectrum of the 21cm signal (dotted green), of the Lyα forest flux (dashed blue) and of their cross-correlation in absolute value (solid magenta). The bottom panel shows the cross-correlation coefficient, $r(k)$, among the two fields.

probing its cosmological amount and its spatial distribution at the same time. This is a clear hint of why the cross-correlation signal 21cm - Lyα constitutes an optimal large scale structure probe.

To quantify the improvement gained by looking for HI information in cross-correlation, next in Section 7.3.1 we make predictions on the accuracy of this measurement.

7.3.1 *Error estimation and forecasts for SKA*

In this section we estimate the signal to noise ratio of a measurement of the auto- and cross-power spectrum of the Lyα forest and the 21cm fields from SKA1-MID [24] and a BOSS-like survey [53] for the Lyα flux field[5]. As already pointed out, in this work we just focus on the cosmological signal, i.e. we neglect contributions from residual foreground contamination.

5 As for example DESI, the Dark Energy Spectroscopic Instrument (http://desi.lbl.gov), that has a planned time-line similar to the one of SKA.

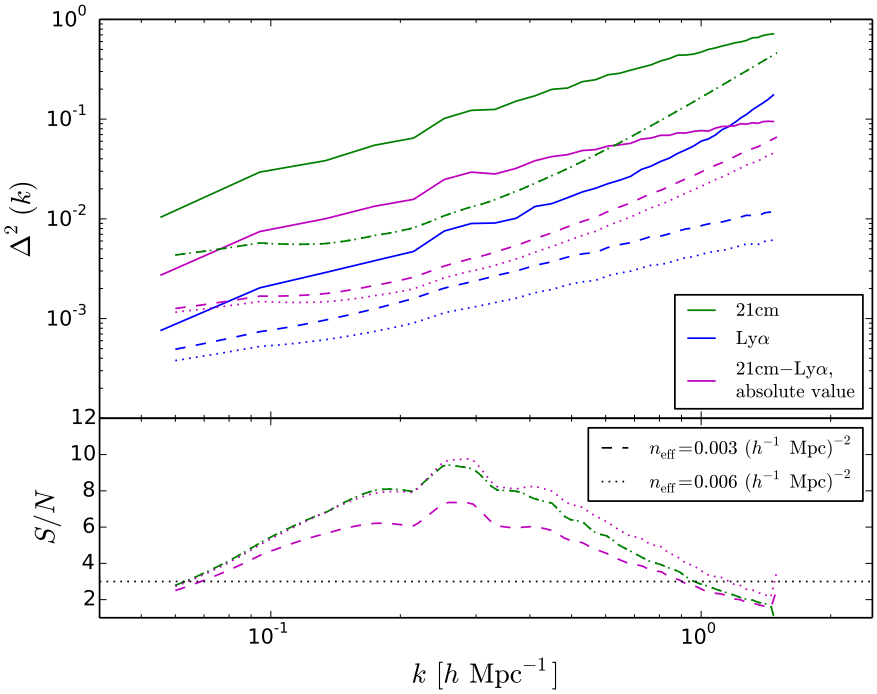


Figure 7.5: *Upper panel:* The dimensionless power spectra $\Delta^2(k) = k^3 P(k)/2\pi^2$ in solid lines for the 21cm radiation (green), the Lyα forest flux (blue) and for their cross-correlation in absolute value (magenta). The non-solid lines are their estimated errors for a SKA1-MID (21cm intensity mapping) and BOSS like (Lyα flux) surveys, the latter with an effective density of lines of sight $n_{\rm eff} = 0.003(h^{-1}\,{\rm Mpc})^{-2}$ in dashed lines or $n_{\rm eff} = 0.006(h^{-1}\,{\rm Mpc})^{-2}$ in dotted lines. *Lower panel:* Signal-to-noise ratio. The horizontal dotted line marks $S/N = 3$.

The accuracy with which one can measure the power spectrum $P_a(\mathbf{k})$, where a stands for 21cm, Lyα or 21cm-Lyα, is generally quantified by the signal to noise ratio S/N:

$$\left(\frac{S}{N}\right)^2_a(\mathbf{k}) = N_k \frac{P_a^2(\mathbf{k})}{\sigma^2[P_a(\mathbf{k})]}, \tag{7.4}$$

where

- N_k number of modes in each given bin centred at $(k, \cos(\theta))$
- θ angle between $\mathbf{k}$ and the line of sight.

For simplicity and clearness, we focus on errors on the monopoles: we average the amplitude of the power spectrum of modes with $k \in [k, k + dk]$, thus, the S/N ratio becomes:

$$\left(\frac{S}{N}\right)^2_a = \frac{2\pi k^2 dk V_{\rm survey}}{(2\pi)^3} \int_0^{\frac{\pi}{2}} \frac{P_a^2(k,\theta)\sin(\theta)d\theta}{\sigma^2[P_a(k,\theta)]}, \tag{7.5}$$

We can cast the error on the 21cm detection $\sigma^2[P_{21cm}(k,\theta)]$ via interferometric observations as already shown in Section 3.4.2 through

Equation 3.15. All the parameters in Equation 3.15 are set by the specifications of SKA1-MID and the redshift of observation (in our case $z = 2.4$). What we can tune is the total observation time t_0 that we conservatively set to $t_0 = 100$ hours (we later show how results do not depend on this choice).

We express the error on the 3D Lyα flux power spectrum as a combination of the noise term with an aliasing term, due to the sparse sampling of the Lyα field made by the discrete lines of sight [63]:

$$\sigma^2[P_{Ly\alpha}(k,\theta)] = \left[P_{Ly\alpha}(k,\theta) + P^{1D}_{Ly\alpha}(k\cos\theta)n_{eff}^{-1}\right]^2 . \tag{7.6}$$

The aliasing term contributes in the line of sight direction with the 1D flux power spectrum $P^{1D}_{Ly\alpha}(k\cos\theta)$ multiplied by a noise-weighted density of lines of sight per unit area n_{eff}. From [141] we know that for a redshift bin to which $z = 2.4$ belongs the lines of sight density for BOSS is $n_{eff} \simeq 15\,deg^{-2}$. We choose to use both this value and its double ($30\,deg^{-2}$) to estimate errors for a BOSS-like and a next generation BOSS-like survey like DESI [6]. Using the conversion table in [94], we obtained the values $n_{eff} = 0.003$ and $0.006\,(h^{-1}\,Mpc)^{-2}$.

Finally, the error on the measurement of the cross-correlation power spectrum can be written as:

$$\sigma^2[P_X(k,\theta)] = \frac{1}{2}\left(P_X^2(k,\theta) + \sigma[P_{21cm}(k,\theta)]\sigma[P_{Ly\alpha}(k,\theta)]\right) . \tag{7.7}$$

On small scales, the error budget of the 21cm power spectrum is dominated by the system noise and, by looking at Equation 3.15, we know it scales with the observing time as $1/t_0$, whereas for the cross-correlation power spectrum (see Equation 7.7) goes as $1/\sqrt{t_0}$, so if we choose a more optimistic survey observing time of $t_0 = 1000$ hours instead of 100 hours, $\sigma[P_{21cm}(k)]$ would improve by a factor 10 and $\sigma[P_X(k)]$ would be ~ 3 times smaller. Anyway, at the scales we are looking at ($k < 0.2h\,Mpc^{-1}$) this is no longer valid: observing for longer does not beat cosmic variance. Hence, we keep $t_0 = 100$ hours for our analysis and we do not investigate for other possible observing times.

We summarise the errors estimation in Figure 7.5, where we plot their magnitudes together with the different auto- and cross-power spectra. In the upper panel of Figure 7.5 we plot the dimensionless power spectrum $\Delta^2_a(k) = k^3P_a(k)/(2\pi^2)$ with solid lines: green for the 21cm radiation, blue for the Lyα flux and magenta for the absolute value of their cross-signal, together with the errors $\sigma[P_a(k)]$ in dashed lines using $n_{eff} = 0.003\,(h^{-1}\,Mpc)^{-2}$ and dotted lines for $n_{eff} = 0.006\,(h^{-1}\,Mpc)^{-2}$. Clearly, the error on the 21cm power spectrum does not depend on the value of n_{eff}, and the improvement on the $P_{21cm-Ly\alpha}(k)$ is evident but smaller than that on $P_{Ly\alpha}(k)$. Nonetheless,

6 http://desi.lbl.gov/

we can notice that the aliasing term in $\sigma^2[P_{Ly\alpha}(k)]$ (and consequently in $\sigma^2[P_X(k)]$) dominates the variance budget, therefore the way to do better would be to conduct more sensitive Lyα spectra surveys, e.g. by increasing the surveyed volume and/or by increasing the quasars number density (up to its true cosmic value).

In the bottom panel of Figure 7.5 we plot the signal to noise ratio S/N for $P_{21cm}(k)$ and $P_{21cm-Ly\alpha}(k)$. We find that S/N ratio for both power spectra peak around $k \simeq 0.25\ h\text{Mpc}^{-1}$. On larger scales cosmic variance dominates the error budget, while on smaller scales instrumental noise kicks in.

We thus conclude that the cross-power spectrum of the Lyα forest from a BOSS like survey with 21cm intensity mapping interferometry observations in the post-reionization era from the SKA1-MID instrument can be detected with a large S/N ratio on scales $k \in [6 \times 10^{-3} - 1]\ h\,\text{Mpc}^{-1}$ with a very conservative total observing time of 100 hours. We notice that these numbers should be regarded as lower limits, since our simplistic model for $\rho_{HI}(r|M,z)$ underestimates the clustering at the halo-size scale (since we lack the HI spatial distribution in halos, see Section 3.4.1).

The amplitude and shape of the cross-power spectrum on large-scales depends on the value of the bias and redshift-space distortion parameters of both fields. In the following section we check whether by analysing our simulated data we are able to retrieve the bias parameters and how well linear theory is able to describe the simulated fields.

7.3.2 *Linear theory comparison and bias parameters estimation*

In this section we introduce the theoretical framework for the auto- and cross-power spectra and use it to blindly estimate the value of the bias parameters of both the Lyα forest and the 21cm fields, by performing fits to the simulated power spectra. We also investigate how much information we gain by using the cross-power spectrum together with the auto-power spectrum measurements.

At linear order, the amplitude and shape of the 21cm, Lyα flux and 21cm-Lyα power spectra in redshift-space can be expressed as:

$$P_{21cm}(k,\mu) = A^2\Omega_{HI}^2 b_{HI}^2 \left(1+\beta_{HI}\mu^2\right)^2 P_m(k), \tag{7.8}$$

$$P_{Ly\alpha}(k,\mu) = b_F^2 \left(1+\beta_F\mu^2\right)^2 P_m(k), \tag{7.9}$$

$$P_X(k,\mu) = A\Omega_{HI} b_{HI} \left(1+\beta_{HI}\mu^2\right) b_F \left(1+\beta_F\mu^2\right) P_m(k), \tag{7.10}$$

where the normalisation factor A is computed from Equation 3.10-3.11, defining $\overline{\delta T}_b = A\Omega_{HI}$, with $\Omega_{HI} = 10^{-3}$ in our modelling, and

b_{HI}	linear bias of the 21cm field
β_{HI}	redshift-space distortion parameter of the 21cm field
b_F	linear bias of the Lyα forest
β_F	redshift-space distortion parameter of the Lyα forest
$P_m(k)$	linear matter power spectrum
μ	cosine of the angle between the Fourier mode vector $\mathbf{k}$ and the line of sight

From the numerically computed $P_{21cm}(\mathbf{k})$, $P_{Ly\alpha}(\mathbf{k})$ and $P_X(\mathbf{k})$, we determine the values of the 4 bias parameters (b_F, β_F, b_{HI}, β_{HI}) using 2 different methods.

METHOD 1: AUTO-POWER SPECTRUM MULTIPOLES. This method consists in determining the value of the 2 bias parameters of each observable through fitting the monopole and quadrupole of each respective field in redshift-space, i. e.we do not employ any cross-correlation info. The power spectrum multipoles can be computed from the 2D power spectrum as:

$$P_l(k) = \frac{2l+1}{2}\int_{-1}^{1} P(k,\mu)L_l(\mu)d\mu, \tag{7.11}$$

where $L_l(\mu)$ are Legendre polynomials. Thus, we can express the monopole and quadrupole as:

$$P_{a,0}(k) = b_a^2\left(1+\frac{2}{3}\beta_a+\frac{1}{5}\beta_a^2\right)P_m(k), \tag{7.12}$$

$$P_{a,2}(k) = b_a^2\beta_a\left(\frac{4}{3}+\frac{4}{7}\beta_a\right)P_m(k), \tag{7.13}$$

where a stands either for 21cm or for Lyα (and here we omitted the normalisation factors for the 21cm case, see above in Equation 7.8). If we were interested in the value of β_a only, we could extract it from the ratio between quadrupole and monopole, which depends only on this parameter; to constrain also b_a, we need to assume a cosmological model through $P_m(k)$[7]. The 21cm redshift-space distortion parameters contain information on the growth rate f, since

$$f = \beta_{HI} b_{HI} \tag{7.14}$$

in linear theory. Hence, by knowing b_{HI} and f (which is specified by the cosmology), it turns out that also β_{HI} is determined. Equation 7.14 does not hold for the Lyα case[8], due to its intrinsically non-linear correspondence to the underlying matter density field [141] and this

7 This is true when dealing with observations. For simulations the cosmology is anyway implicitly set.

8 See also the FGPA paragraph in Section 6.2.1 for an illustration of the Lyα flux being a non-linear tracer of matter.

is the reason why adding information from the cross-power spectrum cannot directly improve the measurement of the linear growth rate, but it tightens the constraints on the (b_{HI}, β_{HI}) parameters as we later show.

We do the best fit of the power spectra derived from the simulation to Equation 7.12-7.13 via a Monte Carlo Markov Chain (MCMC) on the 2 free parameters (b_α and β_α) making use of the *emcee* package [64]. Errors on the monopoles and quadrupoles of the auto-power spectra are computed assuming the modes follow a Gaussian distribution (see Appendix A), which is a good description on large and linear enough scales, and just taking into account the contribution from cosmic variance[9]. We consider a cosmological volume equal to the one probed by our simulations and using only modes with $k < 0.2\ h\mathrm{Mpc}^{-1}$. We account for the correlation between monopoles and quadrupoles of each single field in the covariance matrix, but for simplicity we neglect the correlation between multipoles of the two fields. The best-fit values together with their 1σ errors are shown in the upper row of Table 7.1.

The recovered values for the 21cm field

$$b_{HI} = 1.520^{+0.058}_{-0.060} \quad \text{and} \quad \beta_{HI} = 0.720^{+0.14}_{-0.13}$$

are in agreement with the input ones ($b_{HI} = 1.45$, $\beta_{HI} = 0.67$) at $\sim 1\sigma$. On the Lyα side, we find

$$b_F = -0.144^{+0.007}_{-0.007} \quad \text{and} \quad \beta_F = 1.480^{+0.21}_{-0.20}$$

while from observations it has been measured $b_F(1+\beta_F) = -0.374 \pm 0.007$ and $\beta_F = 1.39 \pm 0.1$ [17], thus in perfect agreement. We obtain a good normalized χ^2 of 13.6/12, showing the model is a good description of the data.

METHOD 2: CROSS P(k) MULTIPOLES. In this second method we exploit also the information contained in the cross-correlation power spectrum $P_\times(\mathbf{k})$. Following Equation 7.11 and Equation 7.10, we can write the cross-power spectrum multipoles as:

$$P_0(k) = b_{HI} b_F \left(1 + \frac{1}{3}(\beta_F + \beta_{HI}) + \frac{1}{5}\beta_F\beta_{HI}\right) P_m(k) \quad (7.15)$$

$$P_2(k) = b_{HI} b_F \left(\frac{2}{3}(\beta_F + \beta_{HI}) + \frac{4}{7}\beta_F\beta_{HI}\right) P_m(k)\,. \quad (7.16)$$

multiplied by a normalisation factor $A\Omega_{HI} = \overline{\delta T}_b$ from Equation 3.10-3.11. We fit simultaneously the monopoles and quadrupoles of the auto- and cross-power spectra, again employing the *emcee* package [64] to perform Monte Carlo Markov Chain (MCMC) on the 4 free parameters (b_F, β_F, b_{HI}, β_{HI}) employing only power spectra measurements for $k < 0.2\ h\mathrm{Mpc}^{-1}$. As in the case of the auto-power spectra,

9 Foreground contamination is neglected everywhere in this analysis and system noise is sub-dominant at these large scales as shown in Section 7.3.1.

b_F	β_F	b_{HI}	β_{HI}	χ^2/dof
$-0.144^{+0.007}_{-0.007}$	$1.480^{+0.21}_{-0.20}$	$1.520^{+0.058}_{-0.060}$	$0.720^{+0.14}_{-0.13}$	13.6/12
$-0.139^{+0.005}_{-0.005}$	$1.579^{+0.16}_{-0.15}$	$1.472^{+0.043}_{-0.044}$	$0.761^{+0.10}_{-0.10}$	30.2/20

Table 7.1: Value of the bias and β parameters derived by fitting the results of the simulations using the auto-power spectrum multipoles alone (upper row) and making a joint fit to all auto- and cross-power spectra of the 2 fields (bottom row).

we estimate the errors on the multipoles of the cross-power spectra assuming the modes follow a Gaussian distribution and accounting for the correlation between monopoles and quadrupoles among the different auto- and cross-power spectra (see Appendix A). The results are shown in the bottom row of Table 7.1 and the degeneracies among parameters are displayed in Figure 7.6.

With this method we obtain a normalized χ^2 of 30.2/20, showing that we may be underestimating some errors[10]. We obtain values of the parameters equal to

$$b_{HI} = 1.472^{+0.043}_{-0.044}, \qquad \beta_{HI} = 0.720^{+0.14}_{-0.13},$$
$$b_F = -0.139^{+0.005}_{-0.005} \quad \text{and} \quad \beta_F = 1.580^{+0.16}_{-0.15}.$$

The derived values for the HI are in perfect agreement with the input ones, and the combination $b_F(1+\beta_F)$ also reproduces the observational constraints [17]. We also notice that the values derived with this method are in agreement with the ones obtained by fitting only the auto-power spectra.

The best-fit values for the monopoles (left panel) and quadrupoles (right panel) of the auto- and cross-power spectra from the joint fitting are summarised in Figure 7.7: we plot the 3 simulation derived spectra $P_{21cm}(k)$ (green dotted), $P_{Ly\alpha}(k)$ (blue dashed) and $P_X(k)$ (magenta dash-dotted), whereas all the fits are plotted with red lines. The amplitude of errors shrinks at small scales merely because we are neglecting the instrumental noise. Even if included, results would not change as here we are looking at (almost) linear scales (see Section 7.3.1: system noise is sub-dominant at those scales) and indeed the fit is performed with the simulated data up to $k = 0.2\ h\text{Mpc}^{-1}$, limit highlighted by the black dotted vertical line in Figure 7.7.

An extra feature to be noticed in Figure 7.7 is that, although using only power spectra measurements with $k < 0.2\ h\text{Mpc}^{-1}$ for the inference of the bias parameters, the fits with such inferred parameters (red lines) are capable of reproducing the amplitude and shape of the cross-power spectrum multipoles to a remarkable accuracy down to the smallest scales we probe. This could be just a coincidence, or it

10 Probably the correlation between different multipoles of the two fields. We leave this check for future work.

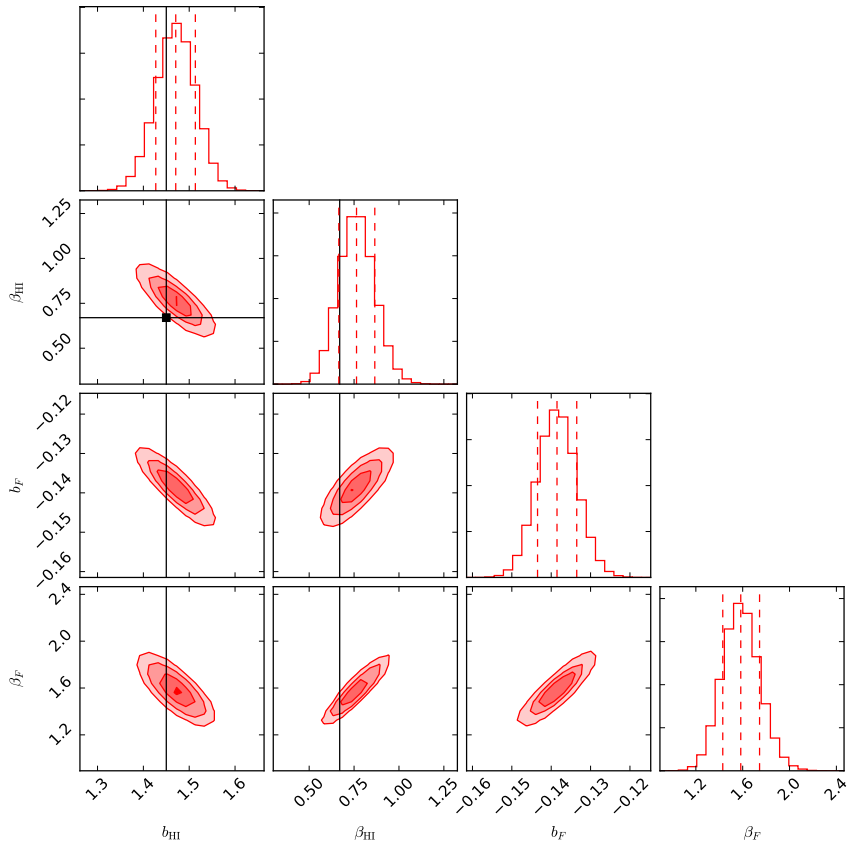


Figure 7.6: Constraints and degeneracies on the bias parameters of the 21cm and Lyα forest obtained by performing a joint fit to the auto- and cross-power spectra of the two fields. The straight lines indicate the expected value for the 21cm, that we know by construction.

may arises because we are not properly modelling HI inside halos, or it may hint to something physically more interesting. We leave this issue to future investigation.

We show in Figure 7.8 a more detailed comparison between the derived bias values of the 21cm from the 2 methods and the input ones. The dotted line in that plot represents the function $\beta_{\rm HI} = (f/b_{\rm HI})$, where $f(z) \simeq [\Omega_{\rm m}(z)]^{0.545}$ from linear theory; for the cosmological set-up employed in this simulation at redshift $z = 2.4$, $f = 0.97$. There is agreement among the 2 $(b_{\rm HI}, \beta_{\rm HI})$ points, and both are compatible with the Kaiser approximation (dotted line) within 1σ.

It is important to notice that by adding the information embedded in the cross-power spectra, together with assuming perfect anti-correlation on linear scales, the errors on the 21cm and Lyα forest parameters decrease by $\sim 30\%$. Therefore cross-power spectra have to be seen as a powerful way to look for systematics in the 21cm field

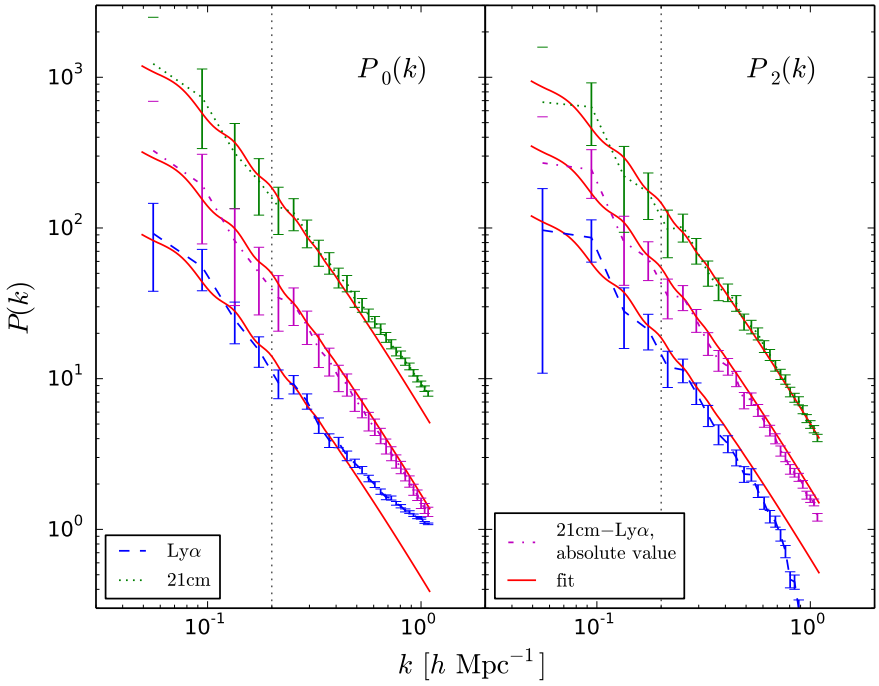


Figure 7.7: The red solid lines are the result of fitting the P(k) monopoles (left panel) and the quadrupoles (right), the green dotted lines refer to the 21cm, the blue dashed to Lyα forest flux and the magenta dashed-dotted to their cross-correlation in absolute value. The black dotted vertical line marks $k = 0.2h\,\mathrm{Mpc}^{-1}$, the mode up to which we perform the fit.

but also as a way to add extra information that can shrink the error on the model parameters.

We want to stress again that our task is

i) to retrieve the input b_{HI} value through our simulated data as a consistency check,

ii) to check the validity of linear theory at describing the modelled field,

iii) to show the error improvement by using the information coming from the cross-power spectrum.

We do not attempt to constraint the absolute b_{HI} since it is tuned in our model by construction.

Since the amplitude of the 21cm power is proportional to $\Omega_{HI}b_{HI}$, 21cm intensity mapping surveys are sensitive to that product: for example it has been found $\Omega_{HI}b_{HI} = 0.62 \times 10^{-3}$ at $z \sim 0.8$ [147]; for comparison, we remind that in our modelling we set $\Omega_{HI} = 10^{-3}$ as an input parameter and using a model with a different value of Ω_{HI} would only shift the amplitude of the 21cm power spectrum, but not its shape, i.e. the values of (b_{HI}, β_{HI}) would not be affected.

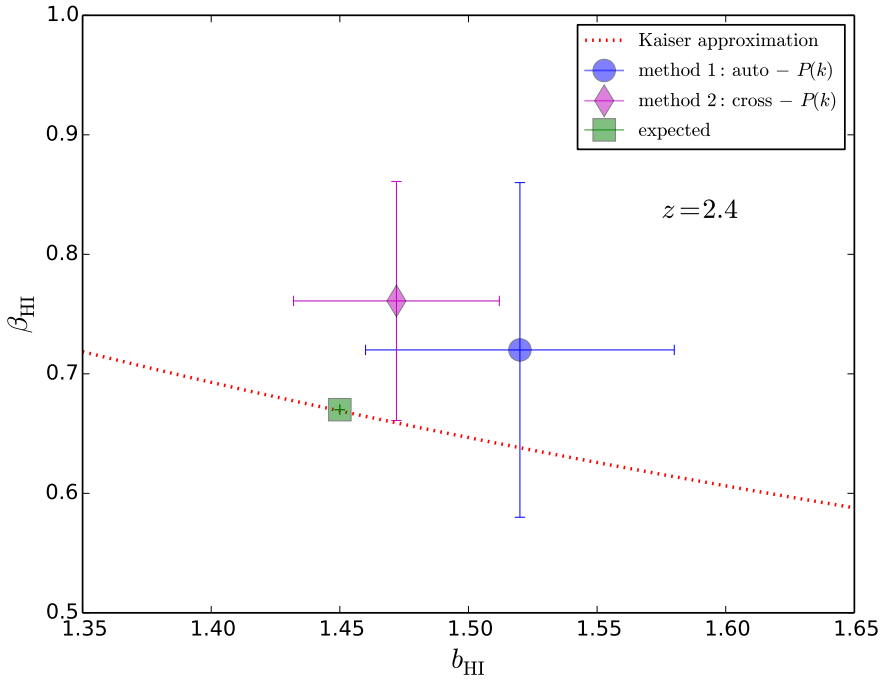


Figure 7.8: Values of the bias parameters b_{HI} and β_{HI} derived with the two procedures described in Section 7.3.2 and summarised in Table 7.1. The green square indicates the expected values, that we know by construction. The red dotted line is the Kaiser approximation: $\beta_{HI} \times b_{HI} = f(\Omega_m) \simeq \Omega_m^{0.55}(z)$.

The degeneracy $\Omega_{HI} - b_{HI}$ can be broken by adding information from other surveys, such as the HI column density distribution function from the Lyα forest (see Section 3.3), that is directly sensitive to Ω_{HI}.

The effect of shifting the 21cm power spectrum amplitude can be produced also by allowing the dark matter particle to have mass of the order of keV (thermal relic) instead of the perfectly cold DM approximation or to belong to another non-cold scenario that displays a similar cut-off in the matter power spectrum (as we show in Chapter 4 and Chapter 5). Hence, to extract information about Ω_{HI} by looking at the observed $P_{21cm}(k)$, one should also consider the nature of the dark matter particle and disentangle the two responses on the $P_{21cm}(k)$, e.g. as we illustrate in Section 4.3.

7.3.2.1 *Degeneracies between the bias b_{HI}, Ω_{HI} and the growth factor f: a Fisher matrix analysis*

Given the results shown in Section 7.3.2, we can infer that the Kaiser approximation $f = \beta_{HI} b_{HI}$ agrees with our 21cm-Lyα modelling (Figure 7.8) over a compelling range of k (Figure 7.7) and thus we can

b_F	β_F	b_{HI}	f	χ^2/dof
$-0.144^{+0.007}_{-0.007}$	$1.478^{+0.21}_{-0.20}$	$1.515^{+0.046}_{-0.047}$	$1.1^{+0.09}_{-0.09}$	24.4/12
$-0.141^{+0.005}_{-0.005}$	$1.508^{+0.18}_{-0.17}$	$1.485^{+0.043}_{-0.044}$	$1.1^{+0.08}_{-0.08}$	44.3/20

Table 7.2: Value of the bias and β parameters and of the cosmological growth factor f derived by fitting within the linear theory framework the results of the simulations using the auto-power spectrum multipoles alone (upper row) and making a joint fit to all auto- and cross-power spectra of the 2 fields (bottom row).

rewrite the linear order 21cm power spectrum of Equation 7.8 replacing $\beta_{HI} = f/b_{HI}$; its monopole and quadrupole become:

$$P_{21cm,0}(k) = A^2\Omega_{HI}^2\left(b_{HI}^2 + \frac{2}{3}fb_{HI} + \frac{1}{5}f^2\right)P_m(k), \quad (7.17)$$

$$P_{21cm,2}(k) = A^2\Omega_{HI}^2 f\left(\frac{4}{3}b_{HI} + \frac{4}{7}f\right)P_m(k). \quad (7.18)$$

Analogously, the cross-power spectrum multipoles of Equation 7.15-Equation 7.16 become:

$$P_0(k) = A\Omega_{HI}b_F\left(b_{HI} + \frac{1}{3}(f + b_{HI}\beta_F) + \frac{1}{5}f\beta_F\right)P_m(k), \quad (7.19)$$

$$P_2(k) = A\Omega_{HI}b_F\left(\frac{2}{3}(\beta_F b_{HI} + f) + \frac{4}{7}\beta_F f\right)P_m(k)\,. \quad (7.20)$$

We thus perform a new MCMC analysis as described in Section 7.3.2, for determining the parameters (b_F, β_F, b_{HI}, f) employing only power spectra measurements for $k < 0.2$ hMpc^{-1}, first fitting together the monopoles and quadrupoles of the auto-power spectra, and then adding also the cross-power spectrum. Cosmic variance errors (no system noise) are estimated again as in Appendix A. The results are shown in Table 7.2. These last χ^2 are larger than those of the previous analysis: having imposed the Kaiser approximation has worsened our fits.

There is agreement between the (b_F, β_F, b_{HI}, f) values determined with the two fits and adding information coming from the cross-power spectrum again shrinks the associated errors. Both fits prefer an unreasonable growth factor value greater than unity, $f \simeq 1.1 \pm 0.1$, although being in agreement with linear theory $f = f(\Omega_m) \simeq \Omega_m^{0.55}(z) = 0.97$ within 1σ. This can be seen already in Figure 7.8, where both points were above the dotted line corresponding to the relation $\beta_{HI} \times b_{HI} = f(\Omega_m)$. We remind that we do not add any prior on the physical value of f.

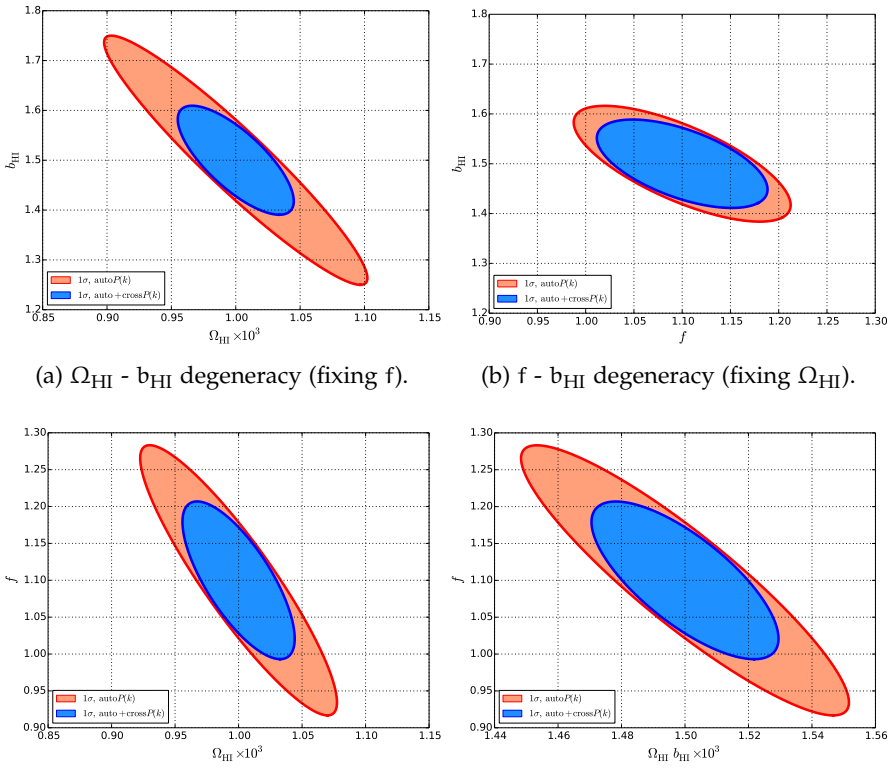
To better understand the degree of degeneracy of parameters involved in the 21cm characterization (Ω_{HI}, b_{HI} and f), we perform a

Fisher matrix analysis using monopoles and quadrupoles of the 21cm auto-power spectrum and the of cross-power spectrum with the Lyα flux. This is a good exercise especially to check the effect of the uncertainty on Ω_{HI}, that is tuned by construction in the previous analysis and in the simulated 21cm field.

As already quoted in Section 4.3, the Fisher matrix analysis quantifies the amount of information that the 21 cm power spectrum as observable carries about the three parameters Ω_{HI} - b_{HI} - f. Practically, we use as prior the values found with the MCMC fit as in Table 7.2 (i.e. $\Omega_{HI} = 10^{-3}$, $b_{HI} = 1.5$ and $f = 1.1$), and we calculate analytically how much the 21cm power spectrum varies by varying the values of those parameters. We make use of Equation 7.12-7.13-7.15-7.16 as templates and of the expressions shown in the Appendix A for building the covariance matrix, i.e. the Gaussian uncertainties linked to these parameters.

In Figure 7.9 we show the results. In the top left panel we show the uncertainty contours for Ω_{HI} - b_{HI} (fixing f), top right the uncertainty contours for f - b_{HI} (fixing Ω_{HI}), bottom left the uncertainty contours for Ω_{HI} - f (fixing b_{HI}) and in last panel on bottom left the uncertainty contours for the product Ω_{HI} b_{HI} - f. The orange ellipses refer to the uncertainty using the auto-power spectrum information only, the blue ellipses using information coming from both auto- and cross- power spectra: in the second case the uncertainties are always reduced, i.e. the cross-power spectrum shrinks our constraints in any case. Especially for the Ω_{HI} - b_{HI} correlation: fixing the growth factor helps reducing the degeneracy and adding the cross- information shrinks the errors by $\sim 50\%$.

Concerning the ability of a 21cm - Lyα joint analysis to constrain the cosmological growth factor f: we will need an independent measurement of Ω_{HI} in order to reach sufficient precision. Anyway, in the redshift range probed by 21cm intensity mapping surveys and Lyα flux experiments ($z \sim 2-3$) we have no other precise f measurement, thus whatever hint would be valuable.



(a) Ω_{HI} - b_{HI} degeneracy (fixing f).

(b) f - b_{HI} degeneracy (fixing Ω_{HI}).

(c) Ω_{HI} - f degeneracy (fixing b_{HI}).

(d) Ω_{HI} b_{HI} - f degeneracy.

Figure 7.9: 1σ contours of the values of the bias b_{HI}, the density parameter Ω_{HI} and the growth factor f determined using either only the 21cm power spectrum (orange areas) or adding the cross 21cm-Lyα power spectrum (blue areas). The Fisher matrix analysis is performed using the theoretical templates of the power spectra multipoles and errors described in Section 7.3.2.1, using $k < 0.2h\,\mathrm{Mpc}^{-1}$ information.

8

CONCLUSIONS AND FUTURE PROSPECTS

The spatial distribution of matter in the Universe embeds a huge amount of information, for example on the fraction that each component contributes to the total energy content of the Universe and on the nature of gravity. Unfortunately, matter is not directly observable, but can be mapped through tracers, as for example cosmic neutral hydrogen (HI).

In this thesis we have focused on HI and its 21cm emission, in particularly considering a relatively new technique called intensity mapping (IM). IM consists of carrying out low angular resolution radio-observations with the goal of measuring 21cm flux perturbations from large patches of the sky, without resolving the galaxies that host HI. This technique will play a major role in cosmology, given the spectroscopic nature of the observed signal and the large volumes possibly sampled. In this work we have assessed the 21cm IM potential for constraining cosmological models.

In this last chapter we summarise the main results of the work illustrated in this thesis and propose how to address questions that remained open. Lastly and more generally, I give an outlook on possible science cases I would like to pursue in the future.

8.1 MAIN RESULTS

HI distribution modelling

HI clouds are hosted in dark matter halos which are biased tracers of the matter density field, hence the use of numerical simulations is needed to provide an accurate description of the HI clustering. In Chapter 3 we address how to model the spatial distribution of HI a-posteriori on cosmological simulations, employing different methods: halo based (HI is assigned to dark matter halos) and particle based (HI is assigned to gas particles according to their physical properties). We test these models by computing the HI column density distribution function at redshifts $z = 3$, 4 and 5 and find fair agreement with observational data.

Dark matter

We study non-cold dark matter (DM) models that, suppressing gravitational clustering at small scales, can alleviate the small scale ten-

sions of cold DM. We consider simulations (hydrodynamic and N-body only) run with 3 families of models:

i) WDM: thermal warm DM, with intrinsic momenta derived from a Fermi-Dirac distribution, in Chapter 4;

ii) ULADM: ultra-light axion DM models, also known as *fuzzy* DM, in Chapter 5;

iii) LFDM: late-forming DM models, in Chapter 5.

All these models are characterised by a suppression of the amplitude of matter density fluctuations at small scales below a characteristic length that for WDM models depends on the thermal relic mass, for ULADM on the axion particle mass, while in the case of LFDM models depends on the phase transition redshift.

We investigate the spatial distribution of HI in these non-standard DM cosmologies and we study their signatures on the shape and amplitude of the 21cm power spectrum. We also forecast with which sensitivity the SKA radio telescope would be able to distinguish between the different DM models.

The most interesting conclusion is that, contrary to naive expectations, the suppression of power present in the linear and non-linear matter power spectra of non-cold DM models, results in an increase of power in terms of HI and 21cm power spectra. The reason lies in having a lack of small mass halos in non-cold DM scenarios with respect to CDM: in order to distribute a total amount of HI within the models, a larger quantity has to be placed in the most massive halos, that are more biased compared to the CDM cosmology.

Taking care of system noise only (i.e. assuming that astrophysical foregrounds, radio frequency interference and others nuisances have been already removed from the observed data) we find that

i) SKA1-LOW array would be able to distinguish the WDM models with 1, 2, and 3 keV from CDM with $t_0 = 1000$ hours of time observation;

ii) 4 keV WDM is consistent with CDM at 1σ confidence level at $z = 5$;

iii) SKA1-MID array would be able to distinguish all of the ULADM and LFDM models considered at redshifts $z = 1$ and 2 at high statistical significance.

There are other effects that could enhance the 21cm signal, as an increase in the HI cosmic abundance, Ω_{HI}. Anyway, a higher value of Ω_{HI} translates into a higher value of the amplitude of the 21cm power spectrum, while the presence of non-cold DM also modifies its shape. We exploit this fact in Chapter 4 where we study the Ω_{HI} - WDM degeneracy with a Fisher matrix analysis.

Dark energy

In Chapter 5 we study the imprint that different dark energy (DE) cosmologies leave on the 21cm power spectrum. We consider a set of large volume N-body simulations of 2 flat DE models:

i) RPCDM: a quintessence model with dynamical equation of state as given by the scalar field evolution in a Ratra-Peebles self-interacting potential,

ii) SUCDM: a quintessence model with supergravity self-interacting potential,

together with the standard cosmological model with a cosmological constant Λ. The cosmological parameters of both models have been calibrated in order to reproduce available cosmological observations, i.e. they are statistically indistinguishable from ΛCDM, yet we find that they leave distinct signatures on the 21cm spectra across a wide range of scales at $z = 1$ and 2.3, which correlates with differences in the halo mass function and the on-set of the non-linear regime of clustering, as these dynamical DE models experience a less accelerated expansion at late times. In particular, the RPCDM model can be distinguished from ΛCDM at high-statistical significance. Even the SUCDM model, characterised by a cosmic expansion and a linear growth rate similar to that of the ΛCDM, can be potentially distinguished by SKA1-MID at more that 1σ at $z = 2.3$ in the range of scales corresponding to $0.02 \lesssim k\,[\mathrm{Mpc}^{-1}\,h] \lesssim 2$.

Cross-correlation with the Lyman-α forest

A big drawback of 21cm intensity mapping is that the cosmological signal is expected to be buried by galactic and extragalactic foregrounds, whose amplitude is several order of magnitude larger.

Motivated by this, in Chapter 7 we study the cross-correlation between 21cm maps and the Lyα forest at redshift $z = 2.4$. We perform this study by means of state-of-the-art hydrodynamic simulations. We find that on large, linear scales, the Lyα forest is completely anti-correlated with the 21cm field. Indeed, the Lyα forest traces low density, highly ionized gas, while the 21cm signal arises from high density regions where hydrogen is neutral because of possible self-shielding.

We use the anisotropy of the power spectra in redshift-space to determine the values of the bias and of the redshift-space distortion parameters of both fields. We find that

i) the errors on the value of the cosmological and astrophysical parameters could decrease by 30% when adding data from the cross-power spectrum;

ii) linear theory is capable of reproducing the shape and amplitude of the cross-power up to rather non-linear scales;

iii) the 21cm-Lyα cross-power spectrum can be detected by combining data from a BOSS-like survey together with 21cm intensity mapping observations by SKA1-MID with a S/N ratio higher than 3 in $k \in [0.06, 1]\ h\,\mathrm{Mpc}^{-1}$.

8.2 FUTURE PROSPECTS

21cm intensity mapping

In this thesis we spatially model HI by running and analysing high resolution hydrodynamical simulations: these computationally expensive runs are needed in order to accurately distribute HI, because very little is known on its clustering properties on mildly non-linear scales. If we want to simulate HI in larger volumes, we need to rely on approximate methods as assigning the HI to the halos of a N-body simulation or of a mock halo catalogues generator.

In the local Universe the HI content of galaxies is well known, thanks to surveys as HIPASS[1] and ALFALFA[2]. Moreover, also on the theoretical side great effort has been made at developing semi-analytic models and HI-mass prescription for describing growth and evolution of galaxies (e.g. [55, 57, 108]). It would be timely to sharpen our HI modelling by including the knowledge we have on galactic scales (assuming galaxies have same properties along different redshifts), and thus develop an HI-HOD model (within the Halo Occupation Distribution framework), to assign HI among each dark matter halo[3].

Furthermore, having the HI-HOD basics, it could be possible to develop a HI halo model, providing the community with analytic tools to compute the 21cm power spectrum for a given cosmology that can then be used to perform intensity mapping forecasts.

In Chapter 7 we compute the cross-correlation signal between 21cm IM and the Lyman-α forest by means of hydrodynamical simulations, since, as extensively discussed, the drawback of 21cm IM is its intrinsic signal weakness compared to the large foregrounds. Similar studies could be performed with other observational probes. I think for example of weak lensing, that estimates as well the matter power spectrum through its shear across a large range of scales and, moreover, it has promising observational effort foreseen (e.g. the Euclid survey).

1 http://www.atnf.csiro.au/research/multibeam/release/
2 http://egg.astro.cornell.edu/index.php/
3 Data analysis as the recent work in [67] would be optimal for constructing such a model.

Nature of dark matter and the epoch of reionization

In Chapter 4-5 we illustrate how, when considering alternative non-cold DM models, the signal in 21cm coming from intermediate redshifts ($z = 1 - 5$) is boosted such to be detectably different. We get such boost in signal because these alternative DM models naturally suppress the abundance of low mass halos and, in order to distribute the same total amount of HI (Ω_{HI}) within all cosmologies, HI is forced to cluster in the most massive (and more biased) halos, increasing the 21cm signal power.

As the cold gas gets more clustered in the fewer halos available, a different star formation histories should be at play. Hence, what kind of impact different DM models have on cosmic reionization? Much of the reionization power comes from the faint galaxies that should reside in these missing small halos. It would be interesting to develop theoretical model predictions of the high redshift evolution of the optical depth in different DM scenarios using results from simulations. These predictions could be tested against measurements from Planck, being able to constrain a variety of DM models.

8.3 GENERAL OUTLOOK

Here I introduce some projects not directly related to the topic of this thesis, yet emerged from my Ph.D. work.

Impact of non-linearities on the cosmological parameter inference from the matter power spectrum measurements of Euclid

In the upcoming years a new generation of experiments will provide accurate measurements of the matter distribution in the Universe. E.g. surveys such as SKA, that we extensively talked about, and also the Euclid[4] satellite mission. Euclid will survey the distribution of galaxies in the universe over an unprecedented range of scales and redshifts.

These measurements are expected to probe the clustering of matter to few percent uncertainty level, thanks to the precise determination of the growth rate of structures using matter power spectrum measurements at different redshifts, as well as weak lensing estimates of the shear power spectrum. Euclid will also provide accurate cosmic distance measurements from the signal of baryon acoustic oscillations (BAO) in the galaxy clustering.

Key to successful unbiased analyses of these measurements is the capacity to account for the effects of non-linear dynamics of the growth of structures at small scales. Such effects are difficult to model due to

4 http://sci.esa.int/euclid

mode couplings that cause deviations from linear perturbation theory predictions. Even on the scale of the BAO (~ 120 Mpc h^{-1}) the clustering of the matter density field when probed at the 1% level is sensitive to non-linear effects, thus requiring expensive numerical simulations to perform sub-percent accurate theoretical predictions [135]. This strongly limits the ability to perform unbiased cosmological parameter likelihood data analysis through standard Markov Chain Monte Carlo methods using linear theory predictions.

To this purpose a variety of approximate models have been introduced in the literature to estimate the non-linear clustering of matter at small scales. Such models are usually calibrated against simulations and include analytical model prescriptions such as the popular HaloFit approach [142], or simulations emulators such as PINOCCHIO [99] and COLA [148].

To which extent Euclid measurements are sensitive to the different predictions of these approximate models? What kind of bias the use of these different methods induce in the Euclid cosmological parameter analyses of matter power spectrum data? Is there a scale where the predictions of such different methods converge to within Euclid expected statistical uncertainties? If so cutting data below such a scale will allow to reach the Euclid goal on the cosmological parameter estimation? These are questions I would like to address.

Mergers as building blocks of structure formation

Another significant outcome of ΛCDM is the universality of the density profile of halos, as we discussed Section 1.1.2 when illustrating their *cuspy* radial profile. In brief, halos look the same, regardless of how we configure the dark matter when we start any simulation [104].

Recent works claim that the first formed proto-halos are lead to display a universal density profile thanks to merger events, that are substantially more frequent at the beginning of structure formation [8, 109]. However, these studies are performed by using numerical simulations only and little effort has been done in trying to theoretically describe this phenomenon. In [36] (a work not included in this thesis), I demonstrated how the ejection of particles due to a minor merger (as result of the subsequent relaxation process) it is not due to two-body interactions, but it is a mean-field effect and as such can be illustrated without the need of running N-body simulation and, more important, can be interpreted by physical first principles (see Figure 8.1). This phenomenon should be further investigated, as I believe it could provide an analytical tool to understand the universality of dark matter halos.

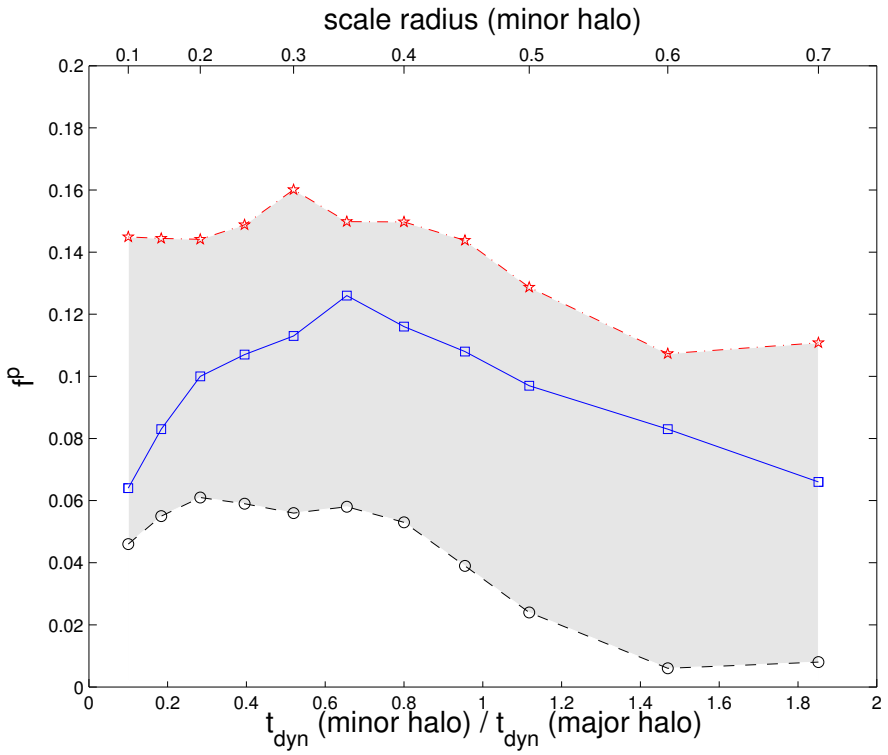


Figure 8.1: With a solid blue line is the fraction of ejected particles in N-body simulations as function of the dynamical properties of the two merging halos; the blue line is well enclosed in the grey area spanned by analytical predictions (the area between the two extreme approximation cases which are meant to bracket all realistic possibilities). Details in [36].

A

GAUSSIAN ERRORS DERIVATION

In this appendix we derive the equations governing the Gaussian errors of the multipoles of auto- and cross- power spectra.

FITTING SIMULTANEOUSLY MONOPOLES AND QUADRUPOLES OF THE AUTO-POWER SPECTRA

If we have measurements of the monopoles and quadrupoles of the auto-power spectra of the two fields that occupy the same volume and we want to fit them simultaneously we need a theoretical model and an estimation of the errors. The theoretical model is given by the Legendre expansion of Equation 7.11, that translates to Equation 7.12-7.13 for the 2 fields we are considering (dubbed with α and β subscripts in what follows). We derive the associated covariance:

$$
\begin{aligned}
\sigma^2 \quad (P_{\ell_1,\alpha} \quad & (k_i), P_{\ell_2,\beta}(k_j)) = \\
&= \quad \left\langle \left(\hat{P}_{\ell_1,\alpha}(k_i) - P_{\ell_1,\alpha}(k_i)\right)\left(\hat{P}_{\ell_2,\beta}(k_j) - P_{\ell_2,\beta}(k_j)\right)\right\rangle \\
&= \quad \langle \hat{P}_{\ell_1,\alpha}(k_i)\hat{P}_{\ell_2,\beta}(k_j)\rangle - P_{\ell_1,\alpha}(k_i)P_{\ell_2,\beta}(k_j) \\
&= \quad \frac{(2\ell_1+2)(2\ell_2+1)}{N_k^2} \times \\
&\quad \times \sum_{l=1}^{N_k}\sum_{m=1}^{N_k} \langle \delta_\alpha(k_{i,l})\delta^*_\alpha(k_{i,l})\delta_\beta(k_{j,m})\delta^*_\beta(k_{j,m})\rangle L_{\ell_1}(\mu_l)L_{\ell_2}(\mu_m) - \\
&\quad -P_{\ell_1,\alpha}(k_i)P_{\ell_2,\beta}(k_j) \\
&= \quad \frac{(2\ell_1+2)(2\ell_2+1)}{N_k^2} \times \\
&\quad \times \sum_{l=1}^{N_k}\sum_{m=1}^{N_k} \langle \delta_\alpha(k_{i,l})\delta^*_\beta(k_{j,m})\rangle\langle \delta^*_\alpha(k_{i,l})\delta_\beta(k_{j,m})\rangle L_{\ell_1}(\mu_l)L_{\ell_2}(\mu_m) \\
&= \quad \frac{(2\ell_1+2)(2\ell_2+1)}{N_k^2} \sum_{l=1}^{N_k}\sum_{m=1}^{N_k} P^2_{\alpha\beta}(k_{i,l})L_{\ell_1}(\mu_l)L_{\ell_2}(\mu_m)\delta_{l,m}\delta_{k_i,k_j} \\
&= \quad \frac{(2\ell_1+2)(2\ell_2+1)}{N_k^2} \sum_{l=1}^{N_k} P^2_{\alpha\beta}(k_{i,j})L_{\ell_1}(\mu_l)L_{\ell_2}(\mu_l)\delta_{k_i,k_j} \qquad \text{(A.1)}
\end{aligned}
$$

where N_k is the number of independent modes in the k-interval $[k, k+dk]$, i.e. $N_k = \frac{1}{2}\frac{4\pi k^2 dk}{k_F^3}$ where $k_F = 2\pi/L$ is the value of the fundamental frequency, with L being the size of the cubic volume and dk is the k-bin size, usually chosen as $dk = k_F$. We notice that

in the previous expression we have taken into account that the imaginary part of the cross-power spectrum is 0[1], i.e.

$$\Im(P_{12}(\mathbf{k})) = \frac{1}{2}\left(\langle\delta_1\delta_2^*\rangle - \langle\delta_1^*\delta_2\rangle\right) = 0 \tag{A.2}$$

thus, $\langle\delta_1(\mathbf{k}_1)\delta_2^*(\mathbf{k}_2)\rangle = P_{12}(\mathbf{k})$. In the continuous limit, the above equation can be expressed as

$$\begin{aligned}\sigma^2 \quad & (P_{\ell_1,\alpha}(k_1), P_{\ell_2,\beta}(k_2)) = \\ = \quad & \delta_{k_1,k_2}\frac{(2\ell_1+2)(2\ell_2+1)}{2N_k}\int_{-1}^{1} P^2_{\alpha\beta}(k,\mu)L_{\ell_1}(\mu)L_{\ell_2}(\mu)d\mu\end{aligned} \tag{A.3}$$

and taking into account that in redshift-space $P_{\alpha\beta}(k,\mu) = b_\alpha b_\beta(1+\beta_\alpha\mu^2)(1+\beta_\beta\mu^2)P_m(k)$ we obtain

$$\begin{aligned}
\sigma^2(P_{0,\alpha}(k_1), P_{0,\beta}(k_2)) \quad &= \gamma(k)\Big[\quad \frac{1}{9}\beta_\alpha^2\beta_\beta^2 + \frac{2}{7}(\beta_\alpha^2\beta_\beta + \beta_\alpha\beta_\beta^2) + \\
&\quad + \frac{1}{5}(\beta_\alpha^2 + 4\beta_\alpha\beta_\beta + \beta_\beta^2) + \\
&\quad \frac{2}{3}(\beta_\alpha + \beta_\beta) + 1\Big]\delta_{k_1,k_2} \\
\sigma^2(P_{2,\alpha}(k_1), P_{2,\beta}(k_2)) \quad &= 5\gamma(k)\Big[\quad \frac{415}{1287}\beta_\alpha^2\beta_\beta^2 + \frac{170}{231}(\beta_\alpha^2\beta_\beta + \beta_\alpha\beta_\beta^2) + \\
&\quad + \frac{3}{7}(\beta_\alpha^2 + 4\beta_\alpha\beta_\beta + \beta_\beta^2) + \\
&\quad \frac{22}{21}(\beta_\alpha + \beta_\alpha) + 1\Big]\delta_{k_1,k_2} \\
\sigma^2(P_{0,\alpha}(k_1), P_{2,\beta}(k_2)) \quad &= \tfrac{4}{693}\gamma(k)\Big[\quad 70\beta_\alpha^2\beta_\beta^2 + 165(\beta_\alpha^2\beta_\beta + \beta_\alpha\beta_\beta^2) + \\
&\quad + 99(\beta_\alpha^2 + 4\beta_\alpha\beta_\beta + \beta_\beta^2) + \\
&\quad 231(\beta_\alpha + \beta_\beta)\Big]\delta_{k_1,k_2}
\end{aligned} \tag{A.4}$$

where

$$\gamma(k) = \frac{b_\alpha^2 b_\beta^2 P_m^2(k)}{N_k} \tag{A.5}$$

FITTING SIMULTANEOUSLY AUTO- AND CROSS-POWER SPECTRA

If we have measurements of monopoles and quadrupoles of the two auto-power spectra and measurements of monopole and quadrupole of their cross-power spectrum and we want to fit all six functions together, we need again a theoretical model and an estimation of the errors. The theoretical model is given again by the expansion in Leg-

1 This arises by assuming that $\xi(\mathbf{r})$ is even, i.e. $\xi(-\mathbf{r}) = \xi(\mathbf{r})$.

endre polinomial as in Equation 7.12-7.13 and Equation 7.15-7.16. It follows that the covariance will be given by

$$\begin{aligned}\sigma^2 \quad (P_{\ell_1,\alpha} \quad (k_1), P_{\ell_2,12}(k_2)) &= \\ &= \quad \delta_{k_1,k_2}\frac{(2\ell_1+2)(2\ell_2+1)}{2N_k}\int_{-1}^{1} P_\alpha(k,\mu)P_{12}(k,\mu)L_{\ell_1}(\mu)L_{\ell_2}(\mu)d\mu\end{aligned} \tag{A.6}$$

and taken into account that $P_\alpha(k,\mu) = b_\alpha^2(1+\beta_\alpha\mu^2)^2P_m(k)$ and $P_{12}(k,\mu) = b_1b_2(1+\beta_1\mu^2)(1+\beta_2\mu^2)P_m(k)$ we get

$$\begin{aligned}\sigma^2 \quad (P_{0,1} \quad (k_1), P_{0,12}(k_2)) &= \gamma(k)\times \\ \times \quad &\left[\frac{1}{9}\beta_1^3\beta_2+\frac{1}{7}(\beta_1^3+3\beta_1^2\beta_2)+\frac{1}{5}(3\beta_1^2+3\beta_1\beta_2)+\frac{1}{3}(3\beta_1+\beta_2)+1\right]\delta_{k_1,k_2} \\ \sigma^2 \quad (P_{2,1} \quad (k_1), P_{2,12}(k_2)) &= 5\gamma(k)\times \\ \times \quad &\left[\frac{415}{1287}\beta_1^3\beta_2+\frac{85}{231}(\beta_1^3+3\beta_1^2\beta_2)+\frac{9}{7}(\beta_1^2+\beta_1\beta_2)+\frac{11}{21}(3\beta_1+\beta_2)+1\right]\delta_{k_1,k_2} \\ \sigma^2 \quad (P_{0,1} \quad (k_1), P_{2,12}(k_2)) &= \frac{2}{693}\gamma(k)\times \\ \times \quad &\left[140\beta_1^3\beta_2+165(\beta_1^3+3\beta_1^2\beta_2)+594(\beta_1^2+\beta_1\beta_2)+231(3\beta_1+\beta_2)\right]\delta_{k_1,k_2}\end{aligned} \tag{A.7}$$

where

$$\gamma(k) = \frac{b_1^3b_2P_m^2(k)}{N_k}\,. \tag{A.8}$$

BIBLIOGRAPHY

[1] R. Adhikari et al. "A White Paper on keV Sterile Neutrino Dark Matter." In: *ArXiv e-prints* (Feb. 2016). arXiv:1602.04816 [hep-ph] (cit. on p. 34).

[2] S. Agarwal and P.-S. Corasaniti. "Structural properties of artificial halos in nonstandard dark matter simulations." In: *Physical Review D* 91.12, 123509 (June 2015), p. 123509. DOI: 10.1103/PhysRevD.91.123509. arXiv:1503.03503 (cit. on pp. 61, 62).

[3] S. Agarwal, P.-S. Corasaniti, S. Das, and Y. Rasera. "Small scale clustering of late forming dark matter." In: *Physical Review D* 92.6, 063502 (Sept. 2015), p. 063502. DOI: 10.1103/PhysRevD.92.063502. arXiv:1412.1103 (cit. on p. 59).

[4] J.-M. Alimi, A. Füzfa, V. Boucher, Y. Rasera, J. Courtin, and P.-S. Corasaniti. "Imprints of dark energy on cosmic structure formation - I. Realistic quintessence models and the non-linear matter power spectrum." In: *Montly Notices of the Royal Astronomical Society* 401 (Jan. 2010), pp. 775–790. DOI: 10.1111/j.1365-2966.2009.15712.x. arXiv:0903.5490 [astro-ph.CO] (cit. on pp. 56–58, 67).

[5] D. Alonso, P. G. Ferreira, and M. G. Santos. "Fast simulations for intensity mapping experiments." In: *Montly Notices of the Royal Astronomical Society* 444 (Nov. 2014), pp. 3183–3197. DOI: 10.1093/mnras/stu1666. arXiv:1405.1751 (cit. on p. 10).

[6] D. Alonso, P. Bull, P. G. Ferreira, and M. G. Santos. "Blind foreground subtraction for intensity mapping experiments." In: *Montly Notices of the Royal Astronomical Society* 447 (Feb. 2015), pp. 400–416. DOI: 10.1093/mnras/stu2474. arXiv:1409.8667 (cit. on p. 11).

[7] R. E. Angulo, O. Hahn, and T. Abel. "The warm dark matter halo mass function below the cut-off scale." In: *Montly Notices of the Royal Astronomical Society* 434 (Oct. 2013), pp. 3337–3347. DOI: 10.1093/mnras/stt1246. arXiv:1304.2406 (cit. on p. 36).

[8] R. E. Angulo, O. Hahn, A. Ludlow, and S. Bonoli. "Earth-mass haloes and the emergence of NFW density profiles." In: *ArXiv e-prints* (Apr. 2016). arXiv:1604.03131 (cit. on p. 106).

[9] J. S. Bagla, N. Khandai, and K. K. Datta. "HI as a probe of the large-scale structure in the post-reionization universe." In: *Montly Notices of the Royal Astronomical Society* 407 (Sept. 2010), pp. 567–580. DOI: 10.1111/j.1365-2966.2010.16933.x. arXiv:0908.3796 [astro-ph.CO] (cit. on pp. 19–22).

[10] L. A. Barnes and M. G. Haehnelt. "Faint extended Lyα emission due to star formation at the centre of high column density QSO absorption systems." In: *Montly Notices of the Royal Astronomical Society* 403 (Apr. 2010), pp. 870–885. DOI: 10.1111/j.1365-2966.2009.16172.x. arXiv:0912.1345 [astro-ph.CO] (cit. on pp. 19, 20).

[11] R. A. Battye, R. D. Davies, and J. Weller. "Neutral hydrogen surveys for high-redshift galaxy clusters and protoclusters." In: *Montly Notices of the Royal Astronomical Society* 355 (Dec. 2004), pp. 1339–1347. DOI: 10.1111/j.1365-2966.2004.08416.x. eprint: astro-ph/0401340 (cit. on p. 8).

[12] G. D. Becker, P. C. Hewett, G. Worseck, and J. X. Prochaska. "A refined measurement of the mean transmitted flux in the Lyα forest over 2 z 5 using composite quasar spectra." In: *Montly Notices of the Royal Astronomical Society* 430 (Apr. 2013), pp. 2067–2081. DOI: 10.1093/mnras/stt031. arXiv:1208.2584 (cit. on pp. 24, 75–77).

[13] R. H. Becker et al. "Evidence for Reionization at z˜6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar." In: *The Astronomical Journal* 122 (Dec. 2001), pp. 2850–2857. DOI: 10.1086/324231. eprint: astro-ph/0108097 (cit. on p. 7).

[14] S. Bharadwaj, B. B. Nath, and S. K. Sethi. "Using HI to Probe Large Scale Structures at $z\tilde{3}$." In: *Journal of Astrophysics and Astronomy* 22 (Mar. 2001), p. 21. DOI: 10.1007/BF02933588. eprint: astro-ph/0003200 (cit. on p. 8).

[15] S. Bharadwaj and S. K. Sethi. "HI Fluctuations at Large Redshifts: I–Visibility correlation." In: *Journal of Astrophysics and Astronomy* 22 (Dec. 2001), pp. 293–307. DOI: 10.1007/BF02702273. eprint: astro-ph/0203269 (cit. on p. 8).

[16] L. Blitz and E. Rosolowsky. "The Role of Pressure in GMC Formation II: The H_2-Pressure Relation." In: *Astrophysical Journal* 650 (Oct. 2006), pp. 933–944. DOI: 10.1086/505417. eprint: astro-ph/0605035 (cit. on p. 24).

[17] M. Blomqvist et al. "Broadband distortion modeling in Lyman-α forest BAO fitting." In: *Journal of Cosmology and Astroparticle Physics* 11, 034 (Nov. 2015), p. 034. DOI: 10.1088/1475-7516/2015/11/034. arXiv:1504.06656 (cit. on pp. 93, 94).

[18] G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees. "Formation of galaxies and large-scale structure with cold dark matter." In: *Nature* 311 (Oct. 1984), pp. 517–525. DOI: 10.1038/311517a0 (cit. on pp. 4, 17).

[19] P. Bode, J. P. Ostriker, and N. Turok. "Halo Formation in Warm Dark Matter Models." In: *Astrophysical Journal* 556 (July 2001), pp. 93–107. DOI: 10.1086/321541. eprint: astro-ph/0010389 (cit. on pp. 33, 36).

[20] J. S. Bolton, E. Puchwein, D. Sijacki, M. G. Haehnelt, T.-S. Kim, A. Meiksin, J. A. Regan, and M. Viel. "The Sherwood simulation suite: overview and data comparisons with the Lyman α forest at redshifts $2 \leqslant z \leqslant 5$." In: *Montly Notices of the Royal Astronomical Society* 464 (Jan. 2017), pp. 897–914. DOI: 10.1093/mnras/stw2397. arXiv:1605.03462 (cit. on pp. 71, 72, 76, 82).

[21] A. Boyarsky, O. Ruchayskiy, and D. Iakubovskyi. "A lower bound on the mass of dark matter particles." In: *Journal of Cosmology and Astroparticle Physics* 3, 005 (Mar. 2009), p. 005. DOI: 10.1088/1475-7516/2009/03/005. arXiv:0808.3902 [hep-ph] (cit. on p. 33).

[22] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat. "Too big to fail? The puzzling darkness of massive Milky Way subhaloes." In: *Montly Notices of the Royal Astronomical Society* 415 (July 2011), pp. L40–L44. DOI: 10.1111/j.1745-3933.2011.01074.x. arXiv:1103.0007 [astro-ph.CO] (cit. on p. 4).

[23] R. Braun. "Cosmological Evolution of Atomic Gas and Implications for 21 cm H I Absorption." In: *Astrophysical Journal* 749, 87 (Apr. 2012), p. 87. DOI: 10.1088/0004-637X/749/1/87. arXiv:1202.1840 (cit. on p. 27).

[24] R. Braun, T. Bourke, J. A. Green, E. Keane, and J. Wagg. "Advancing Astrophysics with the Square Kilometre Array." In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 174 (Apr. 2015), p. 174 (cit. on p. 88).

[25] P. Brax and J. Martin. "Robustness of quintessence." In: *Physical Review D* 61.10, 103502 (May 2000), p. 103502. DOI: 10.1103/PhysRevD.61.103502. eprint: astro-ph/9912046 (cit. on p. 56).

[26] A. Brooks. "Re-examining astrophysical constraints on the dark matter model." In: *Annalen der Physik* 526 (Aug. 2014), pp. 294–308. DOI: 10.1002/andp.201400068. arXiv:1407.7544 (cit. on pp. 4, 33).

[27] P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos. "Late-time Cosmology with 21 cm Intensity Mapping Experiments." In: *Astrophysical Journal* 803, 21 (Apr. 2015), p. 21. DOI: 10.1088/0004-637X/803/1/21. arXiv:1405.1452 (cit. on pp. 8–11, 31).

[28] J. S. Bullock and M. Boylan-Kolchin. "Small-Scale Challenges to the ΛCDM Paradigm." In: *ArXiv e-prints* (July 2017). arXiv:1707.04256 (cit. on pp. 4, 5).

[29] A. Burkert. "The Structure of Dark Matter Halos in Dwarf Galaxies." In: *Astrophysical Journal Letters* 447 (July 1995), pp. L25–L28. DOI: 10.1086/309560. eprint: astro-ph/9504041 (cit. on p. 5).

[30] S. Camera, M. G. Santos, P. G. Ferreira, and L. Ferramacho. "Cosmology on Ultralarge Scales with Intensity Mapping of the Neutral Hydrogen 21 cm Emission: Limits on Primordial Non-Gaussianity." In: *Physical Review Letters* 111.17, 171302 (Oct. 2013), p. 171302. DOI: 10.1103/PhysRevLett.111.171302. arXiv:1305.6928 [astro-ph.CO] (cit. on p. 10).

[31] V. Cardoso, I. P. Carucci, P. Pani, and T. P. Sotiriou. "Black Holes with Surrounding Matter in Scalar-Tensor Theories." In: *Physical Review Letters* 111.11, 111101 (Sept. 2013), p. 111101. DOI: 10.1103/PhysRevLett.111.111101. arXiv:1308.6587 [gr-qc] (cit. on pp. vii, 3).

[32] V. Cardoso, I. P. Carucci, P. Pani, and T. P. Sotiriou. "Matter around Kerr black holes in scalar-tensor theories: Scalarization and superradiant instability." In: *Physical Review D* 88.4, 044056 (Aug. 2013), p. 044056. DOI: 10.1103/PhysRevD.88.044056. arXiv:1305.6936 [gr-qc] (cit. on pp. vii, 3).

[33] S. M. Carroll. "The Cosmological Constant." In: *Living Reviews in Relativity* 4, 1 (Feb. 2001), p. 1. DOI: 10.12942/lrr-2001-1. eprint: astro-ph/0004075 (cit. on p. 2).

[34] I. P. Carucci, P.-S. Corasaniti, and M. Viel. "Imprints of non-standard Dark Energy and Dark Matter Models on the 21cm Intensity Map Power Spectrum." In: *ArXiv e-prints* (June 2017). arXiv:1706.09462 (cit. on p. vii).

[35] I. P. Carucci, F. Villaescusa-Navarro, and M. Viel. "The cross-correlation between 21 cm intensity mapping maps and the Lyα forest in the post-reionization era." In: *Journal of Cosmology and Astroparticle Physics* 4, 001 (Apr. 2017), p. 001. DOI: 10.1088/1475-7516/2017/04/001. arXiv:1611.07527 (cit. on p. vii).

[36] I. P. Carucci, M. Sparre, S. H. Hansen, and M. Joyce. "Particle ejection during mergers of dark matter halos." In: *Journal of Cosmology and Astroparticle Physics* 6, 057 (June 2014), p. 057. DOI: 10.1088/1475-7516/2014/06/057. arXiv:1405.6725 (cit. on pp. vii, 106, 107).

[37] I. P. Carucci, F. Villaescusa-Navarro, M. Viel, and A. Lapi. "Warm dark matter signatures on the 21cm power spectrum: intensity mapping forecasts for SKA." In: *Journal of Cosmology and Astroparticle Physics* 7, 047 (July 2015), p. 047. DOI: 10.1088/1475-7516/2015/07/047. arXiv:1502.06961 (cit. on pp. vii, 20, 84).

[38] E. Castorina and F. Villaescusa-Navarro. "On the spatial distribution of neutral hydrogen in the Universe: bias and shot-noise of the HI Power Spectrum." In: *ArXiv e-prints* (Sept. 2016). arXiv:1609.05157 (cit. on pp. 19, 20, 22, 31, 64).

[39] T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald. "Baryon Acoustic Oscillation Intensity Mapping of Dark Energy." In: *Physical Review Letters* 100.9, 091303 (Mar. 2008), p. 091303. DOI: 10.1103/PhysRevLett.100.091303. arXiv:0709.3672 (cit. on p. 8).

[40] M. Chevallier and D. Polarski. "Accelerating Universes with Scaling Dark Matter." In: *International Journal of Modern Physics D* 10 (2001), pp. 213–223. DOI: 10.1142/S0218271801000822. eprint: gr-qc/0009008 (cit. on p. 57).

[41] B. Ciardi and A. Ferrara. "The First Cosmic Structures and Their Effects." In: *Space Science Reviews* 116 (Feb. 2005), pp. 625–705. DOI: 10.1007/s11214-005-3592-0. eprint: astro-ph/0409018 (cit. on p. 7).

[42] A. M. Cieplak and A. Slosar. "Towards physics responsible for large-scale Lyman-α forest bias parameters." In: *Journal of Cosmology and Astroparticle Physics* 3, 016 (Mar. 2016), p. 016. DOI: 10.1088/1475-7516/2016/03/016. arXiv:1509.07875 (cit. on pp. 77, 78).

[43] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis. "Modified gravity and cosmology." In: *Physics Reports* 513 (Mar. 2012), pp. 1–189. DOI: 10.1016/j.physrep.2012.01.001. arXiv:1106.2476 (cit. on p. 3).

[44] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky. "A Direct Empirical Proof of the Existence of Dark Matter." In: *Astrophysical Journal Letters* 648 (Sept. 2006), pp. L109–L113. DOI: 10.1086/508162. eprint: astro-ph/0608407 (cit. on p. 3).

[45] D. Coe. "Fisher Matrices and Confidence Ellipses: A Quick-Start Guide and Software." In: *ArXiv e-prints* (June 2009). arXiv:0906.4123 [astro-ph.IM] (cit. on p. 51).

[46] A. Cooray and R. Sheth. "Halo models of large scale structure." In: *Physics Reports* 372 (Dec. 2002), pp. 1–129. DOI: 10.1016/S0370-1573(02)00276-4. eprint: astro-ph/0206508 (cit. on p. 18).

[47] P. S. Corasaniti, S. Agarwal, D. J. E. Marsh, and S. Das. "Constraints on dark matter scenarios from measurements of the galaxy luminosity function at high redshifts." In: *Physical Review D* 95.8, 083512 (Apr. 2017), p. 083512. DOI: 10.1103/PhysRevD.95.083512. arXiv:1611.05892 (cit. on pp. 33, 59–61, 70).

[48] N. H. M. Crighton et al. "The neutral hydrogen cosmological mass density at z = 5." In: *Montly Notices of the Royal Astronomical Society* 452 (Sept. 2015), pp. 217–234. DOI: 10.1093/mnras/stv1182. arXiv:1506.02037 (cit. on p. 27).

[49] R. A. C. Croft, D. H. Weinberg, N. Katz, and L. Hernquist. "Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyα Forest." In: *Astrophysical Journal* 495 (Mar. 1998), pp. 44–62. DOI: 10.1086/305289. eprint: astro-ph/9708018 (cit. on p. 77).

[50] S. Das and N. Weiner. "Late forming dark matter in theories of neutrino dark energy." In: *Physical Review D* 84.12, 123511 (Dec. 2011), p. 123511. DOI: 10.1103/PhysRevD.84.123511. eprint: astro-ph/0611353 (cit. on p. 59).

[51] R. Davé, N. Katz, B. D. Oppenheimer, J. A. Kollmeier, and D. H. Weinberg. "The neutral hydrogen content of galaxies in cosmological hydrodynamic simulations." In: *Montly Notices of the Royal Astronomical Society* 434 (Sept. 2013), pp. 2645–2663. DOI: 10.1093/mnras/stt1274. arXiv:1302.3631 [astro-ph.CO] (cit. on pp. 19, 20, 23, 24).

[52] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. "The evolution of large-scale structure in a universe dominated by cold dark matter." In: *Astrophysical Journal* 292 (May 1985), pp. 371–394. DOI: 10.1086/163168 (cit. on pp. 4, 36, 58, 61, 83).

[53] K. S. Dawson et al. "The Baryon Oscillation Spectroscopic Survey of SDSS-III." In: *The Astronomical Journal* 145, 10 (Jan. 2013), p. 10. DOI: 10.1088/0004-6256/145/1/10. arXiv:1208.0022 (cit. on p. 88).

[54] P. Dayal, A. Mesinger, and F. Pacucci. "Early Galaxy Formation in Warm Dark Matter Cosmologies." In: *Astrophysical Journal* 806, 67 (June 2015), p. 67. DOI: 10.1088/0004-637X/806/1/67. arXiv:1408.1102 (cit. on pp. 33, 47, 50).

[55] G. De Lucia and J. Blaizot. "The hierarchical formation of the brightest cluster galaxies." In: *Montly Notices of the Royal Astronomical Society* 375 (Feb. 2007), pp. 2–14. DOI: 10.1111/j.1365-2966.2006.11287.x. eprint: astro-ph/0606519 (cit. on p. 104).

[56] T. Delubac et al. "Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars." In: *Astronomy and Astrophysics* 574, A59 (Feb. 2015), A59. DOI: 10.1051/0004-6361/201423969. arXiv:1404.1801 (cit. on pp. 13, 78).

[57] C. Evoli, P. Salucci, A. Lapi, and L. Danese. "The H I Content of Local Late-type Galaxies." In: *Astrophysical Journal* 743, 45 (Dec. 2011), p. 45. DOI: 10.1088/0004-637X/743/1/45. arXiv:1110.4199 [astro-ph.CO] (cit. on p. 104).

[58] H. I. Ewen and E. M. Purcell. "Observation of a Line in the Galactic Radio Spectrum: Radiation from Galactic Hydrogen at 1,420 Mc./sec." In: *Nature* 168 (Sept. 1951), p. 356. DOI: 10.1038/168356a0 (cit. on p. 8).

[59] C.-A. Faucher-Giguère, J. X. Prochaska, A. Lidz, L. Hernquist, and M. Zaldarriaga. "A Direct Precision Measurement of the Intergalactic Lyα Opacity at $2 < z < 4.2$." In: *Astrophysical Journal* 681, 831-855 (July 2008), pp. 831–855. DOI: 10.1086/588648. arXiv:0709.2382 (cit. on p. 75).

[60] J. L. Feng. "Dark Matter Candidates from Particle Physics and Methods of Detection." In: *Annual Review of Astronomy & Astrophysics* 48 (Sept. 2010), pp. 495–545. DOI: 10.1146/annurev-astro-082708-101659. arXiv:1003.0904 [astro-ph.CO] (cit. on p. 4).

[61] R. A. Flores and J. R. Primack. "Observational and theoretical constraints on singular dark matter halos." In: *Astrophysical Journal Letters* 427 (May 1994), pp. L1–L4. DOI: 10.1086/187350. eprint: astro-ph/9402004 (cit. on p. 4).

[62] A. Font-Ribera et al. "The large-scale cross-correlation of Damped Lyman alpha systems with the Lyman alpha forest: first measurements from BOSS." In: *Journal of Cosmology and Astroparticle Physics* 11, 059 (Nov. 2012), p. 059. DOI: 10.1088/1475-7516/2012/11/059. arXiv:1209.4596 (cit. on p. 19).

[63] A. Font-Ribera et al. "Quasar-Lyman α forest cross-correlation from BOSS DR11: Baryon Acoustic Oscillations." In: *Journal of Cosmology and Astroparticle Physics* 5, 027 (May 2014), p. 027. DOI: 10.1088/1475-7516/2014/05/027. arXiv:1311.1767 (cit. on pp. 19, 90).

[64] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. "emcee: The MCMC Hammer." In: *Publications of the Astronomical Society of the Pacific* 125 (Mar. 2013), pp. 306–312. DOI: 10.1086/670067. arXiv:1202.3665 [astro-ph.IM] (cit. on p. 93).

[65] S. R. Furlanetto, S. P. Oh, and F. H. Briggs. "Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe." In: *Physics Reports* 433 (Oct. 2006), pp. 181–301. DOI: 10.1016/j.physrep.2006.08.002. eprint: astro-ph/0608032 (cit. on pp. 11, 28).

[66] M. Götz and J. Sommer-Larsen. "Galaxy formation: Warm dark matter, missing satellites, and the angular momentum problem." In: *Astrophysics and Space Science* 284 (Apr. 2003), pp. 341–344. DOI: 10.1023/A:1024073909753. eprint: astro-ph/0210599 (cit. on pp. 38, 61).

[67] H. Guo, C. Li, Z. Zheng, H. J. Mo, Y. P. Jing, Y. Zu, S. H. Lim, and H. Xu. "Constraining the HI-Halo Mass Relation from Galaxy Clustering." In: *Astrophysical Journal* 846, 61 (Sept. 2017), p. 61. DOI: 10.3847/1538-4357/aa85e7. arXiv:1707.01999 (cit. on pp. 20, 104).

[68] F. Haardt and P. Madau. "Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background." In: *Astrophysical Journal* 746, 125 (Feb. 2012), p. 125. DOI: 10.1088/0004-637X/746/2/125. arXiv:1105.2039 (cit. on p. 72).

[69] A. Hall, C. Bonvin, and A. Challinor. "Testing general relativity with 21-cm intensity mapping." In: *Physical Review D* 87.6, 064026 (Mar. 2013), p. 064026. DOI: 10.1103/PhysRevD.87.064026. arXiv:1212.0728 [astro-ph.CO] (cit. on p. 11).

[70] S. Horiuchi, P. J. Humphrey, J. Oñorbe, K. N. Abazajian, M. Kaplinghat, and S. Garrison-Kimmel. "Sterile neutrino dark matter bounds from galaxies of the Local Group." In: *Physical Review D* 89.2, 025017 (Jan. 2014), p. 025017. DOI: 10.1103/PhysRevD.89.025017. arXiv:1311.0282 (cit. on p. 33).

[71] V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton, and G. D. Becker. "First constraints on fuzzy dark matter from Lyman-α forest data and hydrodynamical simulations." In: *ArXiv e-prints* (Mar. 2017). arXiv:1703.04683 (cit. on p. 59).

[72] V. Iršič et al. "New Constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman-α forest data." In: *ArXiv e-prints* (Feb. 2017). arXiv:1702.01764 (cit. on pp. 33, 60).

[73] N. Kanekar, R. Morganti, F. H. Briggs, J. Lazio, J. N. Chengalur, and J. X. Prochaska. "Probing Galaxy Evolution with HI 21cm Absorption Spectroscopy." In: *astro2010: The Astronomy and Astrophysics Decadal Survey*. Vol. 2010. ArXiv Astrophysics e-prints. 2009 (cit. on p. 8).

[74] H.-S. Kim, J. S. B. Wyithe, C. M. Baugh, C. d. P. Lagos, C. Power, and J. Park. "The spatial distribution of neutral hydrogen as traced by low H I mass galaxies." In: *Montly Notices of the Royal Astronomical Society* 465 (Feb. 2017), pp. 111–122. DOI: 10.1093/mnras/stw2779. arXiv:1603.02383 (cit. on pp. 19–21).

[75] T.-S. Kim, J. S. Bolton, M. Viel, M. G. Haehnelt, and R. F. Carswell. "An improved measurement of the flux distribution of the Lyα forest in QSO absorption spectra: the effect of continuum fitting, metal contamination and noise properties." In: *Montly Notices of the Royal Astronomical Society* 382 (Dec. 2007), pp. 1657–1674. DOI: 10.1111/j.1365-2966.2007.12406.x. arXiv:0711.1862 (cit. on pp. 75, 82).

[76] A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada. "Where Are the Missing Galactic Satellites?" In: *Astrophysical Journal* 522 (Sept. 1999), pp. 82–92. DOI: 10.1086/307643. eprint: astro-ph/9901240 (cit. on p. 4).

[77] E. Komatsu et al. "Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation." In: *Astrophysical Journal Supplement* 180 (Feb. 2009), pp. 330–376. DOI: 10.1088/0067-0049/180/2/330. arXiv:0803.0547 (cit. on p. 57).

[78] M. Kowalski et al. "Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets." In: *Astrophysical Journal* 686, 749-778 (Oct. 2008), pp. 749–778. DOI: 10.1086/589937. arXiv:0804.4142 (cit. on p. 57).

[79] M. R. Krumholz and N. Y. Gnedin. "A Comparison of Methods for Determining the Molecular Content of Model Galaxies." In: *Astrophysical Journal* 729, 36 (Mar. 2011), p. 36. DOI: 10.1088/0004-637X/729/1/36. arXiv:1011.4065 (cit. on p. 24).

[80] A. Lapi and L. Danese. "Cold or warm? Constraining dark matter with primeval galaxies and cosmic reionization after Planck." In: *Journal of Cosmology and Astroparticle Physics* 9, 003 (Sept. 2015), p. 003. DOI: 10.1088/1475-7516/2015/09/003. arXiv:1508.02147 (cit. on p. 33).

[81] A. Lewis, A. Challinor, and A. Lasenby. "Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models." In: *Astrophysical Journal* 538 (Aug. 2000), pp. 473–476. DOI: 10.1086/309179. eprint: arXiv:astro-ph/9911177 (cit. on pp. 17, 35).

[82] E. V. Linder. "Exploring the Expansion History of the Universe." In: *Physical Review Letters* 90.9, 091301 (Mar. 2003), p. 091301. DOI: 10.1103/PhysRevLett.90.091301. eprint: astro-ph/0208512 (cit. on p. 57).

[83] A.S. Lipatov. *The Hybrid Multiscale Simulation Technology: An Introduction with Application to Astrophysical and Laboratory Plasmas*. Scientific Computation. Springer Berlin Heidelberg, 2013. ISBN: 9783662050125. URL: https://books.google.it/books?id=w6fvCAAAQBAJ (cit. on p. 15).

[84] A. Loeb and J. S. B. Wyithe. "Possibility of Precise Measurement of the Cosmological Power Spectrum with a Dedicated Survey of 21cm Emission after Reionization." In: *Physical Review Letters* 100.16, 161301 (Apr. 2008), p. 161301. DOI: 10.1103/PhysRevLett.100.161301. arXiv:0801.1677 (cit. on p. 8).

[85] M. R. Lovell, C. S. Frenk, V. R. Eke, A. Jenkins, L. Gao, and T. Theuns. "The properties of warm dark matter haloes." In: *Montly Notices of the Royal Astronomical Society* 439 (Mar. 2014), pp. 300–317. DOI: 10.1093/mnras/stt2431. arXiv:1308.1399 (cit. on p. 61).

[86] Z. Lukić, C. W. Stark, P. Nugent, M. White, A. A. Meiksin, and A. Almgren. "The Lyman α forest in optically thin hydrodynamical simulations." In: *Montly Notices of the Royal Astronomical Society* 446 (Feb. 2015), pp. 3697–3724. DOI: 10.1093/mnras/stu2377. arXiv:1406.6361 (cit. on p. 76).

[87] A. V. Macciò and F. Fontanot. "How cold is dark matter? Constraints from Milky Way satellites." In: *Montly Notices of the Royal Astronomical Society* 404 (May 2010), pp. L16–L20. DOI: 10.1111/j.1745-3933.2010.00825.x. arXiv:0910.2460 (cit. on p. 33).

[88] P. Madau, A. Meiksin, and M. J. Rees. "21 Centimeter Tomography of the Intergalactic Medium at High Redshift." In: *Astrophysical Journal* 475 (Feb. 1997), pp. 429–444. DOI: 10.1086/303549. eprint: astro-ph/9608010 (cit. on p. 11).

[89] U. Maio and M. Viel. "The first billion years of a warm dark matter universe." In: *Montly Notices of the Royal Astronomical Society* 446 (Jan. 2015), pp. 2760–2775. DOI: 10.1093/mnras/stu2304. arXiv:1409.6718 (cit. on pp. 33, 47).

[90] D. J. E. Marsh. "Axion cosmology." In: *Physics Reports* 643 (July 2016), pp. 1–79. DOI: 10.1016/j.physrep.2016.06.005. arXiv:1510.07633 (cit. on p. 59).

[91] A. M. Martin, R. Giovanelli, M. P. Haynes, and L. Guzzo. "The Clustering Characteristics of H I-selected Galaxies from the 40% ALFALFA Survey." In: *Astrophysical Journal* 750, 38 (May 2012), p. 38. DOI: 10.1088/0004-637X/750/1/38. arXiv:1202.6005 [astro-ph.CO] (cit. on p. 19).

[92] R. Massey et al. "Dark matter maps reveal cosmic scaffolding." In: *Nature* 445 (Jan. 2007), pp. 286–290. DOI: 10.1038/nature05497. eprint: astro-ph/0701594 (cit. on p. 3).

[93] P. McDonald, J. Miralda-Escudé, M. Rauch, W. L. W. Sargent, T. A. Barlow, and R. Cen. "A Measurement of the Temperature-Density Relation in the Intergalactic Medium Using a New Lyα Absorption-Line Fitting Method." In: *Astrophysical Journal* 562 (Nov. 2001), pp. 52–75. DOI: 10.1086/323426. eprint: astro-ph/0005553 (cit. on p. 82).

[94] M. McQuinn and M. White. "On estimating Lyα forest correlations between multiple sightlines." In: *Montly Notices of the Royal Astronomical Society* 415 (Aug. 2011), pp. 2257–2269. DOI:

10.1111/j.1365-2966.2011.18855.x. arXiv:1102.1752 (cit. on p. 90).

[95] M. McQuinn, O. Zahn, M. Zaldarriaga, L. Hernquist, and S. R. Furlanetto. "Cosmological Parameter Estimation Using 21 cm Radiation from the Epoch of Reionization." In: *Astrophysical Journal* 653 (Dec. 2006), pp. 815–834. DOI: 10.1086/505167. eprint: astro-ph/0512263 (cit. on p. 8).

[96] A. A. Meiksin. "The physics of the intergalactic medium." In: *Reviews of Modern Physics* 81 (Oct. 2009), pp. 1405–1469. DOI: 10.1103/RevModPhys.81.1405. arXiv:0711.3358 (cit. on p. 71).

[97] N. Menci, A. Merle, M. Totzauer, A. Schneider, A. Grazian, M. Castellano, and N. G. Sanchez. "Fundamental Physics with the Hubble Frontier Fields: Constraining Dark Matter Models with the Abundance of Extremely Faint and Distant Galaxies." In: *Astrophysical Journal* 836, 61 (Feb. 2017), p. 61. DOI: 10.3847/1538-4357/836/1/61. arXiv:1701.01339 (cit. on p. 33).

[98] M. Miranda and A. V. Macciò. "Constraining warm dark matter using QSO gravitational lensing." In: *Montly Notices of the Royal Astronomical Society* 382 (Dec. 2007), pp. 1225–1232. DOI: 10.1111/j.1365-2966.2007.12440.x. arXiv:0706.0896 (cit. on p. 33).

[99] P. Monaco, E. Sefusatti, S. Borgani, M. Crocce, P. Fosalba, R. K. Sheth, and T. Theuns. "An accurate tool for the fast generation of dark matter halo catalogues." In: *Montly Notices of the Royal Astronomical Society* 433 (Aug. 2013), pp. 2389–2402. DOI: 10.1093/mnras/stt907. arXiv:1305.1505 (cit. on p. 106).

[100] B. Moore. "Evidence against dissipation-less dark matter from observations of galaxy haloes." In: *Nature* 370 (Aug. 1994), pp. 629–631. DOI: 10.1038/370629a0 (cit. on p. 4).

[101] B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi. "Dark Matter Substructure within Galactic Halos." In: *Astrophysical Journal Letters* 524 (Oct. 1999), pp. L19–L22. DOI: 10.1086/312287. eprint: astro-ph/9907411 (cit. on p. 4).

[102] S. Mukae, M. Ouchi, K. Kakiichi, N. Suzuki, Y. Ono, Z. Cai, A. K. Inoue, Y.-K. Chiang, T. Shibuya, and Y. Matsuda. "Cosmic Galaxy-IGM HI Relation at $z \sim 2-3$ Probed in the COSMOS/UltraVISTA 1.6 Deg2 Field." In: *Astrophysical Journal* 835, 281 (Feb. 2017), p. 281. DOI: 10.3847/1538-4357/835/2/281. arXiv:1605.00379 (cit. on p. 86).

[103] R. Murgia, A. Merle, M. Viel, M. Totzauer, and A. Schneider. ""Non-cold" dark matter at small scales: a general approach." In: *ArXiv e-prints* (Apr. 2017). arXiv:1704.07838 (cit. on p. 34).

[104] J. F. Navarro, C. S. Frenk, and S. D. M. White. "A Universal Density Profile from Hierarchical Clustering." In: *Astrophysical Journal* 490 (Dec. 1997), pp. 493–508. DOI: 10.1086/304888. eprint: astro-ph/9611107 (cit. on pp. 4, 106).

[105] P. Noterdaeme, P. Petitjean, C. Ledoux, and R. Srianand. "Evolution of the cosmological mass density of neutral gas from Sloan Digital Sky Survey II - Data Release 7." In: *Astronomy and Astrophysics* 505 (Oct. 2009), pp. 1087–1098. DOI: 10.1051/0004-6361/200912768. arXiv:0908.1574 [astro-ph.CO] (cit. on pp. 23, 50, 51).

[106] P. Noterdaeme et al. "Column density distribution and cosmological mass density of neutral gas: Sloan Digital Sky Survey-III Data Release 9." In: *Astronomy and Astrophysics* 547, L1 (Nov. 2012), p. L1. DOI: 10.1051/0004-6361/201220259. arXiv:1210.1213 [astro-ph.CO] (cit. on pp. 19, 25, 44).

[107] J. Oñorbe, M. Boylan-Kolchin, J. S. Bullock, P. F. Hopkins, D. Kereš, C.-A. Faucher-Giguère, E. Quataert, and N. Murray. "Forged in FIRE: cusps, cores and baryons in low-mass dwarf galaxies." In: *Montly Notices of the Royal Astronomical Society* 454 (Dec. 2015), pp. 2092–2106. DOI: 10.1093/mnras/stv2072. arXiv:1502.02036 (cit. on p. 4).

[108] D. Obreschkow, D. Croton, G. De Lucia, S. Khochfar, and S. Rawlings. "Simulation of the Cosmic Evolution of Atomic and Molecular Hydrogen in Galaxies." In: *Astrophysical Journal* 698 (June 2009), pp. 1467–1484. DOI: 10.1088/0004-637X/698/2/1467. arXiv:0904.2221 [astro-ph.CO] (cit. on p. 104).

[109] G. Ogiya, D. Nagai, and T. Ishiyama. "Dynamical evolution of primordial dark matter haloes through mergers." In: *MNRAS* 461 (Sept. 2016), pp. 3385–3396. DOI: 10.1093/mnras/stw1551. arXiv:1604.02866 (cit. on p. 106).

[110] S.-H. Oh et al. "High-resolution Mass Models of Dwarf Galaxies from LITTLE THINGS." In: *The Astronomical Journal* 149, 180 (June 2015), p. 180. DOI: 10.1088/0004-6256/149/6/180. arXiv:1502.01281 (cit. on p. 5).

[111] F. Pacucci, A. Mesinger, and Z. Haiman. "Focusing on warm dark matter with lensed high-redshift galaxies." In: *Montly Notices of the Royal Astronomical Society* 435 (Aug. 2013), pp. L53–L57. DOI: 10.1093/mnrasl/slt093. arXiv:1306.0009 [astro-ph.CO] (cit. on p. 33).

[112] H. Padmanabhan, T. R. Choudhury, and A. Refregier. "Modelling the cosmic neutral hydrogen from DLAs and 21-cm observations." In: *Montly Notices of the Royal Astronomical Society* 458 (May 2016), pp. 781–788. DOI: 10.1093/mnras/stw353. arXiv:1505.00008 (cit. on p. 19).

[113] H. Padmanabhan and A. Refregier. "Constraining a halo model for cosmological neutral hydrogen." In: *Montly Notices of the Royal Astronomical Society* 464 (Feb. 2017), pp. 4008–4017. DOI: 10.1093/mnras/stw2706. arXiv:1607.01021 (cit. on pp. 19, 20).

[114] O. H. Parry, V. R. Eke, C. S. Frenk, and T. Okamoto. "The baryons in the Milky Way satellites." In: *Montly Notices of the Royal Astronomical Society* 419 (Feb. 2012), pp. 3304–3318. DOI: 10.1111/j.1365-2966.2011.19971.x. arXiv:1105.3474 (cit. on p. 4).

[115] A. H. Patil et al. "Upper Limits on the 21 cm Epoch of Reionization Power Spectrum from One Night with LOFAR." In: *Astrophysical Journal* 838, 65 (Mar. 2017), p. 65. DOI: 10.3847/1538-4357/aa63e7. arXiv:1702.08679 (cit. on p. 12).

[116] P. J. E. Peebles. *The large-scale structure of the universe*. 1980 (cit. on p. 15).

[117] P. J. E. Peebles. "Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations." In: *Astrophysical Journal Letters* 263 (Dec. 1982), pp. L1–L5. DOI: 10.1086/183911 (cit. on p. 4).

[118] A. Pénin, O. Umeh, and M. Santos. "A scale dependent bias on linear scales: the case for HI intensity mapping at z=1." In: *ArXiv e-prints* (June 2017). arXiv:1706.08763 (cit. on p. 22).

[119] S. Perlmutter et al. "Measurements of Ω and Λ from 42 High-Redshift Supernovae." In: *Astrophysical Journal* 517 (June 1999), pp. 565–586. DOI: 10.1086/307221. eprint: astro-ph/9812133 (cit. on p. 2).

[120] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters." In: *Astronomy and Astrophysics* 594, A13 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv:1502.01589 (cit. on pp. 2, 3, 29, 34, 72).

[121] A. Pontzen and F. Governato. "How supernova feedback turns dark matter cusps into cores." In: *Montly Notices of the Royal Astronomical Society* 421 (Apr. 2012), pp. 3464–3471. DOI: 10.1111/j.1365-2966.2012.20571.x. arXiv:1106.0499 (cit. on p. 4).

[122] A. Pontzen, F. Governato, M. Pettini, C. M. Booth, G. Stinson, J. Wadsley, A. Brooks, T. Quinn, and M. Haehnelt. "Damped Lyman α systems in galaxy formation simulations." In: *Montly Notices of the Royal Astronomical Society* 390 (Nov. 2008), pp. 1349–1371. DOI: 10.1111/j.1365-2966.2008.13782.x. arXiv:0804.4474 (cit. on p. 21).

[123] J. Pritchard et al. "Cosmology from EoR/Cosmic Dawn with the SKA." In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 12 (Apr. 2015), p. 12. arXiv:1501.04291 (cit. on p. 50).

[124] J. X. Prochaska and A. M. Wolfe. "On the (Non)Evolution of H I Gas in Galaxies Over Cosmic Time." In: *Astrophysical Journal* 696 (May 2009), pp. 1543–1547. DOI: 10.1088/0004-637X/696/2/1543. arXiv:0811.2003 (cit. on pp. 23, 50, 51).

[125] A. Rahmati and J. Schaye. "Predictions for the relation between strong HI absorbers and galaxies at redshift 3." In: *Montly Notices of the Royal Astronomical Society* 438 (Feb. 2014), pp. 529–547. DOI: 10.1093/mnras/stt2235. arXiv:1310.3317 (cit. on p. 21).

[126] B. Ratra and P. J. E. Peebles. "Cosmological consequences of a rolling homogeneous scalar field." In: *Physical Review D* 37 (June 1988), pp. 3406–3427. DOI: 10.1103/PhysRevD.37.3406 (cit. on p. 56).

[127] J. I. Read, O. Agertz, and M. L. M. Collins. "Dark matter cores all the way down." In: *Montly Notices of the Royal Astronomical Society* 459 (July 2016), pp. 2573–2590. DOI: 10.1093/mnras/stw713. arXiv:1508.04143 (cit. on p. 4).

[128] A. G. Riess et al. "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant." In: *The Astronomical Journal* 116 (Sept. 1998), pp. 1009–1038. DOI: 10.1086/300499. eprint: astro-ph/9805201 (cit. on p. 2).

[129] V. C. Rubin. "The rotation of spiral galaxies." In: *Science* 220 (June 1983), pp. 1339–1344. DOI: 10.1126/science.220.4604.1339 (cit. on p. 3).

[130] G. B. Rybicki and A. P. Lightman. *Radiative Processes in Astrophysics*. June 1986, p. 400 (cit. on p. 28).

[131] C. L. Sarazin. "X-ray emission from clusters of galaxies." In: *Reviews of Modern Physics* 58 (Jan. 1986), pp. 1–115. DOI: 10.1103/RevModPhys.58.1 (cit. on p. 3).

[132] A. Schneider, R. E. Smith, and D. Reed. "Halo mass function and the free streaming scale." In: *Montly Notices of the Royal Astronomical Society* 433 (Aug. 2013), pp. 1573–1587. DOI: 10.1093/mnras/stt829. arXiv:1303.0839 [astro-ph.CO] (cit. on pp. 36, 39, 40, 61).

[133] A. Schneider, R. E. Smith, A. V. Macciò, and B. Moore. "Nonlinear evolution of cosmological structures in warm dark matter models." In: *Montly Notices of the Royal Astronomical Society* 424 (July 2012), pp. 684–698. DOI: 10.1111/j.1365-2966.2012.21252.x. arXiv:1112.0330 [astro-ph.CO] (cit. on pp. 38, 61).

[134] A. Schneider, D. Anderhalden, A. V. Macciò, and J. Diemand. "Warm dark matter does not do better than cold dark matter in solving small-scale inconsistencies." In: *Montly Notices of the Royal Astronomical Society* 441 (June 2014), pp. L6–L10. DOI: 10.1093/mnrasl/slu034. arXiv:1309.5960 (cit. on p. 33).

[135] A. Schneider, R. Teyssier, D. Potter, J. Stadel, J. Onions, D. S. Reed, R. E. Smith, V. Springel, F. R. Pearce, and R. Scoccimarro. "Matter power spectrum and the challenge of percent accuracy." In: *Journal of Cosmology and Astroparticle Physics* 4, 047 (Apr. 2016), p. 047. DOI: 10.1088/1475-7516/2016/04/047. arXiv:1503.05920 (cit. on p. 106).

[136] R. K. Sheth, H. J. Mo, and G. Tormen. "Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes." In: *Montly Notices of the Royal Astronomical Society* 323 (May 2001), pp. 1–12. DOI: 10.1046/j.1365-8711.2001.04006.x. eprint: astro-ph/9907024 (cit. on pp. 23, 86).

[137] R. K. Sheth and G. Tormen. "Large-scale bias and the peak background split." In: *Montly Notices of the Royal Astronomical Society* 308 (Sept. 1999), pp. 119–126. DOI: 10.1046/j.1365-8711.1999.02692.x. eprint: astro-ph/9901122 (cit. on pp. 18, 23).

[138] R. K. Sheth and G. Tormen. "An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier." In: *Montly Notices of the Royal Astronomical Society* 329 (Jan. 2002), pp. 61–75. DOI: 10.1046/j.1365-8711.2002.04950.x. eprint: arXiv:astro-ph/0105113 (cit. on pp. 40, 83, 86).

[139] M. Sitwell, A. Mesinger, Y.-Z. Ma, and K. Sigurdson. "The imprint of warm dark matter on the cosmological 21-cm signal." In: *Montly Notices of the Royal Astronomical Society* 438 (Mar. 2014), pp. 2664–2671. DOI: 10.1093/mnras/stt2392. arXiv:1310.0029 (cit. on p. 33).

[140] A. Slosar, S. Ho, M. White, and T. Louis. "The acoustic peak in the Lyman alpha forest." In: *Journal of Cosmology and Astroparticle Physics* 10, 019 (Oct. 2009), p. 019. DOI: 10.1088/1475-7516/2009/10/019. arXiv:0906.2414 [astro-ph.CO] (cit. on p. 78).

[141] A. Slosar et al. "The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data." In: *Journal of Cosmology and Astroparticle Physics* 9, 001 (Sept. 2011), p. 001. DOI: 10.1088/1475-7516/2011/09/001. arXiv:1104.5244 (cit. on pp. 13, 75, 82, 90, 92).

[142] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman. "Stable clustering, the halo model and non-linear

cosmological power spectra." In: *Montly Notices of the Royal Astronomical Society* 341 (June 2003), pp. 1311–1332. DOI: 10.1046/j.1365-8711.2003.06503.x. eprint: astro-ph/0207664 (cit. on p. 106).

[143] S. Smith. "The Mass of the Virgo Cluster." In: *Astrophysical Journal* 83 (Jan. 1936), p. 23. DOI: 10.1086/143697 (cit. on p. 3).

[144] V. Springel. "The cosmological simulation code GADGET-2." In: *Montly Notices of the Royal Astronomical Society* 364 (Dec. 2005), pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x. eprint: arXiv:astro-ph/0505010 (cit. on pp. 29, 34, 71).

[145] V. Springel and L. Hernquist. "Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation." In: *Montly Notices of the Royal Astronomical Society* 339 (Feb. 2003), pp. 289–311. DOI: 10.1046/j.1365-8711.2003.06206.x. eprint: astro-ph/0206393 (cit. on pp. 30, 34).

[146] V. Springel et al. "Simulations of the formation, evolution and clustering of galaxies and quasars." In: *Nature* 435 (June 2005), pp. 629–636. DOI: 10.1038/nature03597. eprint: astro-ph/0504097 (cit. on p. 4).

[147] E. R. Switzer et al. "Determination of z ~ 0.8 neutral hydrogen fluctuations using the 21 cm intensity mapping autocorrelation." In: *Montly Notices of the Royal Astronomical Society* 434 (July 2013), pp. L46–L50. DOI: 10.1093/mnrasl/slt074. arXiv:1304.3712 [astro-ph.CO] (cit. on pp. 19, 96).

[148] S. Tassev, M. Zaldarriaga, and D. J. Eisenstein. "Solving large scale structure in ten easy steps with COLA." In: *Journal of Cosmology and Astroparticle Physics* 6, 036 (June 2013), p. 036. DOI: 10.1088/1475-7516/2013/06/036. arXiv:1301.0322 [astro-ph.CO] (cit. on p. 106).

[149] M. Tegmark and M. Zaldarriaga. "Separating the early universe from the late universe: Cosmological parameter estimation beyond the black box." In: *Physical Review D* 66.10, 103508 (Nov. 2002), p. 103508. DOI: 10.1103/PhysRevD.66.103508. eprint: astro-ph/0207047 (cit. on pp. 6, 7).

[150] M. Tegmark et al. "Cosmological parameters from SDSS and WMAP." In: *Physical Review D* 69.10, 103501 (May 2004), p. 103501. DOI: 10.1103/PhysRevD.69.103501. eprint: astro-ph/0310723 (cit. on p. 3).

[151] T. Tepper-García. "Voigt profile fitting to quasar absorption lines: an analytic approximation to the Voigt-Hjerting function." In: *Montly Notices of the Royal Astronomical Society* 369 (July 2006), pp. 2025–2035. DOI: 10.1111/j.1365-2966.2006.10450.x. eprint: astro-ph/0602124 (cit. on p. 73).

[152] R. Teyssier. "Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES." In: *Astronomy and Astrophysics* 385 (Apr. 2002), pp. 337–364. DOI: 10.1051/0004-6361:20011817. eprint: astro-ph/0111367 (cit. on p. 59).

[153] T. Theuns, A. Leonard, G. Efstathiou, F. R. Pearce, and P. A. Thomas. "P^3M-SPH simulations of the Lyalpha forest." In: *Montly Notices of the Royal Astronomical Society* 301 (Dec. 1998), pp. 478–502. DOI: 10.1046/j.1365-8711.1998.02040.x. eprint: astro-ph/9805119 (cit. on pp. 73, 76).

[154] E. J. Tollerud et al. "The SPLASH Survey: Spectroscopy of 15 M31 Dwarf Spheroidal Satellite Galaxies." In: *Astrophysical Journal* 752, 45 (June 2012), p. 45. DOI: 10.1088/0004-637X/752/1/45. arXiv:1112.1067 [astro-ph.CO] (cit. on p. 4).

[155] S. Tsujikawa. "Quintessence: a review." In: *Classical and Quantum Gravity* 30.21, 214003 (Nov. 2013), p. 214003. DOI: 10.1088/0264-9381/30/21/214003. arXiv:1304.1961 [gr-qc] (cit. on p. 56).

[156] M. Viel, S. Matarrese, H. J. Mo, M. G. Haehnelt, and T. Theuns. "Probing the intergalactic medium with the Lyα forest along multiple lines of sight to distant QSOs." In: *Montly Notices of the Royal Astronomical Society* 329 (Feb. 2002), pp. 848–862. DOI: 10.1046/j.1365-8711.2002.05060.x. eprint: astro-ph/0105233 (cit. on p. 12).

[157] M. Viel, K. Markovič, M. Baldi, and J. Weller. "The non-linear matter power spectrum in warm dark matter cosmologies." In: *Montly Notices of the Royal Astronomical Society* 421 (Mar. 2012), pp. 50–62. DOI: 10.1111/j.1365-2966.2011.19910.x. arXiv:1107.4094 [astro-ph.CO] (cit. on p. 35).

[158] M. Viel, G. D. Becker, J. S. Bolton, and M. G. Haehnelt. "Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data." In: *Physical Review D* 88.4, 043502 (Aug. 2013), p. 043502. DOI: 10.1103/PhysRevD.88.043502. arXiv:1306.2314 [astro-ph.CO] (cit. on pp. 33, 34, 60, 72).

[159] F. Villaescusa-Navarro, D. Alonso, and M. Viel. "Baryonic acoustic oscillations from 21 cm intensity mapping: the Square Kilometre Array case." In: *Montly Notices of the Royal Astronomical Society* 466 (Apr. 2017), pp. 2736–2751. DOI: 10.1093/mnras/stw3224. arXiv:1609.00019 (cit. on p. 11).

[160] F. Villaescusa-Navarro, P. Bull, and M. Viel. "Weighing Neutrinos with Cosmic Neutral Hydrogen." In: *Astrophysical Journal* 814, 146 (Dec. 2015), p. 146. DOI: 10.1088/0004-637X/814/2/146. arXiv:1507.05102 (cit. on p. 52).

[161] F. Villaescusa-Navarro, M. Viel, K. K. Datta, and T. R. Choudhury. "Modeling the neutral hydrogen distribution in the post-reionization Universe: intensity mapping." In: *Journal of Cosmology and Astroparticle Physics* 9, 050 (Sept. 2014), p. 50. DOI: 10.1088/1475-7516/2014/09/050. arXiv:1405.6713 (cit. on pp. 20, 22, 24, 26, 31, 32, 84, 86).

[162] F. Villaescusa-Navarro, M. Viel, D. Alonso, K. K. Datta, P. Bull, and M. G. Santos. "Cross-correlating 21cm intensity maps with Lyman Break Galaxies in the post-reionization era." In: *Journal of Cosmology and Astroparticle Physics* 3, 034 (Mar. 2015), p. 034. DOI: 10.1088/1475-7516/2015/03/034. arXiv:1410.7393 (cit. on pp. 10, 20, 32).

[163] F. Villaescusa-Navarro et al. "Neutral hydrogen in galaxy clusters: impact of AGN feedback and implications for intensity mapping." In: *Montly Notices of the Royal Astronomical Society* 456 (Mar. 2016), pp. 3553–3570. DOI: 10.1093/mnras/stv2904. arXiv:1510.04277 (cit. on pp. 19–22, 84).

[164] K. Vonnegut. *Cat's Cradle: A Novel*. Random House Publishing Group, 2009. ISBN: 9780307567277. URL: https://books.google.fr/books?id=w25sx0G6nRsC (cit. on p. 1).

[165] J. Wang and S. D. M. White. "Discreteness effects in simulations of hot/warm dark matter." In: *Montly Notices of the Royal Astronomical Society* 380 (Sept. 2007), pp. 93–103. DOI: 10.1111/j.1365-2966.2007.12053.x. eprint: astro-ph/0702575 (cit. on pp. 38, 61).

[166] D. H. Weinberg, J. S. Bullock, F. Governato, R. Kuzio de Naray, and A. H. G. Peter. "Cold dark matter: Controversies on small scales." In: *Proceedings of the National Academy of Science* 112 (Oct. 2015), pp. 12249–12255. DOI: 10.1073/pnas.1308716112. arXiv:1306.0913 (cit. on pp. 4, 5).

[167] S. Weinberg. "The Cosmological Constant Problems (Talk given at Dark Matter 2000, February, 2000)." In: *ArXiv Astrophysics e-prints* (May 2000). eprint: astro-ph/0005265 (cit. on p. 2).

[168] M. White, A. Pope, J. Carlson, K. Heitmann, S. Habib, P. Fasel, D. Daniel, and Z. Lukic. "Particle Mesh Simulations of the Lyα Forest and the Signature of Baryon Acoustic Oscillations in the Intergalactic Medium." In: *Astrophysical Journal* 713 (Apr. 2010), pp. 383–393. DOI: 10.1088/0004-637X/713/1/383. arXiv:0911.5341 (cit. on p. 76).

[169] A. M. Wolfe, E. Gawiser, and J. X. Prochaska. "Damped Ly α Systems." In: *Annual Review of Astronomy & Astrophysics* 43 (Sept. 2005), pp. 861–918. DOI: 10.1146/annurev.astro.42.053102.133950. eprint: astro-ph/0509481 (cit. on p. 8).

[170] L. Wolz, F. B. Abdalla, C. Blake, J. R. Shaw, E. Chapman, and S. Rawlings. "The effect of foreground subtraction on cosmological measurements from intensity mapping." In: *Montly Notices of the Royal Astronomical Society* 441 (July 2014), pp. 3271–3283. DOI: 10.1093/mnras/stu792. arXiv:1310.8144 (cit. on p. 11).

[171] T. Zafar, C. Péroux, A. Popping, B. Milliard, J.-M. Deharveng, and S. Frank. "The ESO UVES advanced data products quasar sample. II. Cosmological evolution of the neutral gas mass density." In: *Astronomy and Astrophysics* 556, A141 (Aug. 2013), A141. DOI: 10.1051/0004-6361/201321154. arXiv:1307.0602 [astro-ph.CO] (cit. on pp. 19, 25, 44).

[172] S. Zaroubi. "The Epoch of Reionization." In: *The First Galaxies*. Ed. by T. Wiklind, B. Mobasher, and V. Bromm. Vol. 396. Astrophysics and Space Science Library. 2013, p. 45. DOI: 10.1007/978-3-642-32362-1_2. arXiv:1206.0267 (cit. on pp. 7, 9, 12).

[173] M. A. Zwaan, M. J. Meyer, L. Staveley-Smith, and R. L. Webster. "The HIPASS catalogue: $\Omega_{\rm HI}$ and environmental effects on the HI mass function of galaxies." In: *Montly Notices of the Royal Astronomical Society* 359 (May 2005), pp. L30–L34. DOI: 10.1111/j.1745-3933.2005.00029.x. eprint: astro-ph/0502257 (cit. on p. 27).

[174] F. Zwicky. "Die Rotverschiebung von extragalaktischen Nebeln." In: *Helvetica Physica Acta* 6 (1933), pp. 110–127 (cit. on p. 3).

ACKNOWLEDGEMENTS

To begin with, I want to warmly thank my supervisor, Matteo Viel, for his patient guidance and for having being always supportive.

I am deeply grateful to SISSA, especially to the people of the astrophysics sector, but I need to acknowledge the whole SISSA community for making SISSA a great place where to be and where to work.

I really enjoyed interacting with students and researchers of the Astronomical Observatory of Trieste, where I felt always welcome and where I received great feedbacks through their journal clubs and the N/C meeting series.

Special thanks go to Francisco Villaescusa-Navarro, also known as Paco: the improvements of my technical and numerical skills are due to having being able to work with him side by side, and that is not all of it. Paco always found the time to discuss with me and to give me his scientific counselling, even after he left Trieste. ¡Muchas gracias!

I owe my gratitude to Pier Stefano Corasaniti, for having welcomed me in his research group in Paris, giving me opportunities to learn a lot of physics and to be exposed to a scientifically rich environment.

I have had the honour and the pleasure to have met and discussed with many more scientists and fellow students who somehow and differently participated in this journey of mine. I take the occasion to thank some of them in particular that devoted a bit of their time to me and eventually to the work described in this thesis: Jingjing Shi, Giulio Fabbian, Camilla Penzo and Andrej Obuljen.

Final thanks to David Alonso and Lauro Moscardini, who kindly accepted to read this manuscript and be part of my Ph.D. defence committee.