Scuola Internazionale Superiore di Studi Avanzati[1] - Trieste

Doctoral Thesis

Large-Scale Structure with 21cm Intensity Mapping

Author:
Andrej Obuljen

Supervisor:
Prof. Matteo Viel

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in

Astrophysics

September 2018

SISSA - Via Bonomea 265 - 34136 TRIESTE - ITALY

Abstract

Andrej Obuljen

Large-Scale Structure with 21cm Intensity Mapping

We are witnessing exciting times in the field of cosmology. Current and future experiments and surveys will provide us with tight constraints on the key cosmological parameters. A new and promising technique of mapping the Large-Scale Structure (LSS) in our Universe is the 21cm Intensity Mapping (IM) in which one uses the emission of the neutral hydrogen as a tracer of the underlying matter field. In principle this technique can be used to map huge portions of our Universe and deliver 3D structure measurements providing us with the information that is complementary to the information extracted from the Cosmic Microwave Background (CMB) experiments. However, the field of 21cm IM cosmology is still in its raising and is severely limited by the foreground issues and problems.

In this thesis we will consider several aspects of using the 21cm IM as an LSS probe in order to better constrain the cosmological parameters.

First, we present and analyse a Baryon Acoustic Oscillation (BAO) reconstruction method that consists of displacing pixels instead of galaxies and whose implementation is easier than the standard reconstruction method. We show that this method is equivalent to the standard reconstruction technique in the limit where the number of pixels becomes very large. This method is particularly useful in surveys where individual galaxies are not resolved, as in 21cm IM observations. We validate this method by reconstructing mock pixelated maps, that we build from the distribution of matter and halos in real- and redshift-space, from a large set of numerical simulations. We find that this method is able to decrease the uncertainty in the BAO peak position by 30-50% over the typical angular resolution scales of 21 cm IM experiments.

Second, we investigate the possibility of performing cosmological studies in the redshift range $2.5 < z < 5$ through suitable extensions of existing and upcoming radio-telescopes like CHIME, HIRAX and FAST. We use the Fisher matrix technique to forecast the bounds that those instruments can place on the growth rate, the BAO distance scale parameters, the sum of the neutrino masses and the number of relativistic degrees of freedom at decoupling, $N_{\rm eff}$. We point out that quantities that depend on the amplitude of the 21cm power spectrum, like $f\sigma_8$, are completely degenerate with $\Omega_{\rm HI}$ and $b_{\rm HI}$. Then, we propose several strategies to independently constrain them through cross-correlations with other probes. We study in detail the dependence of our results on the instrument, amplitude of the HI bias, the foreground wedge coverage, the nonlinear scale used in the analysis, uncertainties in the theoretical modeling and the priors on $b_{\rm HI}$ and $\Omega_{\rm HI}$. We conclude that 21cm IM surveys operating in this redshift range can provide extremely competitive constraints on key cosmological parameters.

Thridly, we have used TNG100, a large state-of-the-art magneto-hydrodynamic simulation of a $75\,h^{-1}$ Mpc box size, which is part of the IllustrisTNG Project, to study the neutral hydrogen density profiles in dark matter halos. We find that while the density profiles of HI exhibit a large halo-to-halo scatter, the mean profiles are universal across mass and redshift.

Finally, we combine information from the clustering of HI galaxies in the 100% data release of the Arecibo Legacy Fast ALFA survey (ALFALFA), and from the HI content of optically-selected galaxy groups found in the Sloan Digital Sky Survey (SDSS) to constrain the relation between halo mass M_h and its average total HI mass content $M_{\rm HI}$. We model the abundance and clustering of neutral hydrogen through a halo-model-based approach, parametrizing the $M_{\rm HI}(M_h)$ relation as a power law with an exponential mass cutoff. To break the degeneracy between the amplitude and low-mass cutoff of the $M_{\rm HI}(M_h)$ relation, we also include a recent measurement of the cosmic HI abundance from the 100% ALFALFA sample. We find that all datasets are consistent with a power-law index $\alpha = 0.44 \pm 0.08$ and a cutoff halo mass $\log_{10} M_{\rm min}/(h^{-1}M_\odot) = 11.27^{+0.24}_{-0.30}$. We compare these results with predictions from state-of-the-art magneto-hydrodynamical simulations, and find both to be in good qualitative agreement, although the data favours a significantly larger cutoff mass that is consistent with the higher cosmic HI abundance found in simulations. Both data and simulations seem to predict a similar value for the HI bias ($b_{\rm HI} = 0.875 \pm 0.022$) and shot-noise power ($P_{\rm SN} = 92^{+20}_{-18}\,[h^{-1}{\rm Mpc}]^3$) at redshift $z = 0$.

List of Publications

This thesis is a result of the work I have done while being a PhD student in Astrophysics at SISSA under the supervision of Prof. Matteo Viel (SISSA). During my PhD studies I have mainly collaborated with Prof. Matteo Viel, Dr. Francisco Villaescusa-Navarro, Dr. Emanuele Castorina and Dr. David Alonso, but also others. The work presented here is based upon the following publications:

- Chapter 3:
  Baryon Acoustic Oscillations reconstruction with pixels,
  Obuljen, A.; Villaescusa-Navarro, F.; Castorina, E. and Viel, M. [arXiv:1610.05768].
  Published in JCAP **1709** (2017) 012.

- Chapter 4:
  High-redshift post-reionization cosmology with 21cm intensity mapping,
  Obuljen, A.; Castorina, E.; Villaescusa-Navarro, F. and Viel, M. [arXiv:1709.07893].
  Published in JCAP **1805** (2018) 004.

- Chapter 5:
  The HI content of dark matter halos at $z \approx 0$ from ALFALFA,
  Obuljen, A.; Alonso, D.; Villaescusa-Navarro, F.; Yoon, I.; Jones, M. [arXiv:1805.00934].
  Submitted to MNRAS.

and a partial contribution from my side to the following work:

- Chapter 2, section 2.1.2:
  Ingredients for 21cm intensity mapping,
  Villaescusa-Navarro, F.; Genel, S.; Castorina, E.; Obuljen, A.; Spergel, D. N.; Hernquist, L.; Nelson, D.; Carucci, I. P.; Pillepich, A.; Marinacci, F.; Diemer, B.; Vogelsberger, M.; Weinberger, R.; Pakmor, R. [arXiv:1804.09180]. Submitted to ApJ.

Acknowledgements

I would like to start by thanking SISSA and INFN for all the financial support to travel and attend many conferences that have been of great importance in my work. I also acknowledge the support by the INFN grant PD 51 INDARK. All of my numerical simulations and most of the numerical calculations have been carried out at the Ulysses cluster in SISSA. I would also like to acknowledge the great hospitality of the Center for Computational Astrophysics of the Simons Foundation in NY.

There are many people that have contributed to the work presented here in their own special ways.

First of all I would like to thank my supervisor Matteo Viel for all the help, encouragement, drive and good spirit that he has provided. Most of all, I thank him for introducing me into the field of 21cm cosmology and letting me feel free to work on what was interesting for me. It has been a real pleasure to work and talk to him. Through him I have got to chance to work with his post-doc — Francisco Villaescusa-Navarro (a.k.a. Paco) and his PhD student — Emanuele Castorina. Paco was present in all of the work I have done so far and his help and perspiration to get things done correctly and efficiently has been nothing but a good influence and an enormous help for me. His availability and good spirit are amazing. I thank you Paco. I also thank him for inviting me to spend two months at the Flatiron Institute which has been great experience. I would like to thank Emanuele who provided great ideas for our work, involved me in projects that would turn out to be very useful for me and has been a great help throughout my PhD studies. Thank you Ema.

I have had the chance to discuss science and collaborate with many other excellent people from whom I learned a lot during my PhD. I would like to thank: Marko Simonović, David Alonso, Anže Slosar, Emiliano Sefusatti, David Spergel, Giulio Fabian, Paolo Cremineli, Ravi Sheth, Carlo Baccigalupi, Ariel Sanchez, David Hogg, Alkistis Pourtsidou for useful conversations and discussions. I would also like to thank Phil Bull for useful comments on the thesis.

My SISSA colleagues — Elias Kammoun, Francesca Lepori, Gor Oganesyan and Riccardo Murgia have been great coworkers and have become great friends. I thank them all for all the discussions and memorable times.

On a more personal side, there is a group of my hometown best friends that has supported me and encouraged me to always achieve more, but also to have a lot of fun while doing it. I thank Miša Popović, Dušan Đilas, Radovan Dojčilović and Pavle Avramović for great times.

I don't think I would even be doing this if it was not for Petnica Science Center in my high-school, undergraduate, master and PhD years. It has changed me in many ways for better. I have met some people there that have been a big part of my life. I wish to thank especially Ivan Milić, Mateja Bošković, Ivan Grahek-Jovanović, Stanislav Milošević, Nemanja Martinović, Nikola Božić.

My family has always been very supportive, caring and never doubted nor questioned my strive to continue doing research. I would like to thank my parents —Ljiljana and Viktor Obuljen, and my big brother Aljoša for everything.

Finally, I would like to thank my love and wife Ri for always being there for me and for making me feel more happy than I thought it can be done. I love you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main goal of cosmology is to describe the content, geometry, dynamics and evolution of our Universe. In the last decades the field of cosmology has gone through a tremendous progress. We now have a well established theoretical model (ΛCDM model) with only 6 parameters which provides us with a very detailed picture of how the Universe evolved since its beginning. This model is tested at different scales, energies and epochs by observations at the percent (and in many cases at the sub-percent) level.

The era of precision cosmology started with the measurement of the anisotropies in the temperature of the Cosmic Microwave Background (CMB) in the beginning of the 1990's done with the COBE satellite [1]. Since then, the CMB anisotropies have been measured with more precision by other experiments, e.g. Boomerang, WMAP, Planck etc, which have provided us with a wealth of information [2, 3, 4, 5, 6]. We are still in the regime where the most of the constraining power on the ΛCDM model comes from CMB.

In parallel, we now also have very precise measurements of the Large Scale Structure (LSS) in the late Universe. This was achieved by increasingly deeper and more precise galaxy surveys studying the spatial distribution of galaxies, weak lensing or the properties of the Lyα forest, e.g. [7, 8]. The way galaxies are distributed is tightly related to the initial fluctuations seen in the CMB and thus we expect the same model can describe both observations. All of these datasets are consistent with the ΛCDM model and in synergy provide us with even tighter constraints on the cosmological parameters.

The main difference between the LSS and CMB is the number of scales one can measure. While the CMB is coming from a thin shell known as the last scattering surface, LSS is three-dimensional. We thus expect LSS to have more constraining power than CMB. However, only recently LSS has become comparable to the constraints from CMB. This is due to the different physical regimes that these two approaches probe. However, with galaxy surveys we are still very far from using all the available volume. In fact, we are only using a small fraction of it. This is mainly due to the fact that distant galaxies become fainter and harder to detect.

A promising new tracer to use in order to map the rest of the Universe is the neutral hydrogen (HI). HI can be seen in absorption of the CMB photons up to very early cosmic times but also in emission in the post-reionization epoch. It could in principle allow us to map out most of the volume of the observable Universe and extract the information we are still largely missing. In the post-reionization Universe, HI can be mapped with radio telescopes with a technique called intensity mapping (IM). In this thesis we will discuss and give contribution to this field and in particular in its use as a LSS probe of the Universe in the post-reionization epoch.

Standard Cosmological ΛCDM Model

In the standard cosmological model, our Universe is assumed to be isotropic and homogeneous on very large scales (>100 Mpc). This assumption is what is called *the cosmological principle* (CP) and it is crucial in order to solve the Einstein equations and make predictions. In this paradigm, the Universes' background evolution is governed by its content. The main constituents are assumed to be the following:

- Baryonic Matter – representing the regular atomic matter visible by EM radiation, e.g. stars, planets, gas etc. Current measurements show that this components constitutes around 5% of the total energy density of the Universe.
- Cold Dark Matter (CDM) – currently unknown form of matter making up roughly 25% of the total energy density of the Universe and 5 times more abundant than baryonic matter. It is called dark because it does not emit or absorb EM spectrum in the way that "normal" baryonic matter does. Its existence is inferred by the gravitational effect it has on the dynamics of the Universe as a whole, the kinematics of objects and the light propagation inside our Universe.
- Dark Energy (DE, Λ) – component making up roughly 70% of the total energy density at present time discovered in the late 1990's [9, 10]. Current measurements point towards it behaving like the energy density of vacuum and being constant with time, in contrast to standard matter whose energy density decays with the expansion.
- Relativistic species – this includes radiation made in the early Universe whose energy density was dominant early on, but has decayed below the energy density of all other components very quickly.
- Neutrinos – almost massless non-relativistic particles at present time, while relativistic early on (at the time of recombination).

Aside from that, the ΛCDM model also assumes that the spatial curvature of our Universe is very small.

1.1 Homogeneous Universe

1.1.1 Friedmann equations

Assuming CP and General Relativity (GR), the most general spacetime metric one can write in terms of the comoving coordinates also known as the Friedmann-Robertson-Walker (FRW) metric[1]:

$$ds^2 = -c^2dt^2 + a(t)^2\left[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right], \tag{1.1}$$

where the function $a(t)$ is known as the *scale factor* and it describes the expansion/contraction. It also describes the distances between objects at fixed comoving coordinates as a function of time. This function is usually normalised to its value at the present time $a(t_0) = 1$.

The dynamics of the expansion is described by the following Friedmann equations:

---

[1] Throughout the thesis we will assume a spatially flat Universe, an assumption that is consistent with current observations [11] and predicted by inflation.

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho, \; \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p), \tag{1.2}$$

where ρ is the total energy density, p is the pressure of the fluid and the dots denote the derivatives with respect to cosmic time t. These equations are obtained by solving the Einstein equations assuming a flat Universe[2]. The quantity H is called the Hubble parameter.

By combining equations 1.2 we can obtain the continuity equation:

$$\dot{\rho} + 3H(\rho + p) = 0. \tag{1.3}$$

This equation holds for any present fluid separately. Assuming a simple form of the equation of state $p = w\rho$, where w is a constant parameter, one can integrate the previous equation to obtain the evolution of the energy density of a given component:

$$\rho \propto a^{-3(1+w)}. \tag{1.4}$$

Furthermore, it is also straightforward to find a solution for the scale factor as a function of time:

$$a(t) \propto \begin{cases} t^{\frac{2}{3(1+w)}}, & w \neq 1 \\ e^{Ht}, & w = 1 \end{cases} \tag{1.5}$$

Usually, we consider three types of fluids: non-relativistic matter, e.g. ordinary and dark matter, with $w_\mathrm{m} = 0$, relativistic matter, e.g. radiation, with $w_\mathrm{r} = 1/3$ and dark energy with $w_\Lambda = -1$. We can thus see that the corresponding energy densities and the scale factor scale as $\rho_\mathrm{m} \propto a^3$ and $a(t) \propto t^{2/3}$, $\rho_\mathrm{r} \propto a^4$ and $a(t) \propto t^{1/2}$, while ρ_Λ is constant and the Universe expands exponentially.

The present time critical energy density required for the flat space geometry is defined as $\rho_\mathrm{crit} = \frac{3H_0^2}{8\pi G}$. It is now useful to also define the dimensionless *cosmological parameters*[3] as ratios of different components to the critical density $\Omega_i(t) = \rho_i(t)/\rho_\mathrm{crit}$ and write the first Friedmann equation as:

$$\frac{H^2(t)}{H_0^2} = \Omega_\mathrm{m} a^{-3} + \Omega_\mathrm{r} a^{-4} + \Omega_\Lambda. \tag{1.6}$$

1.1.2 Redshift

The redshift of an emitting source is defined as the fractional difference of the observed and the emitted wavelength with respect to the emitted wavelength:

$$z = \frac{\lambda_\mathrm{o}}{\lambda_\mathrm{e}} - 1. \tag{1.7}$$

In the homogeneous and isotropic Universe this shift is caused by the expansion of the Universe and is a measurable quantity. This is why it is useful to relate it to the scale factor $a(t)$:

$$\frac{a(t_\mathrm{o})}{a(t_\mathrm{e})} = \frac{1}{a(t)} = 1 + z. \tag{1.8}$$

---

[2] For a complete derivation of Friedmann equations from GR we refer the reader to classic textbooks, e.g. [12].

[3] These form a subset of the full ΛCDM model parameters.

We see that when we observe objects at high redshift, we are observing the Universe at early times ($a(t) \ll 1$), while the objects in the present Universe have low redshifts. We can now write the equation 1.6 in terms of redshift:

$$\frac{H^2(z)}{H_0^2} = \Omega_{\rm m}(1+z)^3 + \Omega_{\rm r}(1+z)^4 + \Omega_\Lambda. \tag{1.9}$$

1.1.3 Distances in cosmology

One of the most important endeavours in astronomy and cosmology is to measure both accurately and precisely the distances. In the expanding Universe distance measurements need to be taken with special care since distances change with time (for a pedagogical and complete review see [13]). Given the cosmological parameters we can use the geodesic equation and compute what is the proper distance travelled by a photon moving along the radial direction towards an observer centred at the origin of coordinate system ($d\theta = d\phi = 0$):

$$ds = a(t)dr = cdt. \tag{1.10}$$

The proper distance is then

$$r(z) = \int_{t_{\rm e}}^{t_{\rm o}} \frac{cdt}{a(t)} = \int_0^z \frac{cdz}{H(z)} = D_{\rm H} \int_0^z \frac{dz}{E(z)}, \tag{1.11}$$

where $D_{\rm H} \equiv c/H_0$ is the Hubble distance and we have used $H(z) = H_0 E(z)$ with:

$$E(z) \equiv \sqrt{\Omega_{\rm M}(1+z)^3 + \Omega_{\rm r}(1+z)^4 + \Omega_\Lambda}. \tag{1.12}$$

We can see that the proper distances are determined by the content of the Universe and thus one can use the measured distances to constrain the cosmological parameters. However, the proper distance is not a measurable quantity. Therefore we need a way to relate the proper distance to quantities we measure – redshifts, fluxes and angles.

- **Luminosity distance** One way is to consider the case in which we measure the redshifts and fluxes of objects, e.g. galaxies or supernovae (SN). Given an intrinsic luminosity of the source (L), the measured flux (f) depends on the distance to an object by:

  $$f = \frac{L}{4\pi d_{\rm L}^2}, \tag{1.13}$$

  where $d_{\rm L}$ is called the *luminosity distance*. Due to the expansion of the Universe, the observed wavelength of the photon is stretched and its energy is reduced by a factor of $E_{\rm o} = E_{\rm e}(1+z)$. Additionally, the rate of arrival of photons is reduced by a factor of $\delta t_{\rm o} = \delta t_{\rm e}(1+z)$. The net result is that the measured flux is smaller by a factor of $(1+z)^2$

  $$f = \frac{L}{4\pi(1+z)^2 r(z)^2}. \tag{1.14}$$

  The luminosity distance can then be be related to the proper distance:

  $$d_{\rm L} = (1+z)r(z). \tag{1.15}$$

  Luminosity distance is actually something we can measure. For objects that have known luminosity, also known as *standard candles*, we can measure the flux and infer d_{L}. One example of such objects are SN Ia. In fact, measurements of the distances to nearby SN Ia have been used to constrain the cosmological parameters which has led to the discovery of the accelerated expansion [9, 10].

- **Angular diameter distance** Another way to measure distances is purely geometrical. If one knows the physical size l that subtends an angle θ in the sky then in the static and Euclidian Universe this angle is related to the distance by $d_{\mathrm{A}} \equiv l/\theta$. In the expanding Universe, the physical size is related to the comoving distance by the scale factor at the time of the emission $l = a(t_{\mathrm{e}})r(z)\theta = r(z)\theta/(1+z)$. Thus the angular comoving distance is defined as $d_{\mathrm{A}} \equiv r(z)/(1+z)$. Cases where we know the intrinsic physical size are called *standard rulers* and we can use them to measure d_{A}. An example of a standard ruler is the physical scale $\approx 150\,\mathrm{Mpc}$ of the baryon acoustic oscillations (BAO) peak in the two-point correlation function of, e.g. galaxies (see section 1.2.7) that has been used extensively to constrain the distance-redshift relation at a range of distances/redshifts not easily accessible by other means of measuring distances.

1.2 Inhomogeneous Universe

In the previous section we have shown how a perfectly homogeneous Universe expands, how its components evolve on average and how one goes about constraining the cosmological parameters by measuring the distances. However, the Universe is obviously not homogeneous on all scales and all times. We have large inhomogeneities in our Local Universe (galaxy clusters, groups of galaxies, galaxies, voids) which are even more pronounced on small scales (different galactic components, globular clusters, interstellar medium, stars, planets etc.). We now also have measurements that give us a glimpse at how the Universe looked at very early stages and from CMB we see a completely different picture than the one today. The Universe back then was much more homogeneous with average relative inhomegeneities of $\approx 10^{-5}$. In contrast, in the local Universe this number is many orders of magnitude larger. As an example, the typical relative inhomogenity value necessary to form a DM halo is ≈ 200. In order to understand how has all of this structure formed and evolved with time, we need to consider inhomogeneities in the primordial density field. While the details of these primordial inhomogeneities is still unclear and some inflationary models do predict primordial non-gaussianities, it appears that the initial conditions (IC) could not have been far from those of the Gaussian random field. In fact all of the current observations are consistent with completely Gaussian primordial field [14] and in this thesis we will assume Gaussian IC. These IC will eventually evolve into the non-linear structure we observe in the local Universe and become less and less Gaussian.

1.2.1 Linear Perturbation Theory

We are now interested in describing the evolution of the matter field $\rho(\vec{x})$ using the linear perturbation theory (PT). We will do this in terms of the overdensity field defined as:

$$\delta(\vec{x},t) = \frac{\rho(\vec{x},t)}{\bar{\rho}(t)} - 1, \tag{1.16}$$

where $\bar{\rho}$ is the average matter field density over all the space. We can see that the overdensity field has to satisfy $-1 \leq \delta(\vec{x},t) < \infty$. At large enough scales and at early enough times, δ is a small quantity so it is justified to use it as the expansion parameter in PT.

To describe the evolution of $\rho(\vec{x})$ we will describe the main results under a few assumptions. First one is that all dark matter is cold. Second is that we are considering scales smaller than the Hubble horizon ($\lambda \ll H^{-1}$) where the GR effects become more important and that the peculiar velocities of particles are non-relativistic ($v_\mathrm{p} \ll c$) – also known as the Newtonian approximation. Furthermore, we will first consider the matter dominated era where $\Omega_\mathrm{M} = 1$, while later on we will generalise the solution for the full ΛCDM.

The linearised equations for a perfect fluid can be written in terms of $\delta(\vec{x})$ as:

$$\dot{\delta} + \frac{1}{a}\nabla \cdot \mathbf{v} = 0, \tag{1.17}$$

$$\dot{\mathbf{v}} + H\mathrm{v} + \frac{\nabla P}{\rho a} + \frac{1}{a}\nabla\Phi = 0, \tag{1.18}$$

$$\nabla^2\Phi = 4\pi G a^2 \overline{\rho}\delta, \tag{1.19}$$

where Φ is the gravitational potential, P is the pressure and the spatial derivatives are taken with respect to the comoving coordinates x. Equations 1.17, 1.18 and 1.19 are derived from the continuity, Euler's and Poisson equations, respectively, when working up to first-order in δ.

These three equations can be combined into one second-order equation of δ:

$$\frac{\partial^2\delta}{\partial t^2} + 2H\frac{\partial\delta}{\partial t} - 4\pi G\overline{\rho}\delta = \frac{\nabla^2 P}{\overline{\rho}a^2}. \tag{1.20}$$

Under the assumption of pressureless fluid and neglecting entropy perturbations, this equation is reduced to:

$$\frac{\partial^2\delta}{\partial t^2} + 2H\frac{\partial\delta}{\partial t} - 4\pi G\overline{\rho}\delta = 0. \tag{1.21}$$

The second term in 1.21 acts as a friction term and is slowing down the growth of perturbations due to the Hubble expansion, while the last term is the gravitational source term. We can solve this equation by requesting $\delta(\vec{x}, t) = D(a)\delta_\mathrm{in}(\vec{x})$, where $D(a)$ is known as the linear growth factor and $\delta_\mathrm{in}(\vec{x})$ is the initial overdensity field. The solution for a given initial field in the matter dominated era is given by two independent growth factor functions:

$$\begin{aligned} &D_+(t) \propto a(t), \\ &D_-(t) \propto a(t)^{-3/2}. \end{aligned} \tag{1.22}$$

We call D_+ and D_- the growing and decaying mode, respectively. The final solution is given by the linear combination of the two modes:

$$\delta(\vec{x}, t) = A(\vec{x})D_+(t) + B(\vec{x})D_-(t). \tag{1.23}$$

Relaxing the constraint of $\Omega_\mathrm{M} = 1$ and considering the full ΛCDM, the linear growth factor is given by:

$$D_+(z) = \frac{5}{2}\Omega_m H(z)\int_z^\infty dz\frac{1+z}{H(z)^3}. \tag{1.24}$$

We conclude this section by pointing out that in linear theory the growth of initial perturbations is self-similar. The overdense regions ($\delta > 0$) will grow, and the underdense ($\delta < 0$) will become more underdense. At some point δ will become a quantity larger than 1 where the linear PT breaks down.

1.2.2 Correlation function

In order to describe and model the observed clustering of matter, galaxies or gas, it is useful to look at the correlation function of the overdensity field of these tracers. Assuming that the overdensity is a Gaussian random field with mean value of zero, all the information is contained in the 2-point (auto)correlation function. However, the non-linear structure formation causes a part of this information to move into higher n-point correlation functions, though the majority of the information is still in 2-point correlation function[4]. The correlation function $\xi(r)$ can be defined as an excess probability of finding a pair of galaxies at a given separation r_{12}:

$$dP = n^2 \delta V_1 \delta V_2 [1 + \xi(r_{12})], \tag{1.25}$$

where n is the average number density, δV_i are the volume elements at separation r_{12}. Because of the assumed homogeneity and isotropy, the correlation function is invariant to translations and rotations, and thus must only depend on the relative distance. The correlation function can also be defined as $\langle \delta(x_1)\delta(x_2)\rangle \equiv \xi(r_{12})$.

1.2.3 Power spectrum

The overdensity field $\delta(\vec{x})$ can be also transformed into the Fourier space. The quantities in Fourier space have the same information as the ones in the configuration space (space of $\vec{x}$), though sometimes it is easier to work in rather that the other. We will use the following notation to move in and out of Fourier space in the example of $\delta(\vec{x})$ and $\delta(\vec{k})$:

$$\delta(\vec{x}) = \int \frac{d^3k}{(2\pi)^3} \delta(\vec{k}) e^{-i\vec{k}\cdot\vec{x}}, \ \delta(\vec{k}) = \int d^3x \delta(\vec{x}) e^{i\vec{k}\cdot\vec{x}}. \tag{1.26}$$

where the wavenumber (or mode) k is related to the wavelength λ of the fluctuation by $k = 2\pi/\lambda$.

The power spectrum $P(k)$ of the overdensity field is defined as:

$$\left\langle \delta(\vec{k}_1)\delta^*(\vec{k}_2) \right\rangle \equiv (2\pi)^3 \delta_{\rm D}(\vec{k}_1 + \vec{k}_2) P(k_1), \tag{1.27}$$

and can be identified as the Fourier transform of the correlation function:

$$\xi(r) = \int \frac{d^3k}{(2\pi)^3} P(k) e^{-i\vec{k}\cdot\vec{x}}. \tag{1.28}$$

One advantage of working at the level of the power spectrum is that it is the quantity predicted by theories of inflation and structure formation. Another one is that at the level of the linear density field, modes evolve independently and the covariance matrix is purely diagonal. As the density filed becomes more non-linear with time, modes become coupled (see [15] and references therein).

---

[4] For brevity, from now on we will call the 2-point correlation function just the correlation function as in this thesis we do not deal with the higher correlation functions.

1.2.4 Linear matter power spectrum

As we saw in section 1.2.1, we can use the linear theory to predict the evolution of an overdensity field. However, those results hold only for the matter field. We know that the Universe has went through three main epochs. At early stages the radiation was the dominant component, followed by the matter domination and finally the DE era we are living in now. In order to compute the matter power spectrum in any of these stages, one has to account for all the various components present (radiation field, baryons, DM, DE) and evolve the full fluid equations. These different components are interacting and thus affect the evolution of growth of structures. For example, at early epochs the baryons are tightly coupled with the radiation field which is preventing the growth of baryon structures. On the other hand, DM perturbations grow in this regime. After recombination the perturbations in DM are different that the ones of baryons. Baryons infall into the gravitational wells of DM and form galaxies. On the other hand, the perturbations due to sound waves in the photon-baryon fluid remains imprinted in the distribution of baryons which in turn affects the distribution of DM also.

We also need to know the primordial spectrum of perturbations. Even before the inflationary paradigm emerged as a way to explain the initial conditions, it was proposed that the power spectrum of initial perturbations in the gravitational potential needs to have a scale-invariant power spectrum $k^3 P_\Phi \propto const.$ To describe any deviation from the scale-invariance it is conventional to write the primordial power spectrum as $P_\Phi \propto k^{n_s-1}/k^3$ where n_s is the spectral index with a value close to unity. A case of $n_s = 1$ is known as the Harrison-Zeĺdovich power spectrum. Inflationary paradigms also predict a scale-free spectrum with a small deviation from $n_s = 1$.

Since we are interested in the matter power spectrum, a way to relate the potential perturbations power spectrum – $P_\Phi \propto \langle|\Phi(k)|^2\rangle$ to the matter power spectrum – $P_\mathrm{m} \propto \langle|\delta(k)|^2\rangle$, is by using the Poisson equation which relates the gravitational potential and the matter overdensity field $\delta(k) \propto k^2\Phi(k)$. Finally, we see that the predicted initial matter power spectrum has the following scale dependence:

$$P_\mathrm{m}(k, z) \propto k^{n_\mathrm{s}} \tag{1.29}$$

Spectral index n_s is also one of the main cosmological parameters in the ΛCDM. Current measurement of this parameter from the Planck CMB experiment is $n_\mathrm{s} = 0.968 \pm 0.006$ [6], shows a small deviation from unity, as predicted by inflationary paradigm.

Another important effect to be taken into account is the effect of the size of the Hubble horizon on the growth of different modes. The Hubble horizon ($\propto H^{-1}$) grows with time. At early times, all relevant scales are outside the horizon and do not grow. Once the mode enters the horizon, it starts growing depending on the era in which this happens. Since the growth is different in eras in which radiation or matter dominates, modes entering at different times will have different histories. This information is encoded in the transfer function $T(k, z)$.

We write the linear matter power spectrum as a function of scale and time as:

$$P_\mathrm{m}(k, z) \propto D^2(z) T^2(k, z) k^{n_\mathrm{s}} \tag{1.30}$$

In order to obtain the prediction of the power spectrum by including all different components, one needs to solve the full Boltzmann equations numerically. There are two publicly available numerical codes that are mostly used: `CAMB` [16] which is the one used in this thesis and `CLASS` [17].

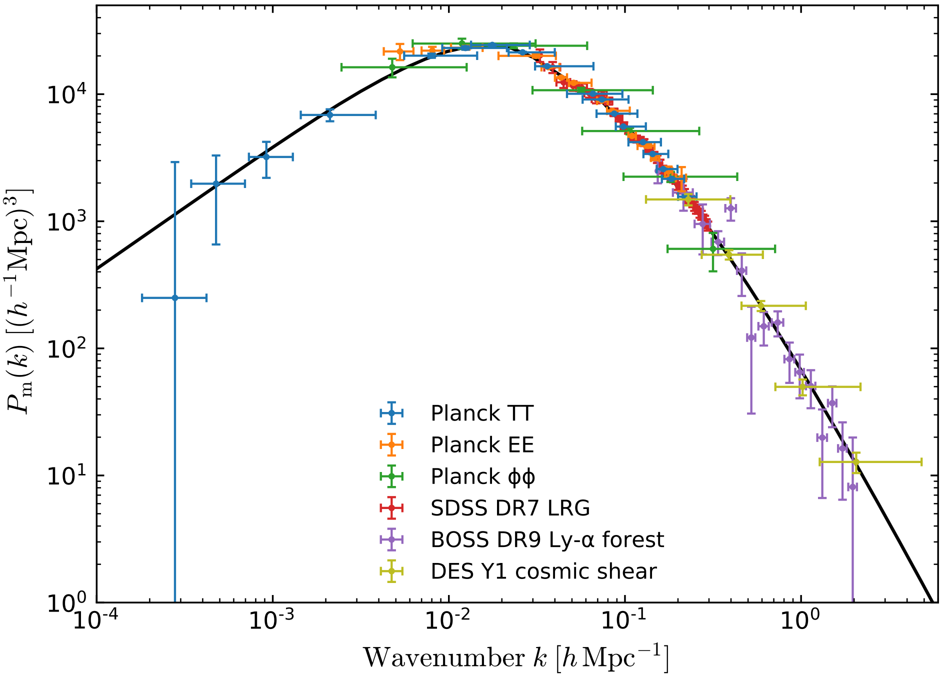

FIGURE 1.1: The linear matter power spectrum (at $z = 0$) obtained from different probes along with the best-fit model (solid black line). This plot is taken from [18] where all the details about datasets used can be found.

The shape of the power spectrum is very sensitive to the cosmological parameters and different cosmological probes allow us to measure it at different scales. Furthermore, we can also measure the temporal evolution of the shape at different redshifts with galaxy surveys and lensing studies. By combining these different probes we have been able to measure the cosmological parameters to a level of high precision we have today. To illustrate this, in figure 1.1 we show the reconstructed linear matter power spectrum at $z = 0$ obtained using different cosmological probes [18]. We can see how different experiments probe different scales and the results agree remarkably well across over a huge range of scales with the linear theory prediction for the standard ΛCDM model.

1.2.5 Galaxy power spectrum

While in the previous sections we focused on the power spectrum of the total matter field (baryons + dark matter), in reality this is not something we can directly observe. What we do observe are tracers of the underlying matter field which can be: galaxies, gas in emission or absorption lines in distant quasar spectra etc. The task of inferring the clustering properties of the total matter field from the observed tracers is complicated due to the unknown relation between the clustering properties of the two fields. A simple approximation is that galaxies follow the distribution of total matter up to factor called the *linear galaxy bias* $b_{\rm g}$:

$$P_{\rm g}(k,z) = b_{\rm g}^2(z)P_{\rm m}(k,z). \tag{1.31}$$

This assumption holds true only at large scales and it becomes much more complicated when going to smaller scales [19].

1.2.6 Redshift space distortions

So far we have assumed that we can observe objects at their comoving coordinates $\vec{x}$, also known as the *real space* coordinates. However, what we measure is actually the redshift from which we infer the positions. Furthermore, objects have peculiar velocities, i.e. they are not moving relative to us only due to the Hubble expansion (aka Hubble flow). If an object has a peculiar velocity and is moving towards us, we will measure a lower redshift and naively conclude that the object is closer to us. The reverse holds for objects moving away from us. On large scales we expect the peculiar velocities to be uncorrelated and the main contribution are the coherent flows of matter. These coherent flows will increase the clustering on large scales as they tend to move galaxies towards the overdense regions. On the contrary, going to small scales, for example the surroundings of a massive DM halo, the peculiar velocities become more random and large. This randomness will erase any correlation at those scales. Thus we expect on large scales to have more power, while at small scales to have less power with respect to the real space. These effects of peculiar velocities on the clustering are called redshift space distortions (RSD) we have to account for them in our model of the observed overdensity field. We will first start showing the effect on large scales and then move to modelling on smaller scales.

We start by differentiating the proper distance $\vec{r}(t) = a(t)\vec{x}(t)$ with respect to time:

$$\vec{v}(t) = \frac{d\vec{r}(t)}{dt} = H(t)\vec{r}(t) + \vec{u}(t), \tag{1.32}$$

where we have defined the peculiar velocities field $\vec{u}(t) = a(t)\frac{d\vec{x}}{dt}$. We see that in the radial direction which we denote with $\hat{z}$ throughout the thesis (though not to be confused with the redshift): $v_z = Hr_z + u_z = H(r_z + u_z/H)$. Thus we can define a new comoving coordinate system that includes the effect of the peculiar velocities: $\vec{s} = \vec{r} + u_{\hat{z}}\hat{z}/H$. We will call this new system the *redshift space* and use subscripts s for quantities dealing with redshift space, while the one with only $\vec{r}$ is called the *real space* with subscript r.

By using the linearised continuity equation 1.17, moving to Fourier space and under the assumption that the velocity field is irrotational, we can express the peculiar velocity field:

$$v_{\mathbf{k}} = -i\frac{aH}{k}f\delta_{\mathbf{k}}, \tag{1.33}$$

where f is the growth rate defined

$$f \equiv \frac{d\ln\delta}{d\ln a}. \tag{1.34}$$

Another requirement is that the number of objects in real and redshift space must be conserved. Using this, the redshift space overdensity field can be expressed in terms of the real space overdensity field:

$$\delta_s(\vec{k}) = (1 + \beta\mu^2)\delta_r(\vec{k}), \tag{1.35}$$

where $\mu = k_{\|}/k$ is the cosine of the angle between $\vec{k}$ and the line-of-sight, while we define $\beta \equiv f/b$ with b being the linear bias of a tracer considered.

The linear prediction for the matter power spectrum in redshift space is then [20, 21]:

$$P_s(k,\mu,z) = (1 + \beta(z)\mu^2)^2 P_r(k,z). \tag{1.36}$$

One way to parametrize the growth rate is to use an approximation $f(z) = \Omega_{\mathrm{M}}(z)^{\gamma}$ [22], where γ is the growth index. In the case of GR+ΛCDM Universe the growth index is equal to 0.545 [23] and we will use this approximation throughout this thesis. From equation 1.36 we can see that the amplitude of the power spectrum has an angular dependance but that it also depends on the growth rate. Thus by measuring the angular dependence of the power spectrum we can constrain the growth rate but also test the validity of GR [24].

Moving to small scales the physics of the RSD becomes more complicated since there the non-linear effects are larger. What is generally expected is the suppression of the power spectrum in the radial direction due to the random peculiar velocities inside collapsed objects, i.e. clusters and halos. Due to these large velocities, objects in redshift space appear stretched along the line-of-sight, an effect called the *fingers-of-god* (FoG). This effect is difficult to model and one of the models used in the literature is the Gaussian streaming model [21]:

$$F(k, \mu, \Sigma_{\mathrm{s}}) = e^{-(k\mu\Sigma_{\mathrm{s}})^2} \tag{1.37}$$

where Σ_{s} is the streaming scale describing the dispersion of random peculiar velocities along the line-of-sight direction that washes out the information on small scales. In principle this scale is also redshift dependant, though in chapter 3 we will fix it to a certain value.

The total effect of RSD on the power spectrum is then given by:

$$P_s(k, \mu, z) = (1 + \beta(z)\mu^2)^2 F(k, \mu, \Sigma_{\mathrm{s}}) P_r(k, z). \tag{1.38}$$

In contrast to real space, the redshift space power spectrum introduces anisotropy. This anisotropy can then be used to measure the growth rate. In chapter 4 we will see how well the futuristic 21cm IM surveys (discussed in chapter 2) can constrain the growth rate at $2.5 \leq z \leq 5$.

One can then decompose the 2D power spectrum into Legendre polynomials $L_\ell(\mu)$:

$$\begin{aligned} P(k, \mu, z) &= \Sigma_{\ell=0}^{\infty} P_\ell(k, z) L_\ell(\mu) && (1.39) \\ P_\ell(k, z) &= \frac{2\ell + 1}{2} \int_{-1}^{1} P(k, \mu, z) L_\ell(\mu) d\mu && (1.40) \end{aligned}$$

and also get the multipoles of the 2D correlation function:

$$\xi_\ell(r) = i^\ell \int \frac{k^3 d\log k}{2\pi^2} P_\ell(k) j_\ell(kr), \tag{1.41}$$

where j_ℓ is the spherical Bessel function of the l-th order. Without the FoG effect, the only non-zero multipoles in linear theory are the monopole ($\ell = 0$), quadrupole ($\ell = 2$) and hexadecapole ($\ell = 4$). All other even multipoles are zero, along with all odd multipoles. Adding the FoG term, higher multipoles also become non-zero, while the odd ones still remain zero.

1.2.7 Baryon Acoustic Oscillations

We now turn our attention to the behaviour of the photon-baryon fluid prior to recombination. In this epoch, the photons are coupled to free electrons by the Compton scattering and together act as a single fluid. The sound speed in such a fluid is given by:

$$c_{\mathrm{s}} = \sqrt{\frac{1}{3(1+R)}}, \tag{1.42}$$

where $R = 3\rho_{\mathrm{b}}/(4\rho_{\gamma})$. We see that this reduces to the standard $c_s = \sqrt{1/3}$ when baryons are absent. With baryons present and contributing with their mass to the fluid, the sound speed becomes lower.

Now let us assume an initial overdensity in some region of this fluid. The increased energy density of photons in this region will exert radiation pressure on the surrounding medium causing an spherical acoustic wave to propagate. This wave will travel with a speed c_{s} until the decoupling of the photons. The distance which such a wave can travel in time from the Big-Bang to the decoupling is given by the sound horizon:

$$r_{\mathrm{s}}(t_{\mathrm{dec}}) = \int_0^{t_{\mathrm{dec}}} \frac{c_{\mathrm{s}}(t)}{a(t)} dt. \tag{1.43}$$

From this simple consideration we expect that, by the time recombination ends, there will be an overdensity of baryons at a scale of r_{s} which for a standard cosmology is $\approx 150\,\mathrm{Mpc}$. This will imprint a peak in the correlation function at the scales of the sound horizon at recombination known as the baryon acoustic oscillations (BAO) peak. In Fourier space at a given cosmic time this single acoustic wave will produce harmonic oscillations. Thus in the power spectrum we expect to observe oscillatory behaviour at wavenumbers corresponding to r_{s}.

Let us consider the total matter field. Perturbations in DM were growing even before recombination and after recombination happens, baryons will interact with DM gravitationally and will infall into the potential wells. This will cause baryons to infall into the overdensities of DM, but also the DM to be affected by the distribution of baryons. Eventually, galaxies will start forming around centres of dark matter halos. Statistically there will be a scale at which the number of galaxies will be more correlated. Thus we expect the BAO peak to be imprinted in both the matter field and the galaxies.

While this simplified explanation of the origin of the BAO considered a single initial overdensity, in reality there will be a continuous field of over- and underdensities and the acoustic waves in the fluid will be a superposition of many different waves. The end effect remains unchanged – statistically there will be an increased probability of finding galaxies at a separation of $\approx 150\,\mathrm{Mpc}$.

One of the advantages of BAOs is its robustness against systematic effects, that can impact more strongly other cosmological observables such as those making use of the shape of the galaxy clustering pattern. BAOs produce a peak in the correlation function at $r \sim 100\ h^{-1}\mathrm{Mpc}$ while in the power spectrum it produces a set of wiggles at $k \gtrsim 0.01\ h\mathrm{Mpc}^{-1}$.

This feature has been first seen in the power spectrum of the temperature fluctuations in the CMB. It appears as a series of peaks in the CMB temperature angular power spectrum. The first peak corresponds to the angle of the sound horizon at the time of recombination and was first detected at the end of the 20th century with instruments TOCO [2], Boomerang [3] and Maxima [4]. Other peaks have been detected in the following years with other more advanced satellite instruments, e.g. WMAP, and the Planck satellite has measured 18 acoustic peaks in the temperature and polarization angular power spectra [18].

We thus expect this feature, imprinted at very early times, to be present in the matter density field even today. In order to detect it one needs to find a tracer (e.g. galaxies, Lyα forest or the 21cm emission) of the matter density field and compute its correlation function in a given redshift bin. The position of the BAO peak in the correlation function corresponds to the sound horizon which is well constrained with the CMB measurements. The BAO peak can be measured in the transverse direction or along the line of sight. By knowing the subtended angle of this known scale one can determine the comoving angular diameter distance $D_\mathrm{A}(z)$ up to the given redshift. On the other hand using the l.o.s. separation one can measure the Hubble parameter as a function of redshift $H(z)$. By doing this at several redshift bins, one can measure distances as a function of redshift even at redshifts where other ways of measuring distances fail. However, seeing this feature in the LSS data required deeper galaxy surveys that probe larger volumes since the scale of the BAO peak is very large. Eventually, in 2005 using the data gathered by 2dF Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS), the BAO has also been detected in the galaxy correlation function [25, 26]. Consequently, the BAO position has been measured more precisely and at different redshifts with better galaxy surveys. Nowadays, BAOs are routinely used for quantitative cosmological investigations (e.g. [27, 7, 28, 29]).

1.2.8 Non-linear effects on the BAO

Since the BAO scale is large, we would expect linear theory to be sufficient enough for the description of the evolution of the BAO even at late times, i.e. low redshifts. Unfortunately, non-linear gravitational evolution produces a damping, broadening [30, 31, 32] and also induces a shift [33, 34, 35] in these features that 1) makes more difficult the task of measuring the sound horizon and 2) could bias the inferred cosmological quantities. While the position of the BAO peak is almost unaffected by the non-linearities, the broadening of the BAO peak increases with time and the effect is largest at recent times, i.e. low-redshifts. In figure 1.2 we show how the BAO peak is affected at different times, i.e. redshifts both at the level of the matter correlation function and the power spectrum. We see that at high redshifts the BAO is peak is more pronounced and the power spectrum wiggles are close to the linear theory ones. With decreasing redshift the wiggles in the power spectrum get damped while the BAO peak gets broader. These non-linear effects on the BAO proved to be modelled fairly well as an additional exponential damping term of the galaxy power spectrum in redshift-space in the following way [36]:

$$P_s^{\mathrm{nl}}(k,\mu) \propto P_s^{\mathrm{lin}}(k,\mu) \exp\left(-\frac{k_\perp^2 \Sigma_\perp^2}{2} - \frac{k_\parallel^2 \Sigma_\parallel^2}{2}\right), \tag{1.44}$$

where $\Sigma_\perp$ and $\Sigma_\parallel$ are the non-linear damping scales in the transverse and the radial direction, respectively. The damping scales roughly correspond to the rms of the displacements and depend on the redshift.

The main reason for the BAO damping has been understood to be caused by the bulk motions on large scales, rather than the random motions of galaxies on small scales [37]. A technique to overcome, or at least to mitigate this problem, has been recently developed and it is called *BAO reconstruction* [38, 15, 37, 39, 40, 41]. The underlying idea is that non-linear gravitational clustering on BAO scales can be accurately modeled by perturbation theory, and in particular, with Lagrangian perturbation theory: the Zel'dovich approximation [42, 43] in its simplest version. In figure 1.3 we show the effect of applying this method on real data. The BAO peak clearly gets sharper and the wiggles more pronounced.

In chapter 3 we will review and summarise the main points of the standard BAO reconstruction method. Then we will present a detailed analysis of a mesh-grid BAO

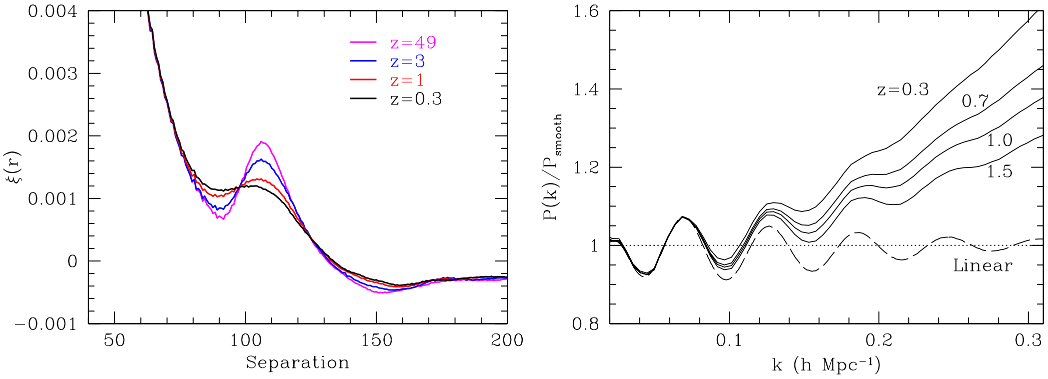

FIGURE 1.2: The effect of non-linearities on the shape of the BAO peak as a function of redshift (taken from [44]). Left panel shows the redshift-space matter correlation function obtained at four different redshifts using numerical simulations. The right panel shows the matter power spectrum divided by the smooth, no-wiggle power spectrum to show more clearly the acoustic oscillations.

reconstruction method with pixels. While the discussion in this section was mainly based on using galaxies as tracers and the information from LSS have mostly came from galaxy surveys, there are other tracers available. For example, the 21cm emission (see chapter 3) from the neutral hydrogen is thought to be very promising tracer of the underlying matter density field.

1.2.9 Massive neutrinos

In the Standard Model of Particle Physics, neutrinos are massless and chargeless particles interacting through the weak force and there are three flavours of neutrinos – ν_e, ν_τ and ν_μ. However, observations of neutrinos from the Sun and from our atmosphere have shown that neutrinos can change their flavours through a process called *neutrino oscillations* [46, 47]. This result can only be explained if neutrinos have non-zero mass. These experiments measure only the squared mass difference between neutrinos and hence the absolute mass difference information is unknown. There are two possible hierarchy schemes that are both consistent with the observations – normal ($m_1 < m_2 < m_3$) and inverted ($m_1 \simeq m_2 \gg m_3$). Neutrino oscillations experiments can only put the lower bounds on the total neutrino mass – Σm_ν and they indicate two possible scenarios for the sum of neutrino masses, the normal hierarchy where $\sum m_\nu \geq 0.06$ eV and the inverted one where $\sum m_\nu \geq 0.1$ eV [48].

Massive neutrinos have several effects on the cosmological observables. While these effects are in general small, with current and future surveys they are important for two main reasons. First is that by measuring these effects we can put upper bounds or potentially measure their total mass. Second is that in the era of precision cosmology these effects, if unaccounted for, can bias the results of the cosmological parameters. There are several excellent review papers on the effects of massive neutrinos in cosmology [49, 47, 50] and here we only briefly describe the main points and effects at the linear level of perturbations used in this thesis.

In the cosmological context, neutrinos were generated in the early Universe and were once kept in thermal equilibrium with the cosmic plasma. Due to their weak interaction, they decoupled from the plasma early on when the temperature of the Universe was $T \sim 1$MeV and behaved like ultra-relativistic particles. At that stage neutrinos are contributing to the radiation energy density along with photons and

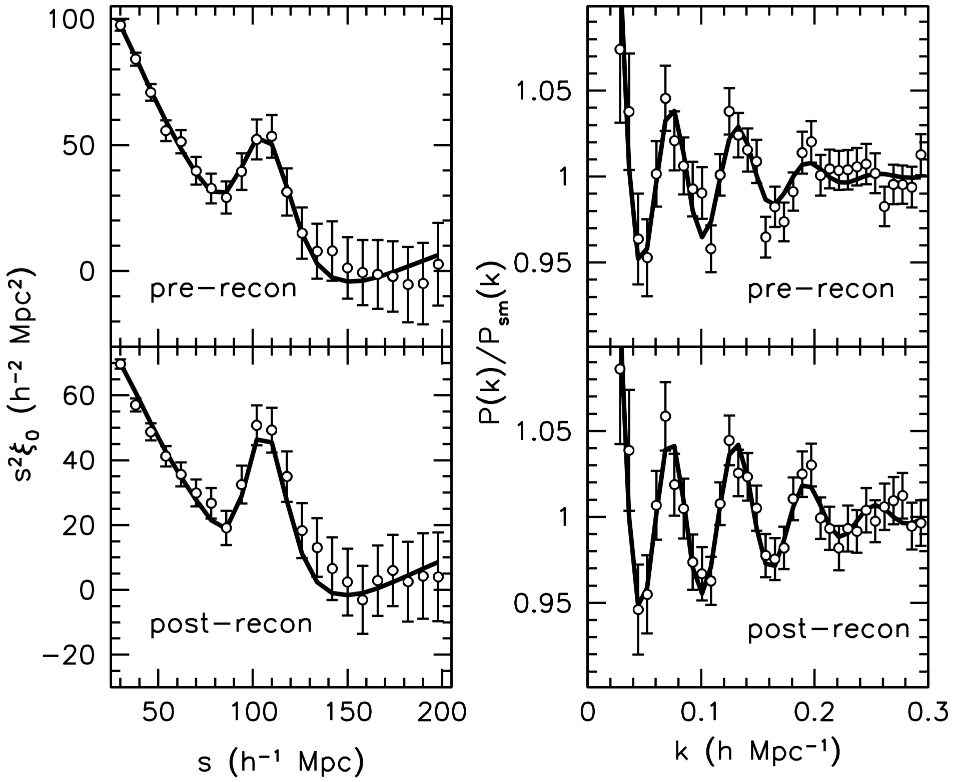

FIGURE 1.3: The effect of BAO reconstruction on the shape of the measured BAO peak in the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS) at the level of the correlation function (left panels) and the power spectrum (right panels) before (upper panels) and after (lower panels) applying the reconstruction method (taken from [45]).

their perturbations do not grow. Once their thermal energy drops below their mass m_ν, they become non-relativistic. This happens at redshift $z_{\rm nr}$ given by [51]:

$$1 + z_{\rm nr}\left(m_\nu\right) \simeq 1890 \left(\frac{m_\nu}{1\mathrm{eV}}\right). \tag{1.45}$$

After that neutrinos act as matter component and contribute to the total matter energy density with Ω_ν [49]:

$$\Omega_{\rm m} = \Omega_{\rm c} + \Omega_{\rm b} + \Omega_\nu, \quad \Omega_\nu = \frac{\sum m_\nu}{93.14h^2\mathrm{eV}}. \tag{1.46}$$

Even after becoming non-relativistic, neutrinos have large thermal velocities $v_{\rm th}$ which introduce a free-streaming scale defined as [49]:

$$\lambda_{\rm FS}\left(m_\nu, z\right) = a(z)\left(\frac{2\pi}{k_{\rm FS}}\right) \simeq 7.7\frac{H_0(1+z)}{H(z)}\left(\frac{1\mathrm{eV}}{m_\nu}\right) h^{-1}\mathrm{Mpc}, \tag{1.47}$$

where $k_{\rm FS}$ is given by [49]

$$k_{\rm FS}\left(z_{\rm nr}\right) \simeq 0.018\,\Omega_m^{1/2}\left(\frac{m_\nu}{1\mathrm{eV}}\right) h\mathrm{Mpc}^{-1}. \tag{1.48}$$

These remnant thermal velocities have a physical effect on the neutrino and consequently on the matter perturbations. Due to free-streaming neutrinos erase density

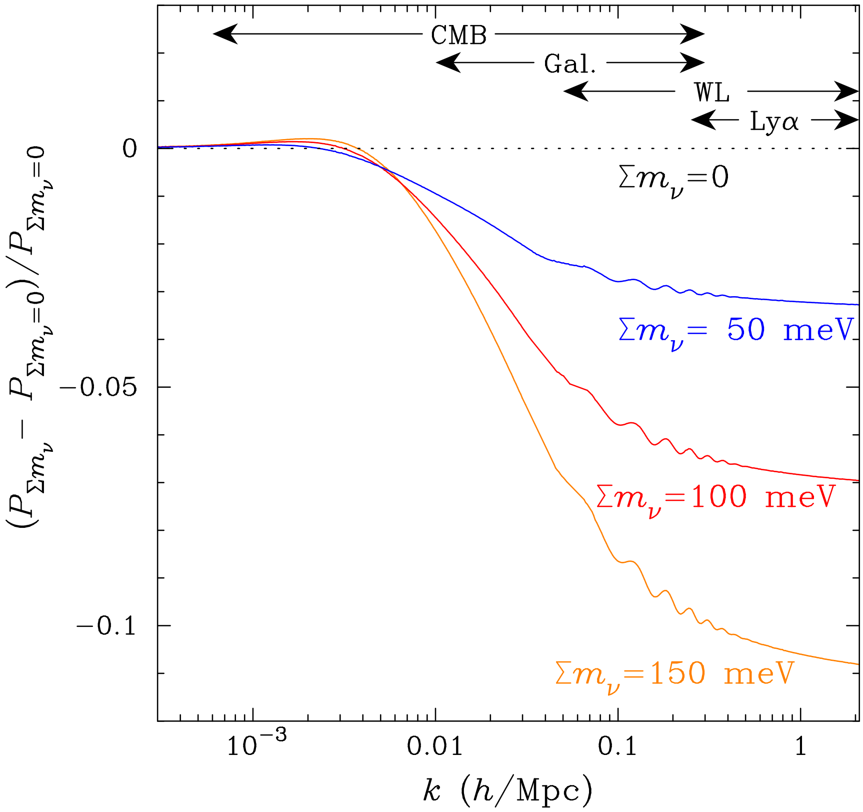

FIGURE 1.4: The effect of massive neutrinos on the matter power spectrum as a function of scale for different values of Σm_ν (taken from [50]).

fluctuations below $\lambda_{\rm FS}$, while on larger scales they behave as CDM. This has an effect on the shape and the amplitude of the power spectrum we observe. In figure 1.4 (taken from [50]) we show this effect of massive neutrinos on the total matter power spectrum as a function of scale for different values of Σm_ν. This effect is shown as a fractional difference between power spectra with non-zero and the power spectra with zero neutrino mass. Also shown are the ranges of scales different cosmological probes are sensitive to. Constraining neutrino masses is one of the main goals of current and future cosmological surveys by combining probes of large scales, like CMB anisotropies, with data from intermediate/small scales. Currently the best constraints are $\sum m_\nu < 0.11$ eV [52], but we expect the next generation of galaxy surveys and CMB probes to significantly improve those bounds. The benchmark to achieve for the error on neutrino masses is set by the laboratory measurements of neutrino oscillations.

In chapter 4, section 4.3.1, we will study further the effects of neutrinos in the case of the expected signal from the neutral hydrogen in the post-reionization Universe and show how the constraints on their total mass can be improved when adding information from 21cm IM surveys covering the redshift range of $2.5 < z < 5$.

Chapter 2

21cm Cosmology

Hydrogen atom is the most abundant element in the Universe both by number and the mass. After it is produced in the first three minutes of Universe, it comes in three forms: ionized (HII), neutral (HI) and molecular (H_2) hydrogen. Early on hydrogen is completely ionized by the radiation field. After recombination ($z_{\rm rec} \approx 1100$) happens it is mostly in its neutral form. As the gas clouds start to condense and cool, molecular hydrogen is also formed. This lasts through the dark ages and until the first stars form ($400 \lesssim z \lesssim 20$). During reionization the hydrogen neutral fraction $x_{\rm HI}$ is decreased, due to the ultraviolet (UV) and X-ray radiation from the new born stars, reaching almost $x_{\rm HI} \approx 0$ when reionization ends. After reionization ends we can assume most of the neutral hydrogen lives in halos and galaxies, where it remains self-shielded from the background UV ionizing radiation [53]. In this thesis we will focus on post-reionization HI as a probe of the LSS.

HI has a characteristic spin-flip transition resulting in the emission/absorption of a spectral line at $\lambda_{21} = 21.11\,{\rm cm}$ or $\nu_{21} = 1420.4\,{\rm MHz}$. The typical temperatures of the clouds hosting neutral hydrogen in galaxies is on the order of $10 - 100{\rm K}$ [54], larger than the energy difference between the hyperfine levels while lower than the energy needed for the Lyα transition. This makes the excited level more populated and hence gives rise to emission of 21cm photons. This hyperfine transition has a very low probability and in turn makes this line very narrow. Fortunately, there is still a lot of HI inside galaxies which makes it possible to observe this line in emission. Owing to the small line width it makes it easier to measure the velocities of the emitting source. This is what has been used in order to map the HI distribution and velocities in our, but also in other nearby galaxies, by using the Doppler shift to infer the radial velocity of the emitting HI. Moving to extragalactic scales, the 21cm line is redshifted due to Hubble expansion $\lambda(z) = (1 + z)\lambda_{21}$. Therefore, by measuring the position of this line from observations we can infer the redshift of the emitting source and perform the 21cm tomography. This can then be used to map out the distribution of HI throughout the Universe by detecting 21cm line at different frequencies, i.e. different cosmic times. In principle, this allows us to probe most of the observable Universe: from the dark ages to current day (for a review see [55]). Figure 2.1 (taken from [56]) shows the amount of observable volume available through the 21cm tomography compared to CMB and galaxy surveys[1]. However, we are currently drastically limited by issues we will discuss below.

The distribution of HI roughly follows the distribution of gas and stars in galaxies, though it also extends to larger distances from the galactic centre [54]. Numerical simulations show that almost all the HI in the post-reionization Universe is inside

---

[1] Note that this figure is of 2008. Meanwhile galaxy surveys (red) did extend to higher redshifts ($z \approx 2$) and future surveys will go even higher. The main point however still remains, there is a large portion of the observable Universe we are still not probing and that is in principle available through the 21cm tomography.

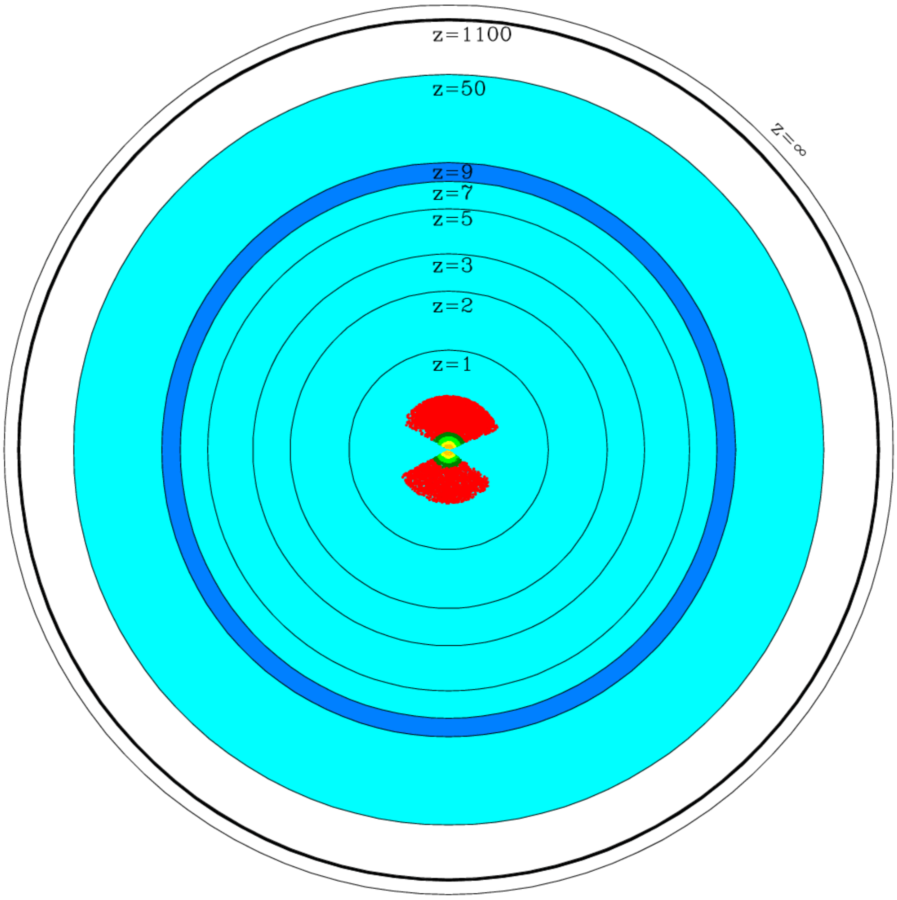

Figure 2.1: Observable volume with the 21cm surveys in cyan (taken from [56]). Also shown is a thin slice probed by the CMB (at $z \approx 1100$) and the galaxy surveys (red). The area of the circles shown is proportional to the comoving volume.

dark matter halos [57, 53]. At low redshift more than 99% of all HI is inside DM halos, while at $z = 5$ this amount drops to 80%. For the rest of this thesis we will assume that all of the neutral hydrogen is inside galaxies. Under that assumption, HI follows the distribution of DM and can be used as a probe of the LSS in a similar fashion as the galaxies.

There are two ways to go about observing the HI. One is to resolve individual galaxies. Similar to galaxy surveys, one uses the radio telescopes to resolve individual galaxies. The redshift comes for free, since it depends on which frequency one observes at. Since the flux is low, one needs long exposure times per galaxy. While plausible in the nearby Universe, going at larger distances one loses the angular resolution and the integration times needed to resolve a galaxy increase drastically due to the increased foregrounds.

Another, more promising technique is the IM [58, 59, 60, 61, 62, 63, 64, 65, 66] – a technique that consists in performing a low angular resolution survey to map the 21cm emission from unresolved galaxies or HI clouds. IM for the Lyα emission line has already been successfully applied for large scale clustering studies at high redshift using BOSS data [67]. The advantages of this technique over traditional methods, such as galaxy redshift surveys, are many. Firstly, IM can survey efficiently very large cosmological volumes. Secondly, IM is spectroscopic in nature and thus offers high radial resolution. Thirdly, it can efficiently probe the spatial distribution of neutral hydrogen from the local Universe to the dark ages. However, its practical implementation is affected by some major challenges which we will discuss below. Many IM surveys (CHIME[2] [68], HIRAX[3][69], Tianlai[4], SKA[5] [65], FAST[6] [70]) are currently being built and taking data with the goal of measuring the BAO scale between $0.4 \lesssim z \lesssim 3$ with unprecedented precision. On the other hand, the Ooty Wide Field Array (OWFA)[7][71] is intended to measure the 21cm HI power spectrum at redshift 3.35 [72, 73, 74, 75].

In this thesis we will focus on using IM as an observing strategy. What one measures with 21cm IM is the flux sourced by the HI emitting galaxies. This flux corresponds to a brightness temperature T_b at various redshifts and angles in the sky and the end result is a series of T_b sky maps. This temperature will depend on the spatial distribution of HI and the observed redshift, while its amplitude will depend on the total amount of HI in the Universe. Similar to the galaxy surveys, we can then use these temperature fluctuations as a probe of the underlying DM field. In the following section we will describe the model we use to predict and describe the relation between the matter power spectrum and the power spectrum of 21cm signal.

2.1 The signal power spectrum

The properties of HI, especially in terms of clustering, are still not fully understood. This is due to a number of reasons: the early stage of IM as an observational probe, the difficulty of detecting the faint 21cm line for a sufficiently large number of sources at high redshifts, and the possibly conflicting evidence [76] coming from observations of low-redshift HI surveys [77, 78], the Lyα forest [79, 80, 81, 82] and the clustering of damped Lyα systems [83, 84]. Understanding HI is vital both for cosmology and astrophysics, since it also plays a vital role in understanding the star formation history [85].

At linear order, the amplitude of the 21cm power spectrum at redshift z is proportional to the product of the HI bias $b_\mathrm{HI}(z)$ and its cosmic abundance $\Omega_\mathrm{HI}(z) = \rho_\mathrm{HI}(z)/\rho_c(z=0)$, where $\rho_\mathrm{HI}(z)$ is the mean HI density at redshift z and $\rho_c(z=0)$ is the critical density at $z=0$. While the value of $\Omega_\mathrm{HI}(z)$ is relatively well constrained in the redshift range $z \in [0,5]$ by several observations [86, 87, 88, 89, 90, 91, 92, 93, 94, 82], the value of the HI bias is poorly known [95, 78, 96].

As stated above, the main assumption is that the distribution of HI follows the distribution of DM. In this scenario at the linear level, the signal power spectrum can be related to the linear matter power spectrum

$$P_{21}(k,z,\mu) = \overline{T}_\mathrm{b}^2(z)\left(b_\mathrm{HI}(z) + f(z)\mu^2\right)^2 P_\mathrm{m}(k,z), \tag{2.1}$$

where the mean brightness temperature as function of redshift is given by (see e.g. [97]):

$$\overline{T}_\mathrm{b}(z) = 180\frac{H_0(1+z)^2}{H(z)}\Omega_\mathrm{HI}(z)h\,\mathrm{mK}. \tag{2.2}$$

---

[2] http://chime.phas.ubc.ca/

[3] http://www.acru.ukzn.ac.za/~hirax/

[4] http://tianlai.bao.ac.cn

[5] https://www.skatelescope.org/

[6] http://fast.bao.ac.cn/en/

[7] http://rac.ncra.tifr.res.in

One important difference with respect to the standard galaxy survey power spectrum is the prefactor $\overline{T}_{\rm b}^2$. This makes any amplitude measurement completely degenerate with the amount of HI in the Universe. In chapter 4 we will show how we can overcome this problem by using external measurements of the amount of HI and the $b_{\rm HI}$.

2.1.1 HI halo model

In order to proceed, we need a model for the abundance and spatial distribution of HI. We use the halo-model formalism [98, 99, 100, 101, 102] which offers an alternative method to predict the abundance and clustering of HI after including two extra ingredients: a relation between total halo mass M_h and the total HI mass inside the halo $M_{\rm HI}$, and a model for the distribution of HI within each halo $\rho_{\rm HI}(r|M_h)$.

We start by assuming that, on average, the HI content of halos depends solely on their mass, and we parametrize the $M_{\rm HI}(M_h)$ relation as [57, 76, 103]:

$$M_{\rm HI}(M_h) = M_0 \left(\frac{M_h}{M_{\rm min}}\right)^{\alpha} \exp\left(-\frac{M_{\rm min}}{M_h}\right). \tag{2.3}$$

where the two free parameters of the model are α, which describes the scaling of $M_{\rm HI}$ with halo mass, and the low-mass cutoff $M_{\rm min}$, which represents the threshold mass needed for a halo to host HI. This mass cut-off is expected, since the gravitational potential of small halos is not deep enough to trigger the clustering and cooling of the hot gas heated by the UV background [57]. The overall normalization M_0 can be immediately associated with the cosmic HI fraction. Both quantities are related through:

$$\Omega_{\rm HI} = \frac{1}{\rho_{\rm c}} \int_0^{\infty} dM_h \, n(M_h) \, M_{\rm HI}(M_h), \tag{2.4}$$

where $n(M_h)$ is the halo mass function.

The HI bias is then fully determined by the model and is given by

$$b_{\rm HI}(z) = \frac{\int_0^{\infty} b(M_h,z) n(M_h,z) M_{\rm HI}(M_h,z) dM_h}{\int_0^{\infty} n(M_h,z) M_{\rm HI}(M_h,z) dM_h}, \tag{2.5}$$

where $b(M_h, z)$ represents the bias of halos of mass M_h at redshift z. For both the theoretical halo mass function and the halo bias we use the fitting functions calibrated from N-body simulations presented in [104]. Similar halo models for the HI are also presented in [105, 103].

The halo model prediction [76, 103, 57] for the HI power spectrum at a fixed redshift is given by the sum of a 1-halo and a 2-halo term:

$$P_{21,{\rm 1h}}(k) = F_2^0(k), \quad P_{21,{\rm 2h}}(k) = P(k) \left[F_1^1(k)\right]^2, \tag{2.6}$$

$$F_\beta^\alpha(k) \equiv \int_0^{\infty} n(M_h) b^\alpha(M_h) \left[\frac{M_{\rm HI}(M_h)}{\bar{\rho}_{\rm HI}} u_{\rm HI}(k|M_h)\right]^{\beta} dM_h, \tag{2.7}$$

where $u_{\rm HI}(k|M_h)$ is the normalized HI density profile in Fourier space.

We can notice that written in this form, $b_{\rm HI} = F_1^1(k=0)$. At large scales the shape of the power spectrum is determined by the 2-halo term, i.e. the linear matter power spectrum multiplied by the standard $b_{\rm HI}^2$. The normalized HI density profile is constant at large scales and the 1-halo term thus acts as a constant contribution to the total power spectrum. This contribution is related to the shot-noise of the HI power spectrum coming from the discreteness of the HI sources is also predicted in the halo model as a limit of the 1-halo term at largest scales, i.e. $k \to 0$, and is given by $P_{\rm SN} = F_2^0(k=0)$. Even though the level of this shot-noise depends on parameters $\alpha, M_{\rm min}$ in equation 2.3, it has been shown that the level of the HI shot-noise is always sub-dominant compared to the signal power spectrum [76]. This is fortunate since it makes the HI power spectrum high signal-to-noise measurement in the post-reionization epoch, in the absence of other noise contributions (instrumental noise, foreground contamination, etc).

On small scales, the clustering of HI is dominated by its distribution within the halo (1-halo term), i.e. the HI density profile. This is very hard to determine from the observations. Instead high-resolution hydrodynamic numerical simulations are needed and are currently being used for this purpose.

The HI halo model, described here, can be used to predict the 21cm signal and the shot-noise. The ingredients needed to make prediction are the free parameters of the $M_{\rm HI}(M_h)$ relation and the HI density profile. In this thesis we will focus and constrain both of these quantities. In the next section we will show results of determining the mean HI density profile by using a large state-of-the-art magneto-hydrodynamic simulations. In chapter 5 we will show the results of measurement of the free parameters in $M_{\rm HI}(M_h)$ relation – $\alpha, M_{\rm min}$, at $z \approx 0$ by using the information from the clustering of HI galaxies from the ALFALFA survey and the HI content of optically-selected galaxy groups found in the SDSS catalog. As stated above, this also allows us to determine the $b_{\rm HI}$ and $P_{\rm SN}$ at $z \approx 0$ which we will also show.

2.1.2 HI density profiles from numerical simulations

As we have seen in the previous section, an important ingredient in describing the spatial distribution of cosmic neutral hydrogen using HI halo models is the HI density profile inside halos. In this section we show the results from [57] where we investigate the spatial distribution of HI inside simulated dark matter halos using the state-of-the-art hydrodynamic simulations TNG100 – part of the IllustrisTNG Project [106, 107, 108, 109, 110]. These simulation follow the evolution of billions of resolution elements representing CDM, gas, black holes and stars in the largest volumes ever explored at such mass and spatial resolution. The simulations are ran using the `AREPO` code [111] which uses tree-particle-mesh method to calculate the gravitational interaction, Godunov method on moving Voronoi mesh for the magneto-hydrodynamics and and sub-grid models to include a range of astrophysical processes (for the full characteristics of these simulations see [57]). The comoving size of the cosmological box evolved in TNG100 is $75\,h^{-1}$ Mpc comoving on the side and the mean mass of the baryonic resolution element is $1.4 \times 10^6\,{\rm M}_\odot$. The dark matter halos have been identified using the Friends-of-Friends algorithm [112] with a linking length parameter $b = 0.2$.

For each halo we have computed the HI mass within narrow spherical shells up to the virial radius, and from them the HI profile. In figure 2.2 we show individual HI profiles for halos in a narrow mass bin at different redshifts with grey lines. The blue lines with errorbars show the mean and the standard deviation of HI profiles from all halos in a given redshift and mass bin, while the red lines show the median. The large halo-to-halo scatter is surprising, and highlights that individual HI profiles, as opposed to dark matter ones, are far from universal.

The scatter is particularly large towards the centers of massive halos which is expected as the halo HI mass function obtained using these simulations also exhibits large scatter in that range as found in [57]. Furthermore, the clustering of halos in that mass range depends significantly on their HI mass. Thus, it is likely that the

HI content of these halos is influenced by their environment, so small halos around more massive ones may have lose or gain a significant fraction of their HI mass due to related effects.

The scatter generally tends to be lower at higher redshifts, and, in particular, is small in halos with masses above $10^{10}\ h^{-1}M_\odot$ at redshift $z = 5$. This is related to the lower scatter found at high redshift in the halo HI mass function in these simulations, $M_{\rm HI}(M_h, z)$ [57]. We speculate that this originates from a reduced role that AGN feedback and environmental gas stripping play at earlier times.

As shown in figure 2.2, in some cases the mean and the median profiles differ substantially. This behavior can be partially attributed to the HI profiles arising from two distinct populations: i.e. HI-rich blue galaxies versus HI-poor red ones [108]. This can clearly be seen in the panel in figure 2.2 corresponding to halos in the mass range $M \in [1-2]\times 10^{12}\ h^{-1}M_\odot$ at $z = 0$. In this range, some halos have a core in their HI profiles while others do not. The reason is that the central galaxy of some halos is experiencing AGN feedback (those with holes in the profile) and are therefore becoming red, while the galaxies in the other halos are not yet being affected by AGN feedback, remaining blue [108].

We find that the HI density profiles of small halos ($M \lesssim 10^{12}\ h^{-1}M_\odot$) increase towards their halo center. We note, however, that the amplitude of the HI profile tends to saturate; i.e. the slope of the profiles declines significantly towards the halo center. For example, at $z = 0$ and $z = 1$ and for halos with masses larger than $10^{11}\ h^{-1}M_\odot$, the mean HI profiles change slope around $\sim 20\ h^{-1}$kpc. This is expected since neutral hydrogen at high densities will turn into molecular hydrogen and stars on short time scales. For higher halo masses ($M \simeq 10^{13}\ h^{-1}M_\odot$) the HI density profile exhibits a hole in the center. This is caused by AGN feedback in the central galaxy of those halos. We note that higher densities in the center of halos can give rise to the formation of molecular hydrogen, that can produce a similar effect [110]. Holes, which extend even further than in groups, are also found in the HI profiles of galaxy clusters, which we however do not show here since there are only a few of them and only at low redshift.

In order to quantitatively investigate what is the effective average HI density profile across different halo masses and redshift, we use the mean measured HI density profile and test two models of HI density that both include an exponential cutoff on small scales.

First we consider a simple power law with an exponential cutoff on small scales — Model 1:

$$\rho_{\rm HI}(r) = \frac{\rho_0}{r^{\alpha_\star}}\exp(-r_0/r), \tag{2.8}$$

where ρ_0 is the overall normalisation, $\alpha_\star$ is the slope parameter and r_0 is the inner radius at which the density drops and the profile changes its slope.

Second, we consider an altered NFW profile [113, 114], found to be a good fit to the multiphase gas distribution at high redshifts in hydrodynamical simulations, with an exponential cutoff on small scales — Model 2:

$$\rho_{\rm HI}(r) = \frac{\rho_0 r_s^3}{(r+3/4r_s)(r+r_s)^2}\exp(-r_0/r), \tag{2.9}$$

where ρ_0 is the overall normalisation and r_s is the scale radius of the HI cloud. In both cases the overall normalisation — ρ_0, is fixed such that the volume integral of the model density profile integrated up to the virial radius of a given halo matches the mean total HI mass obtained from the density profile found in simulations (blue lines in figure 5). We are then left with two free parameters for each model: $\{\alpha_\star, r_0\}$ and

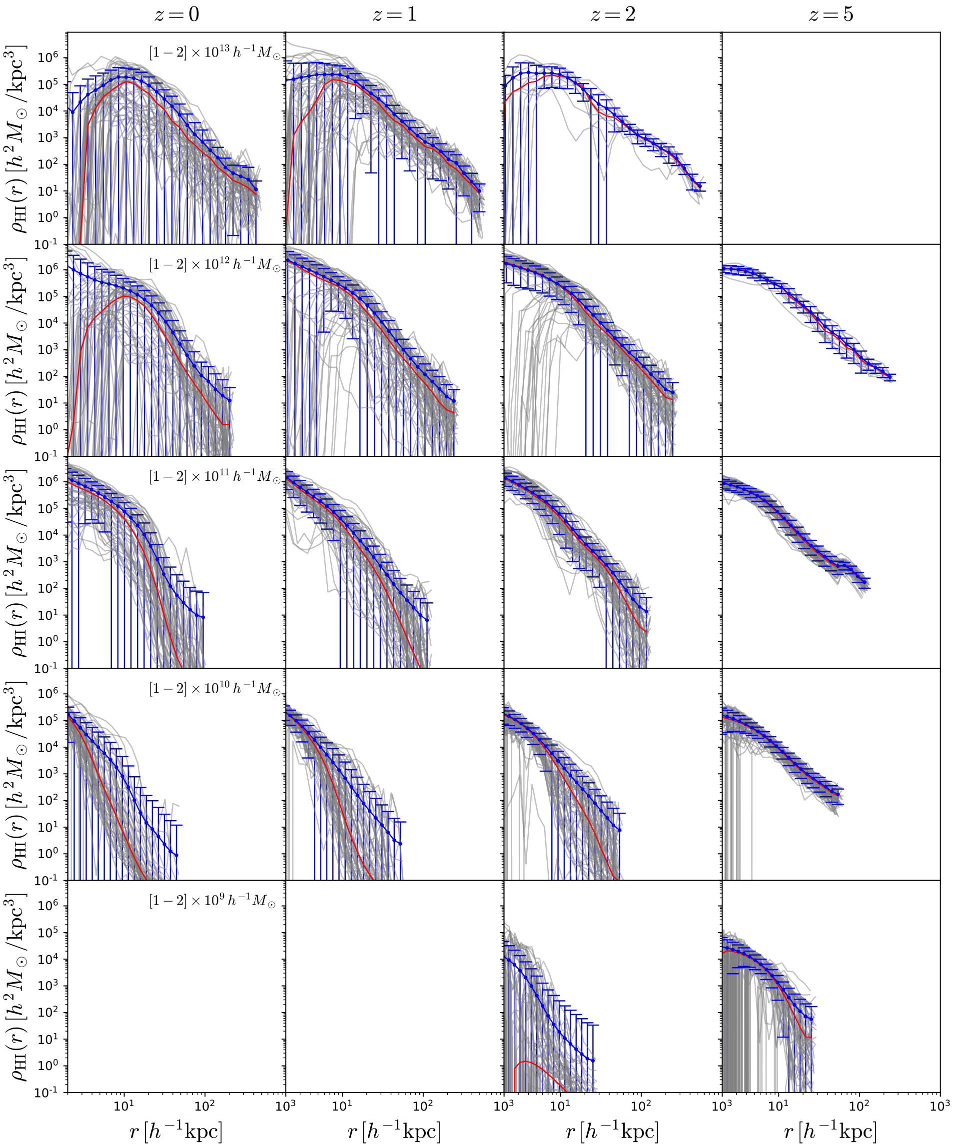

FIGURE 2.2: Density profiles of HI for halos of different masses in different rows (see labels in the left column) at redshifts $z = 0$ (left), $z = 1$ (middle-left), $z = 2$ (middle-right) and $z = 5$ (right). In each panel we display up to 50 individual profiles (grey lines), the mean profile and the standard deviation (blue lines) and the median profile (red lines). Empty panels correspond to situations with either no halos (top-right) or with halos far below the cutoff mass $M_{\rm min}$. In contrast to dark matter, HI density profiles are not universal, and they exhibit, in most of the cases, a very large scatter. The HI-H$_2$ transition saturates the amplitude of the profiles in the core, while processes such as AGN feedback create HI holes in the core of the most massive halos. The mean and the median can be quite different, indicating that the distribution is asymmetric. In some cases, that asymmetry is due to the presence of two different populations such as blue and red galaxies.

$\{r_s, r_0\}$. We fit these models to the measured mean HI density profiles limiting our analysis only to the scales above $r \geq 2h^{-1}$kpc. For the uncertainties in the density profiles we use the scatter among different galaxies (blue error-bars in figure 5) and assume that these uncertainties are uncorrelated between different scales.

The best-fit values along with the 68% confidence intervals are presented in table A.1. In figure 2.3 we show the best-fit results (solid lines) for the Model 1 and in each panel of this plot we also show the best-fit value of r_s, r_0 and the value of the reduced χ^2. Based on the resulting best-fit χ^2, we find that both Model 1 and 2 are good fits for all the considered redshifts and halo masses, except for the most massive halo bin $M_h = 10^{14}\,[h^{-1}M_\odot]$ at $z = 0$. We find that the difference in the best-fit χ^2 between the two models to be negligible. This is to be expected since the models are rather similar and have the same slope on large scales. In the case of Model 1, we find the HI density profile slope to be consistent with a value of $\alpha_\star = 3$ for all the halo masses and redshifts. The inner radius r_0 depends on the halo mass and is larger for larger halo masses at a fixed redshift, while at a fixed halo mass, it increases with increasing redshift. For example, for halos with $M_h \leq 10^{11}h^{1-}$Mpc and $z \leq 2$, r_0 is below the minimum scale considered and the uncertainties are rather large. In the case of Model 2, we find a similar behaviour. The inferred values of r_0 are consistent between two models, with Model 2 having larger uncertainties which is due to the degeneracy between parameters r_0 and r_s.

We note that other observational and simulation studies have found that the HI surface density profile of galaxies can be reproduced by an exponential profile [115, 116]. Based on these studies, other spherically averaged density models have been used in the literature, e.g. an exponential profile [103]. We find that using an exponential profile for the spherically averaged profile does not reproduce our mean data very well.

In summary, we find that the HI density profiles inside halos exhibit a large halo-to-halo variation, while the average HI density profiles are universal. We will use this result as a motivation for the choice of HI profile in HI halo model in chapter 5 where we try to model the observed HI clustering.

2.2 The foregrounds

The major challenge in detecting the HI signal is the presence of strong and complex radio foregrounds [117, 118, 119, 120, 121] which are several orders of magnitude larger than the signal we are after. This is illustrated in figure 2.4 (taken from [122]) where different simulated foreground components are shown as a function of frequency along LOS at different galactic latitudes.

At the frequencies of HI emission in the post-reionization epoch there are several foregrounds that contribute to the measured radio flux. These foregrounds include: Galactic synchrotron radiation, Galactic and extra-galactic free-free emission and the emission of the extra-galactic point sources.

Without dwelling into details of these foregrounds, we state that most of the foreground components have a smooth frequency dependence in contrast to the frequency dependence of the HI emission. Thus, the hope is that one can remove these smooth foregrounds and extract the non-smooth part coming from the HI. In a way this is a similar situation to the one of early CMB measurements and the idea of measuring the CMB polarization. The CMB temperature signal is weaker than the synchrotron and dust emission at most of the frequencies, while the fluctuations in the polarization are even weaker and were considered to be hard to observe. However, we must say that in the case of CMB an advantage is that the underlying signal is accurately described by the linear theory at all scales. Unfortunately, with 21cm this is only the

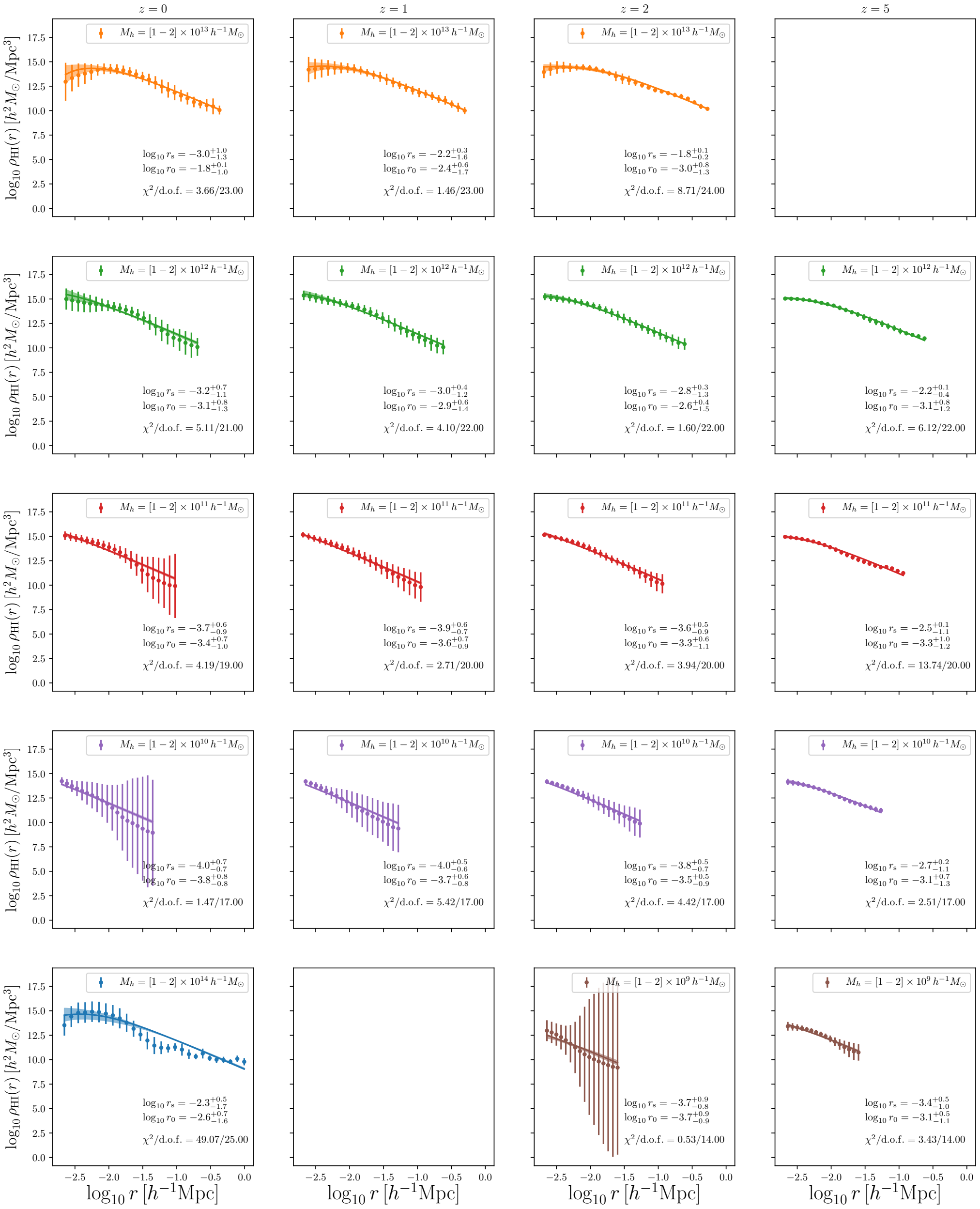

FIGURE 2.3: Each panel shows the mean and standard deviation of the HI profiles for halos in the mass range indicated in the upper-right part. We fit the results using the form $\rho_{\rm HI}(r) = \exp(-r_0/r)\rho_0 r_s^3/[(r + 3/4r_s)(r + r_s)^2]$, where ρ_0, r_s and r_0 are free parameters. The best-fit is shown with a solid line. The dashed region represents the error on the fit. Each panel shows the best-fit values of r_0 and r_s and the value of χ^2.

case at large scales. At small scales, the model too becomes uncertain and, to say at least, hard to compute and predict.

In this thesis we will assume a somewhat ideal situation concerning the foreground contamination. In chapter 3 we will assume no system noise and foreground contaminations when discussing the BAO reconstruction in the case of the 21cm IM observations. In chapter 4 we will consider the system noise and assume that all the

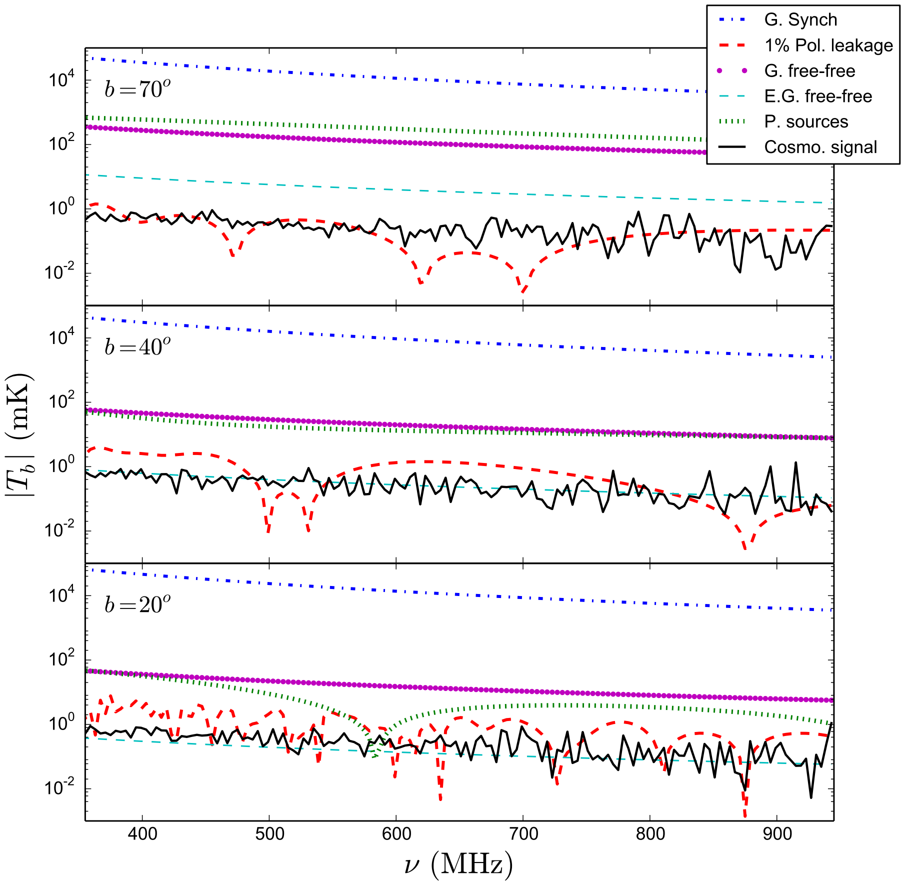

Figure 2.4: Frequency dependence along the LOS of different simulated foreground components at different galactic latitudes (taken from [122]).

astrophysical foregrounds, extragalactic point-sources and radio frequency interference (RFI) have been removed perfectly.

Chapter 3

BAO reconstruction with pixels

Introduction

In section 1.2.8 we have introduced the BAO reconstruction method. The BAO reconstruction technique can be applied to galaxy surveys, where the goal is to try to move back galaxies to their initial positions (or equivalently to move information embedded into higher order correlations back to the two-point function [123]). However, there are some observables, like 21cm IM maps where the output of observations does not consist of a catalogue with the positions of galaxies on the sky, but pixelated maps. For this type of observations the standard reconstruction technique can not be applied, although one possibility would be to use Eulerian reconstruction techniques (see [123]). Recently, a mesh-based method has been proposed [124] to carry out BAO reconstruction from 21cm interferometry observations in the presence of foreground contaminations and has been tested at the level of the propagator.

In this chapter we provide a more complete and detailed study of the mesh-based reconstruction technique that can be applied to both galaxy surveys and 21cm single-dish IM observations. This method is similar to the standard reconstruction technique, but relies on moving pixels instead of points (galaxies). By using a large set of numerical simulations we create mock pixelated maps from the distribution of both matter and halos in both real- and redshift-space. We then apply our method to those maps and investigate the performance of the method. We also demonstrate that in the case of galaxy surveys, this method is equivalent to standard reconstruction in the limit of very small pixels.

This chapter is organized as follows. In section 3.1 we investigate the impact of instrumental effects on the amplitude and shape of the BAO peak from observations consisting on pixelated maps, focusing our attention for concreteness on the case of 21cm IM. In section 3.2 we outline the simulations used in this work, while in section 3.3 we describe the reconstruction algorithm. Our theoretical model and the methods we use to fit the results from the simulations is described in section 3.4. We show and discuss the main results of this chapter in section 3.5. Finally, the main conclusions of this work are presented in section 3.6.

3.1 Pixelated maps observations and instrumental effects

We begin this section by describing briefly one type of cosmological survey that produces as output pixelated maps: single-dish 21cm IM observations. We then investigate the impact of instrumental effects, that determine the resolution of the pixelated maps, onto the shape and amplitude of the BAO peak as inferred from those maps. The goal of this section is to study the impact of maps resolution on the monopole and quadrupole of the correlation function and incorporate those effects into our theoretical template that we will use to determine the position of the BAO peak from our mock maps.

3.1.1 Pixelated observations: IM

An example of surveys producing pixelated maps is given by 21cm IM observations [64, 125, 126]. The IM technique consists in carrying out a low angular resolution survey, where individual galaxies or HI blobs are not resolved, to measure the 21cm radiation from cosmic neutral hydrogen from large patches of the sky. The idea is the same as with galaxy surveys: HI perturbations on large-scales will trace the underlying matter perturbations. There are two types of observations that can be carried out with radio-telescopes: single-dish or interferometry. In this chapter we focus our analysis on single-dish autocorrelation observations[1], where the resolution of the maps depends on the size of the antennae and where the maximum angular transversal scales that can be probed are not limited by the field-of-view (FoV) of the radio-telescope. However, we stress that low angular resolution is a limiting factor also for interferometry. A detailed description of the pros and cons of the two different techniques can be found in [64, 125]. A way to perform reconstruction by combining observations from 21cm IM and galaxy redshift surveys has also been suggested in order to fill in the missing modes lost due to the foregrounds contaminations [127].

In single-dish radio surveys the angular resolution of the 21cm maps is given by $\theta_{\mathrm{FWHM}} \cong \lambda/D$, where $\lambda = 0.21(1+z)$ m is the wavelength (in meters) of the 21cm radiation and D is the diameter of the antenna. We assume for simplicity that the primary beam of the telescope is well described by a Gaussian, thus the measured temperature on the $\hat{n}$ direction of the sky is

$$\tilde{T}_b(\hat{n},\nu) = \int d\check{s}_\perp T_b(\nu,\vec{s}_\perp)\frac{1}{2\pi\sigma^2}e^{-s_\perp^2/2\sigma^2}\,; \tag{3.1}$$

in Fourier space the amplitude of the modes will be given by

$$\delta\tilde{T}_b(k_\parallel,\vec{k}_\perp) = e^{-k_\perp^2\sigma^2/2}\delta T_b(k_\parallel,\vec{k}_\perp)\,, \tag{3.2}$$

where $\delta\tilde{T}_b$ and δT_b represent the observed and cosmological modes and the comoving angular smoothing scale, σ, is given by

$$\sigma = \frac{\chi(z)\theta_{\mathrm{FWHM}}}{2\sqrt{2\log 2}} \tag{3.3}$$

with $\chi(z)$ being the comoving distance to redshift z and the factor $2\sqrt{2\log 2}$ is due to the relation between the FWHM and the standard deviation in the Gaussian function. We notice that while in real observations the density of pixels in a map is closely related to the map resolution, in this chapter we consider these two quantities separately. The reason, as we will see clearly on section 3.3.2, is that the number of pixels can be taken arbitrarily high and this has some benefits for reconstruction. We emphasize that the important quantity in our study is the angular resolution of the maps, parametrized by the parameter σ.

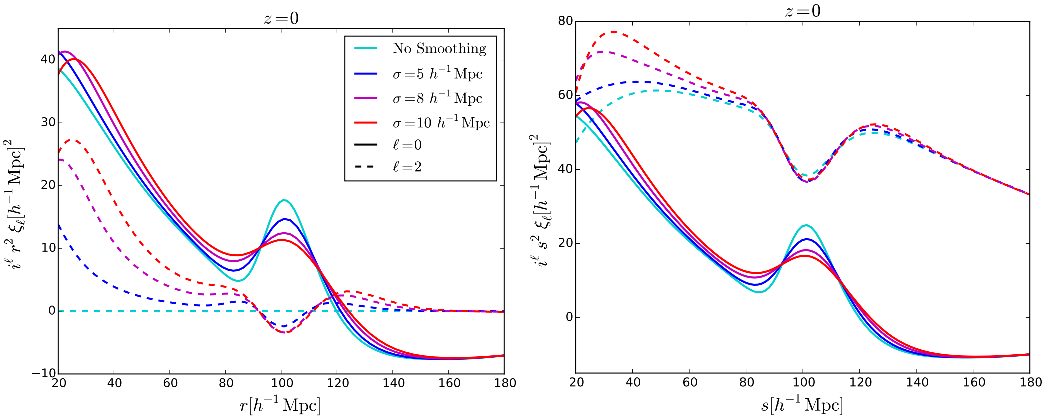

FIGURE 3.1: Linear theory prediction for matter monopole (solid) and quadrupole (dashed) of the correlation function for various smoothing scales in real- (left panel) and redshift-space (right panel) at $z = 0$.

3.1.2 Instrumental effects

The amplitude and shape of the BAO peak is affected by both non-linear gravitational evolution and instrumental effects. While the goal of this work is to develop a method to undo, at least partially, the effect of non-linearities, the effects induced by the instrument may not be taken out. An example is given by the maps resolution from single-dish 21cm IM observations, an effect induced by the antenna diameter and that the presence of system noise avoids the possibility of deconvolving the signal to recover to underlying field (see for instance [125]). In this situation, it is important to incorporate the instrumental effects on the BAO peak when building up the theoretical template. We limit our analysis to the impact of resolution on the shape and amplitude of the BAO peak, but we notice that other effects, such as system noise and foregrounds contamination can also affect it (see [125] for a study where these effects were taking into account when detecting the radial BAO). The aim of this subsection is to show the effects induced by the resolution of the pixelated maps on the BAO feature.

Low angular resolution of the radio telescopes is one of the instrumental effects that must be taken into account when fitting for the position of BAO. The 21cm power spectrum in redshift-space from single-dish observations in linear theory is given by (see Section 4.1.1):

$$\tilde{P}_{21\mathrm{cm}}(k,\mu) = b_{21\mathrm{cm}}^2(1+\beta\mu^2)^2 e^{-(1-\mu^2)k^2\sigma^2} P_{\mathrm{m}}(k), \tag{3.4}$$

where $b_{21\mathrm{cm}}$ is the bias of the 21cm signal, which is given by $b_{21\mathrm{cm}} = \bar{T}_b b_{\mathrm{HI}}$, with b_{HI} being the HI bias and $\bar{T}_b$ is the mean brightness temperature.

We notice that even in real-space ($\beta = 0$), the measured 21cm power spectrum is not isotropic, since this symmetry is broken by the angular smoothing in the angular direction. The multipoles of the observed 21cm power spectrum are given by

$$\tilde{P}_{21\mathrm{cm},\ell}(k) = \frac{2\ell+1}{2}\int_{-1}^{1} L_\ell(\mu)\tilde{P}_{21\mathrm{cm}}(k,\mu)d\mu, \tag{3.5}$$

---

[1] We refer the reader to [124] for a study in reconstruction with interferometry observations.

where $L_\ell(x)$ is the Legendre polinomial of order ℓ. The multipoles of the observed 21cm correlation function can be written as a function of their power spectrum counterparts

$$\tilde{\xi}_{21\mathrm{cm},\ell}(r) = i^\ell \int_0^\infty \frac{k^2 dk}{2\pi^2} \tilde{P}_{21\mathrm{cm},\ell}(k) j_\ell(kr), \tag{3.6}$$

with $j_\ell(x)$ being the spherical Bessel function of order ℓ. In figure 3.1 we plot the monopole and quadrupole of the observed 21cm correlation function in real- and redshift-space at linear order for 21cm maps having different resolutions (characterized by the parameter σ). For simplicity in figure 3.1 we have taken $b_{21\mathrm{cm}} = 1$ mK and $b_{\mathrm{HI}} = 1$. It can be seen that the shape of the BAO peak gets distorted by the map resolution; the effect is similar to the one induced by non-linearities, i.e. the BAO peak gets damped and broader. This distortion increases with σ, both in real- and redshift-space. This happens because the smoothing in the transverse direction erases correlations on angular scales smaller than $\sim \sigma$. For angular smoothing scales $\sigma \gtrsim 10\ h^{-1}$Mpc the BAO feature will be almost completely erased in the monopole of the correlation function, however the BAO peak can still be seen in the radial 21cm power spectrum [125], although the amount of information embedded there is much smaller.

As expected in linear theory, the quadrupole in real-space is zero when no angular smoothing is applied. On the other hand, for values of σ larger than zero the quadrupole deviates significantly from zero in real-space. The reason is that the angular smoothing breaks the isotropy present in real-space, inducing a non-negligible quadrupole that increases with σ. In redshift-space the amplitude of the quadrupole on large scales arising from the Kaiser term is much larger that the one induced by the angular smoothing, so the impact of the telescope angular resolution does not modify significantly the shape and amplitude of the quadrupole on scales $r \gtrsim 80\ h^{-1}$Mpc. On smaller scales the angular smoothing becomes more important, with the amplitude of the quadrupole increasing with σ.

3.2 Simulations

Generating mock 21cm maps is computationally very challenging since large box-size high-resolution hydrodynamic simulations, coupled to radiative transfer calculations, are needed to properly simulate the spatial distribution of neutral hydrogen in the post-reionzation era. A computationally less expensive alternative, although less precise, way consists in populating dark matter halos with neutral hydrogen a-posteriori [53]. The way dark matter halos are populated with HI can be calibrated using hydrodynamic simulations with small box sizes or by means of analytic models that reproduce the observational data [128, 76]. The idea of this method is thus to run a standard N-body simulation to obtain a catalogue of dark matter halos which are populated with HI during the post-processing. While N-body simulations are much faster than hydrodynamic simulations, the simulation set this work requires (500 simulations) made this method computationally unfeasible given the computational resources we have access to.

Many different methods have been developed such as PTHALOS [129], Augmented Lagrangian Perturbation Theory (ALPT) [130], PerturbAtion Theory Catalog generator of Halo and galaxY distributions (PATCHY) [131], Comoving Lagrangian Acceleration method (COLA) [132, 133, 134], Effective Zel'dovich approximation mocks (EZmocks) [135], FastPM [136] and PINOCCHIO [137, 138, 139, 140] (see [141] for a comparison among the different methods and N-body simulations) that are able to either generate a mock distribution of dark matter halos or to evolve directly the matter distribution in an N-body manner. These methods make use of different approximations that determine, in most cases, the accuracy they can reach when comparing results against N-body simulations.

The rationale behind the use of the above methods is to generate halo catalogues and simulate the spatial distribution of matter into the fully non-linear regime in a faster way than an N-body simulation. Among the previous methods, we have chosen COLA to run our numerical simulations. COLA is basically a particle-mesh (PM) code and therefore can be considered as an N-body code. The difference with respect to a fully N-body is the number of times steps and the way COLA deals with small scales. In comparison with the fully non-linear N-body simulations, COLA implementation can achieve an agreement at the level of the dark matter power spectrum to within 2% to 5% at $z = 0$ on scales up to $k = 0.3h\,\mathrm{Mpc}^{-1}$ – relevant for BAO and RSD studies [134]. Given this high accuracy COLA can reach and the fact that it is computationally much faster than an N-body simulation, we decide to use this code to run our numerical simulations.

We have run 500 simulations using the publicly available L-PICOLA code [134], a version of the original COLA code [132]. In the simulations we follow the evolution of 512^3 dark matter particles within a box of side 1 h^{-1}Gpc from $z = 9$ to $z = 0$ using a grid with a number of cells equal to the number of particles. The values of the cosmological parameters we use for all simulations are: $\Omega_{\mathrm{m}} = 0.3175$, $\Omega_\Lambda = 0.6825$, $\Omega_{\mathrm{b}} = 0.049$, $\Omega_\nu = 0.0$, $h = 0.6711$, $n_s = 0.9624$, $\sigma_8 = 0.834$. We save snapshots at $z = 1$ and $z = 0$. The outputs at $z = 1$ are obtained using 10 time-steps, while we use 50 time-steps linearly spaced in scale factor a for outputs at $z = 0$. The mass resolution of the dark matter particles is $6.56 \times 10^{11}\ h^{-1}\mathrm{M}_\odot$. We identify dark matter halos using the Friends-of-Friends algorithm [112] with a linking length parameter $b = 0.2$. Halos containing less than 32 dark matter particles ($M_{\mathrm{halo}} \lesssim 2 \times 10^{13} h^{-1}\mathrm{M}_\odot$) are discarded from our catalogues.

3.2.1 Creating mock maps

Here we explain how we simulate the intrinsic resolution of the 21cm maps in our simulations. From the output of the numerical simulations we build mock pixelated maps using the distribution of matter or halos in both real- and redshift-space. We compute the overdensity field $\delta(\vec{x})$ of particles in a simulation on a regular grid using cloud-in-cell (CIC) scheme. We Fourier transform $\delta(\vec{x})$ to obtain $\tilde{\delta}(\vec{k})$ and we correct for the CIC mass assignment scheme. We then apply a transverse 2D Gaussian filter to the density field with an angular smoothing scale σ:

$$\tilde{\delta}_{\mathrm{sm}}(\vec{k}) = \tilde{\delta}(\vec{k}) e^{-k^2(1-\mu^2)\sigma^2/2}. \tag{3.7}$$

We varied the angular smoothing scale within a reasonable range of values $\sigma = 5, 8, 10\, h^{-1}$Mpc, that we use throughout the chapter. We call the resulting fields – matter and halo maps.

We emphasize that the matter and halo maps constructed following the above procedure do not correspond to actual 21cm IM maps (see for instance [53]). The goal is to create pixelated maps with different levels of complexity, i.e. incorporating redshift-space distortions, halo bias...etc, in order to investigate the robustness of our method against these processes.

When analysing the matter maps in redshift-space, in each realisation we measure the monopole and the quadrupole along three different axes of our simulation and use the average monopole and quadrupole.

To create a pixelated map from a galaxy survey, in which the angular resolution effects are negligible, we follow the procedure described above and set the angular smoothing scale $\sigma = 0$.

3.3 Reconstruction algorithm

We start this section by explaining how the standard reconstruction method works. We then describe in detail our pixelated BAO reconstruction algorithm together with its practical implementation.

3.3.1 Standard reconstruction

The density field reconstruction method was first presented in [38] and it has proved very successful in both data [142, 143, 144] and simulations [15, 145, 39, 146]. Here we summarise the method briefly to set up notation and outline the general idea.

A position of a particle in Eulerian coordinates $\vec{x}$ after time t can be mapped to the initial Lagrangian position $\vec{q}$ using the displacement field $\vec{\Psi}(\vec{q}, t)$:

$$\vec{x}(\vec{q}, t) = \vec{q} + \vec{\Psi}(\vec{q}, t). \tag{3.8}$$

Lagrangian Perturbation Theory (LPT) gives a perturbative solution for this displacement field and the first order solution is the Zel'dovich Approximation (ZA) [147]. In ZA we can express the overdensity field in Eulerian coordinates in terms of the displacement field:

$$\delta(\mathbf{x}) = -\nabla_{\vec{x}} \cdot \vec{\Psi}(\vec{x}). \tag{3.9}$$

In Fourier space the displacement field is thus given by:

$$\tilde{\vec{\Psi}}(\vec{k}) = \frac{i\vec{k}}{k^2}\tilde{\delta}(\vec{k}). \tag{3.10}$$

The idea of BAO reconstruction is to get an estimate of the large scale displacement field from the observed non-linear density field and then use this field to displace the galaxies back to their initial positions. This results in removing the displacements of galaxies on large scales that contribute the most to the smearing of the acoustic peak. When considering also the redshift-space distortions, there are two main ways to do the reconstruction: *anisotropic* and *isotropic* (see [148]). In this work we focus on the anisotropic reconstruction in which the redshift-space distortions are kept in the final density field. Following the convention of [148], the algorithm proceeds in the following way:

1. The observed density field is convolved with a smoothing kernel $S(k)$ to reduce the small-scale non-linearities: $\tilde{\delta}(\vec{k}) \to \tilde{\delta}^s_{\rm nl}(\vec{k})S(k)$, where $S(k)$ is usually a Gaussian filter $S(k) = \exp[-0.5k^2R_\Psi^2]$ with R_Ψ the displacement smoothing scale and $\tilde{\delta}^s_{\rm nl}$ is the observed redshift-space density field.

2. We estimate the negative real-space displacement field from the smoothed density field:
$$\tilde{\vec{s}}^r(\vec{k}) = -\frac{i\vec{k}}{k^2}\frac{\tilde{\delta}^s_{\rm nl}(\vec{k})}{b}S(k),$$

where b is the linear galaxy bias.

3. We displace the galaxies by:

$$\vec{s}^{s}(\vec{x}) = \vec{s}^{r}(\vec{x}) + \frac{f-\beta}{1+\beta}(\vec{s}^{r}(\vec{x})\cdot\vec{z})\vec{z}$$

   to obtain the displaced density field $\delta_d(\vec{x})$, where f is the growth rate and β is the redshift-space distortion parameter: $\beta = f/b$ and $\vec{s}^{s}(\vec{x})$ is the negative redshift-space displacement field.

4. We shift a uniformly distributed grid of particles by the same $\vec{s}^{s}$ to obtain the shifted density field $\delta_s(\vec{x})$.

5. The reconstructed density field is then defined as $\delta_r(\vec{x}) \equiv \delta_d(\vec{x}) - \delta_s(\vec{x})$.

While this method is intended for observations of galaxies in redshift-space, one can also apply it to a particle set such as the matter density field from an N-body simulation by setting $b = 1$. If the galaxy/matter catalogue is in real-space, redshift-space distortions can be switched off by setting $\beta = f = 0$. When applying this method to halo catalogues, instead of galaxies, we use linear halo bias b_h.

3.3.2 Pixelated BAO reconstruction

The standard reconstruction improves the significance of the BAO peak position in the power spectrum or the correlation function of the observed galaxy distribution. However, in the description of the algorithm in the previous section, the fact that the density field was estimated from a discrete number of tracers never played any role[2]. Moreover the ZA, or higher order LPT, are thought to effectively describe the motion of dark matter fluid elements, which could end up containing more than one galaxy. It is therefore worth to see how BAO reconstruction performs on mesh-based fields, and in this section we define the relevant modifications to the original method required when dealing with pixels. A similar method was presented in [124] to derive the reconstructed density field in the presence of the foreground wedge effect.

The main modification compared to the standard reconstruction technique is that we work at the level of a regular grid and we treat the grid cells in the simulations as galaxies in the standard reconstruction algorithm. The grid cells we use are the same as the ones we used to produce the matter and halo maps, as described in 3.2.1. Once we have the matter and halo maps, we proceed to compute the displacement field using the already smoothed density field $\tilde{\delta}_\mathrm{sm}(\vec{k})$ We do this by first applying a Gaussian smoothing kernel to $\tilde{\delta}_\mathrm{sm}(\vec{k})$ such that it is effectively smoothed isotropically with a displacement smoothing scale R_Ψ:

$$\tilde{\delta}_{\vec{\Psi}}(\vec{k}) = \tilde{\delta}_\mathrm{sm}(\vec{k}) \exp\left[-\frac{k^2}{2}((1-\mu^2)R_\perp^2 + \mu^2 R_\Psi^2)\right], \tag{3.11}$$

where $R_\perp = \sqrt{R_\Psi^2 - \sigma^2}$ is the transverse smoothing scale which is smaller than R_Ψ to take into account the fact that we have already smoothed the field in the transverse direction. The choice of R_Ψ will be discussed below.

---

[2] Although note that estimates of the displacement field from very sparse samples affect the performance of BAO reconstruction [149].

Using this overdensity field we calculate the negative displacement field at the centres of grid cells using:

$$\vec{s}(\vec{x}) = \mathrm{IFFT}\left[-\frac{i\vec{k}}{k^2}\tilde{\delta}_{\Psi}(\vec{k})\right]. \tag{3.12}$$

We then use this displacement field to move the centres of grid cells according to the derived displacement field. Next, we compute the displaced field $\delta_d(\vec{x})$ of displaced grid cells on a regular grid using CIC scheme and weighting each grid cell by $(1+\delta_d)$. When computing the shifted field we only need to modify the weights in the CIC scheme since in both the displaced and the shifted field case the initial grid cells have been displaced by the same displacement field. We thus use the positions of displaced grid cells and apply equal weights using the CIC scheme to compute the shifted field $\delta_s(\vec{x})$.

In the last step we subtract shifted field from the displaced field to obtain the reconstructed density field:

$$\delta_r(\vec{x}) = \delta_d(\vec{x}) - \delta_s(\vec{x}).$$

In the case of $\sigma = 0$, we tested several sizes of grid cells and we find that the reconstruction improves as we increase the resolution, converging when the size of the grid cells approaches the mean particle separation in the simulation. In our case this separation is $1000\,h^{-1}\mathrm{Mpc}/512 \approx 2\,h^{-1}\mathrm{Mpc}$, and we use this size for the rest of the chapter. We also find that the choice of grid cell size that we use for performing reconstruction does not depend on the angular resolution σ. This is mainly due to the fact that the radial direction is unaffected by the angular resolution and having smaller grid cell sizes along the radial direction improves the reconstructed density field.

We have tested this method using the matter and halo maps in real- and redshift-space at $z = 0$ and $z = 1$ created from 500 COLA simulations. In Figure 3.2 we show the average monopole and quadrupole at $z = 0$ in real- and redshift-space before and after reconstruction of the matter maps. We show the results at $z = 1$ in Figure 3.3. In Figure 3.4 we show the average monopole and quadrupole at $z = 0$ in real- and redshift-space before and after reconstruction of the halo maps. We would like to note that in Figures 3.2, 3.3 and 3.4 the position of the linear point at roughly $90\,h^{-1}\mathrm{Mpc}$, as proposed recently in [150], remains unchanged with varying the angular resolution. Perhaps more importantly is that it appears invariant after reconstruction, both in scale and height.

We also apply our reconstruction method to matter maps in real-space at $z = 0$ that correspond to a galaxy survey ($\sigma = 0$). We measure the correlation function in each of the simulations before and after performing both standard and our reconstruction method. In Figure 3.5 we show the comparison between the average measured correlation function using the standard and our reconstruction method. In both cases we use the same displacement smoothing scale $R_{\Psi} = 20\,h^{-1}\mathrm{Mpc}$. We see that the two methods basically overlap inside the uncertainty on the mean. In section 3.5.1 we show a more quantitative comparison and agreement between the two methods. In the case of a galaxy survey, we find that this way of performing reconstruction is almost identical to the standard reconstruction as long as the cell sizes are small and the CIC correction is properly accounted for. Furthermore, it is less computationally expensive, since there is no need to 1) interpolate the displacement field for every particle and 2) generate uniform random field of particles, interpolate the displacement

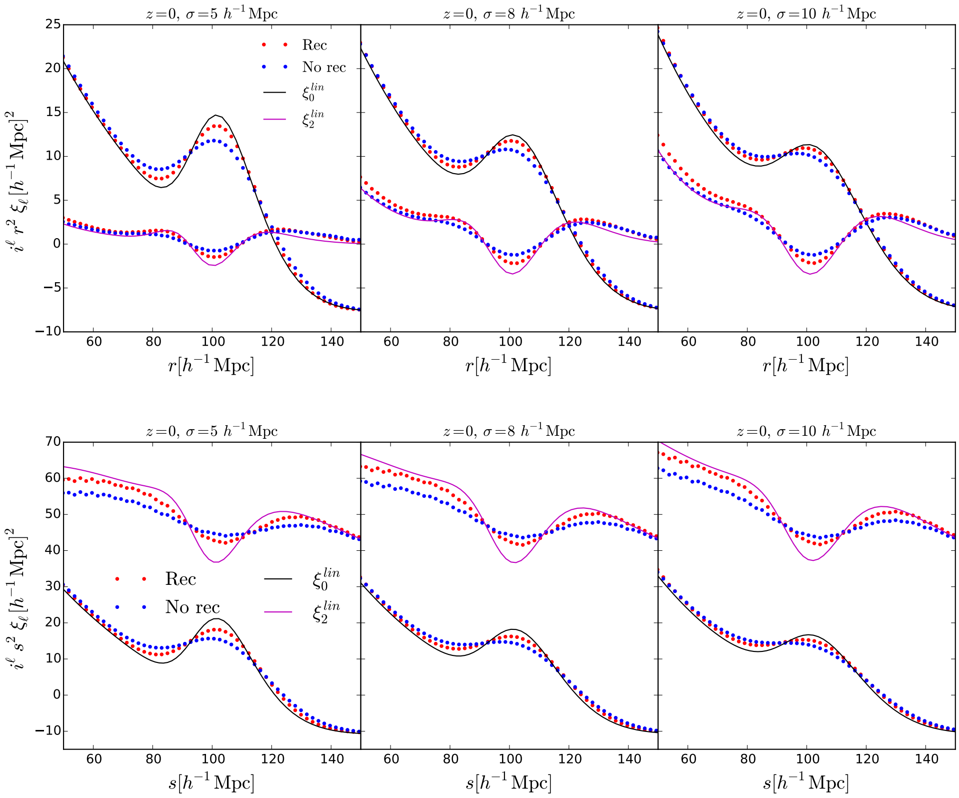

FIGURE 3.2: Average monopoles and quadrupoles of the pixelated matter maps in real-space (top) and redshift-space (bottom) at $z = 0$ before (blue) and after (red) reconstruction in cases of different angular resolution. Overplotted is the linear theory prediction in solid lines.

field and move the particles.

3.3.3 Smoothing scale for the displacement field

The choice of the smoothing scale R_Ψ should be made such as to tame the non-linearities at very small scales, while at the same time keep the valuable information of the mildly non-linear regime. The first requirement means making this scale larger, while the second requires it to be not too large. The impact of the smoothing scale R_Ψ on the standard reconstruction performance has been previously studied in detail both in mocks and data, e.g. [151, 152, 153, 148]. The choice depends on the level of non-linearity in the density field and in the shot-noise contribution [149, 124, 127]. Optimal choice of the scale depends on the case in study and has a broad range of values, ranging from $5 - 15$ h^{-1}Mpc. We are facing a somewhat different situation when we study the observables with low angular resolution. Therefore we tested the performance of our reconstruction method using different smoothing scales. In Figure 3.6 we show mean measured monopole and quadrupole of matter correlation function in real- and redshift-space at $z = 0$ after reconstruction for several different smoothing scales R_Ψ. Using $R_\Psi = 20h^{-1}$Mpc we find better agreement with the linear theory both in real- and redshift-space. We use this value in the rest of the chapter and leave the full analysis of the impact of this choice for future work.

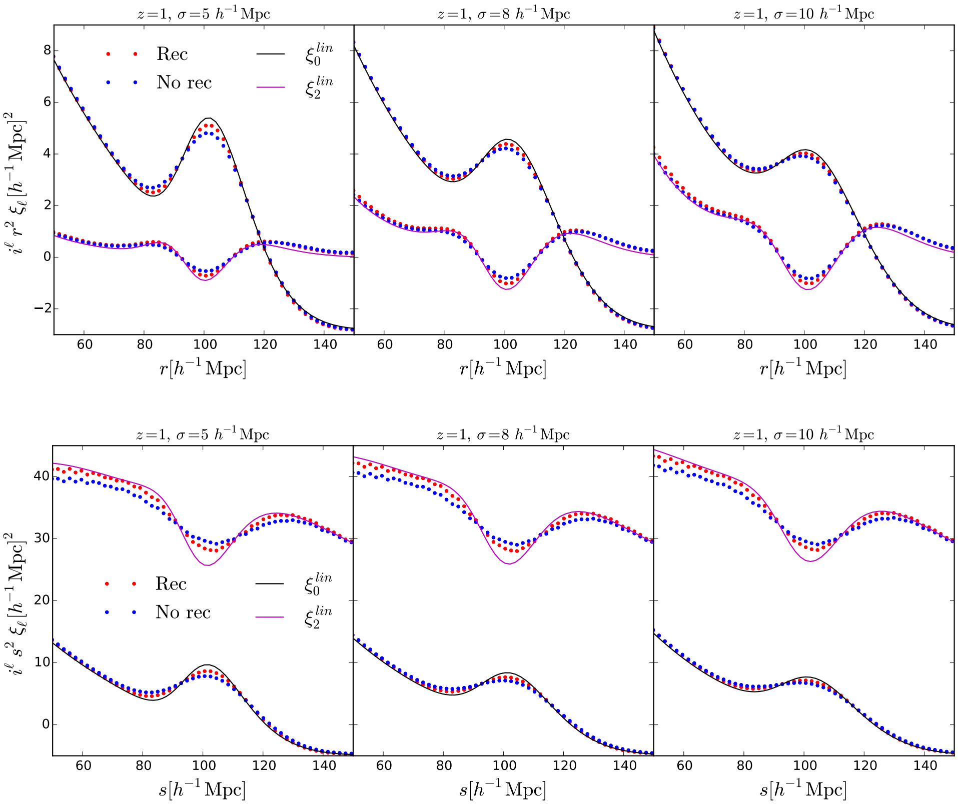

FIGURE 3.3: Average monopoles and quadrupoles of the pixelated matter maps in real-space (top) and redshift-space (bottom) at $z = 1$ before (blue) and after (red) reconstruction in cases of different angular resolution. Overplotted is the linear theory prediction in solid lines.

3.4 Analysis - Fitting the BAO peak

In this section we describe the models we use to obtain the templates for the non-linear correlation functions we are measuring. We then use these templates to build up a model that we use to fit the measured correlation functions in several cases. Our analysis is based on previous BAO analyses, which aim at measuring and put constraints on the position of the BAO peak [154, 143, 155, 144]. In the isotropic case, the BAO peak position we measure in the correlation function provides a measure of spherically averaged distance $D_V(z)$. We also need to take into account in our model the fact that, even if our template is a good approximation, assuming a fiducial cosmology can shift the measured BAO peak and therefore affect the distance measurement. This shift can be parametrised by:

$$\alpha = \frac{D_V(z)/r_d}{D_{V,f}(z)/r_{d,f}} \tag{3.13}$$

where r_d is the sound horizon at the drag epoch. Subscript f corresponds to the assumed fiducial cosmology, while the quantities without subscript refer to the true cosmology.

Once we have anisotropic clustering, like observations in redshift-space or with angular smoothing, we can measure the BAO position both along the line-of-sight and in the transverse direction, corresponding to separately measuring the Hubble parameter $H(z)$ and the angular diameter distance $D_A(z)$, respectively. In this case,

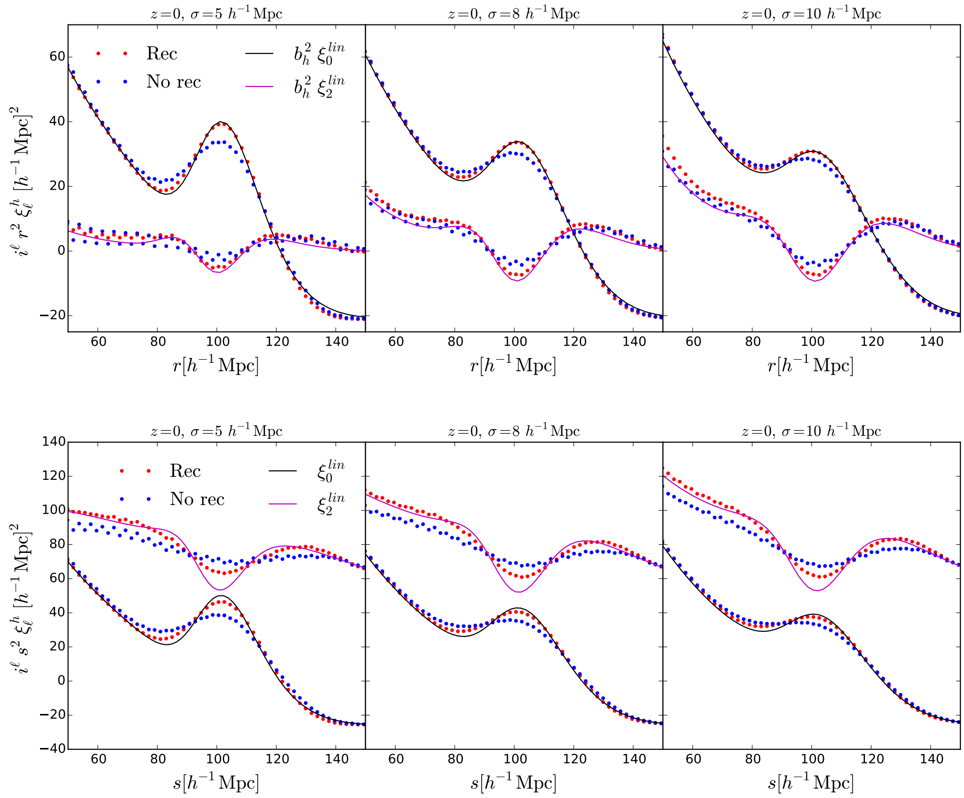

FIGURE 3.4: Average monopoles and quadrupoles of the pixelated halos maps in real-space (top) and redshift-space (bottom) at $z = 0$ before (blue) and after (red) reconstruction in cases of different angular resolution. Overplotted is the linear theory prediction in solid lines.

assuming a fiducial cosmology different from the true one will shift the measured BAO position differently along the line-of-sight and transverse directions. To account for this we will follow the method based on [154, 155]. In this formalism the isotropic shift α in the BAO positions is defined as:

$$\alpha = \left[\frac{D_A^2(z)}{D_{A,f}^2(z)}\frac{H_f(z)}{H(z)}\right]^{1/3}\frac{r_{d,f}}{r_d} \tag{3.14}$$

and the anisotropic shift ϵ:

$$1+\epsilon = \left[\frac{D_{A,f}(z)}{D_A(z)}\frac{H_f(z)}{H(z)}\right]^{1/3}. \tag{3.15}$$

Since we are using numerical simulations with a known cosmology, we expect $\alpha = 1$ and $\epsilon = 0$.

3.4.1 Isotropic case

In order to compare the standard (ST) and pixelated (PM) reconstruction methods we need a theoretical model, for the measured matter correlation function in real-space. In real-space the correlation function is isotropic and we use the following template:

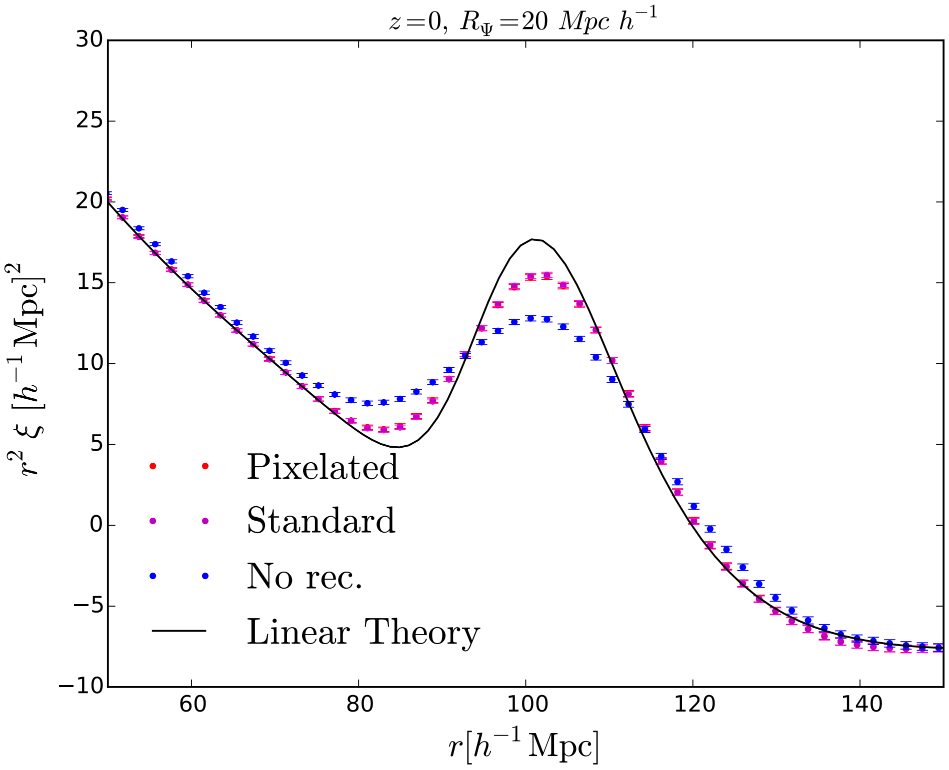

FIGURE 3.5: Measured real-space matter density field correlation function at $z = 0$ after using the standard reconstruction (magenta) and our method (red) to a set of 500 simulations of matter in real-space. Also shown is the non-reconstructed correlation function (blue) and the linear theory prediction (black). The data points show the mean values and the error bars show the scatter of the mean values. Standard and our reconstruction method overlap within the error bars of the mean.

$$\xi_t(r) = \int \frac{k^3 d\log k}{2\pi^2} P_{\mathrm{dw}}(k) j_0(kr), \tag{3.16}$$

where $P_{\mathrm{dw}}(k)$ is the de-wiggled power spectrum [31]. The de-wiggled power spectrum is designed to account for the damping of the power spectrum due to non-linear effects and is given by:

$$P_{\mathrm{dw}}(k) = [P_{\mathrm{lin}}(k) - P_{\mathrm{nw}}(k)] \exp\left[-\frac{k^2\Sigma_{\mathrm{nl}}^2}{2}\right] + P_{\mathrm{nw}}(k), \tag{3.17}$$

where $P_{\mathrm{lin}}(k)$ is the linear theory power spectrum computed using CAMB [16], $P_{\mathrm{nw}}(k)$ is the linear power spectrum without the BAO wiggles computed using a fitting formula in [156] and Σ_{nl} is the Gaussian damping scale. The final model we use to perform the fit is:

$$\xi^m(r) = B_0^2 \xi_t(\alpha r) + A(r) \tag{3.18}$$

where $A(r)$ is a polynomial function:

$$A(r) = \frac{a_1}{r^2} + \frac{a_2}{r} + a_3 \tag{3.19}$$

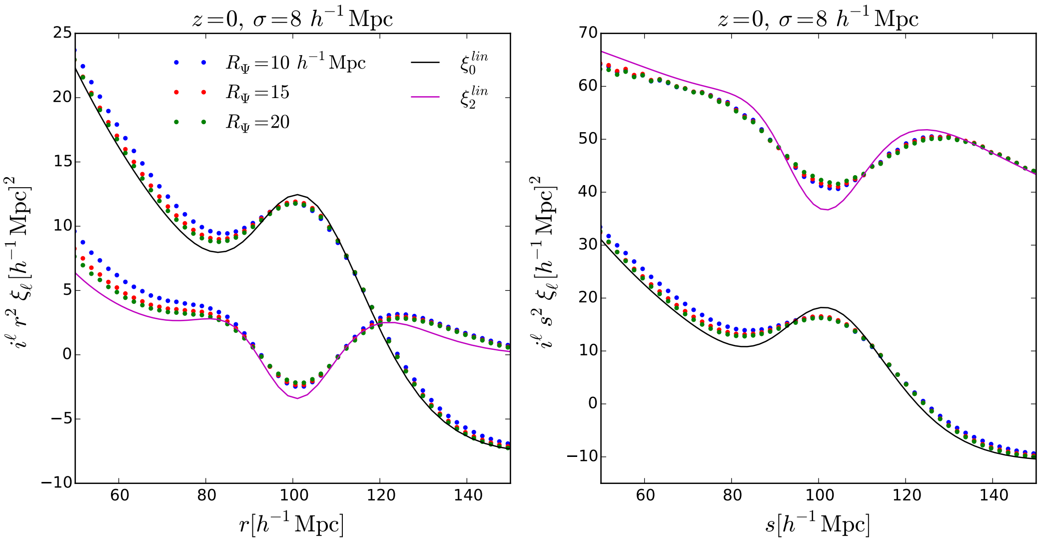

FIGURE 3.6: Mean monopoles and quadrupoles of pixelated matter maps in real-space (left) and redshift-space (right) after reconstruction using different smoothing scales for the displacement field: $R_\Psi = 10, 15, 20\, h^{-1}$Mpc. Here we are only showing results using $\sigma = 8h^{-1}$Mpc while similar results hold for other values of σ.

introduced to model effects that modify the broadband shape of the measured correlation function such as redshift-space distortions, halo bias and so on. The term B_0^2 controls the overall amplitude of the monopole template and, as in the case of the polynomial coefficients, represents a nuisance parameter.

3.4.2 Matter maps

In case of low angular resolution observables, the correlation function becomes anisotropic even in real-space. We write our template for the 2D smoothed non-linear power spectrum of matter field in real-space as:

$$P_t(k,\mu) = P_{\mathrm{dw}}(k,\mu)e^{-k^2(1-\mu^2)\sigma^2} \tag{3.20}$$

where the exponential term represents our 2D smoothing of the density field. At $z = 0$ for the matter density field we fix $\Sigma_{\mathrm{nl}} = 7.5\ h^{-1}$Mpc for non-reconstructed model and $\Sigma_{\mathrm{nl}} = 4.5\ h^{-1}$Mpc for the reconstructed model, while at $z = 1$ we fix $\Sigma_{\mathrm{nl}} = 5\ h^{-1}$Mpc for non- and $\Sigma_{\mathrm{nl}} = 3\ h^{-1}$Mpc for the reconstructed model. We choose these values based on a best fit to the average measured monopole and quadrupole.

In redshift-space we model the 2D smoothed non-linear power spectrum as:

$$P_t(k,\mu) = (1+\beta\mu^2)^2 F(k,\mu,\Sigma_s) P_{\mathrm{dw}}(k,\mu)e^{-k^2(1-\mu^2)\sigma^2}. \tag{3.21}$$

The term $(1+\beta\mu^2)^2$ is the Kaiser factor [20], that models redshift-space distortions on very large scales. We model the finger-of-God (FoG) effect [21] using a Gaussian form [64]:

$$F(k,\mu,\Sigma_s) = e^{-k^2\mu^2\Sigma_s^2}, \tag{3.22}$$

where Σ_s is the streaming scale describing the dispersion of random peculiar velocities along the line-of-sight direction that washes out the information on small scales.

Another form usually used is a Lorentzian [155] with a streaming scale Σ_s' which is different from ours by $\Sigma_s = \Sigma_s'\sqrt{2}$.

In redshift-space, the non-linear damping is not isotropic anymore. To take the anisotropy into account we use the de-wiggled power spectrum $P_{\mathrm{dw}}(k,\mu)$ given by [31]:

$$P_{\mathrm{dw}}(k,\mu) = [P_{\mathrm{lin}}(k) - P_{\mathrm{nw}}(k)] \exp\left[-\frac{k^2(1-\mu^2)\Sigma_\perp^2 + k^2\mu^2\Sigma_\parallel^2}{2}\right] + P_{\mathrm{nw}}(k). \tag{3.23}$$

Non-linear effects that cause the smearing of the BAO peak are modelled by a Gaussian with damping scale $\Sigma_{\mathrm{nl}}^2 = (\Sigma_\parallel^2 + \Sigma_\perp^2)/2$, with components $\Sigma_\parallel$ along and $\Sigma_\perp$ perpendicular to the line-of-sight.

We fix the components of the damping scale to the best-fit of the average measured monopole and quadrupole over all simulations. For the non-reconstructed case we set $\Sigma_\parallel = (1+f)\Sigma_\perp$. At $z = 0$ for the matter density field we fix $\Sigma_\perp = 6.5\ h^{-1}$Mpc, $\Sigma_\parallel = 9.96\ h^{-1}$Mpc and $\Sigma_s = 4.1\sqrt{2}\ h^{-1}$Mpc. For the reconstructed case we fix $\Sigma_\perp = \Sigma_\parallel = 4\ h^{-1}$Mpc and $\Sigma_s = 3.5\sqrt{2}\ h^{-1}$Mpc. At $z = 1$ for the non-reconstructed case we fix $\Sigma_\perp = 5\ h^{-1}$Mpc, $\Sigma_\parallel = 9.39\ h^{-1}$Mpc and $\Sigma_s = 2\sqrt{2}\ h^{-1}$Mpc. For the reconstructed case we fix $\Sigma_\perp = \Sigma_\parallel = 3\ h^{-1}$Mpc and $\Sigma_s = 2\sqrt{2}\ h^{-1}$Mpc.

The power spectrum multipoles of the matter maps templates in real- and redshift-space are given by:

$$P_{\ell,t}(k) = \frac{2\ell+1}{2}\int_{-1}^{1} P_t(k,\mu) L_\ell(\mu) d\mu, \tag{3.24}$$

where L_ℓ is the Legendre polynomial of order ℓ. The multipoles of the correlation function are then given by:

$$\xi_{\ell,t}(r) = i^\ell \int \frac{k^3 d\log k}{2\pi^2} P_{\ell,t}(k) j_\ell(kr). \tag{3.25}$$

We use the perturbative expansion in terms of α and ϵ to construct models for the monopole and quadrupole of the matter correlation function [155]:

$$\begin{aligned} \xi_0^m(r) &= B_0^2\xi_{0,t}(\alpha r) + \frac{2}{5}\epsilon\left[3\xi_{2,t}(\alpha r) + \frac{d\xi_{2,t}(\alpha r)}{d\log(r)}\right] + A_0(r), && (3.26)\\ \xi_2^m(r) &= 2B_0^2\epsilon\frac{d\xi_{0,t}(\alpha r)}{d\log(r)} + \left(1+\frac{6}{7}\epsilon\right)\xi_{2,t}(\alpha r) + \frac{4}{7}\epsilon\frac{d\xi_{2,t}(\alpha r)}{d\log(r)} \\ &\quad + \frac{4}{7}\epsilon\left[5\xi_{4,t}(\alpha r) + \frac{d\xi_{4,t}(\alpha r)}{d\log(r)}\right] + A_2(r), && (3.27) \end{aligned}$$

where

$$A_\ell(r) = \frac{a_{\ell,1}}{r^2} + \frac{a_{\ell,2}}{r} + a_{\ell,3}. \tag{3.28}$$

The polynomials $A_\ell(r)$ are standardly added in these analysis to account for systematics and in general to model effects not included in the template as non-linear redshift-space distortions and scale-dependent bias, that are expected to affect the broadband shape of the correlation function but not the position of the BAO peak.

When performing the best-fit analysis, we keep the following parameters free: B_0, coefficients of the A_ℓ polynomials, α and ϵ.

3.4.3 Halo maps

For the analysis of halo maps we restrict ourselves to maps at $z = 0$. Due to the low mass resolution in our simulations, halos we identify at $z = 1$ are not dense enough tracers of the density field and the shot noise is high enough that we do not see any improvement with reconstruction. At $z = 1$ the mean number density of halos is $\bar{n} \approx 5 \times 10^{-5} (h\,\mathrm{Mpc}^{-1})^3$ which is below the limit ($\sim 10^{-4} (h\,\mathrm{Mpc}^{-1})^3$) at which the standard reconstruction gains saturate [149]. The mean number density of halos at $z = 0$ is $\bar{n} \approx 1.4 \times 10^{-4} (h\,\mathrm{Mpc}^{-1})^3$ which is above this limit.

In addition, we find that the measured halo quadrupole in our simulations is dominated by noise both in real- and redshift-space (see Figure 3.4). For these reasons we focus on fitting only the monopole of the correlation function of the halo maps at $z = 0$.

We use monopole templates for matter maps in real- and redshift-space. The final model we use to perform the fit to the monopole of the halo maps is:

$$\xi_h^m(r) = B_0^2 \xi_{0,t}(\alpha r) + A(r). \tag{3.29}$$

Before performing a fit to the halo monopole, we normalise our template for the halo monopole to the halo bias b_h^2 measured from the simulations. We measure the halo bias as the ratio of halo and matter power spectrum over 500 simulations and take the average value on large scales.

When fitting the results, we fix the non-linear damping scale Σ_nl to the value we find is the best fit to the average measured monopole from the halo maps. In real-space, for the non-reconstructed case we set $\Sigma_\mathrm{nl} = 6.5\ h^{-1}$Mpc. After reconstruction we fix $\Sigma_\mathrm{nl} = 2.5\ h^{-1}$Mpc. In redshift-space we set $\Sigma_\parallel = \Sigma_\perp = \Sigma_\mathrm{nl}$. For the non-reconstructed case we set $\Sigma_\mathrm{nl} = 6.5\ h^{-1}$Mpc and $\Sigma_s = 3.5\sqrt{2}\ h^{-1}$Mpc, while after reconstruction we fix $\Sigma_\mathrm{nl} = 2.5\ h^{-1}$Mpc and $\Sigma_s = 2.8\sqrt{2}\ h^{-1}$Mpc.

When performing the best-fit analysis, we keep the following parameters free: B_0, coefficients of the $A(r)$ polynomial and α.

3.4.4 Fitting procedure

We assume that the measured correlation function follows a Gaussian distribution. Thus, finding the best model that describe the data is equivalent to minimizing

$$\chi^2 = (\vec{m} - \vec{d})^T C^{-1} (\vec{m} - \vec{d}), \tag{3.30}$$

where $\vec{m}$ and $\vec{d}$ are vectors containing the values of correlation function of model and data, respectively. In the anisotropic case of matter maps, the vectors of the model and the data contain both the results of the monopole and quadrupole, while in the case of halos we use only the monopole values. C^{-1} is the inverse covariance matrix described below.

We fit the results in the range $50\ h^{-1}\mathrm{Mpc} \leqslant r \leqslant 150\ h^{-1}\mathrm{Mpc}$. In the case in which we compare standard and pixelated reconstruction in real-space we use bin sizes of $2\ h^{-1}$Mpc and we thus employ 51 data points for the correlation function. In the anisotropic case of matter maps we use bin sizes of $4\ h^{-1}$Mpc and in the fitting range we have 25 data points for monopole and the same for quadrupole, a total of 50 data points. When performing a fit to the monopole of the halo maps, we also use bin sizes $4\ h^{-1}$Mpc and we have 25 data points.

We minimise the χ^2 and sample the model parameter space using Monte Caro Markov Chain (MCMC) using publicly available code `emcee` [157]. We apply a 20%

tophat prior on α and $1+\epsilon$ in order to avoid unphysical values for shift parameters in simulations where the BAO peak is less pronounced. We leave all nuisance parameters free.

3.4.5 Covariance matrices

We calculate the covariance matrix directly from the simulations:

$$C_{ij} = \frac{1}{N_\mathrm{s}-1}\sum_{n=1}^{N_\mathrm{s}}[d_n(r_i)-\bar{d}(r_i)][d_n(r_j)-\bar{d}(r_j)], \tag{3.31}$$

where N_s is the number of simulations, $d_n(r)$ is a vector containing the values of correlation function calculated from nth simulation at radius r and $\bar{d}(r)$ is the vector containing the mean values of correlation function at radius r over all simulations. In anisotropic case, the vectors $d_n(r)$ and $\bar{d}(r)$ contain both the monopole and quadrupole of the correlation function, while for halo maps we use only monopole values.

Since we are using a finite number of simulations, the estimated covariance matrix will be affected by sample noise. This results in a biased estimate for the inverse covariance matrix. This bias can be removed when estimating the inverse covariance matrix by multiplying the inverse estimate by [158]:

$$C^{-1} = C^{-1}_\mathrm{original}\frac{N_\mathrm{s}-N_\mathrm{b}-2}{N_\mathrm{s}-1}, \tag{3.32}$$

where N_b is the number of bins we are using. In the case of comparing standard and pixelated reconstruction method we have 51 data points. For the halo maps we have 25 data bins and for the matter maps we have 50 data bins: 25 from the monopole and 25 from the quadrupole.

Even with this correction, it has been shown that the noise still affects the constraints of the fitting parameters and this has to be accounted for [159]. We account for this by multiplying all the measured variances of the fitting parameters by a factor that depends on N_b, N_s and the number of fitting parameters N_p (see equation 22 in [159]).

3.5 Results

In this section we present the constraints we derive, in terms of the position of the BAO peak, before and after reconstruction. First, we show the results of the comparison between standard reconstruction and our method for a standard galaxy survey, which in our method we simplify to the distribution of pixelated matter in real-space from numerical simulations. Then we show the results of applying our reconstruction method to matter and halos maps in real- and redshift-space.

3.5.1 Standard versus pixelated reconstruction method

We first present the results of comparing the standard (ST) and pixelated (PM) reconstruction methods when applying them on the spatial distribution of matter in real-space at $z = 0$ from our numerical simulations. We have performed the analysis using different smoothing scales for the displacement field R_Ψ. We choose the nonlinear damping parameters Σ_nl for each case by a fit to the average of the measured correlation function. As expected, we find these parameters to be smaller for both reconstruction methods compared to the non-reconstructed case and on average their values decrease by a $\sim 50\%$ after reconstruction. We also find that the obtained values depend on the smoothing scale for the displacement field used. However, we find no significant difference of the reconstructed $\Sigma_{\rm nl}$ between the two reconstruction methods.

We measure the BAO shift parameter α in each of our 500 simulations before and after reconstruction (using the two different methods). In Table 3.1 we give the summary of best-fit results of the BAO shift parameter α as a function of the smoothing scale for the displacement field R_Ψ and used $\Sigma_{\rm nl}$ for each case.

R_Ψ $[h^{-1}\mathrm{Mpc}]$	Reconstruction method	$\Sigma_{\rm nl}$ $[h^{-1}\mathrm{Mpc}]$	α	$\langle\chi^2\rangle$ /dof
–	No	8.0	1.001 ± 0.018	42.1/45
10	ST	3.4	1.000 ± 0.008	42.5/45
	PM	3.5	0.999 ± 0.008	48.6/45
15	ST	4.3	1.000 ± 0.010	42.6/45
	PM	4.4	0.999 ± 0.009	43.4/45
20	ST	5.0	0.999 ± 0.011	42.4/45
	PM	5.0	0.999 ± 0.011	42.4/45

Table 3.1: Constraints on the BAO shift parameter (column 4) using the matter density field in real space – corresponding to a galaxy-survey with $\sigma = 0$. Results shown are obtained from simulations without applying reconstruction (first row) and after applying reconstruction (rows 2-7). The non-linear damping parameter $\Sigma_{\rm nl}$ used in the fit is given in Column 3. Results are shown for different smoothing scales for the displacement field R_Ψ (column 1) and the mean χ^2/dof is given in column 5. The errors shown for α are the standard deviations over 500 simulations. The actual errors on the means of the parameter α are the listed values divided by $\sqrt{500}$.

In Figure 3.7 (left panel) we plot these results for various smoothing scales for the displacement field R_Ψ. The points shown are the best-fit values and the error bars shown are standard deviations divided by $\sqrt{50}$ to show the expected uncertainty in a survey of $50\,[h^{-1}\mathrm{Gpc}]^3$ volume size. Best fits to the reconstructed correlation function in ST and PM case are shown in middle and right panel of Figure 3.7, respectively, in the case of $R_\Psi = 20\,h^{-1}\mathrm{Mpc}$.

We find the obtained values of α to be consistent with the expected value $\alpha = 1$ at the level of uncertainties in both ST and PM methods. The uncertainty on α after reconstruction decreases by $40 - 60\%$, depending on which R_Ψ is used. We find no significant difference in ST and PM methods.

3.5.2 Matter maps

We now focus our attention on the case of matter maps in both real- and redshift-space, considering maps with different resolutions, σ. We measure the BAO shift parameters α and ϵ in each of our 500 simulations before and after reconstruction using the template models outlined in Section 3.4.2. A summary of the best-fit results both in real- and redshift-space at $z = 0$ and $z = 1$ is presented in Table 3.2.

In Figure 3.8 we show the best-fit BAO shift parameters α (left) and ϵ (right) in real- and redshift-space as a function of the smoothing scale σ at $z = 0$ and $z = 1$. Blue points correspond to non- and red to reconstructed maps fits. The error bars

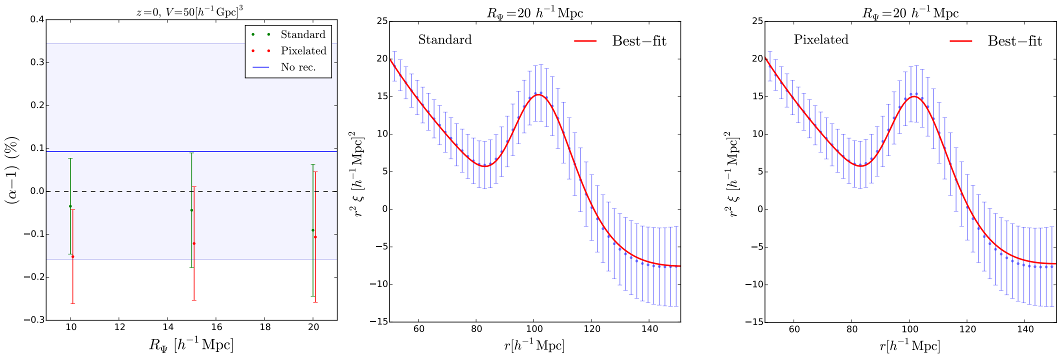

FIGURE 3.7: Reconstruction of the matter density field in real-space – corresponding to a galaxy-survey with $\sigma = 0$, at $z = 0$. *Left panel*: Mean best-fit values of BAO shift parameters α as a function of smoothing scale of the displacement field R_Ψ showing the comparison between standard and pixelated reconstruction algorithm. The error bars shown are the standard deviation from 500 simulations divided by $\sqrt{50}$ (i.e. the expected error for a survey covering a volume of ~ 50 $(h^{-1}\mathrm{Gpc})^3$). *Middle and right panel*: The best-fits to the reconstructed matter correlation function in case of standard and pixelated reconstruction method, using $R_\Psi = 20h^{-1}\mathrm{Mpc}$.

correspond to standard deviations of BAO shift parameters divided by $\sqrt{50}$ to show the expected uncertainty we would expect in a survey of ~ 50 $[h^{-1}\mathrm{Gpc}]^3$ volume size.

Real-space

We find that the uncertainties in both α and ϵ decrease by $\sim 50\,\%$ after reconstruction across the considered range of map angular resolutions σ. The 2D smoothing is smearing the BAO peak in the monopole with increasing σ so one would expect the uncertainty in α to increase as well. On the other hand, larger smoothing scales make the quadrupole more pronounced and therefore more constraining for the shift parameters. After reconstruction, the recovered values of α are closer and consistent with the expected $\alpha = 1$ at the single simulation level. Even by considering the error on the mean, i.e. combining the results of all the simulations to probe a volume equal to 500 $(h^{-1}\mathrm{Gpc})^3$, we find that α is consistent with 1 at 5σ level. We however find a $\sim 0.2\%$ shift in the recovered values of ϵ.

Similar to $z = 0$, at $z = 1$ we again find that the uncertainties in both α and ϵ decrease after reconstruction. These decreases are however smaller compared to $z = 0$, and reach $\sim 35\%$ in α and ϵ. This is expected since the non-linear effects are smaller at higher redshifts. Recovered means of reconstructed α are within 0.2% of the expected values, while the reconstructed ϵ stays within 0.1% of the expected value and within the uncertainties of the full simulated volume.

In Figures 3.9 and 3.10 we show the best-fit to the monopole and quadrupole of the matter maps in real-space after (*top*) and before (*bottom*) reconstruction at $z = 0$ and $z = 1$, respectively. Comparing the monopole and quadrupole before and after reconstruction, we find the acoustic peak gets more pronounced after reconstruction. This is more evident at $z = 0$ than at $z = 1$ since the non-linear effects are smaller at higher redshift. We also find that for smaller smoothing scales reconstruction makes the BAO peak more pronounced both in monopole and quadrupole. This improvement of reconstruction decreases as we move to lower resolution maps. Still,

Matter maps – real-space					
	$\sigma[h^{-1}\mathrm{Mpc}]$	Reconstruction	α	ϵ	$\langle\chi^2\rangle/\mathrm{dof}$
$z = 0$	5	no	1.004 ± 0.018	0.001 ± 0.022	45.3/41
		yes	1.0005 ± 0.0096	0.002 ± 0.010	44.1/41
	8	no	1.002 ± 0.021	0.003 ± 0.023	45.5/41
		yes	1.0008 ± 0.0096	0.003 ± 0.010	46.2/41
	10	no	1.004 ± 0.019	0.004 ± 0.021	45.4/41
		yes	1.002 ± 0.010	0.004 ± 0.011	47.8/41
$z = 1$	5	no	1.0000 ± 0.0098	-0.001 ± 0.012	43.8/41
		yes	0.9983 ± 0.0070	0.0008 ± 0.0080	41.8/41
	8	no	1.003 ± 0.011	-0.001 ± 0.013	45.1/41
		yes	0.9999 ± 0.0071	0.0006 ± 0.0088	44.7/41
	10	no	1.005 ± 0.011	0.001 ± 0.013	46.3/41
		yes	1.0021 ± 0.0095	0.0003 ± 0.0095	50.4/41
Matter maps – redshift-space					
	$\sigma[h^{-1}\mathrm{Mpc}]$	Reconstruction	α	ϵ	$\langle\chi^2\rangle/\mathrm{dof}$
$z = 0$	5	no	1.000 ± 0.023	0.0021 ± 0.0075	39.2/41
		yes	0.998 ± 0.011	0.0002 ± 0.0042	40.0/41
	8	no	1.001 ± 0.025	0.0045 ± 0.0097	38.6/41
		yes	0.996 ± 0.013	0.0011 ± 0.0060	40.4/41
	10	no	1.001 ± 0.026	0.005 ± 0.012	38.9/41
		yes	0.995 ± 0.015	0.0016 ± 0.0076	41.8/41
$z = 1$	5	no	0.997 ± 0.014	0.0012 ± 0.0077	30.4/41
		yes	1.000 ± 0.011	0.0000 ± 0.0040	31.9/41
	8	no	0.998 ± 0.018	0.001 ± 0.010	30.2/41
		yes	1.000 ± 0.013	0.0000 ± 0.0055	31.6/41
	10	no	1.000 ± 0.022	0.000 ± 0.013	31.2/41
		yes	1.001 ± 0.015	0.0000 ± 0.0067	32.4/41

TABLE 3.2: Constraints on BAO shift parameters α and ϵ for matter maps with different angular resolutions (column 2) before and after reconstruction. Columns 3 and 4 show the mean and the standard deviation of BAO shift parameters α and ϵ, respectively. The mean χ^2/dof is given in column 5. The errors shown for α and ϵ are the standard deviations over 500 simulations. The actual errors on the means of the parameters α and ϵ are the listed values divided by $\sqrt{500}$.

one should compare the reconstructed results with the linear theory prediction (solid lines in Figure 3.2, upper panel) and notice that even the linear theory prediction monopole is getting less pronounced as we move to larger values of smoothing scales.

Another test we perform is to check which is the significance of the BAO detection in the case of different angular smoothing scales we are considering and to quantify the improvement after performing reconstruction. We do this by performing another fit to the measured matter monopole and quadrupole with a model that has no BAO feature. This model is constructed by taking $\Sigma_{\rm nl} \to \infty$. Having obtained the $\chi^2_{\rm NO\,BAO}$ for each of our 500 simulations, we quantify the significance as the square root of $\Delta\chi^2 = \chi^2_{\rm NO\,BAO} - \chi^2_{\rm BAO}$. In Figure 3.11 we show the histograms of $\sqrt{\Delta\chi^2}$ before (blue) and after (red) reconstruction for different map resolutions σ at $z = 0$ (top) and $z = 1$ (bottom).

We find that performing reconstruction greatly improves the significance of BAO detection. At $z = 0$, using our matter maps in real-space covering a volume equal to

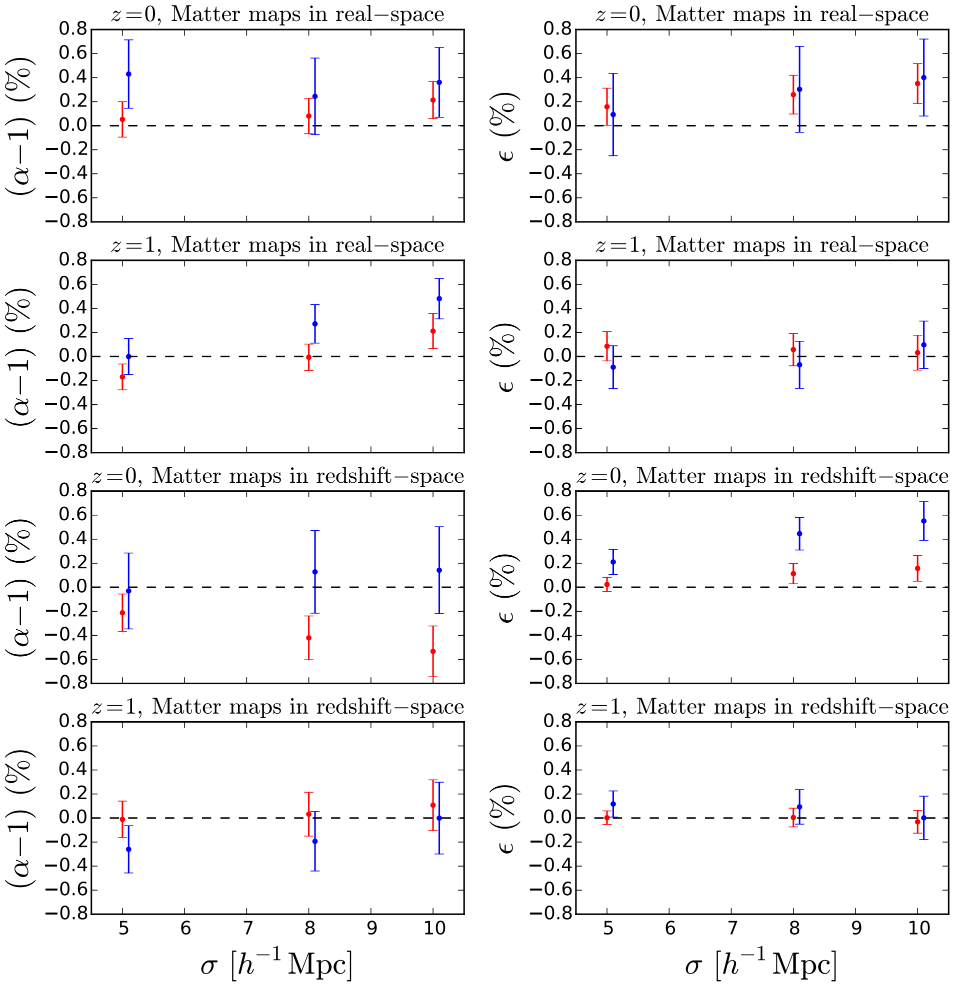

FIGURE 3.8: Mean best-fit values of BAO shift parameters α (left) and ϵ (right) as a function of map spatial resolution, σ, at $z = 0$ and $z = 1$ for matter maps. Blue points represent the non-reconstructed, while red points represent the reconstructed density field results. The error bars shown are the standard deviation from 500 simulations divided by $\sqrt{50}$ to show the uncertainty we would expect in a survey of $\sim 50\,[h^{-1}\mathrm{Gpc}]^3$ volume size

$1\,h^{-1}\mathrm{Gpc}^3$, we find that 100% (99%) of our mocks shows better than 3σ (5σ) significance of detecting BAO after reconstruction. This result holds for all the considered map resolutions. We find similar results for reconstruction at $z = 1$: 100% (98.6%) of our mocks shows a detection of the BAO with a significance above 3σ (5σ).

We also find that the improvement over the significance of BAO detection after reconstruction is greater for smaller values of the angular resolution, while the improvement decreases as we use larger values of angular resolution. Furthermore, the improvement is greater at $z = 0$ than at $z = 1$, as expected, since the non-linear effects that reconstruction partially removes are smaller at higher redshifts.

Redshift-space

By performing reconstruction over matter maps at $z = 0$ in redshift-space we find that uncertainties in α decrease by $\sim 50\,\%$ after reconstruction, while the uncertainties in ϵ decrease by $\sim 40\,\%$ (see Table 3.2). Recovered mean values of ϵ after reconstruction

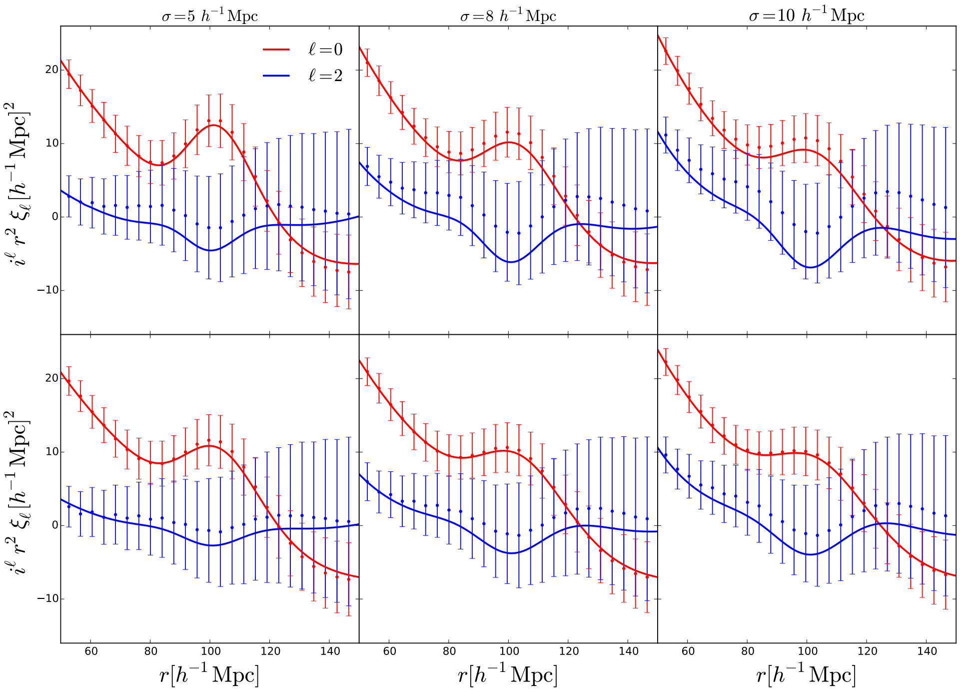

FIGURE 3.9: The best-fit to monopole and quadrupole correlation function for matter maps in real-space at $z = 0$ – reconstructed (top) and unreconstructed (bottom).

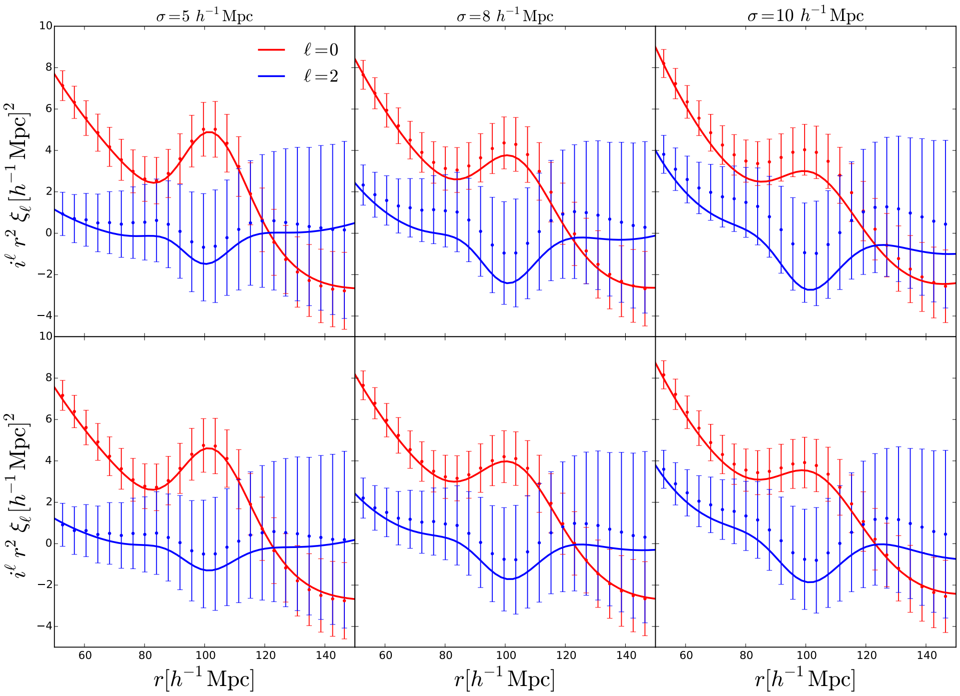

FIGURE 3.10: The best-fit to monopole and quadrupole correlation function for matter maps in real-space at $z = 1$ – reconstructed (top) and unreconstructed (bottom).

are at most 0.2% away from the expected $\epsilon = 0$ value and are within 5σ uncertainties considering the error on the mean, i.e. a total volume of 500 $(h^{-1}\mathrm{Gpc})^3$. On the other hand, we find a biased estimate of the recovered values in α that increases up

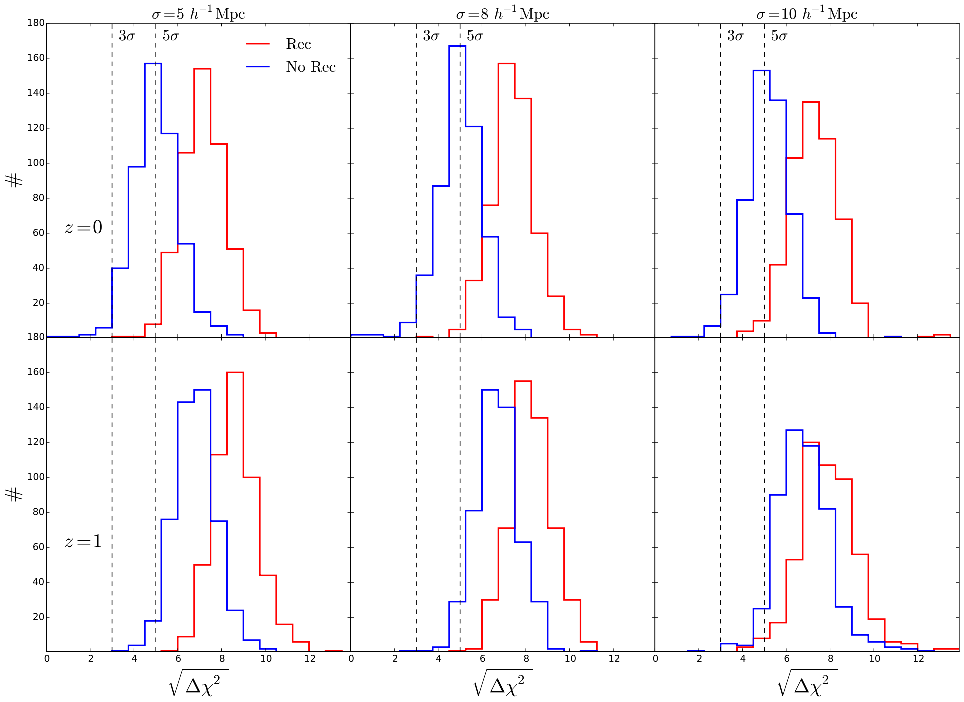

Figure 3.11: Significance of detecting a BAO peak with (red) and without (blue) reconstruction at $z = 0$ (top) and $z = 1$ (bottom). Histograms show the distribution of the square root of the absolute difference between best-fit χ^2 values with and without the BAO feature in a template for matter maps in real-space. Vertical dashed lines show the 3σ and 5σ detection significance.

to 0.5% for larger smoothing scales σ. Even though this bias is statistically significant at a level of more than 5σ for a total simulated volume, it is still compatible with the expected value considering the typical volume of a future 21cm survey.

At $z = 1$ we find that the uncertainties after reconstruction in α decrease by $\sim 30\,\%$, while for ϵ we find $\sim 50\,\%$ decrease. All the recovered values of both α and ϵ are consistent with the expected values. Small biases we find are within 5σ uncertainties for the full simulation volume.

As can be seen from Table 3.2 and Figure 3.8 the errors on ϵ parameter are significantly smaller than the errors on α. This result is in contrast to the results from real-space where the errors on α and ϵ are similar. The reason why this happens in redshift-space is due to the following fact. As described in Section 3.2, for matter maps in redshift-space we measure the monopole and the quadrupole along three different axes of our simulation. We then compute the covariance matrix by taking the average monopole and quadrupole along three different axes for each realisation. Since the scatter of quadrupole is high along three different axes in a particular realisation, taking the average reduces the variance and in effect makes the covariance matrix values smaller. Since the quadrupole is more sensitive to ϵ, in turn this makes the uncertainties on ϵ smaller by roughly a factor of 3 compared to the case we use only one axis. On the other hand, since the monopole is using the spherically averaged information, the measured scatter between different axes is much smaller. In turn the scatter is not affected by averaging over three different axes. Being that the monopole is more sensitive to the α parameter, the constraints are very similar when considering only one axis or average over three axes. To summarise, if we use only one axis, which is a more realistic scenario and what we have done in real-space, we get the errors on ϵ to be comparable and larger than the errors on α. In this case, the main reason why the constraints are similar for alpha and epsilon is the angular resolution, similar to the real-space consideration.

We show the best-fit model for matter maps in redshift space before and after reconstruction at $z = 0$ in Figure 3.12 and at $z = 1$ in Figure 3.13. Similar to real-space, monopole is more broad for maps with larger angular smoothing scale, while in the quadrupole this effect is reversed. We find both monopole and quadrupole to be more pronounced after reconstruction. The effect is not so evident for the monopole, but we emphasize again that these results should be compared with the linear theory prediction (solid lines in Figure 3.2, bottom panel).

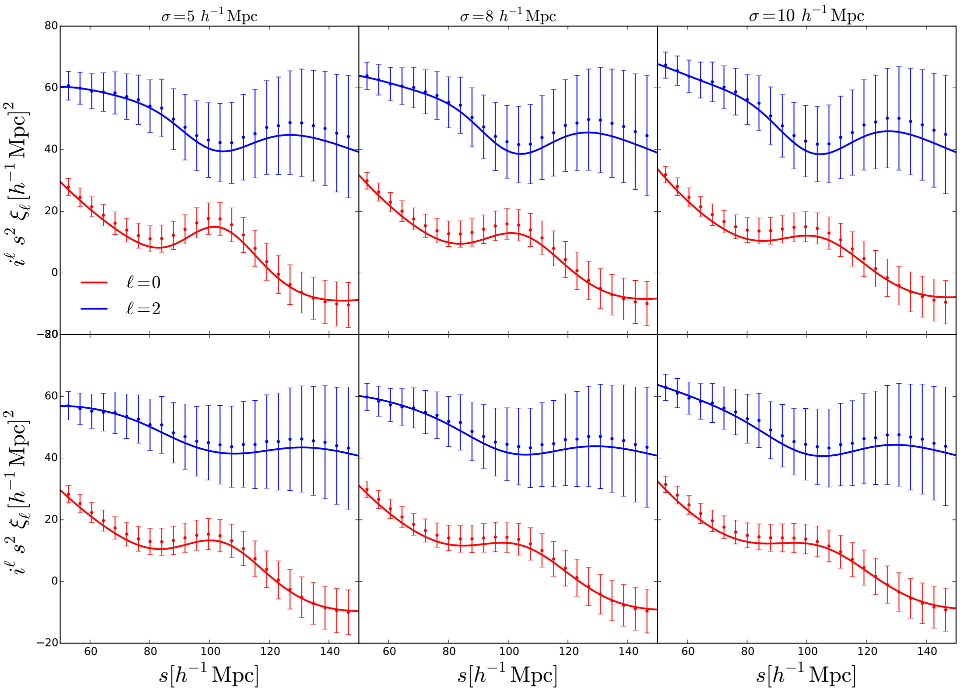

FIGURE 3.12: The best-fit to monopole and quadrupole correlation function for matter maps in redshift-space at $z = 0$ – reconstructed (top) and unreconstructed (bottom).

3.5.3 Impact of angular resolution on measured distances

Another useful parametrization of the position of the acoustic scale is in terms of dilations along the line of sight and perpendicular to the line of sight,

$$\alpha_{||} \equiv \frac{H_f r_{d,f}}{H r_d} \quad \text{and} \quad \alpha_{\perp} \equiv \frac{D_A r_{d,f}}{D_{A,f} r_d} \ , \tag{3.33}$$

which in real-space define $r^2 = \alpha_{||}^2 r_{||}^2 + \alpha_{\perp}^2 r_{\perp}^2$. A nice property of this parametrization is that it is linear in the cosmological parameters one wants to measure and therefore easier to interpret. The relation to the α and ϵ previously defined reads (see equation 3.14)

$$\alpha = \alpha_{||}^{1/3} \alpha_{\perp}^{2/3} \quad \text{and} \quad \epsilon = \left(\frac{\alpha_{||}}{\alpha_{\perp}}\right)^{1/3} . \tag{3.34}$$

As discussed in Section 3.4, the effect of angular resolution is to further smooth the field perpendicularly to the line of sight, hence we expect the constraints on the

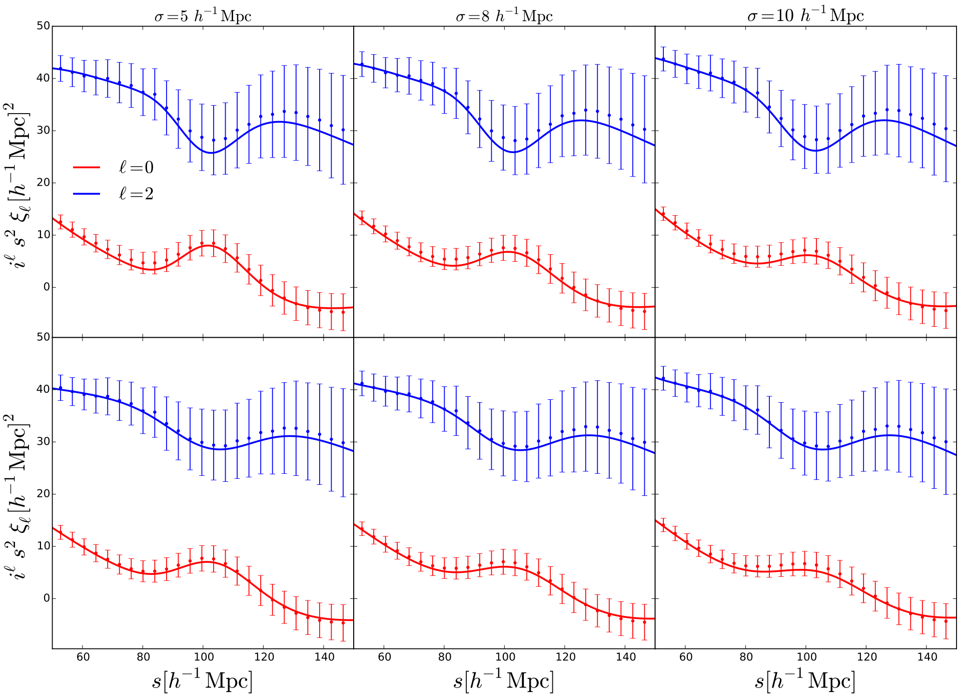

FIGURE 3.13: The best-fit to monopole and quadrupole correlation function for matter maps in redshift-space at $z = 1$ – reconstructed (top) and unreconstructed (bottom).

angular diameter distance to be more affected by the value of σ than the Hubble parameter (this has been extensively discussed in [125]). This is shown in Figure 3.14 where we plot the constraints on $\alpha_{||}$ and $\alpha_{\perp}$ for dark matter in redshift space. While $\alpha_{||}$ benefit from reconstruction, both in terms of central value and 1σ error, almost independently of the additional angular smoothing, the same is not true for $\alpha_{\perp}$. At $z = 0$, and for large values of σ, the best fit value of $\alpha_{\perp}$ is still biased with respect to the true value even after reconstruction. This indicates that non-linear shifts of the BAO are not well captured by the reconstruction procedure when too many modes are missing. We also note that the error on $\alpha_{\perp}$ is not reduced much by reconstruction when the angular resolution is too low. At $z = 1$ the picture is somehow better, since change in the position of the acoustic peak induced by gravity are less important as one moves to higher redshift. However the gain in errorbars after reconstruction is only marginal. This result actually questions how well the BAO could be measured in the transverse direction by a 21 cm IM experiment (in [125] it was shown that SKA1-MID will not even detect the isotropic BAO peak at $z \geq 1$). For instance, an experiment like CHIME has an angular resolution at $z = 1$ comparable to our idealized single-dish case.

3.5.4 Halo maps

We measure the BAO shift parameter α in each of our 500 halo maps before and after reconstruction using the template models outlined in Section 3.4.3. In Table 3.3 we give the summary of the best-fit results for the isotropic BAO shift parameter α in real- and redshift-space at $z = 0$.

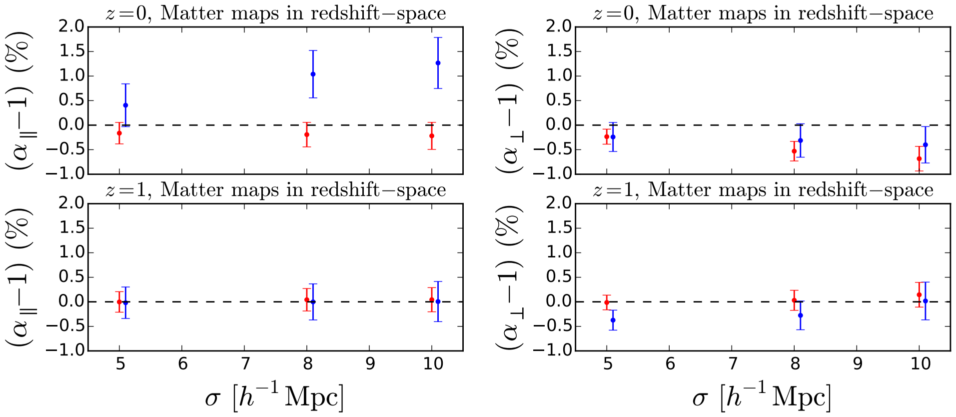

FIGURE 3.14: Mean best-fit values of BAO dilation parameters $\alpha_{||}$ and $\alpha_{\perp}$ in redshift-space as a function of the smoothing scale σ. Blue points represent the non-reconstructed, while red points represent the reconstructed halo maps results. The error bars shown are the standard deviation from 500 simulations divided by $\sqrt{50}$ to show the uncertainty we would expect in a survey of $\sim 50\,[h^{-1}\mathrm{Gpc}]^3$ volume size.

Halos – Real Space			
$\sigma[h^{-1}\mathrm{Mpc}]$	Reconstruction	α	$\langle\chi^2\rangle$ /dof
5	no	1.002 ± 0.026	32.2/20
	yes	1.000 ± 0.016	27.0/20
8	no	1.004 ± 0.028	31.5/20
	yes	1.001 ± 0.018	27.7/20
10	no	1.005 ± 0.031	28.0/20
	yes	1.001 ± 0.020	26.2/20
Halos – Redshift Space			
$\sigma[h^{-1}\mathrm{Mpc}]$	Reconstruction	α	$\langle\chi^2\rangle$ /dof
5	no	0.999 ± 0.032	28.1/20
	yes	0.998 ± 0.019	26.3/20
8	no	0.998 ± 0.034	27.6/20
	yes	0.998 ± 0.022	26.3/20
10	no	0.998 ± 0.036	25.2/20
	yes	0.999 ± 0.024	24.5/20

TABLE 3.3: Fitting results for halo maps in real- and redshift-space at $z = 0$ for different map resolutions σ (column 1). Column 3 show the mean and the standard deviation of BAO shift parameter α. The mean χ^2/dof is given in column 4. The errors shown for α are the standard deviations over 500 simulations. The actual errors on the means of the parameter α are the listed values divided by $\sqrt{500}$.

Real-space

In Figure 3.15 we show the mean best-fit values of the shift parameter α as a function of the angular resolution σ in real-space (left panel). We find that our reconstruction method works well in real-space and decreases the uncertainties on the parameter α by roughly 40% compared to the unreconstructed case. The recovered shift parameter are consistent with the expected $\alpha = 1$ and are within the uncertainties for all angular resolutions considered.

The uncertainty on α shows an increases as we go to larger values of the angular resolution - σ. This is expected as we are here only considering the monopole in which the BAO peak gets less pronounced with larger angular resolution σ. We also expect that the constraints on α could get tighter and less σ-dependent if we were also able to use the information from the halo maps quadrupole.

In Figure 3.16 we show the best-fit to the halo monopole in real-space. The BAO peak in the monopole gets more pronounced after reconstruction, suggesting our reconstruction method is able to partially remove the non-linear effects that cause the smearing of the BAO peak. With higher angular resolution σ used, the monopole gets more broad, even in linear theory (as shown in upper panel in Figure 3.4) and the effect of reconstruction is not so evident anymore.

Redshift-space

Figure 3.15 shows the mean best-fit values of shift parameter α as a function of angular resolution σ in redshift-space (right panel). Similar to the results in real-space, we find the uncertainties on α after reconstruction decrease by 30% – for larger angular resolution, up to 40% – for smaller angular resolution σ. Our recovered mean values of α after reconstruction are within 0.2% and consistent with the expected value 1. The biases we find are at the level of 3σ for the full 500 simulations volume.

In Figure 3.17 we show the best-fit to halo correlation function in redshift space. Similar to the results in real-space, the BAO peak in the monopole gets more pronounced after reconstruction, suggesting our reconstruction method is able to partially remove the non-linear effects that cause the smearing of the BAO peak in redshift-space too. Using higher angular resolution scales, the monopole gets less pronounce.

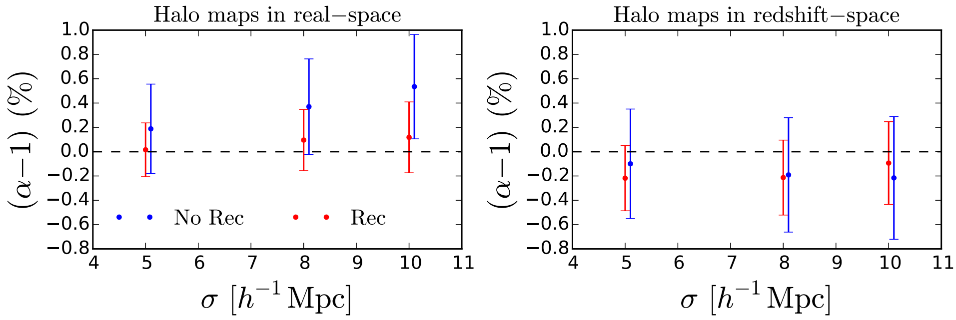

FIGURE 3.15: Halo maps at $z = 0$: Mean best-fit values of BAO shift parameter α in real- (left panel) and redshift-space (right panel) as a function of the angular resolution σ. Blue points represent the non-reconstructed, while red points represent the reconstructed halo maps results. The error bars shown are the standard deviation from 500 simulations divided by $\sqrt{50}$ to show the uncertainty we would expect in a survey of $\sim 50\,[h^{-1}\mathrm{Gpc}]^3$ volume size.

3.6 Summary and Conclusions

Perturbations in the early Universe produced sounds waves, called baryon acoustic oscillations, that propagated in the baryon-photon plasma until the recombination epoch. This phenomenon left its signature on the spatial distribution of matter and galaxies in the Universe as a peak (or set of wiggles) in the correlation function (power

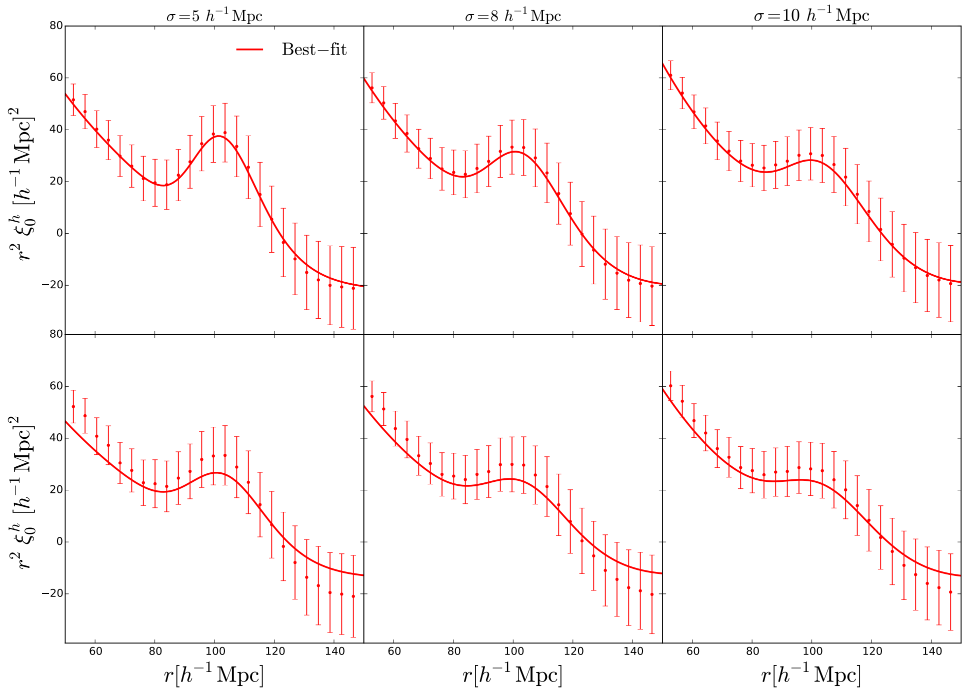

FIGURE 3.16: The best-fit to monopole of the correlation function for halo maps maps in real-space at $z = 0$ – reconstructed (top) and unreconstructed (bottom).

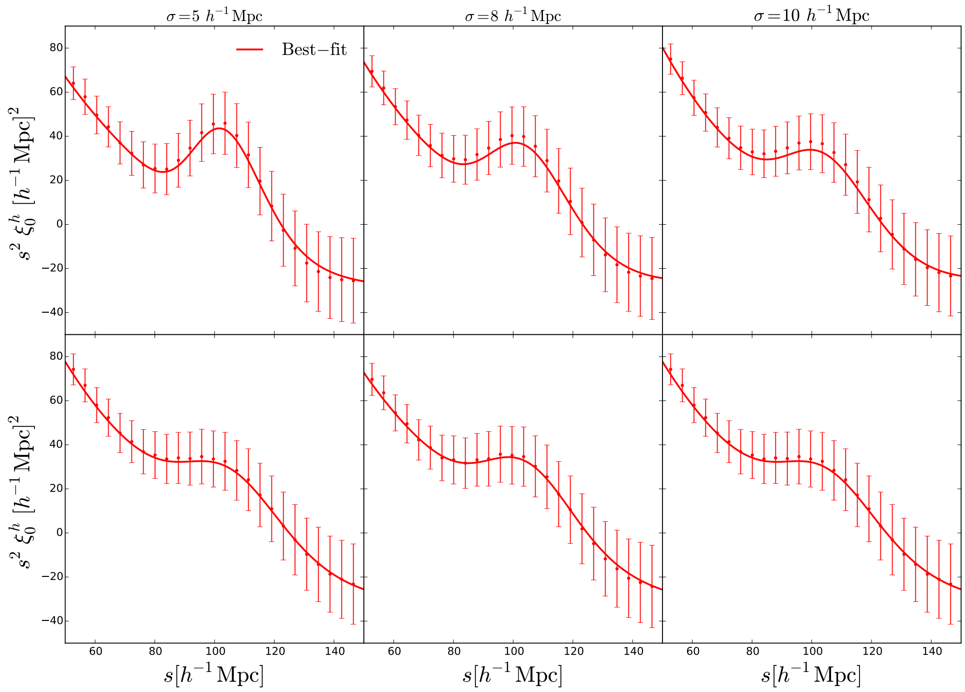

FIGURE 3.17: The best-fit to monopole of the correlation function for halo maps in redshift-space at $z = 0$ – reconstructed (top) and unreconstructed (bottom).

spectrum) that can be used as a standard ruler. The position of the BAO peak is very well constrained by CMB experiments, and by measuring it from low redshift cosmological probes (such as galaxy catalogues or 21cm IM observations) constraints on the value of the cosmological parameters, and therefore on the nature of dark energy, can be set.

BAO are extremely robust cosmological probes with respect to systematic effects. Unfortunately, non-linear gravitational evolution tends to smear out that feature by inducing a damping and broadening on the BAO peak and produces a shift in its position. Those effects will induce a systematic bias on the derived value of the cosmological parameters (due to the peak shift) and will increase the error bars on those since the peak position will be less clear. *Reconstruction* is a technique developed to undo (at least partially) the effects of non-linear gravitational evolution.

The ultimate goal of standard reconstruction methods is to place galaxies in their initial positions. This is partially achieved by moving back galaxies by estimating the amplitude of the underlying density field and using the Zel'dovich approximation to compute displacement field [38]. This procedure has proven to be very successful and the constraints on the value of the cosmological parameters have improved after applying this technique [151, 144].

However, there are cosmological observations that do not produce as output galaxy catalogues, but pixelated maps; an example of this kind of observations is a 21cm IM survey. While the power spectrum or correlation function inferred from these observations will be affected by non-linearities, in the same way galaxy surveys are, it is not obvious, a-priori, the way reconstruction should be performed on those density maps.

In this chapter we have tested a new BAO reconstruction method that consists in moving pixels instead of galaxies. We work on a regular grid to compute the displacement field and then treat the grid cells as galaxies in the standard reconstruction. By doing this we avoid two interpolations of the displacement field – one for the particles/galaxies and one for the uniform field of particles to compute the shifted field. Having the grid cells small enough, we recover the results from the standard method.

The main features of this method are:

- It can be applied to both galaxy surveys and pixelated maps (e.g. 21cm IM observations).
- In the limit of very small pixels it is equivalent to standard reconstruction method.
- It is faster and easier to implement than the standard method.

We have tested this method against the standard one in the case of matter density field in real-space on a large set of large box-size numerical simulations. We varied the smoothing scale for the displacement field across a wide range and we find that the methods agree. We find no significant difference between the methods in the constraints on the position of the BAO peak that we recover.

We have then applied this method to the spatial distribution of matter and halos in both real- and redshift-space using the same numerical simulations. In all cases we take into account the pixelated nature of the observations by creating mock maps that we obtain by convolving the simulating field (matter or halos) with a 2-dimensional Gaussian beam that mimic the effect of the telescope primary beam.

We use a theoretical template that parametrizes the effect of non-linearities on the broadening of the BAO peak, redshift-space distortions, FoG effect and basic instrumental effects such as the telescope beam size (embedded into the map resolution in our analysis) to fully model the measured correlation functions.

Our findings can be summarized as follow:

- By reconstructing maps created from the spatial distribution of matter in real-space at $z = 0$ and $z = 1$ we find that the recovered values on α and ϵ are compatible with those expected, $\alpha = 1$ and $\epsilon = 0$, respectively. At $z = 0$ the errors on α and ϵ decrease by $\sim$ 50% with respect to the case without reconstruction. We find the relative decrease in errors after reconstruction on both shift parameters vary within 5% and shows no evident dependence on the map resolution. At $z = 1$ the error on ϵ decreases by $\sim 30\,\%$, while the errors on α decrease by 30% after reconstruction for smaller values of the smoothing scale, while it decreases by 15% for larger values of the smoothing scale.

- By reconstructing maps created from the spatial distribution of matter in redshift-space at $z = 0$ and $z = 1$, we find that, at the level of expected precision of a future 21cm survey, the recovered values on α and ϵ are compatible with those expected, $\alpha = 1$ and $\epsilon = 0$. At $z = 0$ the error on α error decreases by a 40-50% with respect to the case without reconstruction, while the error of ϵ decreases by 35-45% after reconstruction. We find the relative decrease in errors depends on the angular smoothing scale and the performance of reconstruction is more effective for smaller smoothing scales. At $z = 1$ errors decrease by 30% and 50% for α and ϵ, respectively, after reconstruction, and we find no significant dependence on the angular smoothing scale used.

- Using a different parametrisation of the BAO peak shifts $(\alpha_{||}, \alpha_{\perp})$, we see more clearly the effect of low angular resolution on the constraints on the BAO peak position along the line-of-sight and in the transverse direction. Our reconstruction of matter maps in redshift-space recovers the expected values of $\alpha_{||}$ at the level of expected precision of a future 21cm survey and provides better constraints at both $z = 0$ and $z = 1$ with almost no dependence on the angular smoothing scale σ. In the case of $\alpha_{\perp}$, we find that with increasing angular smoothing scale, relative gains of reconstruction get smaller at $z = 0$, while the situation is somewhat better at $z = 1$.

- By reconstructing maps created from the spatial distribution of halos in real-space at $z = 0$ we find that the recovered value on α is compatible with 1. The error on the α after reconstruction decreases by a 40% for the smallest smoothing scale, while it decreases by 30% for the largest smoothing scale we used. We find the relative improvement on the constraints of α after reconstruction to depend on the smoothing scale.

- By reconstructing maps created from the spatial distribution of halos in redshift-space at $z = 0$ we find that the recovered value on α is compatible with 1. The error on α after reconstruction decreases by 40% with respect to the case without reconstruction for the smallest, while the decrease is 30% for the largest smoothing scale used.

Chapter 4

High-redshift post-reionization cosmology with 21cm IM

Introduction

We have already seen in chapter 3 some of the current and near future 21cm IM surveys. At the same time galaxy surveys such as DESI [160], Euclid [161], LSST [162], WFIRST [163] will detect millions of galaxies up to $z \simeq 2$ and will allow us to tightly constrain the value of the cosmological parameters.

The redshift window $2.5 < z < 5$ remains however vastly unexplored on cosmological scales. DESI will measure the BAO in the three dimensional clustering of the Lyα-forest in the range $2 < z < 3.5$ remarkably well, with a 1-2% precision, but still far from the cosmic variance limit, and HETDEX [164] will observe Lyα emitting galaxies over 400 square degrees between $1.9 < z < 3.5$ probing both BAO and redshift space distortions (RSD). Indeed, carrying out galaxy redshift surveys at high-redshifts becomes increasingly more difficult, as galaxies become fainter and fainter. IM can survey these redshifts without any limitation, and therefore, assuming its technical difficulties can be overcome, it offers a natural solution to map the high-redshift Universe. For these and other reasons, a hypothetical 21cm survey at high-redshift has been considered by the Cosmic Vision program [165] as one the five possible experiment for the next generation of cosmological surveys.

Although in most models the relative contribution of dark energy to the energy content of the Universe decreases with redshift, it is important to carry out observations at redshifts $z \geqslant 2$ in order to probe the consistency of the model. Alternatively, extensions of the models as for example early dark energy still provide viable solutions to the cosmological constant problem and could be exquisitely probed with IM of the 21cm line [166, 167, 168]. The nature of dark matter can be tested by using IM [169]. Another possibility is to constrain extensions of the vanilla ΛCDM cosmology, e.g. to massive neutrinos and curvature. More generally, measurements at high-redshift could increase our lever arm to constrain any deviation from the standard paradigm.

At high-redshift the modeling of RSD is also more robust, as the dark matter field is more linear on a cosmological scales. Unfortunately, a major limitation of all IM measurements is the interpretation of the redshift-space clustering. This is due to the conversion factor from the measured brightness temperature of the 21cm line to the underlying dark matter distribution, which is related to the mean distribution of neutral hydrogen in the Universe. Exploiting the full potential of 21cm IM, would therefore require external information about this astrophysical parameter.

In this work we want to study how much the cosmological constraints of a 21cm survey will be affected by uncertainties in the astrophysics of neutral hydrogen. We will discuss novel ways of tackle this problem, and then focus our attention to the performances of a hypothetical 21cm survey in the redshift range $2.5 < z < 5$. One of the main goals is to find the cosmological parameters that will mostly benefit from a combination of CMB probes, galaxy survey and 21cm observations.

This chapter is organized as follows. In section 4.1 we present the modelling of the 21cm signal and of the noise along with the Fisher matrix formalism employed for this work. The characteristics of the 21cm IM surveys considered are outlined in section 4.1.2. The constraints on the growth rate are shown in section 4.2.1. In sections 4.3.1 and 4.3.2 we discuss the constraints on two possible extensions of the baseline cosmological model: neutrino masses and effective number of neutrinos, respectively. We finally outline the main conclusions of this work in section 4.4.

4.1 Method

Estimates of the performance of a given survey in constraining cosmological parameters are usually carried out using the Fisher Matrix formalism [170]. Its main ingredients are a model for the signal and the noise one would like to measure. In our case we need to define the 21cm power spectrum, $P_{21}(k, z)$, and the noise level of a given configuration of radio antennas.

The fiducial cosmology we use is the Planck 2015 [6] best-fit ΛCDM: $\Omega_{\rm M} = 0.3075$, $h = 0.6774$, $\Omega_{\rm K} = 0$, $n_{\rm s} = 0.9667$, $A_{\rm s} = 2.142 \times 10^{-9}$, $\sigma_{8,0} = 0.828$, $\Omega_\nu = 0$. For massive neutrinos we assume a fiducial value of $\Sigma m_\nu = 0.06$ eV, and for the number of relativistic degrees of freedom $N_{\rm eff} = 3.046$. In all the calculations with non-zero Σm_ν we keep the total matter fraction $\Omega_{\rm M} = \Omega_{\rm cdm}+\Omega_{\rm b}+\Omega_\nu$ fixed when varying Σm_ν. We do this by including the neutrinos as the part of CDM fraction and decreasing the fiducial CDM fraction $\Omega_{\rm cdm}$ by $\Omega_\nu = \Sigma m_\nu/(94.07\,{\rm eV}h^2)$[1]. Throughout the whole analysis we keep the baryon fraction $\Omega_{\rm b}$ fixed. Power spectra have been computed using CAMB [16].

4.1.1 21cm signal model

We use equation 2.1 for the 21cm power spectrum model. In principle, due to the non-linear effects that tend to damp the BAO peak (see section 1.2.8), equation 2.1 should contain an additional damping term given by equation 1.44. However, the damping scales in equation 1.44 decrease with redshift roughly as the linear growth factor. Since we will be considering high-redshift window, the effects of non-linearities are pushed to small scales. Furthermore, a way to ameliorate this effects is to perform a BAO reconstruction method on the observed non-linear galaxy distribution [38] or the HI intensity maps at low-redshifts (see [124] and chapter 3). Applying these techniques can effectively half the damping scales, pushing the effects of non-linearities to even smaller scales. In fact, we have explicitly checked that including these effects in the redshift range we consider does not make a significant difference in our results. For simplicity, we will not include the damping term in the model for the 21cm power spectrum we consider here.

One immediately sees in equation 2.1 the main difference with respect to a cosmological analysis of a galaxy power spectrum: any constraint on the amplitude of the fluctuations, σ_8, of RSD parameters, e.g. f, or a combination of the two, is completely degenerate with the amount of HI in the Universe. This quantity is very uncertain in the theory modeling, so any information on it must come from independent datasets. If no external priors are available then RSD measurements with the 21cm line will

---

[1] Note that this expression is slightly different from equation 1.46. We use it in this chapter in order to be consistent with the way that this quantity is defined in the CAMB code [16, 171]

not be competitive with galaxy surveys. On the other hand, constraints on the BAO scales are still safe, as they are mostly independent of the broadband shape and normalization of the power spectrum.

The good news is that several measurements of the HI mean density are available nowadays [82, 91, 81, 172]. They mostly come from the detection in quasars (QSO) spectra of Damped Lyα (DLA) systems. DLAs are objects with high column density, $N_{\rm HI} > 10^{20.3}\,{\rm cm}^{-2}$, and contain more than the 90% of the total HI in the Universe [91, 76]. The abundance of DLAs gives a direct estimate of $\Omega_{\rm HI}(z)$, with current errorbars around 5% at $z < 3.5$ and around 30% at $z > 3.5$ [see e.g. 82]. In both cases there is a lot of margin for improvement, as existing low redshift measurements are limited by the noise level in the spectra, whereas at high-redshift by the low numbers of QSO spectra publicly available. In the near future surveys like DESI and WEAVE [173] will collect hundreds of thousands of QSO spectra, 10 times more than current surveys, yielding much better estimates of the abundance of HI. Motivated by this, in the remaining of this work we will assume three different priors on $\Omega_{\rm HI}$: 2%, 5%, 10%. We take the above priors to be redshift independent, although one could also imagine a prior degrading with increasing redshift.

The other astrophysical parameter which is quite degenerate with cosmological parameters is the HI linear bias $b_{\rm HI}$. In particular, any measurement that can break the $\Omega_{\rm HI} - b_{\rm HI}$ would be of great value for parameter estimation with the 21cm power spectrum. Naive cross-correlation with other kinds of tracers will not help, and weak lensing is not a particularly viable option since it probes the $k_{||} \to 0$ limit of the density field which for 21cm surveys is the most affected by foreground contaminants [174, 119].

What we propose, based on the arguments in [76], is to measure the clustering of the same objects one uses to measure $\Omega_{\rm HI}$, e.g. Lyman-limit systems (LLS) and DLAs, but weighting each of them by the value of its column density. This weighted density field can then be cross-correlated with the Lyman-α forest or the 21cm field itself to partially break the $\Omega_{\rm HI} - b_{\rm HI}$ degeneracy. We stress that such a measurement can already be done with existing datasets, yielding the first ever measurement of the bias of the HI field. As indeed shown in [76] the bias of the DLAs and of the HI can be written as

$$b_{\rm HI}(z) \;=\; \frac{c}{H_0}\frac{\int_0^\infty b(M,z)n(M,z)dM \int_0^\infty d\sigma N_{\rm HI}}{\int_0^\infty f(z,N_{\rm HI})N_{\rm HI}\; dN_{\rm HI}} \tag{4.1}$$

$$b_{\rm DLA}(z) \;=\; \frac{c}{H_0}\frac{\int_0^\infty b(M,z)n(M,z)dM \int_{\sigma(10^{20.3})}^\infty d\sigma}{\int_{10^{20.3}}^\infty f(z,N_{\rm HI})\; dN_{\rm HI}} \tag{4.2}$$

where $b(M,z)$ is the halo bias, $n(M,z)$ is the halo mass function,

$$\mathrm{d}\sigma \equiv \mathrm{d}N_{\rm HI}\frac{\mathrm{d}\sigma}{\mathrm{d}N_{\rm HI}}(N_{\rm HI}|M,z)$$

is the cross-section of systems with column density $N_{\rm HI}$, i.e. the probability of observing an object with $N_{\rm HI}$ in a halo, and $f(z,N_{\rm HI})$ is the column density distribution function, i.e. the abundance of such systems. We see that the difference between the two is just an extra factor $N_{\rm HI}$. This means that if we weight DLAs by the value of their column density and then cross-correlate with another tracer would yield an estimate of the HI bias. The lower integration limit in equation 4.2 is not important in this discussion, since most of the neutral hydrogen, 95% or more, lives in objects with $N_{\rm HI} \simeq 10^{21}{\rm cm}^{-2}$ [91, 76], and the weighing by $N_{\rm HI}$ will make the integrals above completely dominated by the tail of the column density distribution function. In the

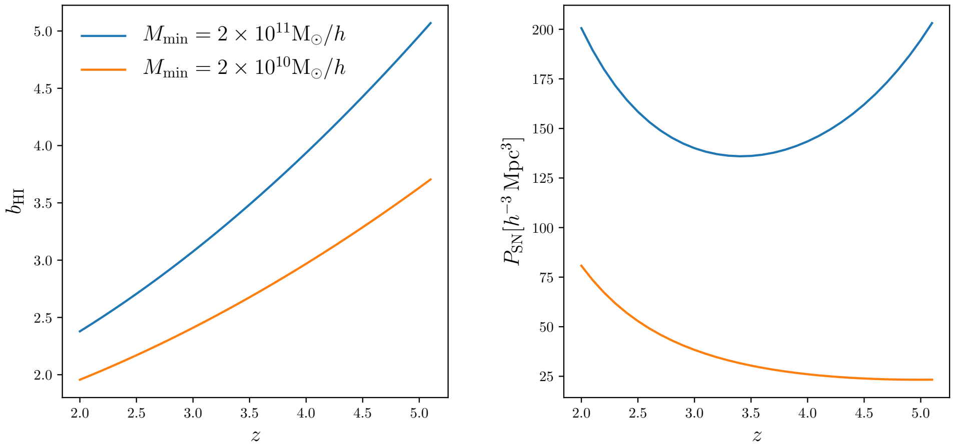

FIGURE 4.1: The figure shows the dependence on redshift of the HI bias (left) and the HI shot-noise (right) for two different models of the HI distribution (see text).

limiting case the bias of DLAs does not depend on this lower threshold, i.e. the DLA cross-section does not depend on halo mass, then $b_{\rm DLA} = b_{\rm HI}$. Recent measurements in [175] show a slight statistical preference for the latter situation.

The cross-correlation of the weighted DLA field, $\delta^w_{\rm DLA}$, with the Lyman-α forest flux, δ_F, will therefore return in the linear regime

$$\langle\delta^w_{\rm DLA}\delta_F\rangle = b_{\rm HI}b_F P(k)(1+f\mu^2/b_{\rm HI})(1+b_v f\mu^2/b_F) \tag{4.3}$$

with b_F and b_v the density and velocity bias of the forest. Projected correlation function will get rid of most of the RSD contribution allowing a clear determination of $b_{\rm HI}$. Recent measurements of the cross-correlation between the forest and DLAs presented in [176, 175] provide an estimate of $b_{\rm DLA} \simeq 2$ with a 5-10% error depending on different data cuts, which in the limit the DLA cross-section does not depend on halo mass exactly yields a measurement of HI bias, see equation 4.2. Another option would be to cross-correlate $\delta^w_{\rm DLA}$ with the 21cm field itself,

$$\langle\delta^w_{\rm DLA}\delta_{21}\rangle = \bar{T}_b b^2_{\rm HI}P(k)(1+f\mu^2/b_{\rm HI})^2, \tag{4.4}$$

and again we see that the degeneracy between $\Omega_{\rm HI}$ and the bias can be partially broken by combining different probes.

We expect next generation of galaxy surveys to drastically improve on these numbers. We will therefore assume, as we did above for the cosmic neutral fraction, three different redshift independent priors on $b_{\rm HI}$: 2%, 5%, 10%

In a Fisher analysis we also need to specify the fiducial value of both the cosmological and astrophysical parameters. For the latter we use the halo model for the HI explained in chapter 3. In this chapter we take $\alpha = 1$ and consider two values of $M_{\rm min} = 2\times10^{10}h^{-1}M_\odot$ and $2\times10^{11}h^{-1}M_\odot$. The overall normalization parameter, M_0, is chosen such that $\Omega_{\rm HI}(z) = 4\times10^{-4}(1+z)^{0.6}$ [82]. The values of the free parameters are motivated by the poorly understood redshift evolution of the HI fraction and the HI bias, both in simulations and data [53, 177, 128, 178, 105, 82]. The HI bias and the HI shot-noise are then computed following the equations given in section 2.1.1. In figure 4.1 we show the fiducial value for the bias and the effective HI shot-noise – $P_{\rm SN}$, for the two different choices of minimum mass $M_{\rm min} = 2\times10^{11}$ $M_\odot/h$ (blue) and $M_{\rm min} = 2\times10^{10}$ $M_\odot/h$ (orange).

4.1.2 Noise model and 21cm IM surveys

Current and planned 21cm IM surveys focus either on the reionization epoch at $z > 6$ (e.g. HERA, PAPER, SKA) or on the low-redshift Universe $z < 2.5$, targeting the BAO feature in the 21cm power spectrum (e.g. CHIME, Tianlai, HIRAX, FAST, SKA1-MID). Some surveys (e.g. SKA1-LOW) will cover both redshift ranges but at the price of low angular resolution [66], while OWFA will focus on a single redshift [75]. Our analysis focuses on the range $2.5 < z < 5$, and therefore we first have to make a few choices on instrument specifications. We consider two approaches: interferometer instruments (similiar to CHIME and HIRAX) and single-dish telescopes (similar to FAST). We modify the configuration of both types of instruments such that they can probe higher redshifts. The main requirement is to keep angular resolution good enough for the scales we want to probe, as a fixed physical distance subtend a smaller angular scale at higher redshift. This usually means we have to increase the size of the longest baseline in an interferometer or the size of the dishes in auto-correlation mode. We dub the interferometers as Ext-CHIME and Ext-HIRAX, while we dub the single-dish with highzFAST. Table 4.1 shows the characteristics of Ext-CHIME, Ext-HIRAX and highzFAST.

	Ext-CHIME	Ext-HIRAX	highzFAST
T_a(K)	10	10	10
L_cyl(m)	200	-	-
W_cyl(m)	20	-	-
N_cyl	10	-	-
N_f	256	-	-
D_dish(m)	-	10	500
N_dish	-	64×64	1
η	0.7	1.0	-
D_max(m)	269	800	-
D_min(m)	$L_\mathrm{cyl}/N_\mathrm{f}$	10	-
N_b	2560	-	-
S_{21} (deg^2)	25000	25000	5000
n_pol	2	2	2
$\Delta\nu$(kHz)	-	-	50

TABLE 4.1: Main characteristics of the considered surveys.

Thermal noise power spectrum - Interferometers

We model the thermal noise power spectrum of an interferometer as in [64],

$$P_\mathrm{N}^\mathrm{th}(z) = \frac{T_\mathrm{sys}^2(z) X^2(z) Y(z) \lambda^4(z) S_{21}}{A_\mathrm{eff}^2 \mathrm{FOV}(z) t_0 n_\mathrm{pol} n(\mathbf{u}, z)}, \tag{4.5}$$

which we use for Ext-CHIME and Ext-HIRAX. The system temperature is the sum of antenna and the background sky temperature $T_\mathrm{sys}(z) = T_\mathrm{a} + T_\mathrm{sky}(z)$. We take the antenna temperature $T_\mathrm{a} = 10\,\mathrm{K}$ and we model the sky temperature as $T_\mathrm{sky}(z) = 60\,\mathrm{K} \times (\nu_{21}(z)/300\mathrm{MHz})^{-2.55}$[64]. While the background sky temperature is usually below antenna temperature at low-redshifts, going to higher redshifts it increases significantly and represents the major noise contribution.

The terms $X(z)$ and $Y(z)$ are used to convert from angular and frequency space to physical space. $X(z)$ is the comoving distance and $Y(z) = \frac{c(1+z)^2}{\nu_{21} H(z)}$, where c is the speed of light, $H(z)$ is the Hubble parameter, and ν_{21} is the frequency of the 21cm line (1420 MHz). The wavelength of the 21cm line at a given redshift is $\lambda(z) = \lambda_{21}(1+z)$. We assume a survey area $S_{21} = 25000\,\mathrm{deg}^2$, the observing time of $t_{\mathrm{obs}} = 10000$ hours and the number of polarisation states $n_{\mathrm{pol}} = 2$ for both Ext-CHIME and Ext-HIRAX.

The last term in the denominator is the number density of baselines in visibility space. We will assume a uniform distribution of $n(u;z) = N_{\mathrm{b}}(N_{\mathrm{b}}-1)/2\pi/(u^2_{\mathrm{max}}(z) - u^2_{\mathrm{min}}(z))$ between $u_{\mathrm{min}} = D_{\mathrm{min}}/\lambda(z)$ and $u_{\mathrm{max}} = D_{\mathrm{max}}/\lambda(z)$. We will use an approximation of this expression $n(u;z) \simeq N_{\mathrm{b}}^2/2\pi/u^2_{\mathrm{max}}(z)$. The total number of beams N_{b} is the total number of antenna feeds $N_{\mathrm{f}} \times N_{\mathrm{cyl}}$ in the case of Ext-CHIME, while for Ext-HIRAX it is the number of dishes N_{dish}.

Ext-CHIME The effective area per feed is computed as $A_{\mathrm{eff}} = \eta L_{\mathrm{cyl}} W_{\mathrm{cyl}}/N_{\mathrm{f}}$, where η is the efficiency factor that we take $\eta = 0.7$; L_{cyl} and W_{cyl} are the length and the width of the cylinders, respectively, and N_{f} is the number of feeds per cylinder. While the primary beam of each element in this set-up is anisotropic, we will assume an isotropic field of view (FOV) for each element and we will use $\mathrm{FOV} \approx 90^\circ \times 1.22\lambda(z)/W_{\mathrm{cyl}}$ [64].

Ext-HIRAX The effective area per beam is $A_{\mathrm{eff}} = \pi(D_{\mathrm{dish}}/2)^2$ where D_{dish} is the diameter of a single dish. The primary beam of each dish is isotropic and defined as $\mathrm{FOV} = \left(1.22\frac{\lambda(z)}{D_{\mathrm{dish}}}\right)^2$ [65].

Window functions for interferometers

In the case we consider – uniform number density of baselines – the noise power spectrum is scale-independent. Nevertheless, this will not be the real case and there will be a range of modes that are going to be (un)observable. To account for this we include sharp cut offs in the k-range probed by the survey, both in $k_{\|}$ and $k_{\perp}$. We take the limits for $k_{\perp}$ according to [64]:

$$k_{\perp}^{min}(z) = k_{D_{min}}(z) = \frac{2\pi D_{min}}{D(z)\lambda(z)}, \tag{4.6}$$

$$k_{\perp}^{max}(z) = k_{D_{max}}(z) = \frac{2\pi D_{max}}{D(z)\lambda(z)}, \tag{4.7}$$

where $D(z)$ is the comoving distance. The values for D_{min} and D_{max} depend on the configuration we consider and are given in table 4.1. For $k_{\|}(\mathrm{z})$ we take the upper limit $k_{\mathrm{max}}(z)$ to be $0.2\,h\mathrm{Mpc}^{-1}$ or the non-linear scale $k_{\mathrm{nl}}(z) = k_{\mathrm{nl},0}(1+z)^{2/3}$, where $k_{\mathrm{nl},0} = 0.2\,h\mathrm{Mpc}^{-1}$ [101]. We take the lower limit $k_{\|}^{\mathrm{min}}(\mathrm{z}) = 2\pi/V_{\mathrm{sur}}^{1/3}(z)$. Based on these, we can now define a window function that defines the range of scales available to a given configuration.

$$W(z,k,\mu) = \Theta(k_{D_{max}} - k_{\perp})\Theta(k_{\perp} - k_{D_{min}})\Theta(k_{\|} - k_{\|}^{\mathrm{min}})\Theta(k_{\mathrm{nl}} - k\mu)\Theta(k_{\mathrm{nl}} - k),$$

where Θ is the Heaviside step function. We show the effect of this window function on the signal power spectrum in figure 4.2 in the case of Ext-HIRAX and in the bottom left panel of figure 4.3 in the case of Ext-HIRAX and Ext-CHIME.

The foreground wedge

The 21cm cosmological signal is several orders of magnitude weaker than the astrophysical foreground emissions. Isolating the cosmological signal will be the major challenge in the data analysis, and it mostly relies on the assumption that foregrounds are spectrally smooth. In principle this means that only the smallest radial wave-numbers $k_{\|}$ are affected by the foregrounds, allowing us to clean the data without much of the cosmological information being lost. However, the chromaticity of the interferometers response function will cause foregrounds to leak into the high-$k_{\perp}$ modes. This will make the foreground cleaning very difficult. This effect is limited to a certain region in $k_{\|} - k_{\perp}$ space, acting on low-$k_{\|}$ and high-$k_{\perp}$, and it is known in the literature as the foreground *wedge* [179, 180, 181, 182, 183, 184, 124]:

$$k_{\|} < \sin(\theta_{\rm FoV})\frac{D(z)H(z)}{c(1+z)}k_{\perp}, \tag{4.8}$$

where $\theta_{\rm FoV}$ is the field-of-view (or the primary beam width) of an interferometer element. We can also express the wedge region in terms of all the modes having $\mu < \mu_{\rm min}$ with:

$$\mu_{\rm min}(z) = \frac{k_{\|}}{\sqrt{k_{\|}^2 + k_{\perp}^2}} = \frac{\sin(\theta_{\rm FoV})D(z)H(z)/(c(1+z))}{\sqrt{1 + [\sin(\theta_{\rm FoV})D(z)H(z)/(c(1+z))]^2}}. \tag{4.9}$$

It is important to stress that the foreground wedge is not an intrinsic limitation of a 21cm interferometric survey, but rather a manifestation of our incomplete knowledge of the calibration of the antennas. Accurate calibration is a difficult task of any data analysis, in IM in particular because of the 10^5 difference in signal strength between the foregrounds and the signal one is trying to measure.

Unfortunately, this contaminated region grows at higher redshifts, as both the comoving distance and the Hubble parameter increase faster than $(1+z)$. For instance, at $z = 2$ the foreground wedge contaminates all wave modes with $\mu \leq \mu_{\rm min}(2) = 0.77$, and the situation gets much worse at $z = 4$ with $\mu \leq 0.9$. The simplest approach is to ignore the modes inside the wedge in data analysis (or in Fisher forecasts). This would mean that we discard 90% of the available modes at $z = 4$. Another possibility is to develop techniques that can recover as much as possible information inside the wedge [174, 119]. While these methods strongly depend on the details of the instrumental set up and of the data analysis procedure, in this chapter, for simplicity we will consider few idealised cases:

1. Wedge – we discard all the modes inside the wedge assuming a horizon limit, i.e. $\sin(\theta_{\rm FoV}) = 1$;
2. Mid-wedge – (i) in the case of Ext-HIRAX we use $\mu = \mu_{\rm min}$; (ii) in the case of Ext-CHIME where the primary beam is anisotropic we assume modes with $\mu = \mu_{\rm min}/1.5$ are unavailable, i.e. one third of the contaminated region is cleaned;
3. No wedge case – the foreground cleaning works perfectly for all the modes.

In figure 4.2 we show the effect of the wedge and the window function on the available modes in the case of Ext-HIRAX. In the left panel we show the effective to survey volume ratio as a function of k and μ. Available modes in the three cases we consider are above the solid lines of constant $\mu_{\rm min}$. We show the $\mu_{\rm min}$ in the wedge and mid-wedge case in solid blue and solid line, respectively. Similarly, in the right panel, we show the signal-to-noise ratio as a function of $k_{\perp}$ and $k_{\|}$, where the effect of

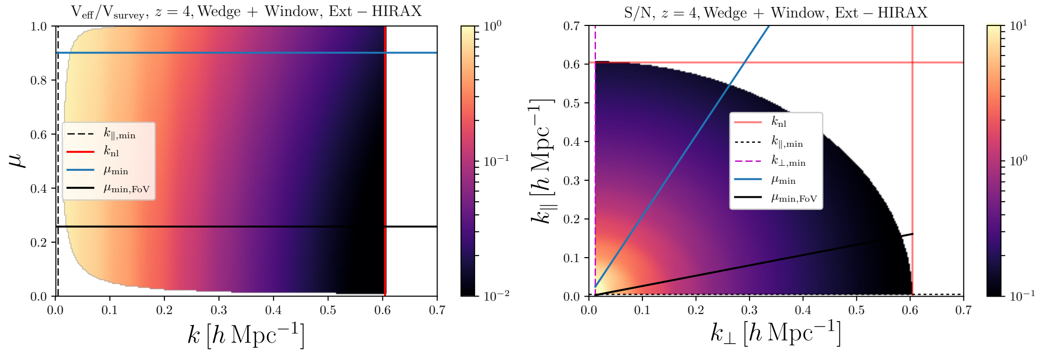

FIGURE 4.2: This plot shows the effect of the window function, wedge and k_{nl} on the number of modes available to extract cosmological information for Ext-HIRAX. The left panel displays the ratio between the effective volume to the survey volume as a function of k and μ. The right panel represents the signal-to-noise ratio as a function of $k_\perp$ and $k_\parallel$. For both panels we have considered a redshift bin $\Delta z = 0.1$ centered at $z = 4$ and assumed $k_{\max} = k_{\mathrm{nl}}(z)$. Unavailable modes are shown as white regions. The effect of the wedge can be seen from the solid black and blue lines. If no wedge is present all modes shown in the panels contribute, while only the modes above the black/blue line contribute in the case of full/mid wedge.

the wedge is more clearly seen. Again, solid blue and black line represent the wedge cases we consider and the available modes are on restricted on the left side of these lines. In both of the panels we use $k_{\max} = k_{\mathrm{nl}}$.

Thermal noise power spectrum - highzFAST

We model the noise power spectrum of a single dish as [185]

$$P_{\mathrm{N}}^{\mathrm{th}}(k, \mu, z) = \sigma_{\mathrm{pix}}^2(z) V_{\mathrm{pix}}(z) W^{-2}(k_\perp, z). \tag{4.10}$$

The first term is the pixel thermal noise given by

$$\sigma_{\mathrm{pix}}^2 = \frac{T_{\mathrm{sys}}^2}{\Delta\nu t_{\mathrm{obs}}(\Omega_{\mathrm{pix}}/S_{21}) N_{\mathrm{dish}} N_{\mathrm{beam}}}, \tag{4.11}$$

where T_{sys} is the system temperature described above, $\Delta\nu$ is the frequency resolution, t_{obs} is the total observing time, Ω_{pix} is the pixel area, S_{21} is the survey sky area, while N_{dish} and N_{beam} are the number of dishes and number of beams per dish, respectively. The pixel area is calculated assuming a Gaussian beam $\Omega_{\mathrm{pix}} = 1.13\theta_{\mathrm{FWHM}}^2$, where $\theta_{\mathrm{FWHM}} = \lambda(z)/D_{\mathrm{dish}}$. The volume of a pixel V_{pix} is computed by integrating the comoving volume element of the pixel area along the line-of-sight in the redshift range corresponding to the frequency resolution $\Delta\nu$:

$$V_{\mathrm{pix}} = \Omega_{\mathrm{pix}} \int_{z-\Delta z/2}^{z+\Delta z/2} dz' \frac{cD^2(z')}{H(z')}. \tag{4.12}$$

The last term in the noise power spectrum is the response function describing the angular resolution of a single-dish telescope. Since the frequency resolution we assume here is very high, we will not take into account the radial response and will only consider the angular one. We use

$$W^2(k_\perp, z) = \exp\left[-k_\perp^2 D_c^2(z)\left(\frac{\theta_{\rm FWHM}}{\sqrt{8\ln 2}}\right)^2\right]. \qquad (4.13)$$

For an instrument like the highzFAST the scales which will be accessible are

$$k_\perp^{min}(z) = \frac{2\pi}{\sqrt{D(z)^2 S_{21}}}, \qquad (4.14)$$

$$k_\perp^{max}(z) = \frac{2\pi D_{\rm dish}}{D(z)\lambda(z)}. \qquad (4.15)$$

We show the noise power spectrum of highzFAST at $z = 4$ in the bottom right panel of figure 4.3 in dashed line. The dashed gray line is the signal power spectrum, while the blue solid line includes the effect of the accessible modes.

4.1.3 Total noise power spectrum

The total noise power spectrum is therefore equal to

$$P_{\rm N}^{\rm tot} = P_{\rm N}^{\rm th}(z) + P_{\rm SN}(z)\overline{T}_b^2(z). \qquad (4.16)$$

We plot the different noise contributions for a survey like Ext-HIRAX at $z = 4$ in the top-left panel of figure 4.3. In the top-right panel we show the expected signal-to-noise ratio for different μ values and different values of $M_{\rm min}$ in the case of Ext-HIRAX. Bottom panels show the total noise power spectra for both interferometers (left panel) and single-dish (right panel) in dashed lines at $z = 4$ and fixed value of $M_{\rm min}$. As expected the thermal noise is the major source of uncertainties, especially going to high-redshift. We see that on cosmological scales smaller than $k \simeq 0.2\,h\,{\rm Mpc}^{-1}$ all surveys we consider become noise dominated. In the upper left panel we plot the signal-to-noise at $k \simeq 0.2\,h\,{\rm Mpc}^{-1}$ for different values of μ

$$nP_{0.2}(z) \equiv \frac{P_{21}(k,\mu;\,z)}{P_{\rm N}^{\rm tot}}. \qquad (4.17)$$

Since the shot-noise is always much smaller than the thermal noise in the antennas, a higher $M_{\rm min}$, i.e. a higher HI bias, results in a higher signal to noise ratio.

4.1.4 Fisher Matrix formalism

Given the signal and noise models discussed in the previous section we can use the Fisher matrix formalism to forecast the constraints on a set of parameters of interest $\{p\}$ from a 21cm IM survey. The Fisher matrix for a single redshift bin is given by [170]

$$F_{ij} = \frac{1}{8\pi^2}\int_{-1}^{1} d\mu \int k^2 dk \frac{\partial \ln P_{21}(k,\mu)}{\partial p_i}\frac{\partial \ln P_{21}(k,\mu)}{\partial p_j} V_{\rm eff}(k,\mu). \qquad (4.18)$$

The effective volume is related to the comoving survey volume in a redshift bin $V_{\rm sur}$ by

$$V_{\rm eff}(k,\mu) = V_{\rm sur}\left(\frac{P_{21}(k,\mu)W(k,\mu)}{P_{21}(k,\mu)W(k,\mu) + P_{\rm N}^{\rm tot}(k,\mu)}\right)^2. \qquad (4.19)$$

The instrument response/window function $W(k,\mu)$ is included to account for the modes that are unavailable to a given instrument. The total noise $P_{\rm N}^{\rm tot}$ accounts

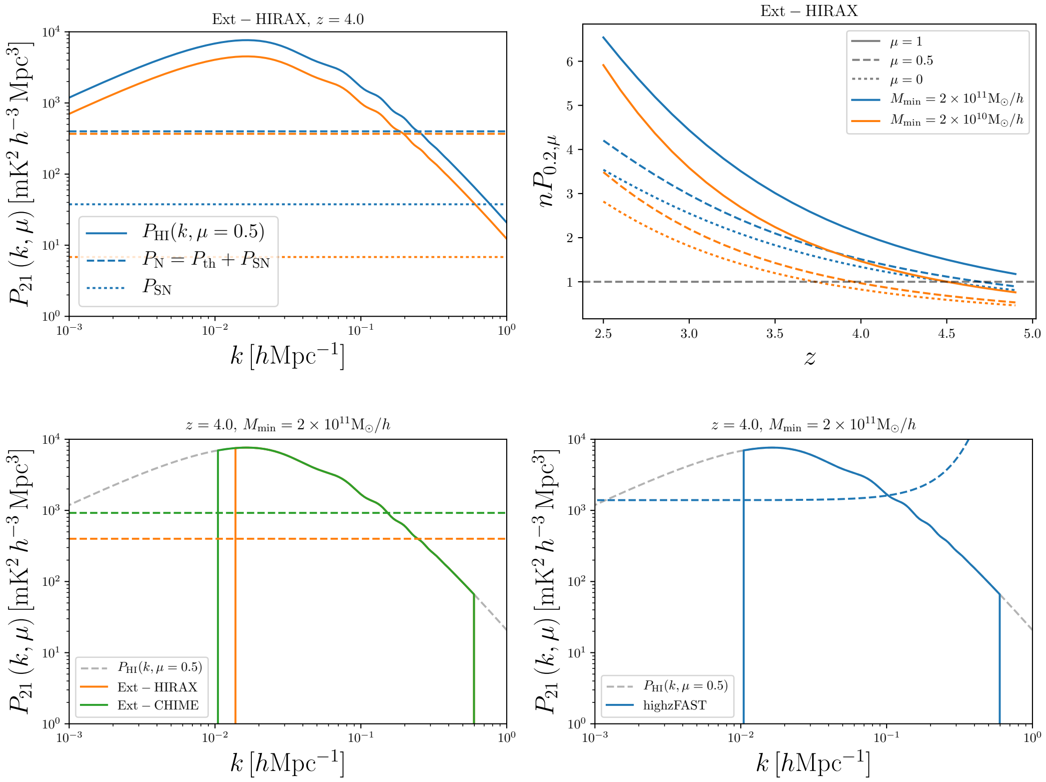

FIGURE 4.3: *Top-left:* HI power spectrum at $\mu = 0.5$ (solid lines) together with the total noise (dashed) and shot-noise (dotted) for Ext-HIRAX at $z = 4$. The different colors show the results for different values of $M_{\rm min} = 2\times10^{10}\ h^{-1}M_\odot$ (orange) and $2\times10^{11}\ h^{-1}M_\odot$ (blue). *Top-right:* Signal-to-noise ratio at $k = 0.2h{\rm Mpc}^{-1}$ for different values of μ, 1 (solid), 0.5 (dashed) and 0 (dotted), for the two models with different values of $M_{\rm min}$. *Bottom-left:* Effects of the window function on the HI power spectrum for $\mu = 0.5$ for Ext-HIRAX (orange) and Ext-CHIME (green) at $z = 4$. The dashed lines show the amplitude of the noise power spectrum. *Bottom-right:* Same as before but for highzFAST.

for the instrument thermal noise and the shot-noise of the HI sources as defined in equation 4.16.

The total redshift window of the 21cm IM surveys considered here, $2.5 < z < 5$, is divided into smaller redshift bins, in each of them we calculate the Fisher matrix for a set of parameters. Under the assumption that each redshift bin is independent, the sum of the Fisher matrices in each redshift bin gives the total Fisher matrix.

Having obtained the Fisher matrix for a given redshift bin or the total redshift window, the forecasted constraint on a given parameter p_i marginalised over all the other parameters is related to the Fisher matrix by $\sigma(p_i) = \sqrt{[F^{-1}]_{ii}}$. We quote these numbers as the 1σ constraints throughout the chapter.

In observations one needs to convert angular coordinates and redshifts to wave vectors in Fourier space. That is done by assuming a fiducial cosmology, which could be different from the true underlying one. By using a cosmology different from the fiducial one, we introduce an additional artificial anisotropy in the clustering measurements. This effect, known as Alcock-Paczynski(AP) effect [186], must be taken into account in the forecast. We include all the geometrical distortions arising from the assumption of a fiducial cosmology [187, 188]. The fiducial wavenumber vector $\vec{k}$ is determined by two numbers, namely, the amplitude k_f and by the cosine of the angle between the $\vec{k}$ and the line-of-sight, $\mu_f = k_{\parallel,f}/k_f$. The component along and transverse to the line-of-sight are related to the true values by $k_\parallel = k_{\parallel,f}(H/H_f)$ and $k_\perp = k_{\perp,f}(D_{A,f}/D_A)$. Furthermore the amplitudes of the wavenumber vectors $\vec{k}$ and $\vec{k}_f$ are related

$$k^2 = \left((1-\mu_f^2)\frac{D_{A,f}^2}{D_A^2} + \mu_f^2 \frac{H^2}{H_f^2} \right) k_f^2. \tag{4.20}$$

We write our full signal power spectrum in a redshift bin centred at redshift z as [189]:

$$P_{21}(k_f, \mu_f, z) = \overline{T}_b^2(z) \frac{D_A(z)_f^2 H(z)}{D_A(z)^2 H(z)_f} \left(b_{\rm HI}(z) + f(k,z)\mu^2 \right)^2 P(k,z). \tag{4.21}$$

We will also use another way of writing the full power spectrum, particularly in the case we are interested in the redshift-space distortions where an important parameter is $f\sigma_8$:

$$P_{21}(k_f, \mu_f, z) = \overline{T}_b^2(z) \frac{D_A(z)_f^2 H(z)}{D_A(z)^2 H(z)_f} \left(b_{\rm HI}\sigma_8(z) + f\sigma_8(z)\mu^2 \right)^2 \frac{P(k,z)}{\sigma_{8,\rm f}^2}. \tag{4.22}$$

Within our model, the derivative of the HI power spectrum with respect to a generic parameter X looks like

$$\begin{aligned} \frac{\partial \ln P_{21}}{\partial X} &= \frac{\partial \ln \overline{T}_b^2}{\partial X} + \frac{\partial \ln D_A^{-2}}{\partial X} + \frac{\partial \ln H}{\partial X} + \frac{\partial \ln P(k)}{\partial X} + \frac{\partial \ln P(k)}{\partial k}\frac{\partial k}{\partial X} \\ &+ \frac{2\mu^2}{b_{\rm HI} + f(k)\mu^2} \left[\frac{\partial f(k)}{\partial X} + \frac{\partial f(k)}{\partial k}\frac{\partial k}{\partial X} + 2f(k)(1-\mu^2)\left(\frac{\partial \ln H}{\partial X} + \frac{\partial \ln D_A}{\partial X} \right) \right], \end{aligned} \tag{4.23}$$

where

$$\frac{\partial k}{\partial X} = k \left[\mu^2 \frac{\partial \ln H}{\partial X} - (1-\mu^2)\frac{\partial \ln D_A}{\partial X} \right]. \tag{4.24}$$

The Fisher matrix formalism can also accommodate theoretical errors in the signal model, for instance due to poor knowledge of non-linearities, along the way described in [190, 191]. We do not include such terms in our analysis, but in Section 4.3.1 we will comment on the effect of marginalising over uncertainties in non-linear models using templates.

4.2 Results from IM alone

In this section we present the constraints on the value of the cosmological parameters that can be achieved by using 21cm data alone. We focus our analysis on the growth rate and on the BAO distance scale parameters.

4.2.1 Growth of structures

We will first consider the case in which we assume that the geometry is fixed, e.g. by CMB measurements, and we can fit for the amplitude of the power spectrum using two parameters only: $b\sigma_8(z)$ and $f\sigma_8(z)$. This choice is motivated by a comparison with the forecasts presented in [192, 160, 193], where BAO distance scale parameters

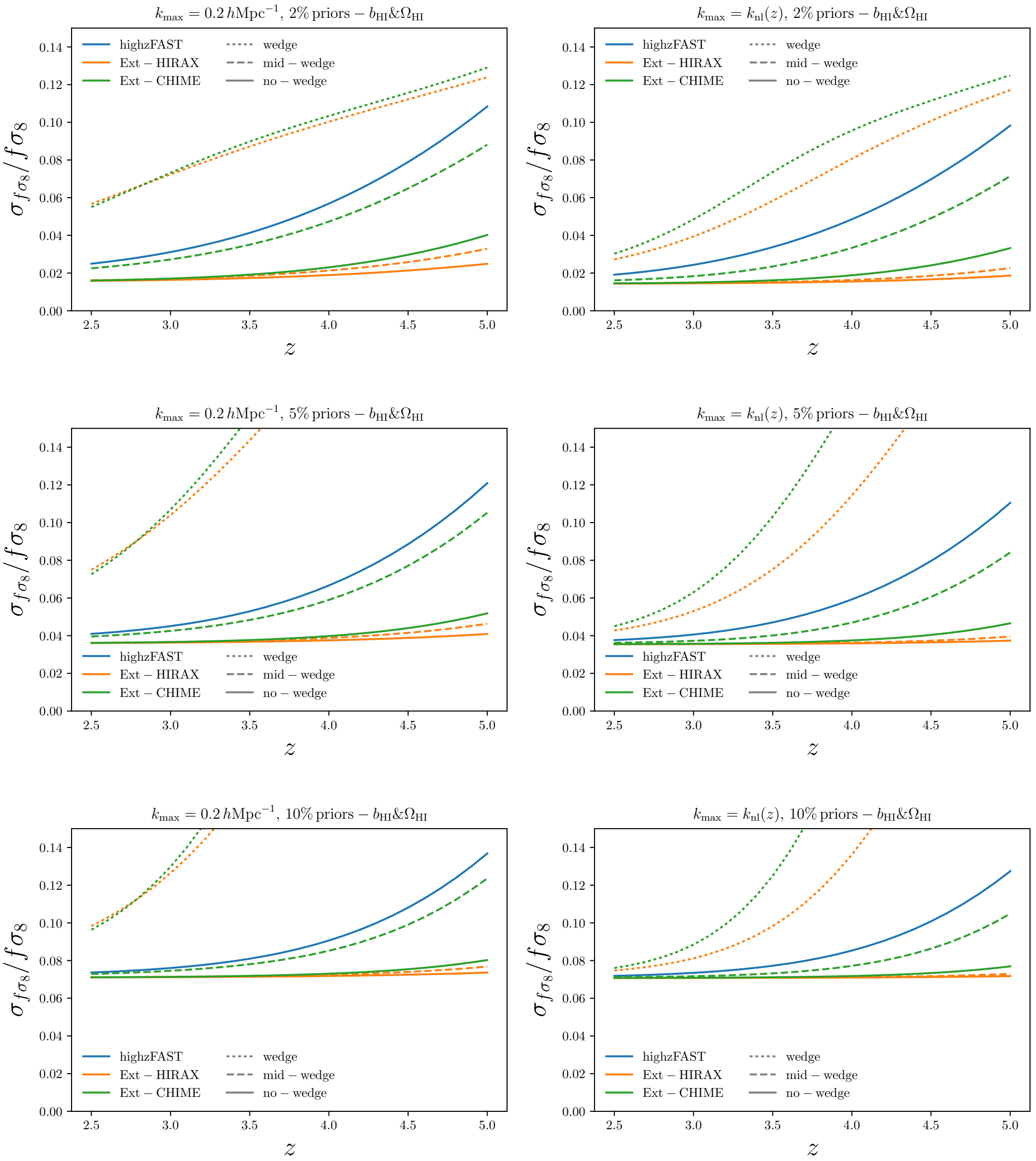

FIGURE 4.4: 1σ constraints on $f\sigma_8$ for highFAST (blue), Ext-HIRAX (orange) and Ext-CHIME (green) for different wedge configurations: no-wedge (solid), mid-wedge (dashed) and full wedge (dotted). Notice that the wedge only apply to interferometers. The different panels show the results for different assumptions of $k_{\rm max}$, 0.2 hMpc^{-1} (left column) and $k_{\rm nl}$ (right column), and for different priors on both $b_{\rm HI}$ and $\Omega_{\rm HI}$, 2% (top row), 5% (middle row) and 10% (bottom row).

have been held fixed when quoting constraints on RSD. Following the discussion in Section 4.1.1, such a measurement in an IM survey is possible only with external information on $\Omega_{\rm HI}$ and $b_{\rm HI}$, that we therefore include as priors in our Fisher forecasts. We present results for 2% 5% and 10% priors in each redshift bin, but it should be kept in mind that, in reality, these priors will be a function of redshift, worsening at higher z. We show the outcome of the Fisher calculation for both a conservative choice of $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ and for $k_{\rm max} = k_{\rm nl}$.

Results are summarized in figure 4.4. Starting from the continuous lines, which show the no-wedge case for Ext-CHIME, Ext-HIRAX, and highzFAST, we see that in

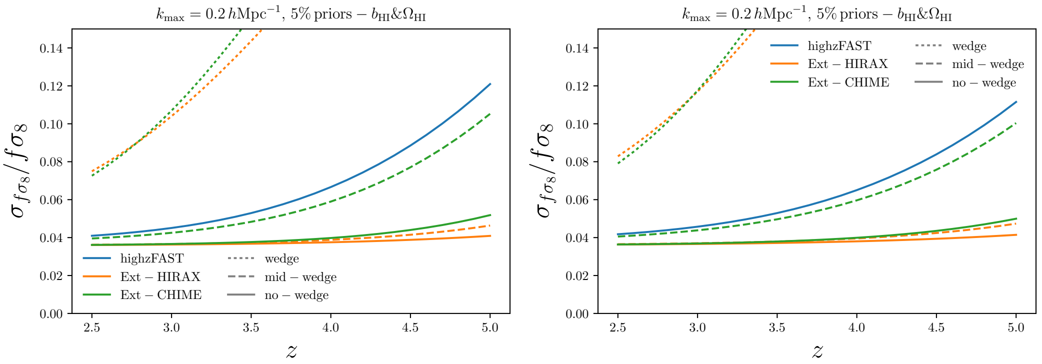

FIGURE 4.5: Dependence of the constraints on $f\sigma_8$ on the fiducial value of the HI bias. The left and right panels show the results when the value of $M_{\rm min}$, a parameter that controls the cut-off mass in the $M_{\rm HI}(M,z)$ function and therefore the amplitude of the HI bias, is set to $2\times10^{10}M_\odot/h$ and $2\times10^{11}M_\odot/h$, respectively. In both panels we consider a 5% prior on the value of both $\Omega_{\rm HI}$ and $b_{\rm HI}$. Different colors and line types represent the different instruments and assumptions about the wedge (see legend), correspondingly.

the absence of foreground contamination 21cm surveys can measure RSD parameters to exquisite precision. Interferometric experiments saturate the priors up to $z=3.5$, and are still able to measure $f\sigma_8$ with 4% precision at $z=5$ in the case of aggresive priors.. In our analysis Ext-HIRAX perfoms better that the other two hypothetical facilities, mostly because of a better angular resolution. Single dish experiments like highzFAST pay the price of low angular resolution, but they are unaffected by the wedge, so the blue line in figure 4.4 can be considered the final result for auto-correlation experiments. If we now turn our attention to the case with partial wedge contamination, the mid-wedge case described in Section 4.1.2, we see a degradation in the constraining power, with surveys with larger FOV yielding larger error bars as expected. For instance we find similar results comparing Ext-CHIME and highzFast. If the wedge extends up to the horizon the picture changes dramatically and interferometer performs much worse than single dishes. As it can be noticed in the right and lower panels, the inclusion of all the modes down to the non-linear scale $k_{\rm nl}$ does not substantially change the results. This is a consequence of the very low S/N ratio on scales smaller than $k\simeq0.2\,h\,{\rm Mpc}^{-1}$ (see figure 4.3). Only in the case where we consider the full wedge, we find marginal improvement from the inclusion of smaller scales.

As discussed in section 4.1.1, the value of the HI bias as a function of redshift is quite uncertain, reflecting our poor knowledge of what halos host neutral hydrogen. Figure 4.5 shows the dependence of our results on the HI bias. We do that by assuming two different values for $M_{\rm min}$: $M_{\rm min}=2\times10^{10}M_\odot/h$ (left panel) and $M_{\rm min}=2\times10^{11}M_\odot/h$ (right panel). The numbers we obtain are very similar in most cases, with the exception of the full wedge scenario, indicating our forecasts should be robust to different choices of fiducial astrophysical parameters. For completeness, the numbers with the errors on $f\sigma_8$ for Ext-HIRAX and various considerations using 5% priors on the values of both $\Omega_{\rm HI}$ and $b_{\rm HI}$ are shown in table B.1.

To better illustrate the synergies of a 21cm IM probe with galaxy surveys, in figure 4.6 we plot, as a function of redshift, the forecasted constraints on $f\sigma_8$ from: DESI (tables 2.3 and 2.5 from [160]), Euclid (table 1.4 from [193], reference case),

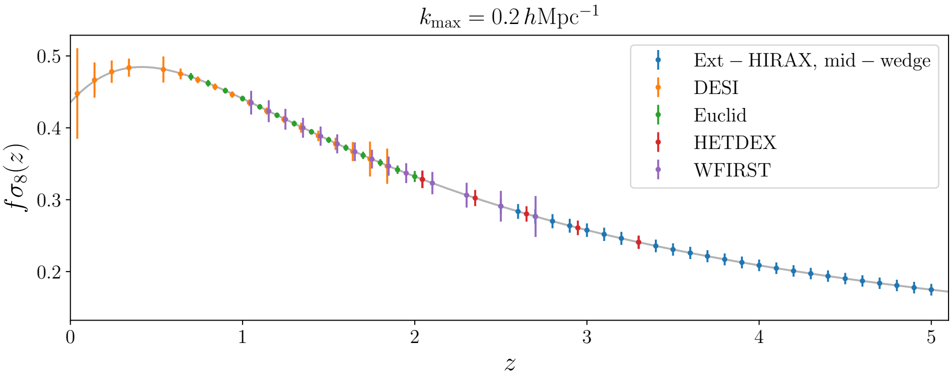

FIGURE 4.6: The evolution of $f\sigma_8$ and the predicted constraints as a function of redshift using Ext-HIRAX considering mid wedge configuration with 5% priors on both $\Omega_{\rm HI}$ and $b_{\rm HI}$ (blue). For comparison we also show predicted measurements (see text for the references) from upcoming galaxy surveys: DESI (orange, data points shifted by -0.01 in redshift), Euclid (green), HETDEX (red) and WFIRST (purple).

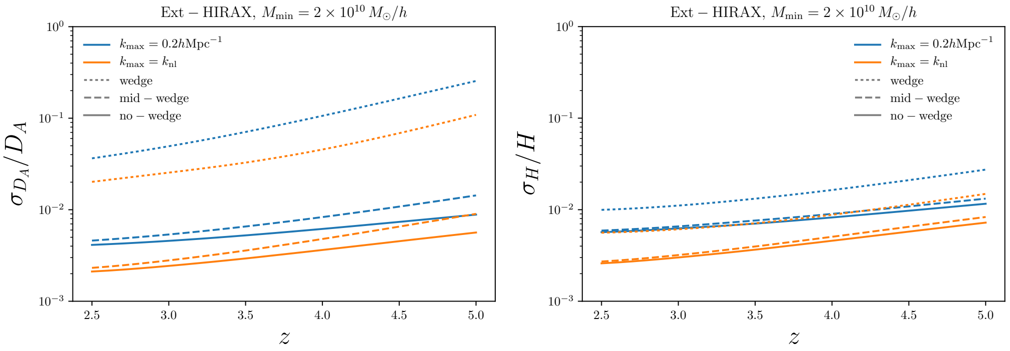

FIGURE 4.7: 1σ constraints on the angular diameter distance (left) and Hubble function (right) from the Ext-HIRAX setup. The forecasts have been derived for two different values of $k_{\rm max} = 0.2\,h{\rm Mpc}^{-1}$ (blue) and $k_{\rm nl}$ (orange). The solid, dashed and dotted lines represent different assumptions over the wedge (see legend). The fiducial value of the HI bias is computed assuming $M_{\rm min} = 2\times10^{10}M_\odot/h$.

HETDEX (figure 6. from [194]), WFIRST (figure C.6 from [163]) and Ext-HIRAX for the mid-wedge case (figure 4.4). We emphasise that the one should not directly compare the precision a survey could possibly achieve in the figure 4.6, since different analysis have different underlying assumptions, but the redshift coverage. 21 cm surveys could be able to fill the gap almost to the end of reionization, in a redshift range difficult to reach for conventional surveys. As a reference we used the case of 5% priors and partial cleaning of the wedge.

4.2.2 BAO distance scale parameters

We now turn our attention to the BAO distance scale parameters. This case is similar to a real analysis if the modelling of the broadband power spectrum has introduced enough nuisance parameters that only an amplitude can be measured with enough precision. In this case any constraint on $f\sigma_8$ makes sense only in combination with CMB data.

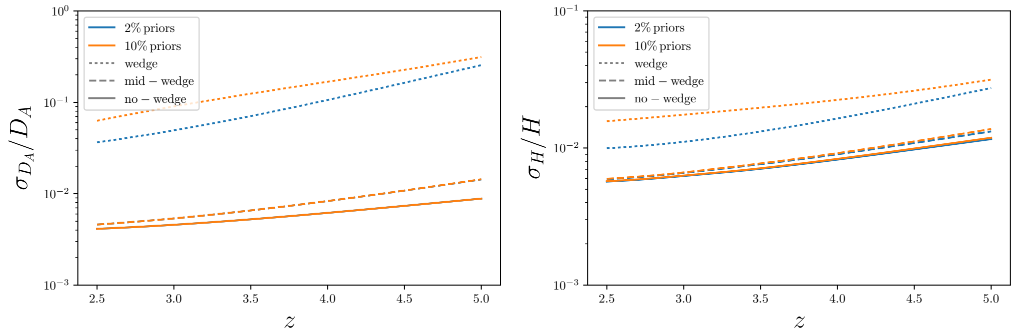

FIGURE 4.8: 1σ constraints on the angular diameter distance (left) and Hubble function (right) from the Ext-HIRAX setup assuming different prior knowledge on both b_{HI} and Ω_{HI}: 2% (blue) and 10% (orange). The forecasts have been derived for $k_{\max}$=0.2 hMpc^{-1}. The solid, dashed and dotted lines represent different assumptions over the wedge (see legend). The fiducial value of the HI bias is computed assuming $M_{\min} = 2 \times 10^{10} M_{\odot}/h$.

Results are shown in figure 4.7. We find that 21cm surveys can provide very accurate measurements of both the Hubble parameter and the angular diameter distance. For most cases, the geometry of the Universe can be constrained at the sub % precision. Similar to the analysis of the growth of structure in the previous section, the presence of the wedge suppresses the amount of information that can be extracted. The Hubble parameter however can still be measured quite accurately. This is expected since purely radial modes are the only ones surviving the wedge.

The precision at which we can measure the BAO distance scale parameters depends weakly on the prior on astrophysical parameters in the cases of low wedge coverage, while the effect is more pronounced in the full wedge case. We show the 1σ constraints on $H(z)$ and $D_A(z)$ dependance on the prior on astrophysical parameters in figure 4.8 in the case of Ext-HIRAX.

Similar to figure 4.6, in figure 4.9 we plot the Hubble function $H(z)$ and the angular diameter distance $D_A(z)$ projected measurements from low to high-redshift combining different probes. The forecasted constraints for DESI are taken from [160], while the numbers for other galaxy surveys – Euclid, HETDEX and WFIRST, have all been taken from the corresponding tables in [192]. We also show the forecasted constraints from Ext-HIRAX for the mid-wedge case and assuming 5% priors on b_{HI} and Ω_{HI}. Once again it is clear the contribution of a 21cm survey in filling up the whole redshift range. In table B.2 we list the constraints for both $H(z)$ and D_A as a function of redshift for different foreground configurations of Ext-HIRAX, fixing the 5% priors on b_{HI} and Ω_{HI}.

4.3 Extension to ΛCDM: Results from probe combination

Cosmological constraints from a single survey usually are limited by the presence of degeneracies between parameters, which can be partially broken by combining different datasets. The purpose of this section is to study the possible benefits of combining data from CMB, galaxy surveys data at low-redshift and 21cm data at high-redshift. Since the three probes we consider do not overlap in the redshift range,

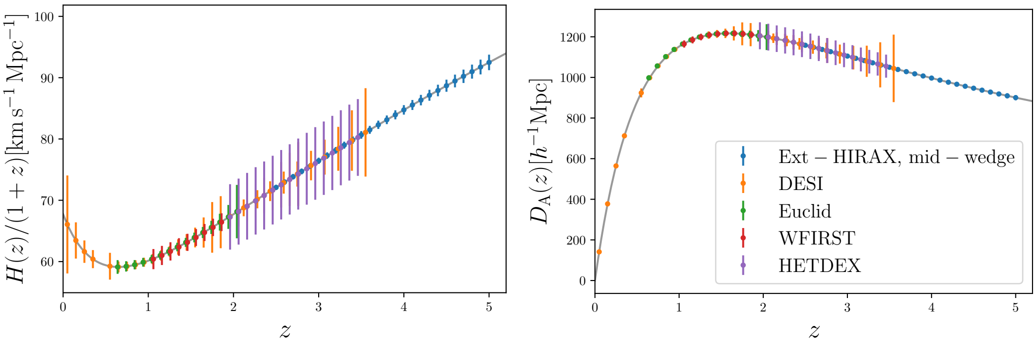

FIGURE 4.9: The evolution of the Hubble function $H(z)$ (left panel) and the angular diameter distance $D_A(z)$ (right panel) with the predicted constraints as a function of redshift using Ext-HIRAX considering mid wedge configuration with 5% priors on both Ω_{HI} and b_{HI} (blue). For comparison we also show predicted measurements (see text for the references) from upcoming galaxy surveys: DESI (orange, Euclid (green), HETDEX (red) and WFIRST (purple). Quoted data points have been shifted in redshift by ±0.01 in order to avoid overlapping.

the Fisher forecasts become very simple as the covariance is diagonal and the different Fisher matrices can be added up.

We include the current constraints on the cosmological parameters measured from Planck 2015 [6] in combination with BAO measurements from BOSS at redshift $z \simeq 0.3$ and $z \simeq 0.5$ [45]. In order to obtain the corresponding Fisher matrix we use the publicly available MCMC chains[2] which are based on Planck temperature (TT) + low TEB polarisation + external BAO constraints. We then compute the covariance matrix for the parameters we consider and invert it.

Additionally, we use the Fisher matrix calculated for a generic CMB Stage 4 experiment[3]. We refer the reader to the CMB S4 Science Book [195] for more details about next generation of CMB experiments. For consistency we remove Planck information if CMB S-4 information is used.

We also separately perform Fisher forecasts for future spectroscopic galaxy redshift surveys such as Euclid [161] or DESI [192], rather than relying on existing results. This choice has been made in order to avoid the final constraints to be affected by different modeling assumptions in different analysis. It is also motivated by the need for the full Fisher matrix rather than the marginalized error on each individual parameter.

To model the galaxy power spectrum, we employ a similar expression to equation 4.22, but without the brightness temperature factor and using the galaxy bias instead of the HI bias. The effect of non-linearities on the galaxy power spectrum is larger at lower redshifts and effectively damps the BAO oscillations in the galaxy power spectrum. To include it in the model we use equation 1.44 where the term $P_{\mathrm{s}}^{\mathrm{lin}}(k,\mu)$ now contains the galaxy bias term – $b(z)^2$. We compute the damping scales using $\Sigma_\perp = 9.4(\sigma_8(z)/0.9)\,h^{-1}\mathrm{Mpc}$ and $\Sigma_\parallel = (1+f(z))\Sigma_\perp$[192]. This effect is particularly

---

[2] We use `base_mnu_plikHM_TT_lowTEB_BAO` chains for the study on the sum of neutrino masses Σm_ν and `base_nnu_plikHM_TT_lowTEB_BAO` chains for the study of the number of relativistic species N_{eff}.

[3] We thank Anže Slosar for providing us the Fisher matrix of CMB S-4 made by Joel Meyers.

important for the BAO measurements at low-redshifts since it deteriorates the constraints on $D_A(z)$ and $H(z)$. In our forecasts for the massive neutrinos we will not assume the BAO reconstruction (see Section 4.1.1) has been applied and we will use the full values of $\Sigma_\perp$ and $\Sigma_\parallel$, while in our forecasts on the $N_{\rm eff}$ we do assume BAO reconstruction and half the damping scales accordingly. The exponential factor in equation 1.44 is taken outside the logarithmic derivatives in the Fisher matrix since we do not want to infer any distance information from these factors. In practice, the exponential only degrades the effective volume.

For Euclid, we assume an area equal to 15000 deg^2 covering the redshift range $0.65 < z < 2.05$. We take the fiducial galaxy bias $b(z) = \sqrt{1+z}$ and treat it as a free parameter that we marginalise over in each redshift bin, $\Delta z = 0.1$. The only noise term in the power spectrum we consider is the shot-noise coming from the discreetness of galaxy sample and equal to the inverse number density of galaxies: $P_{\rm N}^{\rm shot} = \bar{n}^{-1}\,[h^{-3}{\rm Mpc}^3]$. We compute the $\bar{n}$ using the data in table 1.3 (reference case) of [193]. We will show the results for two different values of the maximum wavenumber: Euclid02 for which we take $k_{\rm max} = 0.2\,h{\rm Mpc}^{-1}$ and Euclidnl for which we take $k_{\rm max} = k_{\rm nl}(z)$.

Of all possible extension of the standard cosmological model we focus on the sum of the neutrino masses, $\sum m_\nu$, and on the effective number of relativistic degrees of freedom, $N_{\rm eff}$, since, as we will see, these two have a clear target to be reached.

4.3.1 Massive neutrinos

The goal of this section is to study the improvements on the constraints on the neutrino masses that can be achieved by adding information from 21cm IM surveys in the redshift range $2.5 < z < 5$. Improvements on the constraints coming from the epoch of reionization have been presented in [196, 197, 198].

In the case of massive neutrinos we have to slightly modify our signal model. In fact, as shown [199, 200, 201, 202], halos and galaxies are biased tracers of the CDM+baryons field only, as opposed to the total matter field – CDM+baryons+neutrinos. This means that we have to change the $P(k)$ and $f(k)$ in equation 2.1 to include only the CDM+baryons component. This important physical effect has the consequence of reducing the constraining power of a broadband analysis of $P_{21}(k, z)$, since the total matter power spectrum is more suppressed by the presence of massive neutrinos than the CDM+baryons only. In our case, we find the difference being around 30% in terms of the forecasted σ_{m_ν}, an absolutely non-negligible effect. To the best of our knowledge, existing forecasts on neutrino masses do not include this effect in their analysis (with some notable exceptions as in [66]), suggesting some level of bias in their final answers. Our forecast is therefore more realistic, but it yields worse constraints.

Another important thing to notice is that the effect of neutrinos on the power spectrum become smaller with redshift. This is due to the fact that neutrinos have less time to delay the growth of CDM+baryons perturbations on small scales. This means that the constraint on Σm_ν from two surveys, one at low-redshift and another at high-redshift, covering the same volume and using the same amount of information ($k_{\rm min} < k < k_{\rm max}$) will be different, the high-redshift one being worse. In contrast, going to higher redshifts has two advantages: larger survey volume and larger $k_{\rm nl}$ up to which one can hope to use perturbation theory. Figure 4.10 shows the ratio between the linear CDM+baryons (dashed) and the total matter (solid) power spectrum in a model with massive neutrinos ($\Sigma m_\nu = 0.06$ eV) to the same quantities in the standard ΛCDM model ($\Sigma m_\nu = 0$ eV). Different colors represent different redshift, in blue $z = 1$

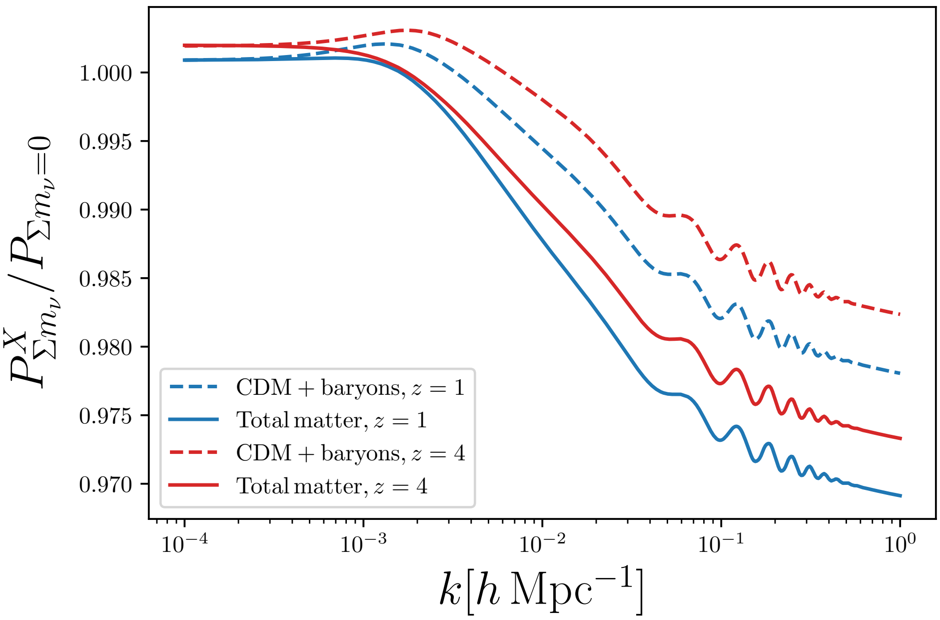

FIGURE 4.10: Ratio between the linear CDM+baryons (dashed) and the total matter (solid) power spectrum in a model with massive neutrinos ($\Sigma m_\nu = 0.06$ eV) to the same quantities in the standard ΛCDM model ($\Sigma m_\nu = 0$ eV). Different colors represent different redshift, in blue $z = 1$ and in red $z = 4$.

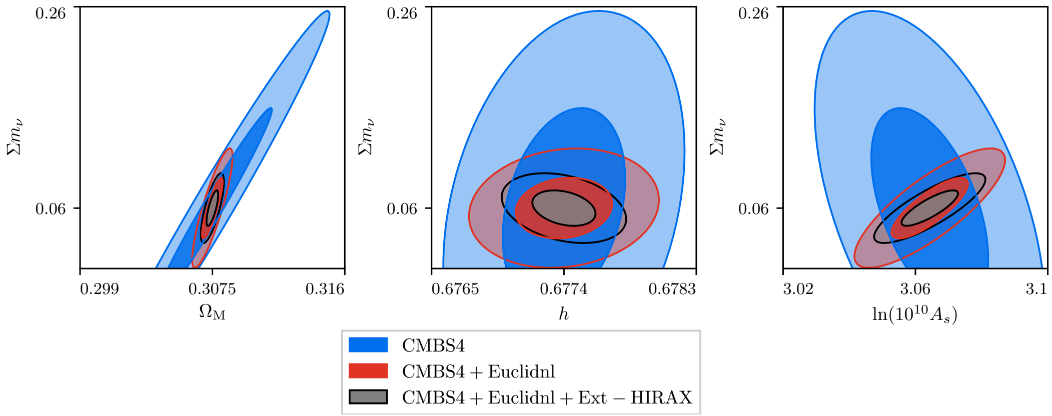

FIGURE 4.11: Constraints on the neutrino masses and their degeneracies with Ω_m (left), h (middle) and A_s (right) from CMBS4 (blue), CMBS4+Euclid (red) and CMBS4+Euclid+Ext-HIRAX (black). For Ext-HIRAX we have assumed $k_\mathrm{max} = k_\mathrm{nl}$, no wedge and 2% priors on both b_HI and Ω_HI.

and in red $z = 4$. We notice that the difference between CDM+baryons and total matter remains constant with redshift, but the overall suppression compared to the vanilla models decrease by almost a factor of two between low and high z.

We show the forecasted constraints on the neutrino masses and other cosmological parameters coming from external datasets alone in table B.3. Constraints on the neutrino masses for various combinations of probes and different foreground assumptions are shown in table B.6 and B.7 for Ext-CHIME and Ext-HIRAX, respectively. We see that in the most pessimistic case, i.e. full wedge and $k_\mathrm{max} = 0.2\ h\,\mathrm{Mpc}^{-1}$, Ext-HIRAX plus Euclid and CMB-S4 is able to pin down neutrino masses to $\sigma_{m_\nu} = 0.028$ eV, with IM bringing a 10% improvement on the errorbars. In the most optimistic case this combination of probes yields $\sigma_{m_\nu} = 0.018$ eV in which case the gain of adding IM can reach 40%. It is important to stress again that the use of CDM+baryons to describe the clustering of galaxies and 21cm is worsening the error bars by roughly 30%. In figure 4.11 we show the constraints and the degeneracies of the neutrino masses with other cosmological parameters when Ext-HIRAX is combined with CMB-S4 and CMB-S4+Euclid in the most optimistic case. Single-dish experiments do not help in constraining neutrinos masses as well as interferometers, as one can see from results presented in tables B.5 for highzFAST alone and table B.8 in which we show the results when combining different probes together.

We notice that the effect of the wedge does not worsen the constraints too much. In Section 4.2.1 we have shown that the presence of the wedge deteriorate significantly the constraints on $f\sigma_8$. Thus, we conclude that most of the improvement brought in by the 21cm data in the combined analysis is due to the very accurate distance measurements, which are less affected by foregrounds. A similar argument applies to external priors on the density and the bias of the HI field, which do not change much the final result on neutrino masses. On the other hand, an even more aggressive, in terms of signal-to-noise, instrument could potentially improve on neutrino masses constraints.

Another advantage of observing the high-redshift Universe is that dark energy is a sub dominant fraction of the energy density. It is in fact well known that in the extended parameter spaces the equation of state of dark energy w is very degenerate with other parameters, e.g. massive neutrinos and curvature [203]. In figure 4.12 we show the degeneracy $\Sigma m_\nu - w$ degeneracy in the case where only low redshift measurements in combination with CMB are available (Euclid02+Planck) and when the information from above $z > 2.5$ (Ext-HIRAX) is added on top. We find a factor of 3 improvement on the error on w when adding Ext-HIRAX, constraining w to the 1% level even if neutrinos are allowed to vary. For Ext-HIRAX in this case we have used $k_{\rm max} = 0.2\, h\,{\rm Mpc}^{-1}$, 2% priors on both $b_{\rm HI}$ and $\Omega_{\rm HI}$, and mid-wedge case.

To conclude this Section we would like to emphasize that even adding high-redshift information the constraint on neutrino masses is primarily limited by the uncertainty in the amplitude of the primordial fluctuation spectra $A_{\rm s}$ and the optical depth of reionization – τ – in CMB data. The parameter τ is weakly constrained by the current Planck results [6], with an error bar almost a factor of two worse than the forecasted value [204]. Improvement in the understanding of systematics in the Planck polarization data will likely improve the constraint on τ, but to be more conservative we decided to adopt the current estimate. This is not the case for pre-Planck forecasts, e.g. DESI [192] or more recent ones [205, 160], who assume in their forecasts the blue book value for σ_τ. This fact, together with our correct choice of the modeling at the beginning of this Section, explains why our constraint on neutrino masses is in general worse than others in the literature. Concerning the future improvements on constraining τ, aside from including all polarization measurements from Planck, another hope is that future reionization epoch IM surveys, e.g. HERA, could bring tight constraints on τ [206], thereby bringing substantial improvements on other cosmological parameters. Figure 4.12 shows how much a better measurement of τ in the CMB could help in constraining the neutrino masses in the best case of Ext-HIRAX with $k_{\rm max} = k_{\rm nl}$, 2% prior on both $b_{\rm HI}$ and $\Omega_{\rm HI}$, and no wedge.

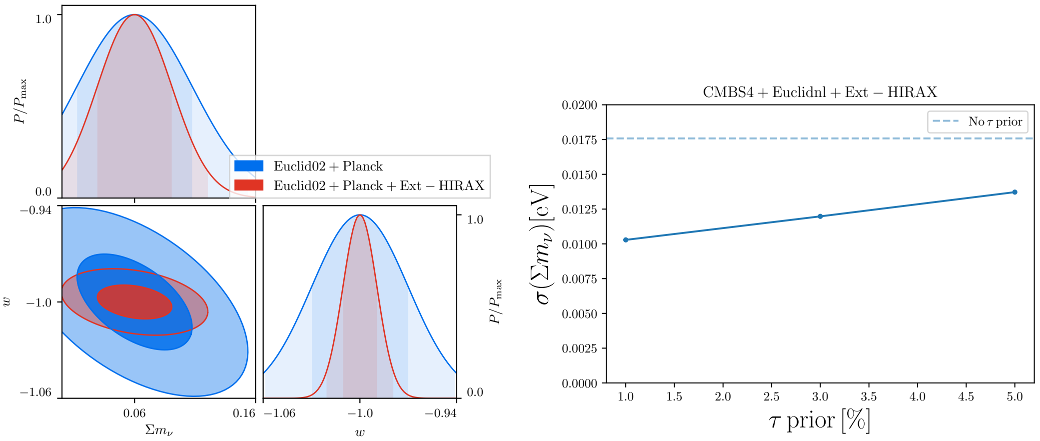

FIGURE 4.12: (Left panel) Degeneracy between Σm_ν and w when considering only Euclid02+Planck (blue) and when adding information from Ext-HIRAX (red). (Right panel) The effect of an additional τ prior on the neutrino masses constraints coming from the combination of CMBS4 + Euclidnl + Ext-HIRAX. The blue dashed line shows the case on no additional prior assumed.

Beyond linear theory

So far we have considered a simple model for the 21cm power spectrum consisting of linear theory and linear bias term. Going beyond linear theory requires new nuisance parameters that need to be marginalised over [191, 190, 205, 207]. This can degrade the constraints on the neutrino masses. In order to estimate the effect of these extra terms we will consider a simple order-of-magnitude extension to our power spectrum model and write the 21cm power spectrum as:

$$P_{\rm HI}(k,z) = \bar{T}_b^2 (b_{\rm HI}^2 + f\mu^2)^2 P_{\rm lin}(k, z=0) D(z)^2 \left(1 + 2k^2R^2 + \alpha P_{\rm lin}(k, z=0) D(z)^2 \frac{k^3}{2\pi^2}\right). \tag{4.25}$$

The second term in the parentheses has the form of the most relevant counter-term in the effective field theory approach to perturbation theory [208, 209]. The third term is included in order to mimic the effects of 1-loop corrections to the power spectrum and can be understood as the “theoretical error” [191] on the linear theory model. This approximation for one-loop contributions is a good order-of-magnitude estimate on scales relevant for constraining the sum of the neutrino masses. The coefficients R and α are free parameters in each redshift bin and we take the fiducial values $R = \alpha = 0$. We marginalise over R and α in each redshift bin. We show the resulting constraints on the cosmological parameters from Ext-HIRAX in table B.4. Compared to the standard case of linear theory and bias, the constraints on the neutrino masses become weaker by a factor of $\sim 25\%$ and 31% in the no wedge case of $k_{\rm max} = 0.2\, h\,{\rm Mpc}^{-1}$ and $k_{\rm max} = k_{\rm nl}(z)$, respectively. Adding other probes (Euclid, CMB S-4) makes this difference smaller and the constraints are weaker by a factor of $5 - 10\%$ (see table B.6). This reinforces our previous argument that most of the constraining power is coming from geometrical measurements of the BAO distances.

4.3.2 The effective numbers of relativistic degrees of freedom

Another extension of the standard ΛCDM model is represented by the presence of extra radiation species. In the standard model, with 3 massive neutrinos, one has

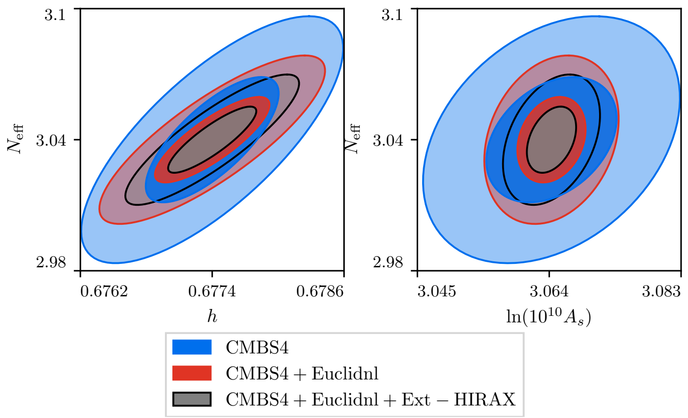

FIGURE 4.13: Constraints on $N_{\rm eff}$ and its degeneracies with h (left) and A_s (right) from CMB-S4 (blue), CMB-S4+Euclid (red) and CMB-S4+Euclid+Ext-HIRAX (black). For Ext-HIRAX we have assumed $k_{\rm max} = k_{\rm nl}$, no wedge and 2% priors on both $b_{\rm HI}$ and $\Omega_{\rm HI}$. We used `cosmicfish` software [210] to make this figure and also figure 4.11.

$N_{\rm eff} = 3.046$ [211], but any light species in thermal equilibrium during the history of the Universe will increase this value. Theoretical estimates in [212, 213] set the minimum $\Delta N_{\rm eff}$ due to any light particles between $z = 0$ and the QCD phase transition to $\Delta N_{\rm eff} = 0.027$. This number is therefore the natural target of any analysis of $N_{\rm eff}$.

In CMB data, $N_{\rm eff}$ is very degenerate with other parameters, and the 1σ forecasted error of CMB S-4 is slightly above the theory benchmark. However, it has been recently realized that a clean signature of $N_{\rm eff}$ exists and can be detected in both CMB and large scale structure [214, 215, 216], as the presence of extra relativistic degrees of freedom introduce a phase shift in the acoustic peaks of the baryon-photon fluid [214]. This effect has already been detected in Planck data [217, 215]. In [218], the authors have shown that this phase shift survives non-linear evolution and therefore could in principle be detected in BAO analysis of galaxy clustering.

Building and testing a template for the phase shift goes beyond the scope of this thesis, but motivated by the above discussion we study the constraints on $N_{\rm eff}$ from various combination of probes. We fit the full broadband power spectrum, which by definition contains also the phase shift, to constraint the number of relativistic degrees of freedom. In a realistic data analysis one would use reconstructed data, which we mimic by dividing by a factor of 2 the damping factor in equation 1.44 [192].

We show the forecasted constraints on $N_{\rm eff}$ and other cosmological parameters coming from external datasets alone in table B.9. Our results for various combinations of probes and different foreground assumptions are presented in tables B.10, B.11, B.12, B.13, B.14 and figure 4.13. Our forecasted error for Ext-HIRAX ranges from $\sigma_{N_{\rm eff}} = 0.015$, for the very optimistic case of no wedge and $k_{\rm max} = k_{\rm nl}$, to $\sigma_{N_{\rm eff}} = 0.021$ for the full wedge and $k_{\rm max} = 0.2\ h\,{\rm Mpc}^{-1}$ case. These numbers are interesting, but on the other hand any deviation from the minimal value will be detected at less than 1σ. It is also important to point out that 21cm surveys will add very little to the combination of CMB data and galaxy surveys. Our results are in broad agreement with the analysis of [216].

4.4 Summary and conclusions

In this chapter we have explored the possibility of studying cosmology through radio-telescopes that operate in the redshift range $2.5 < z < 5$. The reason behind this is that, while this is a redshift-range not considered in current and upcoming setups, the volume it encloses is much larger then the one probed by current and future spectroscopic surveys. The question we try to answer is: how much cosmological information is contained in this redshift window?

We focus our analysis on four key cosmological quantities: 1) the growth rate, $f\sigma_8$, the BAO distance scale parameters, D_A and H, the sum of the neutrino masses, Σm_ν, and the number of relativistic degrees of freedom, $N_{\rm eff}$. We consider four extensions of current or upcoming radio-telescopes like HIRAX, CHIME and FAST, and two different observational strategies: interferometry (for Ext-HIRAX and Ext-CHIME) and single-dish (highzFast).

We carry out our analysis using the Fisher matrix formalism. We model the amplitude and shape of the 21cm signal using the model proposed in [76]. We also account for cosmological and instrumental effects such as the presence of the wedge, the window function, the instrument thermal noise, the angular resolution, the presence of shot-noise...etc.

We point out that measurements that are sensitive to the overall amplitude of the 21cm power spectrum, like $f\sigma_8$, will be completely degenerate with astrophysical parameters like $\Omega_{\rm HI}$ and $b_{\rm HI}$. In order to break that degeneracy it is necessary that we use independent datasets that constrain those quantities. We show how the value of these parameters can be determined through either the Lyα-forest alone or via cross-correlations between 21cm and the Lyα-forest or DLAs.

Under the assumption of the primary beam foreground wedge contamination (mid-wedge case in the text), 5% priors on $b_{\rm HI}$ and $\Omega_{\rm HI}$ and $k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$, that we term the fiducial setup, we find that Ext-HIRAX can constrain the value of $f\sigma_8$ within bins of $\Delta z = 0.1$ at $\simeq 4\%$ in the redshift range $2.5 < z < 5$. A modest improvement is achieved by changing $k_{\rm max}$ from $0.2\ h{\rm Mpc}^{-1}$ to $k_{\rm nl}$. If data from the whole wedge need to be discarded, these constraints degrade between a factor 2 (at $z = 2.5$) and 7 (at $z = 5$).

Under the fiducial setup, we find that Ext-HIRAX will place $\simeq 1\%$ constraints on D_A and H. As with the growth rate, our results point out that going to smaller scales has only a very modest impact on the results. Being able to use a fraction of the modes in the wedge has a huge impact on our results, as removing the information in the whole wedge degrades the constraints between a factor of 10 (at $z = 2.5$) and 20 (at $z = 5$).

We have also studied the impact that the theory model has on the results. By using a theory template that accounts for 1-loop corrections and incorporates 2 free parameters that we marginalise over, we find that the constraints on the cosmological parameters worsen between 10% and 500% when 2% priors on $\Omega_{\rm HI}$ and $b_{\rm HI}$ are used. In the case of the neutrino masses, the constraints worsen between 10% and 30%.

We find that data from Ext-HIRAX in the fiducial setup can constrain the neutrino masses with an error of 0.10 eV. In combination with data from CMB S4 and galaxy clustering from Euclid the errors shrink to $\simeq 20$ meV. Our results are not very sensitive to the wedge coverage, the minimum scale employed and the priors on $b_{\rm HI}$ and $\Omega_{\rm HI}$.

Finally, we find that data from Ext-HIRAX, in the fiducial setup, plus CMB S4 plus Euclid can constrain $N_{\rm eff}$ with a very competitive error of 0.02. As with the neutrino masses, our constraints do not depend much on the $\Omega_{\rm HI}$ and $b_{\rm HI}$ priors, $k_{\rm max}$ and the wedge coverage.

Results for the Ext-CHIME and highzFAST instruments are similar to those of Ext-HIRAX, with the exception of neutrino masses, where highzFAST performs worse than Ext-CHIME or Ext-HIRAX.

We conclude that there is a large amount of cosmological information embedded in the, poorly constrained, redshift range $2.5 < z < 5$. Suitable extensions of existing and upcoming radio-telescopes targeting at this redshift window can provide very tight constraint on key cosmological parameters.

Chapter 5

The HI content of dark matter halos at $z \approx 0$ from ALFALFA

Introduction

In this chapter we will use a self-consistent framework to constrain the $M_{\rm HI}$-M_h relation using the mass-weighed clustering of HI galaxies detected by the Arecibo Legacy Fast ALFA survey (ALFALFA), as well as their abundance in halos extracted from galaxy groups found in the SDSS galaxy survey. We will also explore the possibility of constraining the shape of the HI profile and the impact of modeling assumptions on our results.

This chapter is organized as follows. In section 5.1 we will summarize the theoretical framework already described in section 2.1.1 that we use to characterize the abundance and clustering of HI. We outline the data employed in this work in section 5.2. The methods used to analyze the data and compare with the theory predictions are illustrated in section 5.3. The main results of this work are shown in section 5.4. We discuss the results and summarize the conclusions of this work in section 5.5.

5.1 HI halo model

The purpose of this chapter is to constrain the HI-mass-to-halo-mass relation $M_{\rm HI}(M_h)$ from direct measurements in selected galaxy groups, as well as from the clustering of HI sources. In order to do that we will use the HI halo model described in section 2.1.1. As we saw there the clustering of HI is dominated by its distribution within the halo (i.e. the so-called 1-halo term). Although our constraints will be based solely on the shape of the correlation function on larger scales, we use two different models for the HI density profile, in order to quantify the effect of this assumption on the final results:

- **Altered NFW profile:** this is the model introduced and used in [113, 114, 103]. and assumes the radial profile of the form:

  $$\rho_{\rm HI}(r|M_h) \propto (r + 3/4r_s)^{-1}(r + r_s)^{-2} \tag{5.1}$$

  where $r_{\rm s}$ is the scale radius of the HI cloud, and is related to the halo virial radius $R_{\rm v}(M_h)$ by the concentration parameter – $c_{\rm HI}(M_h, z) \equiv R_{\rm v}(M_h)/r_{\rm s}$. We follow [219, 220] and use a mass-dependent concentration parameter given by:

  $$c_{\rm HI}(M_h, z = 0) = 4\, c_{\rm HI,0} \left(\frac{M_h}{10^{11} M_\odot}\right)^{-0.109}. \tag{5.2}$$

. In section 2.1.2 we have found that a similar form of this profile describes the average HI density profile found in numerical simulations very well. For simplicity, we do not consider the exponential cutoff at small scales (see equations 2.8 and 2.8) in this chapter.

- **Exponential profile:** this is the model implemented in [103], and given by

$$\rho_{\rm HI}(r|M_h) \propto \exp\left(-r/r_{\rm s}\right), \tag{5.3}$$

In both cases the proportionality factors are automatically fixed by requiring that the HI mass be given by the volume integral of the density profile up to the halo virial radius $R_{\rm v}(M)$.

$$M_{\rm HI}(M_h) = 4\pi \int_0^{R_v} dr\, r^2 \rho_{\rm HI}(r|M_h). \tag{5.4}$$

Thus, both profiles are described by one additional free parameter, $c_{\rm HI,0}$. The normalized HI density profile in Fourier space for the altered NFW profile is given in [103] (see their equation A3), while the exponential profile is simply

$$u_{\rm HI}(k|M_h) = \frac{1}{(1+k^2 r_{\rm s}^2)^2}. \tag{5.5}$$

.

As before, for the halo mass function and bias, we use the parametrizations of [104], derived from numerical simulations, but in this chapter we adhere to halo masses defined by a spherical overdensity parameter $\Delta = 180$

$$M_h = \frac{4\pi}{3}\rho_c \Omega_{\rm m} \Delta R_{\rm v}^3. \tag{5.6}$$

Finally, our basic clustering data vector is the 2D projected correlation, given by the projection of the 3D correlation function along the line of sight. This can be computed directly from the power spectrum as:

$$\begin{aligned}\Xi(\sigma) &= \int_{-\infty}^{\infty} d\pi\, \xi(\pi,\sigma) \\ &= \int_0^{\infty} \frac{k\, dk}{2\pi} \left[P_{\rm HI,1h}(k) + P_{\rm HI,2h}(k)\right] J_0(k\sigma),\end{aligned} \tag{5.7}$$

where $J_0(x)$ is the order-0 cylindrical Bessel function. To accelerate the computation of $\Xi(\sigma)$ we made use of FFTLog [221].

Our theoretical model therefore depends on four free parameters $\theta = \{M_0, M_{\rm min}, \alpha, c_{\rm HI,0}\}$. We fix all cosmological parameters to values compatible with the latest ΛCDM constraints measured by Planck [6] ($H_0 = 70\,{\rm km\,s^{-1}Mpc^{-1}}$, $\Omega_{\rm m} = 0.3075$, $n_s = 0.9667$, $\sigma_8 = 0.8159$)[1].

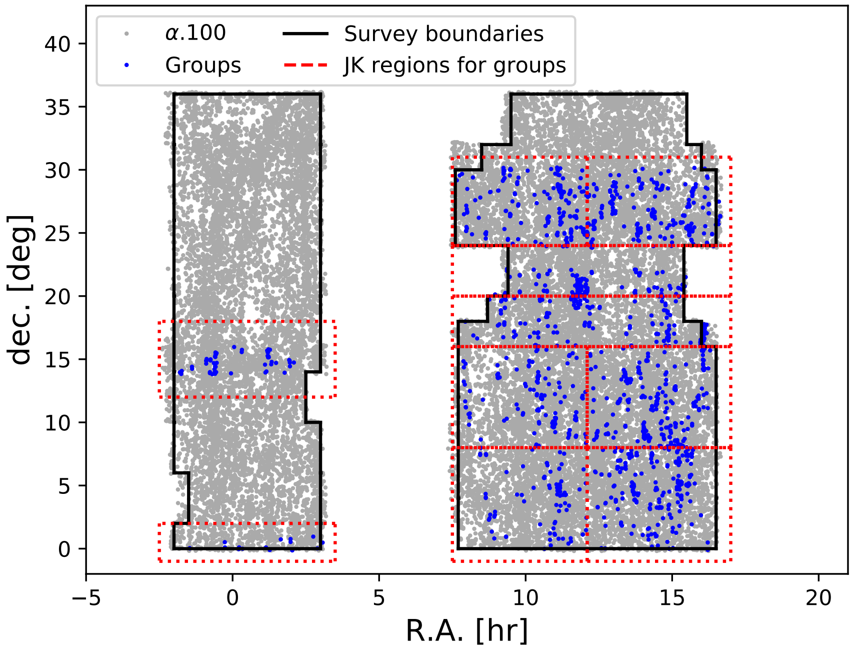

FIGURE 5.1: Sky distribution of the HI selected galaxies from $\alpha.100$ sample (gray dots). The black lines show the survey boundaries used in our clustering analysis (in which all sources outside the boundaries were omitted). The HI sources associated with groups in the SDSS DR7 group catalog are highlighted in blue. The dotted red lines show the jackknife regions used to estimate the cosmic variance uncertainties of the HI mass function in groups (see section 5.3.2).

5.2 Data

5.2.1 The α.100 dataset

The Arecibo Fast Legacy ALFA (Arecibo L-band Feed Array) survey, or ALFALFA[2] [223], is a blind extragalactic HI survey performed using the Arecibo radio telescope. The main goal of ALFALFA is to quantify and study the properties of the HI content of the local Universe ($z \lesssim 0.05$). It represents a significant improvement over previous HI surveys, with a beam FWHM of ~ 3.5 arcmin, an rms noise of $\sim 2.4\,\mathrm{mJy}$ and a spectral resolution of $\sim 10\,\mathrm{km\,s^{-1}}$.

Previous clustering analyses of the ALFALFA samples used the 40% [224, 78, 225], and 70% [96] data releases (labeled α.40 and α.70). Our analysis makes use of the final data release [226], containing ~31500 sources up to a redshift of $z = 0.06$ and covering approximately 7000 square degrees in two continuous regions at either side of the Galactic plane. Sources with good detection significance ($S/N > 6.5$), classified as "code-1", represent the main sample ($\sim 81\%$ of the total 31502 sources). Most of the remaining sources, classified as "code-2", correspond to lower signal-to-noise detections ($S/N > 4.5$) with known optical counterparts. The remaining $\sim 5\%$ of the catalog is mostly composed of high-velocity clouds of galactic provenance. We use only code-1 sources in the clustering analysis described in section 5.3.1, and both code-1 and code-2 objects in the direct measurement of the HI content of galaxy groups (section 5.3.2). For each source, the catalog provides information about their angular coordinates, heliocentric radial velocity, radial velocity in the CMB frame, 21cm flux, line width and HI mass. HI masses for all objects can also be obtained from their distance and 21cm flux as

$$m_{\rm HI} = (2.356 \times 10^5 M_\odot)\, D^2\, S_{21} \tag{5.8}$$

where D is the distance to the source in Mpc, S_{21} is the integrated flux in units of Jy km s^{-1} and $m_{\rm HI}$ is the source's HI mass[3].

In the clustering analysis, the radial velocities $v_{\rm cmb}$ are used to assign radial distances to sources through their redshift $z_{\rm cmb} = v_{\rm cmb}/c$, using the cosmological parameters listed in section 5.1. Due to the radio frequency interference (RFI) we make additional cuts and following [225] we remove the sources outside $700\,\mathrm{km\,s^{-1}} < cz_{\rm cmb} < 15000\,\mathrm{km\,s^{-1}}$. After performing these cuts in the raw data we are left with 24485 code-1 sources and 5365 code-2 sources. Figure 5.1 shows the angular distribution of all sources used in this work. The black lines delineate the survey boundaries used for in the clustering analysis. These cuts further reduce the clustering sample to 23438 objects.

5.2.2 The SDSS group catalog

To assign the HI detected sources to dark matter halos, we cross-match the SDSS galaxies and the ALFALFA sources and determine the group membership of the cross-matched galaxies using a galaxy group catalog, following the procedure described in [227]. We use the SDSS DR7 group catalog[4] updated from the DR4 group catalog

---

[1] We fix the expansion rate to $h = 0.7$ instead of its best-fit measurement $h = 0.6774$ to match the choice in made in [222] to measure $\Omega_{\rm HI}$. We will report our final results as a function of $h_{70} \equiv H_0/70\,\mathrm{km\,s^{-1}Mpc^{-1}}$

[2] `http://egg.astro.cornell.edu/alfalfa/`

[3] To distinguish between the HI mass of ALFALFA sources and the total HI mass associated to a given dark matter halo, we label the latter $M_{\rm HI}$ and the former $m_{\rm HI}$

[4] `http://gax.sjtu.edu.cn/data/Group.html`

[228]. The catalog uses galaxies in the SDSS DR7 spectroscopic sample with $0.01 \leq z \leq 0.2$ and redshift completeness $C > 0.7$. The group finding algorithm has been extensively tested using mock galaxy redshift survey catalogs and has proven to be successful in associating galaxies that reside in a common halo [229]. In particular, this halo-based group finder works well for poor groups and identifies groups with only one member (i.e. isolated galaxies). The group halo masses are determined down to $M_h = 10^{11.8} h^{-1} M_\odot$ using two methods: ranking by luminosity and from the stellar mass of member galaxies. Although we used the luminosity-ranked group halo mass, the results do not change if the stellar-mass-ranked halo mass is used instead. The group finder has been shown to correctly select more than 90% of the true halos with $M_h \geq 10^{12} h^{-1} M_\odot$ [228], which allows us to reliably study our galaxy samples within groups and clusters with halo mass $10^{12.50} h^{-1} M_\odot \leq M_h \leq 10^{15.04} h^{-1} M_\odot$.

For the virial radius of groups with halo mass M_h, we adopt the radius R_{180} that encloses an overdensity $\Delta = 180$ times larger than mean density [228]:

$$R_{180} = 1.26 h^{-1}\text{Mpc} \left(\frac{M_h}{10^{14} h^{-1} M_\odot} \right)^{1/3} (1 + z_{\text{group}})^{-1}, \tag{5.9}$$

which is based on the WMAP3 cosmological model parameters [230], $\Omega_m = 0.238$, $\Omega_\Lambda = 0.762$ and $H_0 = 100h$ km s^{-1}Mpc^{-1}, where $h = 0.73$. While these parameters differ slightly from those used in this study, this does not significantly impact the results at the low redshifts of our sample ($z < 0.055$). We also note that the DR7 group catalog has significant overlap only with the 70% ALFALFA data release, and therefore no new information is gained by using the complete ALFALFA sample (α.100 dataset).

Figure 5.1 shows, in blue, the ALFALFA sources identified as members in the group catalog, as well as the jackknife regions used to compute the cosmic variance uncertainties for our estimate of the HI mass function in groups (dotted red lines, see section 5.3.2).

5.3 Method

We derive constraints on the HI content of dark matter halos by using the clustering properties of HI galaxies weighted by their HI content, as well as direct measurements of the HI content of galaxy groups. We describe the procedures used to compile these two data vectors and their associated covariances here.

As discussed in section 5, our main interest is to quantify the properties of the total HI density inhomogeneities, since these are the relevant proxy of the density fluctuations measured by 21cm IM. To do so, our main assumption will be that the properties of the full HI density field can be inferred from the properties of HI-selected sources as measured by ALFALFA when weighed by their HI mass. This simplifying assumption should be a good approximation as long as the sources detected by ALFALFA account for a significant portion of the total HI mass. The validity of this assumption can be quantified to some extent by examining the measurements of the HI mass function measured by the ALFALFA collaboration in [222], extrapolating it below the detection limit. This calculation shows that, for a conservative threshold of $m_{\text{HI,lo}} = 10^8 \, M_\odot$, less than 5% of the total HI would lie in sources not observed by ALFALFA. Thus, assuming that the tilt of the HI mass function does not vary sharply on smaller masses, the contribution from diffuse or undetected sources to the observables considered here is negligible given the uncertainties in our measurements. This is even more so for measurements of the HI clustering, given that the clustering bias of HI sources has been show to be only weakly dependent on HI mass [225]. Even in the case of the measurement of the HI content in galaxy groups (see section 5.3.2), where this contribution can rise to $\lesssim 30\%$ we will explicitly show that the impact of the missing HI mass on our results is minimal.

5.3.1 The projected 2-point correlation function

Previous studies [78, 225, 96] have measured two-point correlation function (2PCF) of HI-selected galaxies to determine their relation with the underlying dark matter density field. These studies have found that this sample has a low value of the clustering amplitude compared to the dark matter field (i.e. HI-selected galaxies have a low bias - $b_{\mathrm{HI,g}}$). Under the assumption described above, the same measurement can be performed on the 2PCF of HI-selected galaxies weighed by their HI mass to obtain a measurement of the total HI bias b_{HI}, which plays a key role on 21cm IM studies (see section 2.1). We describe the procedure used to estimate the 2PCF and its uncertainty here.

We begin by estimating the 2D 2PCF $\xi(\pi, \sigma)$ as a function of the distance between pairs of objects along the line of sight (π) and in the transverse direction (σ). For this we use the Landy & Szalay estimator [231], given by

$$\xi(\pi, \sigma) = \frac{\mathrm{DD}(\pi, \sigma) - 2\mathrm{DR}(\pi, \sigma) + \mathrm{RR}(\pi, \sigma)}{\mathrm{RR}(\pi, \sigma)}, \tag{5.10}$$

where DD is the normalized histogram of unique weighted pairs of sources separated by a distance (π, σ) found in the data catalog:

$$\mathrm{DD}(\pi, \sigma) = \frac{\sum_{i=1}^{N} \sum_{j>i} w_i w_j \, \Theta(\pi_{ij}; \pi, \Delta\pi) \, \Theta(\sigma'_{ij}; \sigma, \Delta\sigma)}{\sum_{i=1}^{N} \sum_{j>i} w_i w_j}.$$

Here π_{ij} is the distance between the i-th and j-th objects along the line of sight (and similarly for the transverse distance σ_{ij}), and $\Theta(x \in (x_1, x_2)) = 1$ when $x \in (x_1, x_2)$ and 0 otherwise. RR is defined similarly for unique pairs of objects belonging to a random catalog with statistical properties similar to those of the data (e.g. in terms of spatial and weights distribution) but no intrinsic clustering. Finally, DR is given by all pairs of data-random objects. The weights w_i assigned to each object are described below.

Random catalog

The random catalog needed to compute the correlation function should follow the same redshift, angular and weights distribution observed in the data. We use the area cuts reported in [222] to define the survey footprint These are shown in figure 5.1, as black lines, and we discard all sources outside these boundaries. The angular positions of the random objects are then generated by drawing random coordinates with a constant surface density within this area.

We assign redshifts to the random objects by accounting for both the radial selection function described in [225] (see their figure 4) and for RFI incompleteness, using the completeness function presented in the same paper (see their figure 6). Including these two effects is achieved by keeping a point with distance d in the random catalog with a probability corresponding to the product of the selection and RFI completeness functions at d. The final normalized redshift distribution in both the data and the random catalog is shown in the left panel of figure 5.2.

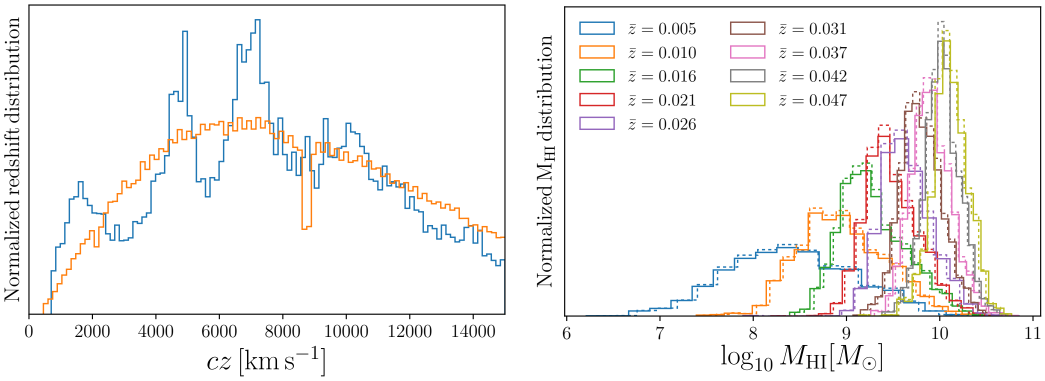

FIGURE 5.2: *Left:* normalized redshift distribution in the data (blue) and the constructed random catalog (orange). *Right:* the HI mass distribution in the data (solid line) and the constructed random catalog (dashed line) in different redshift bins (see legend).

The points in the random catalog must also be assigned mass weights following the same $m_{\rm HI}$ distribution as the data. To achieve this, we split the random and the data set in 10 redshift bins. In each redshift bin we give each random point an HI mass randomly sampled from the data in the same bin. The resulting HI mass distributions are shown in the right panel of figure 5.2.

Weights

The sample we use is not volume-limited, and the objects near the peak of the selection function will dominate the measured correlation function. In order to avoid this, we apply optimal pair-wise weights $w_{i,j} = w_i \times w_j$, where w_i is given by [22, 232]

$$w_i = \frac{m_{\rm HI,i}}{1 + 4\pi n(d_i) J_3(r_{ij})}, \tag{5.11}$$

where $n(d_i)$ is the number density of the sample at the distance d_i to the i-th source, r_{ij} is the comoving separation between both objects and J_3 is an integral over the real-space isotropic correlation function:

$$J_3(r) = \int_0^r r'^2\, \xi(r') dr'. \tag{5.12}$$

Implementing these weights requires an assumption about the shape and amplitude of $\xi(r)$. For these we follow [78] and use $\xi(r) = (r/r_*)^{-1.51}$, with $r_* = 3.3\,h^{-1}\,\mathrm{Mpc}$. In fact, we find that fixing $J_3(r)$ to $J_3(r = 38\,h^{-1}\,\mathrm{Mpc}) = 2962\,\mathrm{Mpc}^3$ is enough to obtain a close-to-optimal correlation function (see figure 5.3). When implementing these weights we approximated the number density as $n(d) = n_0 \exp\left(-(d/d_0)^\gamma\right)$ where $n_0 = 0.23(\,h^{-1}\,\mathrm{Mpc})^{-3}$, $d_0 = 31.18\,h^{-1}\,\mathrm{Mpc}$ and $\gamma = 0.99$. These numbers were obtained by fitting the distance distribution of objects in the random catalog. Note also that equation 5.11 already includes the $m_{\rm HI}$ weights needed to recover the clustering properties of the total HI density.

Using this formalism, the measurement of the correlation function was carried out using the code `CUTE` [233]. We adopted a logarithmic binning in σ in the range $\sigma \in [0.11, 52)\,h^{-1}\,\mathrm{Mpc}$ with $\Delta \log_{10} \sigma/(h^{-1}\,\mathrm{Mpc}) = 0.12$, and we used 59 linear bins of π in the range $\pi \in [0.5, 59.5)\ \ h^{-1}\,\mathrm{Mpc}$. In order to eliminate the effect of redshift-space distortions and be able to compare our measurements with the real-space theoretical

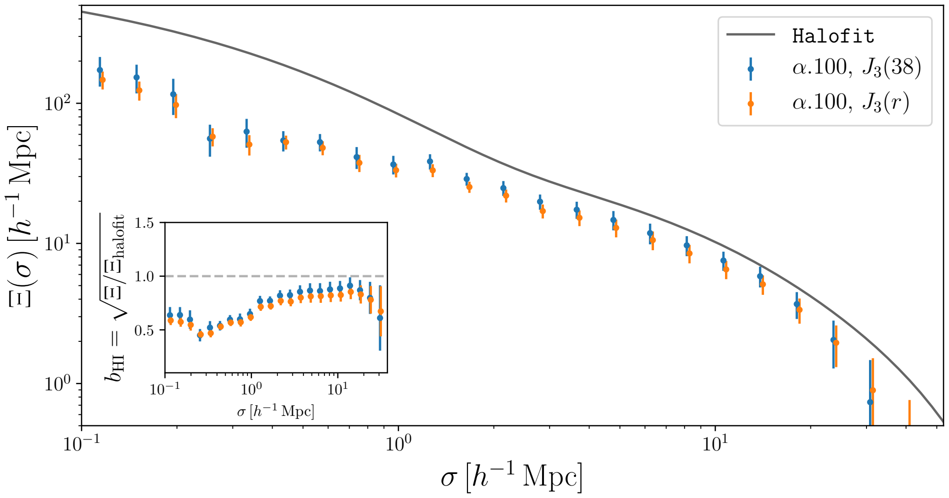

FIGURE 5.3: 2D projected correlation function. The points with error bars show m_{HI}-weighted correlation function computed for the α.100 data set, while the black solid line shows the `HaloFit` prediction for the matter correlation function at $z = 0$. Orange points show the measurements using pair-wise weights that depend explicitly on the pair separation (equation 5.11), while the blue points correspond to the case of fixing $J_3(r)$ to $J_3(38\,h^{-1}\,\mathrm{Mpc})$, independent of separation. The impact of the choice of weighting scheme is found to be negligible. The inset shows the scale-dependent HI bias b_{HI} as a ratio of the measurement with respect to the matter correlation function. Orange points have been slightly shifted to the right.

prediction, we compute the projected correlation function $\Xi(\sigma)$ by integrating $\xi(\pi, \sigma)$ along the line of sight:

$$\Xi(\sigma) = \int_{-\infty}^{\infty} d\pi\, \xi(\sigma, \pi) \simeq 2 \sum_{0}^{\pi_{\max}} \xi(\sigma, \pi)\Delta\pi, \tag{5.13}$$

where, as in [78], we used $\pi_{\max} = 30\,h^{-1}\,\mathrm{Mpc}$.

Figure 5.3 shows the measured HI-mass-weighted, projected correlation function (points with error bars) together with the prediction for the projected correlation function of the total matter overdensity, obtained from the HaloFit model for the matter power spectrum [234]. The scale-dependent HI bias is shown in the inset of the same figure as the square root of the ratio of both quantities. The measured b_{HI} is in good agreement with the measurement of the bias of HI-selected galaxies presented in [78]. This is to be expected, given the observation that the clustering of HI sources shows little or no dependence on HI mass[5].

Covariance matrix

We estimate the uncertainties on the measured projected correlation function using the jackknife resampling method [235, 236]. We divide the survey footprint into $N = 156$ contiguous patches covering $\sim 40\,\mathrm{deg}^2$ each. We remove one patch at a time

---

[5] Note that [96] observe a significant dependence on HI mass above $10^9\,M_\odot$. This possible dependence at high masses, however, does not alter our assumption that the ALFALFA sources can be used to study the properties of the overall HI distribution, including all structures below ALFALFA's detection limit.

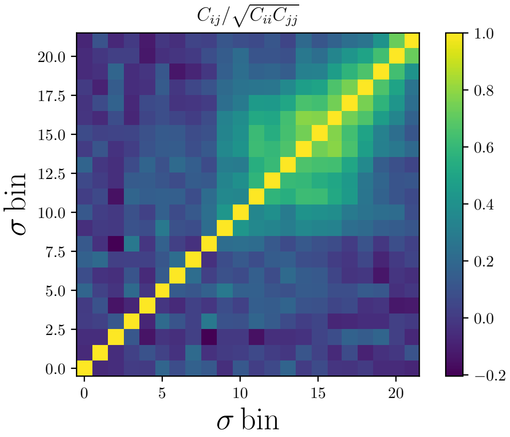

FIGURE 5.4: Jackknife correlation matrix for the projected 2-point correlation function. We use 22 logarithmic bins in the transverse separation σ in the range $\sigma \in [0.11, 30.8)\, h^{-1}$Mpc.

and measure the projected correlation function in the remaining area. The jackknife estimate of the covariance matrix is then given by:

$$C_{ij} = \mathrm{Cov}(\Xi_i, \Xi_j) = \frac{N_s - 1}{N_s} \sum_{p=0}^{N_s} (\Xi_i^p - \bar{\Xi}_i)(\Xi_j^p - \bar{\Xi}_j). \tag{5.14}$$

Here Ξ_i^p is the correlation function measured in the i-th bin after omitting the p-th patch and $\bar{\Xi}_i$ is the average of Ξ_i^p over all patches. Figure 5.4 shows resulting correlation matrix $r_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$.

Ultimately we are interested in the inverse covariance matrix. The inverse of the jackknife covariance is a biased estimate of the true inverse covariance, and we correct for this bias with an overall normalization factor [158]:

$$C^{-1} \to \frac{N_\mathrm{s} - N_\mathrm{b} - 2}{N_\mathrm{s} - 1} C^{-1}, \tag{5.15}$$

where $N_\mathrm{s} = 156$ is the number of jackknife samples and $N_\mathrm{b} = 22$ is the number of σ bins used in the analysis.

5.3.2 HI content in groups

As described in section 5.2.2, we also include direct constraints on the $M_\mathrm{HI}(M_h)$ relation in our analysis, coming from the matching of ALFALFA sources to optical members of galaxy groups with calibrated halo mass detected in the SDSS group catalog. To minimize a potential bias due to the incomplete coverage of the sky-projected area for each group, we estimate volume-correction factors for few large groups near the ALFALFA survey boundary. An estimate of the HI mass of each group is made by directly summing the masses of all ALFALFA member sources and applying the corresponding area correction factor, which is almost negligible for most of the groups. In general, this estimate of the group HI mass would be biased low, since the estimator will miss all ALFALFA sources with no optical counterparts lying in the comoving volume of each group, as well as any diffuse or unresolved HI component. The first cause of this bias (the sources with no optical detections) should have a negligible impact on this study, since it affects only $\sim 6\%$ [224] of

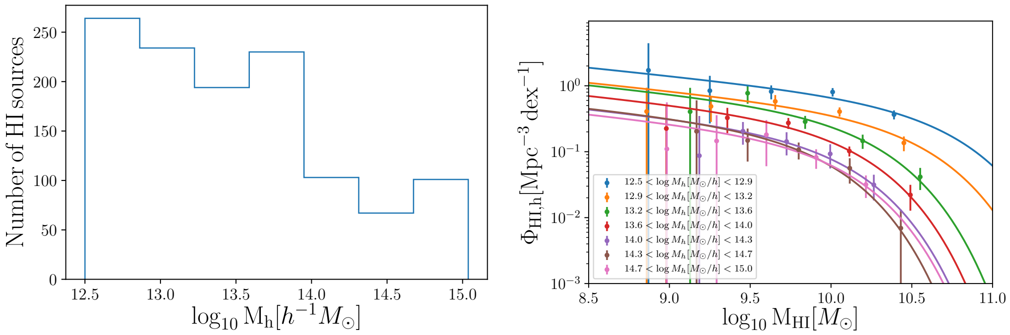

FIGURE 5.5: *Left:* the number of HI sources in the SDSS group catalog lying in each halo mass bin after RFI and 50% completeness cuts. *Right:* HI mass functions estimated from the SDSS group catalog using 2DSWML method in different halo mass bins (see legend).

all ALFALFA sources, and most of those are expected to be galactic high-velocity clouds, and not extragalactic in nature. To quantify and minimize the impact of contributions from undetected HI components, we estimate the HI mass function (i.e. the m_{HI} distribution of ALFALFA sources) in bins of group halo mass. The exact procedure is as follows:

1. We separate the SDSS group catalog into 7 logarithmically spaced bins of halo mass in the interval $\log_{10} M_h/(h^{-1}M_\odot) \in [12.50, 15.04]$. The top panel of figure 5.5 shows the number of HI sources lying in each of these mass bins.

2. In each bin we estimate the HI mass function $\phi(m_{\mathrm{HI}})$ using all the member sources found in the ALFALFA dataset. For this we use the 2D step-wise maximum likelihood (2DSWML) estimator described below.

3. In order to extrapolate below the detection limit, we model the measured mass function as a Schechter function with the form

$$\phi(m_{\mathrm{HI}}) = \ln(10)\,\phi_* \left(\frac{m_{\mathrm{HI}}}{M_*}\right)^{\alpha_s+1} \exp\left(-\frac{m_{\mathrm{HI}}}{M_*}\right). \tag{5.16}$$

4. For each halo mass bin, we compute the corresponding HI mass (and its uncertainty) by integrating over the reconstructed HI mass function, propagating all uncertainties as described below. We also compute a second estimate of the HI mass by integrating over the measured, model-independent 2DWSML mass function. This can only be done within the range of HI masses covered by ALFALFA, and the comparison of these two estimates then allows us to quantify the systematic uncertainty associated with undetected HI sources.

The list of reconstructed HI masses as a function of group halo mass is then appended to the correlation function described in the previous section to form the total data vector.

The 2DSWML mass function estimator

The idea of step-wise maximum-likelihood estimators has been applied in the past to reconstruct the luminosity function from a magnitude-limited sample [237, 238, 239]. The method is non-parametric, modeling the luminosity function as sum of top-hat functions, and finding their amplitudes by maximizing the likelihood of the observed sample. The latter is possible interpreting the luminosity function as a probability distribution. The same logic was applied by [90, 222] to estimate the HI mass function of ALFALFA sources, with the added complication that the completeness of the sample depends on both the HI flux S_{21} and the 21cm line width W_{50}. This gives rise to the 2-dimensional step-wise maximum-likelihood estimator (2DSWML), which we describe briefly here. To simplify the notation we will define here $\mu \equiv \log_{10} m_{\rm HI}/M_\odot$ and $w \equiv \log_{10} W_{50}/{\rm km\,s^{-1}}$.

The probability that a source g is detected with mass μ_g and line width w_g at distance d_g (within an interval $\Delta\mu$, Δw) is given by

$$p_g = \frac{\phi(\mu_g, w_g)\,\Delta\mu\,\Delta w}{\int_{-\infty}^{\infty} dw \int_{\mu_{\rm lim}(d_g,w)}^{\infty} d\mu\,\phi(\mu, w)}, \tag{5.17}$$

where $\phi(\mu, w)$ is the joint distribution of HI masses and line widths. Let us now model $\phi(\mu, w)$ as a 2D step-wise function, taking constant values in intervals of μ and w. Then, maximizing the log-likelihood $\mathcal{L} = \prod_g p_g$, we obtain an expression for the best-fit amplitudes $\phi_{i,j}$ in the i-th interval of μ and the j-th interval of w:

$$\phi_{i,j} = n_{i,j}\left[\sum_g \frac{H_{g,ij}}{\sum_{i',j'} H_{g,i'j'}\phi_{i',j'}}\right]^{-1}, \tag{5.18}$$

where g runs over all sources in the sample, $n_{i,j}$ is the number of galaxies in bin (i, j) and $H_{g,ij}$ is the mean completeness of the sample in that bin for sources at a distance $d = d_g$. The completeness function was determined as described in [90]. We imposed a hard cut on m and w, using only bins with completeness $> 50\%$. We verified that our results did not vary significantly with more stringent completeness cuts.

Note that equation 5.18 gives $\phi_{i,j}$ recursively as a function of itself, and in practice $\phi_{i,j}$ is found through an iterative process. Once a converged solution for $\phi_{i,j}$ has been found, the HI mass function is obtained by marginalizing over W_{50}:

$$\phi_i = \sum_j \phi_{i,j}\Delta w. \tag{5.19}$$

Finally, this method is able to determine $\phi_{i,j}$ up to an overall normalization constant. We fix this by matching the integral of $\phi(m_{\rm HI}, W_{50})$ to the total number of ALFALFA sources in each halo mass bin divided by the comoving volume covered by the corresponding halos, as described in Appendix B of [90].

The bottom panel of figure 5.5 shows the estimated HI mass functions in each halo mass bin used in this analysis, together with their best-fit Schechter models. For this figure, the mass functions were normalized dividing by the total volume enclosed within the virial radii of all groups in each halo mass bins. Note that, since we only use $\phi(m_{\rm HI})$ to estimate the $M_{\rm HI}(M_h)$ relation, our results are independent of this volume, and only depend on the total number of HI sources and galaxy groups in each M_h bin.

Error propagation

The uncertainties in the $M_{\rm HI}(M_h)$ relation inferred from the HI richness of groups, as described above, are driven by the errors in our estimate of the mass function in each M_h bin. Four main sources of uncertainty contribute to these errors [222], and we account for them as follows:

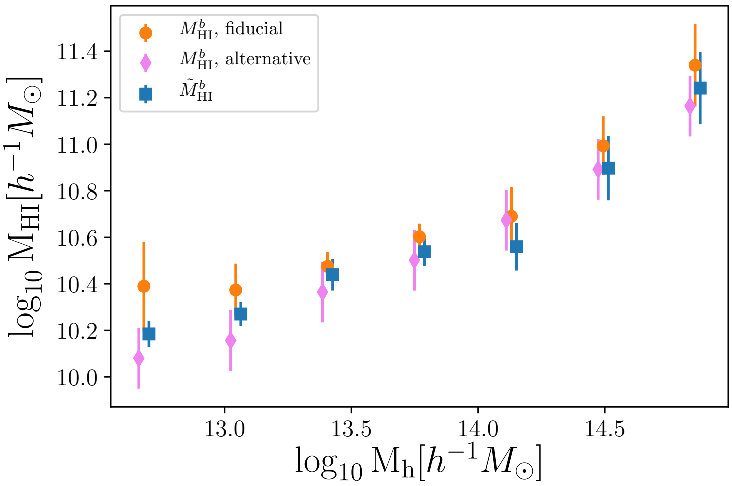

FIGURE 5.6: Estimated total $M_{\rm HI}$ in each halo mass bin obtained from the HI mass functions using three different methods. The total $M_{\rm HI}$ estimated by fitting the HI mass functions using the Schechter parametrization and accounting for the missing HI mass are shown with orange points. The error bars are computed by propagating the Schechter parameter uncertainties. The blue squares show the results of directly integrating the HI mass functions over the available range of HI masses, i.e. without extrapolation. The error bars in this case are computed by propagating the 2DSWML mass function uncertainties in quadrature. The violet diamonds show the second alternative estimate, found by rescaling the best-fit HI mass function in each halo mass bin (see section 5.3.2). The corresponding error bars are computed from the uncertainties in the mass function found by [222].

1. **Poisson:** with each measurement of ϕ_i we associate a Poisson-counting error given by $\sigma(\phi_i) = \phi_i/\sqrt{N_i}$, where N_i is the number of sources contributing to the i-th $m_{\rm HI}$ bin.

2. **Sample variance:** the uncertainty associated with the stochastic variation in ϕ_i induced by the particular density fluctuations covered by the survey volume of ALFALFA was quantified through the jackknife resampling method described in section 5.3.1. In this case we use the 10 jackknife regions shown as red dotted lines in figure 5.1.

3. **Mass measurement errors:** the HI mass of each source is inferred from its 21cm flux and its radial comoving distance. Both quantities have associated measurement uncertainties which propagate into $m_{\rm HI}$, shifting sources between different HI mass bins. To account for this, we generated 100 random realizations of the α.70 catalog by adding a random Gaussian error to the distances and fluxes of all sources (with a standard deviation given by their estimated error). We re-computed the HI masses and corresponding $\phi(m_{\rm HI})$ for each realization (see equation 5.8), and estimate the uncertainty associated to these errors from the scatter of all realizations.

4. **Line width measurement errors:** errors in W_{50} also affect our measurement of the 2DSWML mass function, by shifting sources between different W_{50} bins. The associated uncertainties were estimated from 100 random realizations, following the same procedure described above for mass measurement errors.

We assumed that the errors associated with these 4 sources are uncorrelated and added them in quadrature to find the final uncertainties on ϕ_i.

Once ϕ_i and its uncertainties have been measured, we find the best-fit Schechter models in each M_h bin. To avoid over-fitting, given the relatively small number of points in which we estimate the mass function for each bin, we fix the tilt of the Schechter function to its best-fit value for the overall HI mass function as reported by [222], $\alpha_s = -1.25$. The best-fit Schechter functions in each M_h bin are shown as solid lines in figure 5.5.

To estimate the uncertainties in the Schechter parameters (ϕ_*, M_*), we sample their likelihood running a Markov chain Monte Carlo (MCMC). For any point (ϕ_*, M_*) in these chains, the corresponding HI mass for halos in the b-th M_h bin can be estimated as:

$$\begin{aligned} M^b_{\rm HI} &= \frac{V_b}{N^b_{\rm group}} \int_0^\infty \phi(m_{\rm HI})\, m_{\rm HI}\, d\log_{10} m_{\rm HI} \\ &= \frac{V_b \phi_*\, M_*}{N^b_{\rm group}}\, \Gamma\left(2+\alpha\right), \end{aligned} \tag{5.20}$$

where V_b is the uncorrected volume spanned by all groups in the b-th M_h bin and $N^b_{\rm group}$ is the corresponding number of groups. Our final estimate of the $M_{\rm HI}(M_h)$ relation (and its uncertainty) from the galaxy group data is then given by the mean of $M_{\rm HI}$ (and its scatter) across all points in the MCMC chain. Finally, we correct our results for self-absorption by assuming it has a 10% effect on the measured flux. Note that strictly speaking this is only suitable for the galaxies around the knee mass as described in [222]. The results are shown as orange points with error bars in figure 5.6.

Since our measurement of $M^b_{\rm HI}$ involves extrapolating the HI mass function to very small masses, below the ALFALFA detection limit at the group's redshift, it is worth quantifying the impact of this extrapolation on our results. We do so here by comparing the fiducial measurement of $M_{\rm HI}^b$ described above, with two alternative estimates:

1. The first estimator is given by directly integrating the measured 2DSWML mass function over the available range of HI masses in ALFALFA. Labeling the 2DSWML in the b-th halo mass bin as ϕ_i^b, this alternative estimate is given by

   $$\tilde{M}_{\rm HI}^b = \frac{V^b}{N_{\rm group}^b} \sum_i \phi_i^b \, 10^{\mu_i} \, \Delta\mu. \tag{5.21}$$

   The uncertainty on $\tilde{M}_{\rm HI}^b$ can be estimated trivially from the uncertainties on ϕ_i^b. Since $\tilde{M}_{\rm HI}^b$ and $\hat{\phi}_i^b$ are linearly related, the uncertainties on ϕ_i^b, quantified as described above, can be propagated into $\tilde{M}_{\rm HI}^b$ in quadrature.

2. The second estimator is produced by rescaling the best-fit HI mass function found by [222] in each halo mass bin. The rescaling factor for each group in the bin is estimated as the ratio of the observed number of sources found in that group to the number expected given the 2DSWML estimate of [222] accounting for sample completeness at the distance to the group. $M_{\rm HI}^b$ is then estimated by applying equation 5.20 to the Schechter function found in [222] rescaled by the factor above.

   Unlike our fiducial estimator, this alternative method has no free parameters, and can therefore be used to explore the possible consequences of over-fitting the per-bin mass functions based on a small number of objects. The main drawback of this estimator is that, by constructions, it assumes that the $m_{\rm HI}$ distribution in groups is the same as in the field.

These alternative measurements of the $M_{\rm HI}(M_h)$ relation are shown as blue squares and pink diamonds with error bars in figure 5.6.

As could be expected, the measurements corresponding to the first alternative estimator are consistently below our fiducial estimates generated from the integral of the Schechter functions, with the missing mass corresponding to the contribution of sources below the ALFALFA detection limit. However, the associated mass difference is mostly below ∼25% of our fiducial mass measurements throughout the full mass range. Since this offset is always smaller than the 1σ statistical uncertainties, we find the impact of extrapolating the mass function to lower masses to be minimal. Note also that the blue error bars are consistently smaller than the orange ones. This is also to be expected, since the errors on $\tilde{M}_{\rm HI}^b$ estimated as described above, do not account for the additional uncertainty associated with mass below the detection limit.

The second estimator, based on extrapolating the overall HI mass function, agrees well with our fiducial measurements in general, although it is noticeably lower in the two lowest M_h bins. This is caused by the larger value of M_* preferred by our Schechter fits in the low-M_h bins. This result is consistent with previous measurements of the HI mass function around the region of the Virgo cluster, which suggest that massive ($\sim 10^{15} M_\odot$) halos have a smaller M_* than the field. Although this could be caused by ram pressure or tidal stripping, a better understanding of this result will require a more detailed study of the HI content in low-mass halos in both data and simulations [57]. In any case, both estimates of $M_{\rm HI}^b$ are compatible within present uncertainties, and therefore we conclude that our measurements of this quantity are robust with respect to the method used to estimate it.

5.4 Results

5.4.1 Fiducial results

We produce constraints on the three parameters of the $M_{\rm HI}(M_h)$ relation (equation 2.3), $\theta \equiv \{\log_{10} M_0, \log_{10} M_{\rm min}, \alpha\}$, from a joint data vector composed of three parts:

1. Measurements of the projected correlation function $\Xi(\sigma)$ (see section 5.3.1) in $N_\Xi = 17$ logarithmic bins of σ between $0.43\,h^{-1}{\rm Mpc}$ and $30.8\,h^{-1}{\rm Mpc}$. We use the altered NFW HI density profile described in section 5 as our fiducial model for the small-scale correlation function. We study the impact of this choice, as well as the choice of scale cuts in section 5.4.2.

2. Direct measurements of the $M_{\rm HI}(M_h)$ relation (see section 5.3.2) in the $N_M = 7$ logarithmic bins of halo mass shown in figure 5.6. Our fiducial measurements consist of the $M_{\rm HI}$ estimates derived from the integral of the best-fit Schechter HI mass functions in each M_h bin. We show the impact of extrapolating the HI mass function below ALFALFA's detection limit on our results in section 5.4.3.

3. One measurement of the cosmic HI abundance $\Omega_{\rm HI} = (3.9{\pm}0.1\,({\rm stat.}){\pm}0.6\,({\rm syst}))\times 10^{-4}$ from ALFALFA's α.100 sample, as reported by [222]. In terms of the halo model, the cosmic abundance receives contributions from the HI content of halos with arbitrarily small masses. Since our direct measurements of the $M_{\rm HI}(M_h)$ relation do not go below $\log_{10} M_h/(h^{-1}{\rm Mpc}) \simeq 12.5$, this additional data point allows us to break the degeneracy between the overall amplitude M_0 and the minimum halo mass $M_{\rm min}$ of the $M_{\rm HI}(M_h)$ relation.

Our fiducial data vector $\mathbf{d}$ therefore contains $N_\Xi + N_M + 1 = 25$ elements, which we use to constrain the three-parameter model of the $M_{\rm HI}(M_h)$ relation (in addition to the profile concentration parameter $c_{\rm HI,0}$, which we marginalize over). Assuming Gaussian statistics for $\mathbf{d}$, and in the absence of priors, the posterior distribution of the model parameters θ is given by:

$$\chi^2 \equiv -2\log p(\theta|\mathbf{d}) = [\mathbf{d} - \mathbf{t}(\theta)]^T\,\hat{\mathsf{C}}^{-1}[\mathbf{d} - \mathbf{t}(\theta)], \tag{5.22}$$

where $\mathbf{t}(\theta)$ is the theoretical prediction for $\mathbf{d}$, described in section 5.1, and $\hat{\mathsf{C}}$ is the covariance matrix of our measurements.

We build the covariance matrix $\hat{\mathsf{C}}$ as a block-diagonal matrix, where the first $N_\Xi \times N_\Xi$ block is given by the covariance matrix of the correlation function measurements (see figure 5.4). We assume the remaining $N_M + 1$ elements (corresponding to the HI abundance in groups and the cosmic HI abundance) to be uncorrelated with the correlation function measurements, and that their statistical uncertainties are also uncorrelated among themselves. These measurements are, however, correlated through some of their systematic uncertainties. In particular, the calibration of the absolute flux scale in ALFALFA dominates the systematic error budget in the measurement of $\Omega_{\rm HI}$ and $M^b_{\rm HI}$, and should affect all of these quantities in the same manner, rescaling them by an overall factor. In order to incorporate this correlation in our analysis we add, to the statistical covariance matrix described above, a systematic component that is fully correlated across the last $N_M + 1$ measurements and with an amplitude 0.6×10^{-4} in the $\Omega_{\rm HI}$-$\Omega_{\rm HI}$ component. Note that the measurement of the projected correlation function is immune to the effects of an overall rescaling factor, and therefore the corresponding part of the systematic contribution to the covariance matrix is fixed to 0.

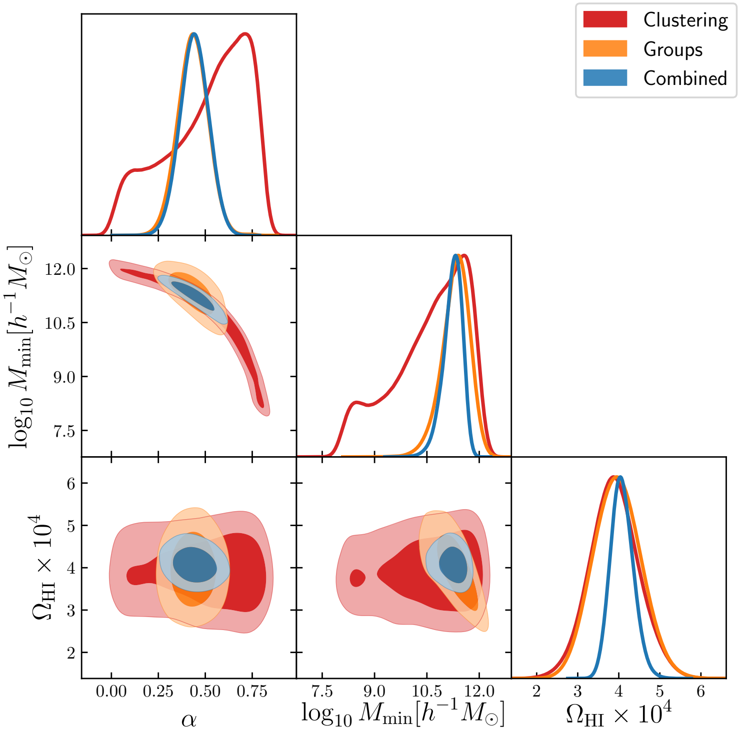

FIGURE 5.7: Final constraints on the parameters of the $M_{\rm HI}(M_h)$ relation. Results are shown for the combination of clustering data and the $\Omega_{\rm HI}$ measurement (red), for the measurements of $M_{\rm HI}$ in groups and $\Omega_{\rm HI}$ (light orange) and for the combination of the three datasets (blue).

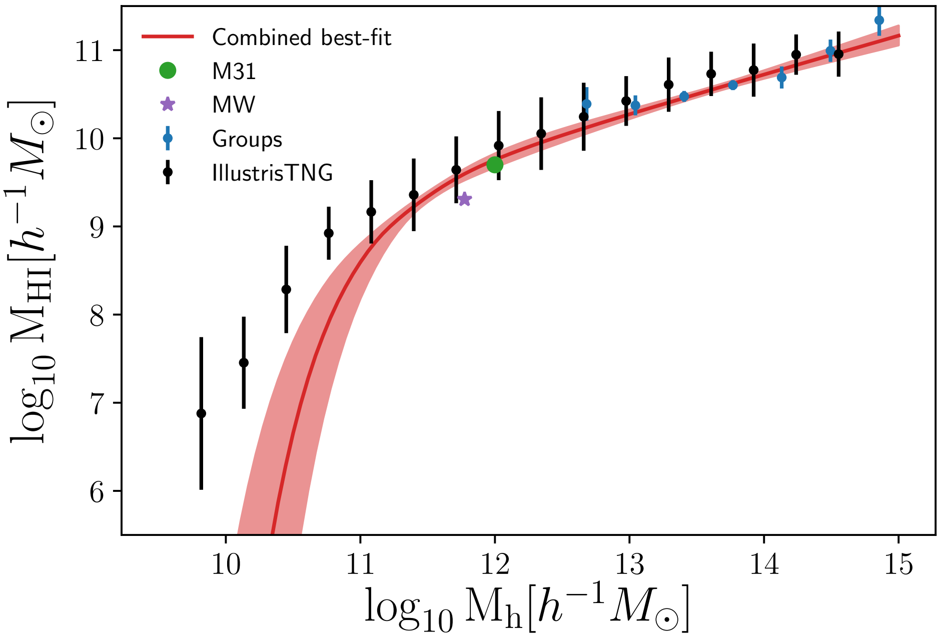

FIGURE 5.8: Combined best-fit $M_{\rm HI}(M_h)$ relation (red solid line) together with the 1σ uncertainty (red shaded region) using three datasets: the projected mass-weighted correlation function $\Xi(\sigma)$, the direct estimates of the $M_{\rm HI}(M_h)$ relation from the galaxy group catalog (shown also as blue points with error bars) and the measurement of the cosmic HI abundance $\Omega_{\rm HI}$ in [222]. For comparison we show the results from the IllustrisTNG magneto-hydrodynamic simulation (black points) [57] with the error bars corresponding to the typical per-halo scatter. We also show measurements of M_h and $M_{\rm HI}$ for individual galaxies: the Milky Way (purple star) and M31 (green circle).

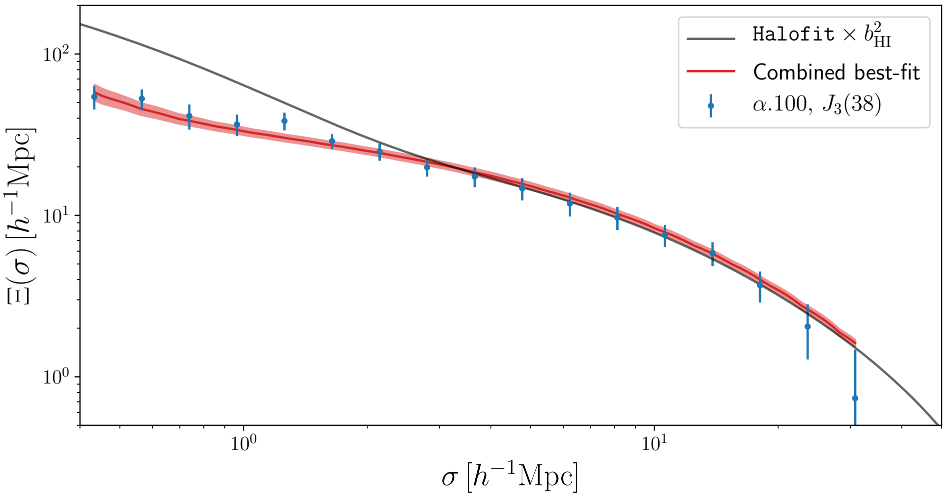

FIGURE 5.9: Predicted projected correlation function for our best-fit $M_{\rm HI}(M_h)$ relation (red solid line) and its 1σ uncertainty (red shaded region). The blue points with error bars show the direct measurements from ALFALFA, and the black solid line corresponds to the HaloFit prediction for the matter correlation function scaled by our our best-fit $b^2_{\rm HI}$ (see section 5.5).

Finally, given the residual degeneracy between $M_{\rm min}$ and M_0 in our parametrization, we choose to show all our results in terms of $(\alpha, \log_{10} M_{\rm min}, \Omega_{\rm HI})$ instead, but will also provide the corresponding best-fit value and uncertainty on $\log_{10} M_0$. We use broad top-hat priors for all parameters, with $c_{\rm HI,0} \in [0, 100]$, $\alpha \in [0, 2]$, $\log_{10} M_{\rm min}/h^{-1}M_\odot \in [8, 13]$ and $\Omega_{\rm HI} \times 10^4 \in [0, 20]$. In all cases we show constraints on $\log_{10} M_0$, $\log_{10} M_{\rm min}$ and α marginalized over the concentration parameter $c_{\rm HI,0}$.

We sample the likelihood in equation 5.22 using the publicly available implementation of the Markov chain Monte Carlo algorithm `emcee` [157]. The resulting constraints on the $M_{\rm HI}(M_h)$ parameters are shown in figure 5.7 for different data combinations. We find compatible constraints from the clustering and groups data separately. Our marginalized constraints on the $M_{\rm HI}(M_h)$ parameters are $\alpha = 0.44 \pm 0.08$, $\log_{10} M_{\rm min}/h^{-1}M_\odot = 11.27^{+0.24}_{-0.30}$, $\log_{10} M_0/h^{-1}M_\odot = 9.52^{+0.27}_{-0.33}$. The maximum-likelihood values are a good fit to the data in all cases, with a $\chi^2 = 13.7$ for 21 degrees of freedom for the full data vector. Although the clustering data is not able to jointly measure α and $\log_{10} M_{\rm min}$, and the groups data dominates the final uncertainties, clustering is still important in tightening the constraints (see e.g. the α-$\log_{10} M_{\rm min}$ plane). In particular, we find that, within this model, the clustering measurements allow us to reduce the uncertainty on $\Omega_{\rm HI}$ with respect to the mass-function measurement of [222], obtaining $\Omega_{\rm HI} = 4.07^{+0.29}_{-0.26} \times 10^{-4}$.

Figure 5.8 shows our best-fit $M_{\rm HI}(M_h)$ relation (red solid line), together with its 1σ uncertainty (shaded area) as well as our fiducial measurements of this relation on galaxy groups (blue points with error bars). The measurements from the DR7 group catalog are shown in blue. In order to jointly reproduce the measured HI content in high-mass halos as well as the measured total HI abundance, the model predicts a sharp drop in HI content below a halo mass $\log M_h/h^{-1}M_\odot \sim 11.5$. The figure also shows, as black points, the $M_{\rm HI}(M_h)$ relation measured in the IllustrisTNG-100 magneto-hydrodynamic simulation [57] from a cosmological volume of $(75\ h^{-1}{\rm Mpc})^3$. The errorbars represent the 1σ halo-to-halo variation on $M_{\rm HI}(M_h)$. Although, overall, we find good agreement between our results and the simulation, for very small halo masses, the amplitude of $M_{\rm HI}(M_h)$ differs significantly between our results and IllustrisTNG. This is however expected, given that the value of $\Omega_{\rm HI}$ in IllustrisTNG is $\simeq 7.5 \times 10^{-4}$, i.e. roughly a factor of 2 larger than the ALFALFA measurement used here. Although our model predicts a larger low-mass cutoff than is found in simulations, existing data on halo masses below the range probed by the SDSS group catalog are not incompatible with this prediction. To illustrate this, figure 5.8 also shows the HI and halo masses measured for the Milky Way and M31 [54, 240].

Finally, figure 5.9 shows our measurement of the projected correlation function (blue points) together with the best-fit prediction and associated uncertainties (red line and shaded area) and the dark matter correlation function from HaloFit (black solid line) scaled by our our best-fit $b^2_{\rm HI}$ (see section 5.5). We also note that in contrast to [96], we are able to reproduce the measured HI clustering without involving assembly bias effects.

5.4.2 Impact of small scales

On small scales, the halo-model prediction of the 2-point correlation function is dominated by the shape of the HI density profile. It is therefore important to evaluate whether our assumptions regarding the distribution of HI within each halo impacts our results on their overall HI content.

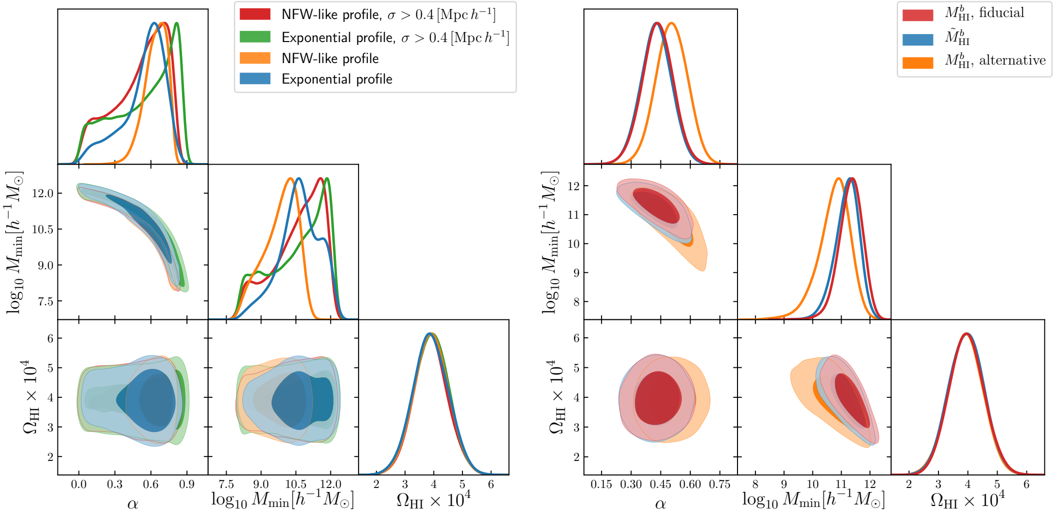

FIGURE 5.10: *Left:* constraints on the $M_{\rm HI}(M_h)$ relation derived from the clustering analysis under different scale cuts and choices of HI density profile. We derive our final scale cuts by demanding final constraints that do not depend on the choice of profile. *Right:* constraints on the $M_{\rm HI}(M_h)$ relation for different estimates of the HI mass in galaxy groups. Our fiducial measurements are shown in blue, while the red and light-orange contours show the results from the two alternative estimates described in section 5.3.2 (see also figure 5.6).

The blue and light-orange contours in the top panel of figure 5.10 show the constraints on the $M_{\rm HI}(M_h)$ relation derived from the measurements of the projected correlation function for the exponential and altered NFW profiles described in section 5.1 respectively. Constraints are shown for the full range of scales ($\sigma \in (0.11, 30.8)\,h^{-1}$Mpc) and combined with the ALFALFA measurement of $\Omega_{\rm HI}$. The figure shows that the constraints on the $M_{\rm HI}(M_h)$ parameters (particularly in terms of uncertainty) depend significantly on the model used to describe the distribution of HI within each halo. This is an undesirable feature, since we aim to constrain the global parameters of the $M_{\rm HI}$-M_h relation, given the large uncertainties in the actual shape of the HI density profile. On sufficiently large scales, in the 2-halo regime, this dependence should become negligible. We have verified this by removing all data points with $\sigma > 0.43\,h^{-1}$Mpc. These results are shown in figure 5.10 in green and red for the exponential and altered NFW profiles respectively. The dependence on the choice of profile, in terms of constraining power, vanishes in this regime. We thus use this restricted range of scales and the altered NFW profile for our fiducial analysis. Although the choice of profile in this regime is not relevant, we note that [57] find that the altered NFW profile with an exponential cut-off on small scales is better able to fit measurements from hydrodynamical simulations.

5.4.3 Low-mass extrapolation

As described in section 5.3.2, our measurement of the HI content of galaxy groups is based on extrapolating the HI mass functions measured in bins of halo mass beyond the detection threshold of ALFALFA. This is a legitimate approach as long as the range of masses covered by our sample constitute the main contribution to the total HI budget, in which case we only incurr in a small systematic effect when extrapolating the abundance of low-HI sources. We have shown that the mass deficit is generally

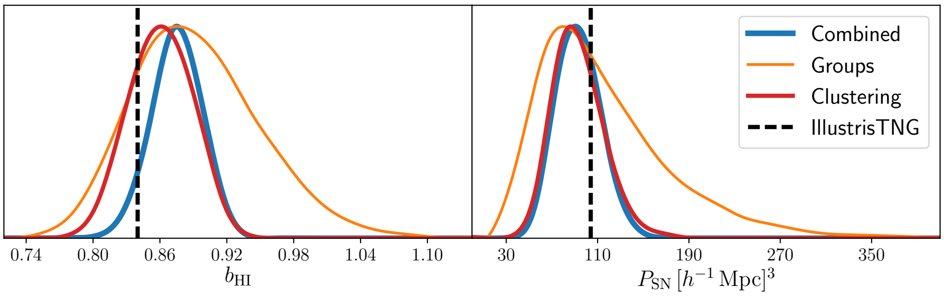

FIGURE 5.11: Posterior distributions for the large-scale HI bias (*left*) and shot-noise power spectrum (*right*) predicted from different combinations of our fiducial data vector: clustering+$\Omega_{\rm HI}$ (red), groups+$\Omega_{\rm HI}$ (light orange) and all data (blue). The vertical dashed line shows the result found in the IllustrisTNG simulation [57].

below $\sim 20\%$ of each individual HI mass measurement, and is always within the 1σ uncertainties. The right panel of figure 5.10 shows the impact of this systematic on our final constraints on the parameters of the $M_{\rm HI}(M_h)$ relation. The figure shows the constraints derived from our fiducial $M_{\rm HI}$ measurements in red, as well as the contours corresponding to our two alternative estimates: summing over the 2DSWML mass function (blue) and re-scaling the global HI mass function (light orange). The constraints derived from both estimates are compatible, with a negligible shift in the best-fit $\log_{10} M_{\rm min}$. We therefore conclude that any residual systematics in the method used to measure the HI content as a function of halo mass in the group catalog is subdominant.

5.5 Discussion

We have placed constraints on the distribution of neutral hydrogen in dark matter halos as a function of halo mass. To do so we have used the HI-weighed clustering of 21cm sources detected by ALFALFA, as well as the abundance of those sources in halos identified in the galaxy group catalog compiled from the SDSS DR7 data. Our results show a power-law relation between $M_{\rm HI}$ and M_h at large halo masses with an exponent $\alpha = 0.44 \pm 0.08$. This relation is exponentially suppressed on masses below $\log_{10} M_{\rm min}/h^{-1}M_\odot = 11.27^{+0.24}_{-0.30}$. Although this suppression is not directly measurable in the data, given the mass range of the group catalog, it can be inferred indirectly by combining the group data with the total HI abundance measured by ALFALFA and our measurement of the 2-point correlation function.

The constraints derived individually from our two datasets are compatible between themselves and with the combined constraints, and in all cases we find the model in equation 2.3 to be a good fit to the data. It is worth emphasizing the fact that, although the clustering data is not able to break the degeneracy between α and $M_{\rm min}$, even when combined with the measurement of $\Omega_{\rm HI}$, it is vital to improve the constraints derived from the combination of the HI abundance in groups and $\Omega_{\rm HI}$. In fact we find that, within our model, clustering information is able to significantly reduce the final uncertainties on $\Omega_{\rm HI}$ compared with direct measurements of this quantity from the HI mass function. Furthermore, the clustering properties of the HI are arguably the most relevant piece of information for future 21cm IM studies, and this information is potentially better summarized by the projected correlation function data used here.

Recently [57] have aimed at characterizing the $M_{\rm HI}(M_h)$ relation from state-of-the-art magneto-hydrodynamic simulations, and it is therefore relevant to explore the level of agreement between these simulated results and our data-driven constraints. In terms of the overall $M_{\rm HI}(M_h)$ relation, this comparison is best summarized in figure 5.6. We find that our results agree well with those of [57] at $z = 0$ for large halo masses ($M_{\rm h} \gtrsim 10^{12.5}\, M_\odot/h$), and that our best-fit model as well, as the simulated data, are in good agreement with individual HI mass measurements. However, we observe that the $M_{\rm HI}(M_h)$ relation derived from simulations departs significantly from our best-fit model on the low-mass end, predicting significantly higher HI masses. This disagreement is correlated with the higher value of $\Omega_{\rm HI} \sim 7 \times 10^{-4}$ measured in IllustrisTNG, which is also the measurement that allows us to place constraints on the cutoff mass scale. The fact that the radiation from the sources is not accounted for in IllustrisTNG may explain the differences in the value of $\Omega_{\rm HI}$ and on the average HI mass inside small halos.

For the purposes of predicting the clustering properties of HI in future 21cm experiments, two quantities are needed beyond $\Omega_{\rm HI}$: the large-scale HI bias $b_{\rm HI}$ and the shot-noise level $P_{\rm SN}$. Given our model for the $M_{\rm HI}(M_h)$ relation, we can make predictions for these two quantities within the halo model ($b_{\rm HI} = F_1^1(k = 0)$, $P_{\rm SN} = F_2^0(k = 0)$, see equation 2.7), which we can then directly compare with the values found by [57]. The results of this comparison are shown in figure 5.11: our constraints on both quantities ($b_{\rm HI} = 0.875 \pm 0.022$, $P_{\rm SN} = 92^{+20}_{-18}\, [h^{-1}{\rm Mpc}]^3$) are in good agreement with the values predicted by IllustrisTNG at $z = 0$. Although this result may seem at odds with the disagreement between data and simulation in terms of the total $\Omega_{\rm HI}$, this can be understood as due to the relatively higher contribution from larger-mass objects to these two quantities, for which our results agree with those of IllustrisTNG. It is also interesting to note that, even though the clustering data alone is not able to break the degeneracies between the $M_{\rm HI}(M_h)$ parameters, they drive the constraints on both $b_{\rm HI}$ and $P_{\rm SN}$.

Our measurement of the $M_{\rm HI}(M_h)$ relation can be translated into a limiting circular velocity to host HI. Defining this as the circular velocity associated with a minimum halo mass such that 98% of the cosmic HI is contained within heavier objects (see [57]), we find $V_{\rm circ} = 53^{+9}_{-11}$ km/s. This is in tension with the value found in [57] ($V_{\rm circ} = 34\,{\rm km/s}$), which is correlated with the higher cutoff halo mass measured in the data and shown in figure 5.6.

The results presented here are also interesting beyond future cosmological 21cm studies, as they provide insight into the distribution of neutral hydrogen across structures of different masses. Furthermore, our direct measurement of the $M_{\rm HI}(M_h)$ relation is based on the characterization of the HI mass function for sources within galaxy groups, and have revealed hints about the relative dependence of the HI mass distribution on halo mass, with higher HI knee masses found on lower-mass halos. In general, the behaviour of the $M_{\rm HI}(M_h)$ relation in the low-mass end ($M_h \lesssim 10^{12}\, M_\odot/h$) is still somewhat uncertain, and its study will benefit in the future from higher-quality data and improved analysis methods.

We must also emphasize that the $M_{\rm HI}(M_h)$ relation contains a huge amount of astrophysical information. In the high-mass end, the strength of processes such as AGN feedback, ram pressure and tidal stripping will leave its signature on the value of α [128], while on the low-mass end the presence of the UV background and the minimum mass to trigger self-shielding will determine the shape and amplitude of $M_{\rm HI}(M_h)$. Our results can be used in combination with hydrodyanmic simulations or semi-analytic models [241, 242] to improve our knowledge on the role of different astrophysical processes.

Chapter 6

Summary

The main goal of current cosmology is to understand the content, geometry, dynamics and evolution of our Universe. There has been many different approaches in order to achieve that goal. They include studying the early Universe through CMB anisotropies and mapping the spatial distribution of galaxies and other structures and objects in the late Universe. Based on these studies and observations we now have a well-established standard ΛCDM model that has only six free parameters. This model describes various observations very well and all of the different observables agree rather well in terms of the inferred values of the cosmological parameters. There are however tensions between different observables, especially in the case of H_0 parameter value obtained from direct local measurements and the value inferred from the CMB measurements [243, 244]. These tensions could give us insight into the new physics or could be due to the unaccounted systematical errors in our measurements. Even though ΛCDM is fairly simple and has only 6 free parameters, there are at least two major components for which we still do not understand the nature of – DM and DE.

In order to learn more, we are at the stage at which we simply need more experimental data. The good news is that there are plenty of experiments taking or planned to be taking data in the near future and that we are starting to use most of the EM spectrum and recently also the gravitational waves [245], to observe objects across the Universe. Galaxy surveys (both spectroscopic and photometric) have proven to be extremely useful and have provided us with a wealth cosmological information at low redshifts in recent years [7, 8]. In the next decade we will witness even better surveys with the hope of learning much more about our Universe [160, 161, 162, 163]. With galaxy surveys we are using galaxies as tracers of the underlying DM density field. However, there are other tracers that could in principle supplement the galaxy surveys and provide additional information on the DM density field. One such tracer is the HI and the 21cm emission from it in the post-reionization Universe.

With the 21cm IM we could in principle map much larger volumes compared to any other approach and short time. Thus the field of 21cm is rather promising as a new LSS probe. However, the field is also rather young still and there are many technical issues that need to be resolved before we can see the promises delivered. The main issue is the relative weakness of the 21cm signal compared to the bright foregrounds. The hope is that we will be able to use the fact that most of the foregrounds have smooth spectral dependance and can in principle be removed and the 21cm signal distinguished and analysed. Until this is achieved with current or upcoming surveys, the theoretical modelling of the HI is also rather important to better understand the signal we are after. Additionally, methods to extract most of the information once the signal is detected need to be developed beforehand.

In this thesis, we have focussed on using the HI as the LSS probe. For that purpose we have:

- developed a method to reconstruct the BAO peak in the 21cm IM surveys.
- studied the gains in performing 21cm IM surveys at redshifts which is currently vastly unobserved
- used already existing data and numerical simulation in order to constrain and improve the HI halo model needed to make accurate and precise predictions on the 21cm signal and the noise.

Chapter 1 is a general introduction into the basic and well known concepts in cosmology. This introduction was intended to give a clearer understanding of the following chapters. We briefly describe homogeneous Universe, its evolution and the evolution of its components. Then we move on to describe how small inhomogeneities, seeded in the very early Universe, grow to form structure that we observe later on in our Universe. We describe the linear theory of perturbations needed in order to model and predict the observations and the structure on linear scales. We then proceed to discuss the origin and application of BAO as a standard ruler as a way to constrain the cosmological parameters by measuring the BAO scale at different epochs of the Universe. We explain the effect of non-linearities on the BAO and engage the BAO reconstruction method which we will discus further in chapter 3.

Chapter 2 is a brief and general introduction into the 21cm cosmology. We explain the advantages of using the IM method to map the spatial distribution of HI across the Universe. We describe the expected signal of the 21cm power spectrum that depends on the total amount of HI in the Universe and the HI bias. We explain the HI halo model we use in order to make prediction of the Hi bias and the shot-noise. We further explain the main issues with observing the 21cm signal, i.e. the foregrounds.

Additionally, in chapter 2 we have also presented the results of determining the HI density profile from numerical simulations IllustrisTNG. We have found that the HI density profiles inside halos exhibit a large halo-to-halo variation, but that the average HI density profiles are universal and can be reproduced by an altered NFW profile with an exponential cutoff on small scales. This result is very useful as it is one of the main ingredients used to describe the HI clustering in the HI halo model.

In chapter 3 we have tested a method in detail that is able to perform BAO reconstruction in both galaxy surveys and pixelated maps as those from 21cm IM surveys. It consists in moving pixels rather than galaxies and it is equivalent to standard reconstruction in the limit of a very fine grid. We have tested this method by using a large set of numerical simulations and find an excellent agreement with theoretical expectations. We find that this method is able to decrease the uncertainty in the BAO peak position by 30-50% over the typical angular resolution scales of 21 cm IM experiments. We believe this method can be particularly useful to tighten the constraints on the value of the cosmological parameters from IM observations in the post-reionization era from surveys such as CHIME and SKA.

In chapter 4 we have performed a detailed study of the advantages of performing cosmological studies with 21cm IM surveys in the redshift range $2.5 < z < 5$. We consider suitable extensions of current or upcoming single-dish radio telescopes and interferometers. By using the Fisher matrix formalism we have obtained forecasted bounds that those instruments could put on several cosmological parameters. We focus our results on the growth rate and the BAO distance scale parameters, but also on two parameters of ΛCDM extension – the sum of the neutrino masses and the number of relativistic degrees of freedom at decoupling, $N_{\rm eff}$. We have also proposed several strategies to independently constrain $\Omega_{\rm HI}$ and $b_{\rm HI}$ through cross-correlations with other probes, in order to be able to measure any other quantity that depends on the amplitude of the 21cm power spectrum. We study in detail how our results depend on the instrument, HI bias, the foreground wedge coverage, the nonlinear scale used in the analysis, uncertainties in the theoretical modeling and the priors on $b_{\rm HI}$ and $\Omega_{\rm HI}$. We find that an instrument performing 21cm IM observations in this redshift range can provide very tight constraints on the considered cosmological parameters.

In chapter 5 we have used the HI-weighed clustering of 21cm sources detected by ALFALFA, as well as the abundance of those sources in halos identified in the galaxy group catalog compiled from the SDSS DR7 data to place constraints on the distribution of neutral hydrogen in dark matter halos as a function of halo mass. We used a halo-model-based approach to model the abundance and clustering of HI by parametrizing the $M_{\rm HI}(M_h)$ relation as a power law with an exponential mass cutoff. We also include a recent measurement of the cosmic HI abundance from the α.100 sample. We find datasets to be consistent in derived parameters and also in agreement with the state-of-the-art magneto-hydrodynamical simulations. Based on these results we make a prediction for the HI bias and the HI shot-noise. These clustering properties of the HI derived in this work are a very relevant piece of information for future 21cm IM studies.

Future perspectives

There are still many interesting and open questions to be addressed. Here we give an outlook on few questions we believe will be useful to study in the future work.

Future galaxy surveys (Euclid, DESI, LSST etc.) will provide extremely precise measurements of the galaxy power spectrum and consequently give tight constraints on the BAO scale, growth rate, neutrino masses etc. This push towards higher precision measurements will at some point require the uncertainty in our theoretical modelling to be better than the experimental precision. In order to extract unbiased and accurate cosmological information from these measurements we will need to be very careful about the systematic uncertainties in our theoretical model [246].

For example, the effect of massive neutrinos is only a few percent at the scales where the theoretical uncertainties of the power spectrum coming from perturbation theory are comparable [191]. Going beyond perturbation theory, it will be interesting to quantify the uncertainties of the HOD models and answer how accurately we must know the HOD parameters in order to be able to detect the minimal sum of the neutrino masses.

In chapter 3 we discussed several BAO reconstruction methods. However, none of these methods take into account the presence of massive neutrinos. As we discussed in 4.3.1, halos and galaxies are biased tracers of the CDM+baryons field only. Thus it is important to take into account the effect of neutrinos on the displacement fields needed to move of galaxies or other tracers when performing BAO reconstruction. Future surveys will measure the BAO peak position at sub-percent level. While the effect of massive neutrinos on the BAO is in principle small, at this level of precision it will be important to quantitatively study and include the effect of neutrinos when performing BAO reconstruction. We would like to investigate such a BAO reconstruction method should be performed in cosmologies with massive neutrinos using state-of-the-art numerical simulations.

Another interesting question to address is the effect of streaming velocities [247] on the BAO peak position of the 21cm IM in the post-reionization era. This effect is small at low-redshift and has been constrained recently using galaxy surveys [248, 7]. However, at higher redshifts this effect should be stronger and future surveys will probe the gas distribution which may be different from the cold dark matter or the total matter one. This can cause an apparent shift of the BAO peak position and we want to investigate this effect using high resolution hydrodynamical simulations.

Appendix A

HI density profiles table

In this appendix we show a table with results from chapter 2 concerning the HI profiles from the numerical simulations.

	Model 1 — power law + exponential cutoff: $\alpha_\star$, $\log_{10} r_0\,[h^{-1}\mathrm{Mpc}]$											
z	$M_h = 10^9\,[h^{-1}M_\odot]$		$M_h = 10^{10}\,[h^{-1}M_\odot]$		$M_h = 10^{11}\,[h^{-1}M_\odot]$		$M_h = 10^{12}\,[h^{-1}M_\odot]$		$M_h = 10^{13}\,[h^{-1}M_\odot]$		$M_h = 10^{14}\,[h^{-1}M_\odot]$	
0	—	—	$3.04^{+0.04}_{-0.03}$,	$-3.59^{+0.85}_{-0.92}$	$3.03^{+0.03}_{-0.02}$,	$-2.8^{+0.5}_{-1.2}$	$3.02^{+0.03}_{-0.03}$,	$-2.32^{+0.33}_{-1.15}$	$3.00^{+0.04}_{-0.04}$,	$-1.71^{+0.09}_{-0.12}$	$2.92^{+0.03}_{-0.03}$,	$-1.91^{+0.11}_{-0.14}$
1	$3.3^{+1.3}_{-0.7}$,	$-2.5^{+1.1}_{-1.6}$	$3.05^{+0.02}_{-0.02}$,	$-3.72^{+0.77}_{-0.84}$	$3.02^{+0.02}_{-0.02}$,	$-3.3^{+0.7}_{-1.1}$	$3.00^{+0.03}_{-0.02}$,	$-2.32^{+0.16}_{-0.28}$	$2.99^{+0.03}_{-0.03}$,	$-1.77^{+0.09}_{-0.11}$	—	—
2	$3.07^{+0.10}_{-0.08}$,	$-3.2^{+0.9}_{-1.2}$	$3.03^{+0.01}_{-0.02}$,	$-3.64^{+0.78}_{-0.89}$	$3.01^{+0.01}_{-0.01}$,	$-2.75^{+0.26}_{-0.68}$	$3.00^{+0.02}_{-0.02}$,	$-2.18^{+0.09}_{-0.12}$	$2.98^{+0.02}_{-0.01}$,	$-1.74^{+0.04}_{-0.05}$	—	—
3	$3.05^{+0.02}_{-0.02}$,	$-3.63^{+0.85}_{-0.93}$	$3.02^{+0.02}_{-0.02}$,	$-3.1^{+0.5}_{-1.1}$	$3.00^{+0.01}_{-0.01}$,	$-2.52^{+0.13}_{-0.20}$	$3.00^{+0.02}_{-0.02}$,	$-2.09^{+0.06}_{-0.07}$	—	—	—	—
4	$3.04^{+0.02}_{-0.02}$,	$-3.3^{+0.7}_{-1.0}$	$3.00^{+0.01}_{-0.01}$,	$-2.46^{+0.15}_{-0.24}$	$3.00^{+0.01}_{-0.01}$,	$-2.32^{+0.07}_{-0.08}$	$2.99^{+0.01}_{-0.01}$,	$-2.04^{+0.03}_{-0.04}$	—	—	—	—
5	$3.03^{+0.02}_{-0.02}$,	$-2.9^{+0.5}_{-1.2}$	$3.00^{+0.01}_{-0.01}$,	$-2.28^{+0.09}_{-0.12}$	$3.00^{+0.01}_{-0.01}$,	$-2.18^{+0.04}_{-0.05}$	$3.00^{+0.01}_{-0.01}$,	$-2.02^{+0.03}_{-0.03}$	—	—	—	—
	Model 2 — altered NFW + exponential cutoff: $\log_{10} r_\mathrm{s}\,[h^{-1}\mathrm{Mpc}]$, $\log_{10} r_0\,[h^{-1}\mathrm{Mpc}]$											
0	—	—	$-4.0^{+0.7}_{-0.7}$,	$-3.8^{+0.8}_{-0.8}$	$-3.7^{+0.6}_{-0.9}$,	$-3.4^{+0.7}_{-1.0}$	$-3.2^{+0.7}_{-1.1}$,	$-3.1^{+0.8}_{-1.3}$	$-3.0^{+1.0}_{-1.3}$,	$-1.8^{+0.1}_{-1.0}$	$-2.3^{+0.5}_{-1.7}$,	$-2.6^{+0.7}_{-1.6}$
1	$-2.8^{+1.8}_{-1.5}$,	$-3.3^{+1.2}_{-1.1}$	$-4.0^{+0.5}_{-0.6}$,	$-3.7^{+0.6}_{-0.8}$	$-3.9^{+0.6}_{-0.7}$,	$-3.6^{+0.7}_{-0.9}$	$-3.0^{+0.4}_{-1.2}$,	$-2.9^{+0.6}_{-1.4}$	$-2.2^{+0.3}_{-1.6}$,	$-2.4^{+0.6}_{-1.7}$	—	—
2	$-3.7^{+0.9}_{-0.8}$,	$-3.7^{+0.9}_{-0.9}$	$-3.8^{+0.5}_{-0.7}$,	$-3.5^{+0.5}_{-0.9}$	$-3.6^{+0.5}_{-0.9}$,	$-3.3^{+0.6}_{-1.1}$	$-2.8^{+0.3}_{-1.3}$,	$-2.6^{+0.4}_{-1.5}$	$-1.8^{+0.1}_{-0.2}$,	$-3.0^{+0.8}_{-1.3}$	—	—
3	$-3.8^{+0.6}_{-0.8}$,	$-3.5^{+0.6}_{-0.9}$	$-3.6^{+0.5}_{-0.8}$,	$-3.3^{+0.5}_{-1.1}$	$-3.3^{+0.4}_{-1.0}$,	$-3.0^{+0.5}_{-1.3}$	$-2.8^{+0.4}_{-1.3}$,	$-2.4^{+0.3}_{-1.4}$	—	—	—	—
4	$-3.6^{+0.5}_{-0.9}$,	$-3.2^{+0.5}_{-1.1}$	$-3.2^{+0.4}_{-1.1}$,	$-3.0^{+0.5}_{-1.3}$	$-2.9^{+0.3}_{-1.3}$,	$-2.7^{+0.4}_{-1.4}$	$-2.6^{+0.3}_{-1.4}$,	$-2.4^{+0.3}_{-1.2}$	—	—	—	—
5	$-3.4^{+0.5}_{-1.0}$,	$-3.1^{+0.5}_{-1.1}$	$-2.7^{+0.2}_{-1.1}$,	$-3.1^{+0.7}_{-1.3}$	$-2.5^{+0.1}_{-1.1}$,	$-3.3^{+1.0}_{-1.2}$	$-2.2^{+0.1}_{-0.4}$,	$-3.1^{+0.8}_{-1.2}$	—	—	—	—

Table A.1: Best-fit values of the parameters determining the HI density profiles. We show the resulting parameters for the two different models considered (see text): an altered NFW profile with an exponential cutoff on small scales (top) and a simple power law with an exponential cutoff on small scales (bottom), as a dependence on the halo mass (columns) and redshift (rows).

Appendix B

Tables from chapter 4

In this appendix we show the tables with results from chapter 4. The tables show the constraints on RSD, BAO distance scale parameters and cosmological parameters coming from various configurations of the proposed 21cm IM surveys or in combination with current and foreseen galaxy surveys and CMB experiments.

Ext-HIRAX $\sigma_{f\sigma_8}/(f\sigma_8)$, 5% $b_{\rm HI}$ & $\Omega_{\rm HI}$						
	$k_{\rm max} = 0.2h{\rm Mpc}^{-1}$			$k_{\rm max} = k_{\rm nl}(z)$		
z	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2.5	0.036	0.036	0.075	0.035	0.035	0.043
2.6	0.036	0.036	0.080	0.035	0.036	0.044
2.7	0.036	0.036	0.085	0.035	0.036	0.046
2.8	0.036	0.036	0.091	0.035	0.036	0.048
2.9	0.036	0.037	0.097	0.036	0.036	0.050
3.0	0.036	0.037	0.104	0.036	0.036	0.053
3.1	0.036	0.037	0.111	0.036	0.036	0.056
3.2	0.036	0.037	0.119	0.036	0.036	0.060
3.3	0.037	0.037	0.127	0.036	0.036	0.065
3.4	0.037	0.037	0.135	0.036	0.036	0.070
3.5	0.037	0.037	0.144	0.036	0.036	0.075
3.6	0.037	0.038	0.153	0.036	0.036	0.082
3.7	0.037	0.038	0.162	0.036	0.036	0.089
3.8	0.037	0.038	0.171	0.036	0.036	0.097
3.9	0.037	0.038	0.181	0.036	0.036	0.105
4.0	0.038	0.039	0.190	0.036	0.036	0.114
4.1	0.038	0.039	0.200	0.036	0.036	0.124
4.2	0.038	0.040	0.209	0.036	0.037	0.135
4.3	0.038	0.040	0.218	0.036	0.037	0.146
4.4	0.039	0.041	0.228	0.036	0.037	0.158
4.5	0.039	0.041	0.237	0.036	0.037	0.170
4.6	0.039	0.042	0.246	0.037	0.038	0.182
4.7	0.040	0.043	0.254	0.037	0.038	0.194
4.8	0.040	0.044	0.263	0.037	0.038	0.206
4.9	0.040	0.045	0.271	0.037	0.039	0.219
5.0	0.041	0.046	0.279	0.037	0.040	0.231

TABLE B.1: Forecasted 1σ constraints on $f\sigma_8$ for Ext-HIRAX as a function of redshift for different wedge configurations and different $k_{\rm max}$. The numbers correspond to 5% priors on the values of both $\Omega_{\rm HI}$ and $b_{\rm HI}$.

Ext-HIRAX, $k_{\max} = 0.2h\mathrm{Mpc}^{-1}$, 5% $b_{\rm HI}$ & $\Omega_{\rm HI}$,						
		σ_{D_A}/D_A			σ_H/H	
z	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2.5	0.004	0.005	0.053	0.006	0.006	0.014
2.6	0.004	0.005	0.057	0.006	0.006	0.014
2.7	0.004	0.005	0.060	0.006	0.006	0.014
2.8	0.004	0.005	0.064	0.006	0.006	0.014
2.9	0.004	0.005	0.067	0.006	0.006	0.014
3.0	0.005	0.005	0.071	0.006	0.007	0.014
3.1	0.005	0.006	0.075	0.006	0.007	0.015
3.2	0.005	0.006	0.080	0.007	0.007	0.015
3.3	0.005	0.006	0.084	0.007	0.007	0.015
3.4	0.005	0.006	0.089	0.007	0.007	0.016
3.5	0.005	0.007	0.095	0.007	0.008	0.016
3.6	0.005	0.007	0.10	0.007	0.008	0.016
3.7	0.006	0.007	0.11	0.008	0.008	0.017
3.8	0.006	0.008	0.11	0.008	0.009	0.017
3.9	0.006	0.008	0.12	0.008	0.009	0.018
4.0	0.006	0.008	0.13	0.008	0.009	0.019
4.1	0.006	0.009	0.14	0.009	0.010	0.019
4.2	0.007	0.009	0.15	0.009	0.010	0.020
4.3	0.007	0.010	0.16	0.009	0.010	0.021
4.4	0.007	0.010	0.17	0.010	0.011	0.022
4.5	0.007	0.011	0.18	0.010	0.011	0.023
4.6	0.008	0.011	0.20	0.010	0.012	0.024
4.7	0.008	0.012	0.21	0.011	0.012	0.025
4.8	0.008	0.013	0.23	0.011	0.013	0.026
4.9	0.009	0.014	0.25	0.011	0.013	0.027
5.0	0.009	0.014	0.27	0.012	0.014	0.029

TABLE B.2: Forecasted 1σ constraints on the expansion rate $H(z)$ and the angular diameter distance $D_A(z)$ for Ext-HIRAX as a function of redshift for different wedge configurations. The numbers correspond to 5% priors on the values of both $\Omega_{\rm HI}$ and $b_{\rm HI}$ and $k_{\max} = 0.2h\mathrm{Mpc}^{-1}$.

External datasets – Σm_ν					
	$\Omega_{\rm M}$	h	Σm_ν[eV]	$\ln(10^{10}A_s)$	n_s
Fiducial values	0.3075	0.6774	0.060	3.064	0.9667
PlanckBAO	0.0082	0.0065	0.067	0.037	0.0048
Euclid02 No Rec.	0.0055	0.0025	0.20	0.091	0.0086
Euclidnl No Rec.	0.0050	0.0021	0.17	0.077	0.0065
CMB-S4	0.0038	0.0004	0.10	0.018	0.0019

TABLE B.3: Forecasted 1σ constraints on cosmological parameters and the Σm_ν considering each external dataset alone. These constraints are obtained using the Fisher matrices explained in Section 4.3. PlanckBAO represents Planck CMB combined with BOSS BAO measurements. We show the projected constraints for Euclid for two different $k_{\max}$ used and the constraints coming from CMB-S4 experiment.

Ext-HIRAX Σm_ν					
	Ω_{M}	h	Σm_ν[eV]	$\ln(10^{10}A_s)$	n_s
Fiducial values	0.3075	0.6774	0.060	3.064	0.9667
No wedge $k_{\mathrm{max}} = 0.2h\mathrm{Mpc}^{-1}$					
2% b_{HI} & Ω_{HI}	0.0016	0.0010	0.059	0.026	0.0036
2% b_{HI} & Ω_{HI}, diff M_{min}	0.0015	0.0009	0.056	0.025	0.0033
2% b_{HI} & Ω_{HI}, 1-loop	0.0020	0.0010	0.081	0.038	0.014
5% b_{HI} & Ω_{HI}	0.0024	0.0015	0.093	0.047	0.0038
10% b_{HI} & Ω_{HI}	0.0029	0.0018	0.11	0.065	0.0040
No wedge $k_{\mathrm{max}} = k_{\mathrm{nl}}(z)$					
2% b_{HI} & Ω_{HI}	0.0010	0.0007	0.040	0.020	0.0016
2% b_{HI} & Ω_{HI}, diff M_{min}	0.0009	0.0006	0.037	0.019	0.0014
2% b_{HI} & Ω_{HI}, 1-loop	0.0011	0.0007	0.058	0.027	0.0060
5% b_{HI} & Ω_{HI}	0.0014	0.0009	0.054	0.032	0.0016
10% b_{HI} & Ω_{HI}	0.0014	0.0009	0.054	0.032	0.0016
Mid wedge $k_{\mathrm{max}} = 0.2h\mathrm{Mpc}^{-1}$					
2% b_{HI} & Ω_{HI}	0.0017	0.0010	0.062	0.027	0.0039
2% b_{HI} & Ω_{HI}, diff M_{min}	0.0016	0.0010	0.060	0.027	0.0037
2% b_{HI} & Ω_{HI}, 1-loop	0.0021	0.0011	0.088	0.041	0.016
5% b_{HI} & Ω_{HI}	0.0025	0.0016	0.098	0.049	0.0042
10% b_{HI} & Ω_{HI}	0.0032	0.0020	0.12	0.071	0.0044
Mid wedge $k_{\mathrm{max}} = k_{\mathrm{nl}}(z)$					
2% b_{HI} & Ω_{HI}	0.0011	0.0007	0.042	0.021	0.0017
2% b_{HI} & Ω_{HI}, diff M_{min}	0.0010	0.0006	0.039	0.020	0.0015
2% b_{HI} & Ω_{HI}, 1-loop	0.0011	0.0008	0.061	0.028	0.0066
5% b_{HI} & Ω_{HI}	0.0015	0.0010	0.058	0.035	0.0017
10% b_{HI} & Ω_{HI}	0.0017	0.0011	0.065	0.051	0.0018
Wedge $k_{\mathrm{max}} = 0.2h\mathrm{Mpc}^{-1}$					
2% b_{HI} & Ω_{HI}	0.0066	0.0048	0.24	0.091	0.010
2% b_{HI} & Ω_{HI}, diff M_{min}	0.0068	0.0050	0.25	0.092	0.010
2% b_{HI} & Ω_{HI}, 1-loop	0.0076	0.0050	0.34	0.138	0.054
5% b_{HI} & Ω_{HI}	0.0072	0.0053	0.27	0.107	0.011
10% b_{HI} & Ω_{HI}	0.0084	0.0062	0.31	0.136	0.011
Wedge $k_{\mathrm{max}} = k_{\mathrm{nl}}(z)$					
2% b_{HI} & Ω_{HI}	0.0027	0.0019	0.11	0.042	0.0044
2% b_{HI} & Ω_{HI}, diff M_{min}	0.0026	0.0019	0.10	0.041	0.0040
2% b_{HI} & Ω_{HI}, 1-loop	0.0029	0.0021	0.16	0.060	0.018
5% b_{HI} & Ω_{HI}	0.0034	0.0025	0.13	0.059	0.0044
10% b_{HI} & Ω_{HI}	0.0048	0.0034	0.17	0.092	0.0045

Table B.4: Fiducial values and 68% confidence intervals on cosmological parameters and the Σm_ν using Ext-HIRAX alone for various considerations. Going from top to bottom, we show the constraints considering different priors on both b_{HI} & Ω_{HI}, different k_{max} we use and different level of foreground wedge contamination. Alongside, focusing only on the case of 2% priors on b_{HI} & Ω_{HI}, we show the results obtained when we use a different value of $M_{\mathrm{min}} = 2 \times 10^{11}\ \mathrm{M}_\odot/h$ (labeled diff M_{min}.) and when we go beyond linear theory and marginalise over 1-loop and counter-terms parameters (labeled 1-loop, see Section 4.3.1).

highzFAST Σm_ν					
	$\Omega_{\rm M}$	h	Σm_ν[eV]	$\ln(10^{10}A_s)$	n_s
Fiducial values	0.3075	0.6774	0.060	3.064	0.9667
$k_{\rm max} = 0.2h{\rm Mpc}^{-1}$					
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0045	0.0026	0.14	0.054	0.013
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0054	0.0032	0.18	0.076	0.013
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0069	0.0041	0.24	0.112	0.013
$k_{\rm max} = k_{\rm nl}(z)$					
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0031	0.0021	0.11	0.044	0.0076
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0041	0.0026	0.15	0.064	0.0078
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0054	0.0034	0.19	0.095	0.0079

TABLE B.5: Fiducial values and 68% confidence intervals on cosmological parameters and the Σm_ν using highzFAST alone for various considerations. We show the constraints considering different astrophysical priors on both $b_{\rm HI}$ & $\Omega_{\rm HI}$, as well as the dependence of the constraints on the $k_{\rm max}$ used.

Ext-HIRAX Σm_ν[eV]						
	$k_{\rm max} = 0.2h{\rm Mpc}^{-1}$			$k_{\rm max} = k_{\rm nl}(z)$		
Euclid No Rec.+PlanckBAO	0.050			0.049		
+21cm	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.037	0.038	0.045	0.030	0.031	0.040
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$, diff $M_{\rm min}$	0.037	0.038	0.044	0.028	0.029	0.039
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$, 1-loop	0.038	0.039	0.046	0.035	0.035	0.042
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.042	0.043	0.048	0.034	0.035	0.043
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.044	0.045	0.049	0.035	0.036	0.044
Euclid No Rec.+CMB-S4	0.031			0.030		
+21cm	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.022	0.022	0.028	0.018	0.018	0.025
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$, diff $M_{\rm min}$	0.021	0.021	0.028	0.017	0.017	0.024
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$, 1-loop	0.023	0.023	0.028	0.020	0.020	0.025
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.023	0.023	0.028	0.018	0.019	0.025
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.023	0.023	0.028	0.019	0.019	0.025

TABLE B.6: Forecasted 1σ constraints on Σm_ν considering Ext-HIRAX instrument combined with CMB+galaxy probes. We show the constraints considering different priors on both $b_{\rm HI}$ & $\Omega_{\rm HI}$, different level of wedge contamination and different $k_{\rm max}$ used. In the top half of the table we show what are the gains of adding Ext-HIRAX measurements on top of combined Fisher matrix for Euclid galaxy survey and PlanckBAO. In the bottom half we do the same only using CMBS-4 Fisher matrix instead of the PlanckBAO one. Additionally, focusing only on the case of 2% priors on $b_{\rm HI}$ & $\Omega_{\rm HI}$, we also show the results obtained when we use a different value of $M_{\rm min} = 2 \times 10^{11}\,{\rm M}_\odot/h$ (labeled diff $M_{\rm min}$.) and when we go beyond linear theory and marginalise over 1-loop and counter-terms parameters (labeled 1-loop, see Section 4.3.1).

Ext-CHIME Σm_ν[eV]						
	$k_{\mathrm{max}} = 0.2h\mathrm{Mpc}^{-1}$			$k_{\mathrm{max}} = k_{\mathrm{nl}}(z)$		
Euclid No Rec.+PlanckBAO	0.050			0.049		
+21cm	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2% b_{HI} & Ω_{HI}	0.038	0.041	0.044	0.033	0.035	0.040
5% b_{HI} & Ω_{HI}	0.043	0.045	0.048	0.037	0.040	0.043
10% b_{HI} & Ω_{HI}	0.044	0.047	0.049	0.037	0.041	0.044
Euclid No Rec.+CMB-S4	0.031			0.030		
+21cm	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2% b_{HI} & Ω_{HI}	0.023	0.026	0.029	0.020	0.022	0.026
5% b_{HI} & Ω_{HI}	0.024	0.026	0.029	0.021	0.023	0.026
10% b_{HI} & Ω_{HI}	0.025	0.026	0.029	0.021	0.023	0.027

TABLE B.7: Forecasted 1σ constraints on Σm_ν considering Ext-CHIME instrument combined with CMB+galaxy probes. We show the constraints considering different priors on both b_{HI} & Ω_{HI}, different level of wedge contamination and different k_{max} used. In the top half of the table we show what are the gains of adding Ext-CHIME measurements on top of combined Fisher matrix for Euclid galaxy survey and PlanckBAO. In the bottom half we do the same only using CMBS-4 Fisher matrix instead of the PlanckBAO one.

highzFAST Σm_ν[eV]		
	$k_{\mathrm{max}} = 0.2h\mathrm{Mpc}^{-1}$	$k_{\mathrm{max}} = k_{\mathrm{nl}}(z)$
Euclid No Rec.+PlanckBAO	0.050	0.049
+21cm		
2% b_{HI} & Ω_{HI}	0.043	0.041
5% b_{HI} & Ω_{HI}	0.048	0.045
10% b_{HI} & Ω_{HI}	0.049	0.047
Euclid No Rec.+CMB-S4	0.031	0.030
+21cm		
2% b_{HI} & Ω_{HI}	0.029	0.028
5% b_{HI} & Ω_{HI}	0.030	0.028
10% b_{HI} & Ω_{HI}	0.030	0.029

TABLE B.8: Forecasted 1σ constraints on Σm_ν considering highzFAST instrument combined with CMB+galaxy probes. We show the constraints considering different priors on both b_{HI} & Ω_{HI} and different k_{max} used. In the top half of the table we show what are the gains of adding highzFAST measurements on top of combined Fisher matrix for Euclid galaxy survey and PlanckBAO. In the bottom half we do the same only using CMBS-4 Fisher matrix instead of the PlanckBAO one.

			External datasets – $N_{\rm eff}$		
	$\Omega_{\rm M}$	h	$\ln(10^{10}A_s)$	n_s	$N_{\rm eff}$
Fiducial values	0.3075	0.6774	3.064	0.9667	3.04
PlanckBAO	0.0088	0.015	0.038	0.0088	0.23
Euclid02	0.0031	0.0064	0.0136	0.0106	0.37
Euclidnl	0.0027	0.0041	0.0096	0.0056	0.23
CMB-S4	0.0012	0.00061	0.0094	0.0028	0.029

TABLE B.9: Forecasted 1σ constraints on cosmological parameters and the $N_{\rm eff}$ considering each external dataset alone. These constraints are obtained using the Fisher matrices explained in Section 4.3. PlanckBAO represents Planck CMB combined with BOSS BAO measurements. We show the projected constraints for Euclid for two different $k_{\rm max}$ used and the constraints coming from CMB-S4 experiment.

			Ext-HIRAX $N_{\rm eff}$		
	$\Omega_{\rm M}$	h	$\ln(10^{10}A_s)$	n_s	$N_{\rm eff}$
Fiducial values	0.3075	0.6774	3.064	0.9667	3.04
			No wedge $k_{\rm max} = 0.2h\text{Mpc}^{-1}$		
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0017	0.0036	0.013	0.0061	0.22
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0019	0.0037	0.023	0.0061	0.23
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0020	0.0037	0.041	0.0061	0.23
			No wedge $k_{\rm max} = k_{\rm nl}(z)$		
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0010	0.0013	0.009	0.0014	0.075
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0011	0.0013	0.020	0.0015	0.077
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0011	0.0013	0.040	0.0015	0.077
			Mid wedge $k_{\rm max} = 0.2h\text{Mpc}^{-1}$		
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0018	0.0039	0.013	0.0067	0.24
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0021	0.0041	0.023	0.0067	0.25
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0022	0.0041	0.042	0.0067	0.25
			Mid wedge $k_{\rm max} = k_{\rm nl}(z)$		
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0011	0.0014	0.010	0.0016	0.081
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0012	0.0014	0.021	0.0016	0.084
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0013	0.0014	0.040	0.0016	0.085
			Wedge $k_{\rm max} = 0.2h\text{Mpc}^{-1}$		
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0057	0.0117	0.03	0.0192	0.73
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0058	0.0118	0.04	0.0192	0.74
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0060	0.0119	0.06	0.0193	0.75
			Wedge $k_{\rm max} = k_{\rm nl}(z)$		
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0027	0.0035	0.01	0.0040	0.21
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0031	0.0035	0.03	0.0041	0.21
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0035	0.0035	0.05	0.0042	0.22

TABLE B.10: Fiducial values and 68% confidence intervals on cosmological parameters and the $N_{\rm eff}$ using Ext-HIRAX alone for various considerations. Going from top to bottom, we show the constraints considering different priors on both $b_{\rm HI}$ & $\Omega_{\rm HI}$, different $k_{\rm max}$ we use and different level of foreground wedge contamination.

highzFAST $N_{\rm eff}$					
	$\Omega_{\rm M}$	h	$\ln(10^{10}A_s)$	n_s	$N_{\rm eff}$
Fiducial values	0.3075	0.6774	3.064	0.9667	3.04
$k_{\rm max} = 0.2h{\rm Mpc}^{-1}$					
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0053	0.011	0.027	0.021	0.67
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0061	0.012	0.036	0.021	0.70
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0067	0.012	0.053	0.021	0.72
$k_{\rm max} = k_{\rm nl}(z)$					
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0038	0.0058	0.018	0.0078	0.34
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0046	0.0059	0.028	0.0079	0.36
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.0052	0.0060	0.047	0.0079	0.37

Table B.11: Fiducial values and 68% confidence intervals on cosmological parameters and the $N_{\rm eff}$ using highzFAST alone for various considerations. We show the constraints considering different astrophysical priors on both $b_{\rm HI}$ & $\Omega_{\rm HI}$, as well as the dependence of the constraints on the $k_{\rm max}$ used.

Ext-HIRAX $N_{\rm eff}$						
	$k_{\rm max} = 0.2h{\rm Mpc}^{-1}$			$k_{\rm max} = k_{\rm nl}(z)$		
Euclid Rec.+PlanckBAO	0.067			0.064		
+21cm	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.046	0.047	0.057	0.038	0.039	0.049
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.047	0.048	0.057	0.039	0.040	0.049
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.047	0.048	0.057	0.039	0.041	0.050
Euclid Rec.+CMB-S4	0.022			0.020		
+21cm	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.019	0.019	0.021	0.015	0.015	0.017
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.020	0.020	0.022	0.015	0.016	0.018
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.020	0.020	0.022	0.015	0.016	0.018

Table B.12: Forecasted 1σ constraints on $N_{\rm eff}$ considering Ext-HIRAX instrument combined with CMB+galaxy probes. We show the constraints considering different priors on both $b_{\rm HI}$ & $\Omega_{\rm HI}$, different level of wedge contamination and different $k_{\rm max}$ used. In the top half of the table we show what are the gains of adding Ext-HIRAX measurements on top of combined Fisher matrix for Euclid galaxy survey and PlanckBAO. In the bottom half we do the same only using CMBS-4 Fisher matrix instead of the PlanckBAO one.

Ext-CHIME $N_{\rm eff}$						
	$k_{\rm max} = 0.2h{\rm Mpc}^{-1}$			$k_{\rm max} = k_{\rm nl}(z)$		
Euclid Rec.+PlanckBAO	0.067			0.064		
+21cm	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.049	0.052	0.058	0.043	0.045	0.052
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.050	0.053	0.058	0.044	0.047	0.052
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.050	0.053	0.058	0.044	0.047	0.052
Euclid Rec.+CMB-S4	0.022			0.020		
+21cm	No wedge	Mid wedge	Wedge	No wedge	Mid wedge	Wedge
2% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.020	0.021	0.021	0.016	0.017	0.018
5% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.020	0.021	0.022	0.016	0.017	0.018
10% $b_{\rm HI}$ & $\Omega_{\rm HI}$	0.020	0.021	0.022	0.016	0.017	0.018

TABLE B.13: Forecasted 1σ constraints on $N_{\rm eff}$ considering Ext-CHIME instrument combined with CMB+galaxy probes. We show the constraints considering different priors on both $b_{\rm HI}$ & $\Omega_{\rm HI}$, different level of wedge contamination and different $k_{\rm max}$ used. In the top half of the table we show what are the gains of adding Ext-CHIME measurements on top of combined Fisher matrix for Euclid galaxy survey and PlanckBAO. In the bottom half we do the same only using CMBS-4 Fisher matrix instead of the PlanckBAO one.

highzFAST N_{eff}		
	$k_{\text{max}} = 0.2h\text{Mpc}^{-1}$	$k_{\text{max}} = k_{\text{nl}}(z)$
Euclid Rec.+PlanckBAO	0.067	0.064
+21cm		
2% b_{HI} & Ω_{HI}	0.062	0.057
5% b_{HI} & Ω_{HI}	0.062	0.058
10% b_{HI} & Ω_{HI}	0.062	0.058
Euclid Rec.+CMB-S4	0.022	0.020
+21cm		
2% b_{HI} & Ω_{HI}	0.022	0.019
5% b_{HI} & Ω_{HI}	0.022	0.019
10% b_{HI} & Ω_{HI}	0.022	0.019

TABLE B.14: Forecasted 1σ constraints on N_{eff} considering highzFAST instrument combined with CMB+galaxy probes. We show the constraints considering different priors on both b_{HI} & Ω_{HI}, different level of wedge contamination and different k_{max} used. In the top half of the table we show what are the gains of adding Ext-CHIME measurements on top of combined Fisher matrix for Euclid galaxy survey and PlanckBAO. In the bottom half we do the same only using CMBS-4 Fisher matrix instead of the PlanckBAO one.

Bibliography

[1] G. F. Smoot et al. "Structure in the COBE differential microwave radiometer first-year maps". In: Astrophys. J. Lett. 396 (Sept. 1992), pp. L1–L5. DOI: 10.1086/186504 (cit. on p. 1).

[2] A. D. Miller et al. "A Measurement of the Angular Power Spectrum of the Cosmic Microwave Background from L = 100 to 400". In: Astrophys. J. Lett. 524 (Oct. 1999), pp. L1–L4. DOI: 10.1086/312293. eprint: astro-ph/9906421 (cit. on pp. 1, 12).

[3] P. de Bernardis et al. "A flat Universe from high-resolution maps of the cosmic microwave background radiation". In: 404 (Apr. 2000), pp. 955–959. DOI: 10.1038/35010035. eprint: astro-ph/0004404 (cit. on pp. 1, 12).

[4] S. Hanany et al. "MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of 10'-5". In: Astrophys. J. Lett. 545 (Dec. 2000), pp. L5–L9. DOI: 10.1086/317322. eprint: astro-ph/0005123 (cit. on pp. 1, 12).

[5] C. L. Bennett et al. "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results". In: Astrophys. J. Suppl. 148 (Sept. 2003), pp. 1–27. DOI: 10.1086/377253. eprint: astro-ph/0302207 (cit. on p. 1).

[6] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: Astron. Astrophys. 594, A13 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 (cit. on pp. 1, 8, 58, 72, 75, 82).

[7] S. Alam et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample". In: Mon. Not. R. Astron. Soc. 470 (Sept. 2017), pp. 2617–2652. DOI: 10.1093/mnras/stx721. arXiv: 1607.03155 (cit. on pp. 1, 13, 103, 106).

[8] Dark Energy Survey Collaboration et al. "The Dark Energy Survey: more than dark energy - an overview". In: Mon. Not. R. Astron. Soc. 460 (Aug. 2016), pp. 1270–1299. DOI: 10.1093/mnras/stw641. arXiv: 1601.00329 (cit. on pp. 1, 103).

[9] A. G. Riess et al. "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: Astron. J. 116 (Sept. 1998), pp. 1009–1038. DOI: 10.1086/300499. eprint: astro-ph/9805201 (cit. on pp. 2, 5).

[10] S. Perlmutter et al. "Measurements of Ω and Λ from 42 High-Redshift Supernovae". In: Astrophys. J. 517 (June 1999), pp. 565–586. DOI: 10.1086/307221. eprint: astro-ph/9812133 (cit. on pp. 2, 5).

[11] Planck Collaboration et al. "Planck 2018 results. VI. Cosmological parameters". In: *ArXiv e-prints* (July 2018). arXiv: 1807.06209 (cit. on p. 2).

[12] S. Weinberg. *Cosmology*. Oxford University Press, 2008 (cit. on p. 3).

[13] D. W. Hogg. “Distance measures in cosmology”. In: *ArXiv Astrophysics e-prints* (May 1999). eprint: `astro-ph/9905116` (cit. on p. 4).

[14] Planck Collaboration et al. “Planck 2015 results. XVII. Constraints on primordial non-Gaussianity”. In: Astron. Astrophys. 594, A17 (Sept. 2016), A17. DOI: `10.1051/0004-6361/201525836`. arXiv: `1502.01592` (cit. on p. 5).

[15] H.-J. Seo et al. “Nonlinear Structure Formation and the Acoustic Scale”. In: Astrophys. J. 686 (Oct. 2008), pp. 13–24. DOI: `10.1086/589921`. arXiv: `0805.0117` (cit. on pp. 7, 13, 32).

[16] A. Lewis, A. Challinor, and A. Lasenby. “Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models”. In: Astrophys. J. 538 (Aug. 2000), pp. 473–476. DOI: `10.1086/309179`. eprint: `astro-ph/9911177` (cit. on pp. 8, 38, 58).

[17] D. Blas, J. Lesgourgues, and T. Tram. “The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes”. In: J. Cosmol. Astropart. Phys. 7, 034 (July 2011), p. 034. DOI: `10.1088/1475-7516/2011/07/034`. arXiv: `1104.2933` (cit. on p. 8).

[18] Planck Collaboration et al. “Planck 2018 results. I. Overview and the cosmological legacy of Planck”. In: *ArXiv e-prints* (July 2018). arXiv: `1807.06205` (cit. on pp. 9, 12).

[19] V. Desjacques, D. Jeong, and F. Schmidt. “Large-scale galaxy bias”. In: Phys. Rep. 733 (Feb. 2018), pp. 1–193. DOI: `10.1016/j.physrep.2017.12.002`. arXiv: `1611.09787` (cit. on p. 9).

[20] N. Kaiser. “Clustering in real space and in redshift space”. In: Mon. Not. R. Astron. Soc. 227 (July 1987), pp. 1–21. DOI: `10.1093/mnras/227.1.1` (cit. on pp. 10, 39).

[21] A. J. S. Hamilton. “Linear Redshift Distortions: a Review”. In: *The Evolving Universe*. Ed. by D. Hamilton. Vol. 231. Astrophysics and Space Science Library. 1998, p. 185. DOI: `10.1007/978-94-011-4960-0_17`. eprint: `astro-ph/9708102` (cit. on pp. 10, 11, 39).

[22] P. J. E. Peebles. *The large-scale structure of the universe*. 1980 (cit. on pp. 11, 87).

[23] E. V. Linder and R. N. Cahn. “Parameterized beyond-Einstein growth”. In: *Astroparticle Physics* 28 (Dec. 2007), pp. 481–488. DOI: `10.1016/j.astropartphys.2007.09.003`. eprint: `astro-ph/0701317` (cit. on p. 11).

[24] L. Guzzo et al. “A test of the nature of cosmic acceleration using galaxy redshift distortions”. In: 451 (Jan. 2008), pp. 541–544. DOI: `10.1038/nature06555`. arXiv: `0802.1944` (cit. on p. 11).

[25] S. Cole et al. “The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications”. In: Mon. Not. R. Astron. Soc. 362 (Sept. 2005), pp. 505–534. DOI: `10.1111/j.1365-2966.2005.09318.x`. eprint: `astro-ph/0501174` (cit. on p. 13).

[26] D. J. Eisenstein et al. “Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies”. In: Astrophys. J. 633 (Nov. 2005), pp. 560–574. DOI: `10.1086/466512`. eprint: `astro-ph/0501171` (cit. on p. 13).

[27] É. Aubourg et al. “Cosmological implications of baryon acoustic oscillation measurements”. In: Phys. Rev. D 92.12, 123516 (Dec. 2015), p. 123516. DOI: `10.1103/PhysRevD.92.123516`. arXiv: `1411.1074` (cit. on p. 13).

[28] G.-B. Zhao et al. “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: tomographic BAO analysis of DR12 combined sample in Fourier space”. In: *ArXiv e-prints* (July 2016). arXiv: `1607.03153` (cit. on p. 13).

[29] F. Beutler et al. “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in Fourier-space”. In: Mon. Not. R. Astron. Soc. (Sept. 2016). DOI: `10.1093/mnras/stw2373`. arXiv: `1607.03149` (cit. on p. 13).

[30] A. Meiksin, M. White, and J. A. Peacock. “Baryonic signatures in large-scale structure”. In: Mon. Not. R. Astron. Soc. 304 (Apr. 1999), pp. 851–864. DOI: `10.1046/j.1365-8711.1999.02369.x`. eprint: `astro-ph/9812214` (cit. on p. 13).

[31] D. J. Eisenstein, H.-J. Seo, and M. White. “On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter”. In: Astrophys. J. 664 (Aug. 2007), pp. 660–674. DOI: `10.1086/518755`. eprint: `astro-ph/0604361` (cit. on pp. 13, 38, 40).

[32] T. Matsubara. “Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture”. In: Phys. Rev. D 78.8, 083519 (Oct. 2008), p. 083519. DOI: `10.1103/PhysRevD.78.083519`. arXiv: `0807.1733` (cit. on p. 13).

[33] Martín Crocce and Román Scoccimarro. “Nonlinear evolution of baryon acoustic oscillations”. In: *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* 77.2 (2008), pp. 1–20. ISSN: 15507998. DOI: `10.1103/PhysRevD.77.023533`. arXiv: `0704.2783` (cit. on p. 13).

[34] N. Padmanabhan and M. White. “Calibrating the baryon oscillation ruler for matter and halos”. In: Phys. Rev. D 80.6, 063508 (Sept. 2009), p. 063508. DOI: `10.1103/PhysRevD.80.063508`. arXiv: `0906.1198 [astro-ph.CO]` (cit. on p. 13).

[35] N. McCullagh and A. S. Szalay. “Nonlinear Behavior of Baryon Acoustic Oscillations from the Zel’dovich Approximation Using a Non-Fourier Perturbation Approach”. In: Astrophys. J. 752, 21 (June 2012), p. 21. DOI: `10.1088/0004-637X/752/1/21`. arXiv: `1202.1306 [astro-ph.CO]` (cit. on p. 13).

[36] H.-J. Seo and D. J. Eisenstein. “Improved Forecasts for the Baryon Acoustic Oscillations and Cosmological Distance Scale”. In: Astrophys. J. 665 (Aug. 2007), pp. 14–24. DOI: `10.1086/519549`. eprint: `astro-ph/0701079` (cit. on p. 13).

[37] N. Padmanabhan, M. White, and J. D. Cohn. “Reconstructing baryon oscillations: A Lagrangian theory perspective”. In: Phys. Rev. D 79.6, 063523 (Mar. 2009), p. 063523. DOI: `10.1103/PhysRevD.79.063523`. arXiv: `0812.2905` (cit. on p. 13).

[38] D. J. Eisenstein et al. “Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak”. In: Astrophys. J. 664 (Aug. 2007), pp. 675–679. DOI: `10.1086/518712`. eprint: `astro-ph/0604362` (cit. on pp. 13, 32, 54, 58).

[39] Yookyung Noh, Martin White, and Nikhil Padmanabhan. "Reconstructing baryon oscillations". In: *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* 80.12 (2009), pp. 1–9. ISSN: 15507998. DOI: 10.1103/PhysRevD.80.123501. arXiv: 0909.1802 (cit. on pp. 13, 32).

[40] S. Tassev and M. Zaldarriaga. "Towards an optimal reconstruction of baryon oscillations". In: J. Cosmol. Astropart. Phys. 10, 006 (Oct. 2012), p. 006. DOI: 10.1088/1475-7516/2012/10/006. arXiv: 1203.6066 (cit. on p. 13).

[41] M. White. "Reconstruction within the Zeldovich approximation". In: Mon. Not. R. Astron. Soc. 450 (July 2015), pp. 3822–3828. DOI: 10.1093/mnras/stv842. arXiv: 1504.03677 (cit. on p. 13).

[42] J. Carlson, B. Reid, and M. White. "Convolution Lagrangian perturbation theory for biased tracers". In: Mon. Not. R. Astron. Soc. 429 (Feb. 2013), pp. 1674–1685. DOI: 10.1093/mnras/sts457. arXiv: 1209.0780 (cit. on p. 13).

[43] M. White. "The Zel'dovich approximation". In: Mon. Not. R. Astron. Soc. 439 (Apr. 2014), pp. 3630–3640. DOI: 10.1093/mnras/stu209. arXiv: 1401.5466 (cit. on p. 13).

[44] D. H. Weinberg et al. "Observational probes of cosmic acceleration". In: Phys. Rep. 530 (Sept. 2013), pp. 87–255. DOI: 10.1016/j.physrep.2013.05.001. arXiv: 1201.2434 (cit. on p. 14).

[45] L. Anderson et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples". In: Mon. Not. R. Astron. Soc. 441 (June 2014), pp. 24–62. DOI: 10.1093/mnras/stu523. arXiv: 1312.4877 (cit. on pp. 15, 72).

[46] Y. Fukuda et al. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Physical Review Letters* 81 (Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. eprint: hep-ex/9807003 (cit. on p. 14).

[47] Julien Lesgourgues et al. *Neutrino Cosmology.* Cambridge University Press, 2013. DOI: 10.1017/CBO9781139012874 (cit. on p. 14).

[48] C. Patrignani et al. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001 (cit. on p. 14).

[49] J. Lesgourgues and S. Pastor. "Massive neutrinos and cosmology". In: Phys. Rep. 429 (July 2006), pp. 307–379. DOI: 10.1016/j.physrep.2006.04.001. eprint: astro-ph/0603494 (cit. on pp. 14, 15).

[50] K. N. Abazajian et al. "Neutrino physics from the cosmic microwave background and large scale structure". In: *Astroparticle Physics* 63 (Mar. 2015), pp. 66–80. DOI: 10.1016/j.astropartphys.2014.05.014. arXiv: 1309.5383 (cit. on pp. 14, 16).

[51] M. Shoji and E. Komatsu. "Erratum: Massive neutrinos in cosmology: Analytic solutions and fluid approximation [Phys. Rev. D 81, 123516 (2010)]". In: Phys. Rev. D 82.8, 089901 (Oct. 2010), p. 089901. DOI: 10.1103/PhysRevD.82.089901. arXiv: 1003.0942 [astro-ph.CO] (cit. on p. 15).

[52] N. Palanque-Delabrouille et al. "Neutrino masses and cosmology with Lyman-alpha forest power spectrum". In: J. Cosmol. Astropart. Phys. 11, 011 (Nov. 2015), p. 011. DOI: 10.1088/1475-7516/2015/11/011. arXiv: 1506.05976 (cit. on p. 16).

[53] F. Villaescusa-Navarro et al. "Modeling the neutral hydrogen distribution in the post-reionization Universe: intensity mapping". In: J. Cosmol. Astropart. Phys. 9, 050 (Sept. 2014), p. 050. DOI: `10.1088/1475-7516/2014/09/050`. arXiv: `1405.6713` (cit. on pp. 17, 18, 30, 31, 60).

[54] B. T. Draine. *Physics of the Interstellar and Intergalactic Medium.* 2011 (cit. on pp. 17, 98).

[55] J. R. Pritchard and A. Loeb. "21 cm cosmology in the 21st century". In: *Reports on Progress in Physics* 75.8, 086901 (Aug. 2012), p. 086901. DOI: `10.1088/0034-4885/75/8/086901`. arXiv: `1109.6012` (cit. on p. 17).

[56] M. Tegmark and M. Zaldarriaga. "Fast Fourier transform telescope". In: Phys. Rev. D 79.8, 083530 (Apr. 2009), p. 083530. DOI: `10.1103/PhysRevD.79.083530`. arXiv: `0805.4414` (cit. on pp. 17, 18).

[57] F. Villaescusa-Navarro et al. "Ingredients for 21cm intensity mapping". In: *ArXiv e-prints* (Apr. 2018). arXiv: `1804.09180` (cit. on pp. 18, 20, 21, 22, 94, 97, 98, 99, 100, 101).

[58] S. Bharadwaj, B. B. Nath, and S. K. Sethi. "Using HI to Probe Large Scale Structures at z ˜ 3". In: *Journal of Astrophysics and Astronomy* 22 (Mar. 2001), p. 21. DOI: `10.1007/BF02933588`. eprint: `astro-ph/0003200` (cit. on p. 18).

[59] 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. 18).

[60] Richard A. Battye, Rod D. Davies, and Jochen Weller. "Neutral hydrogen surveys for high redshift galaxy clusters and proto-clusters". In: *Mon.Not.Roy.Astron.Soc.* 355 (2004), pp. 1339–1347. DOI: `10.1111/j.1365-2966.2004.08416.x`. arXiv: `astro-ph/0401340 [astro-ph]` (cit. on p. 18).

[61] M. McQuinn et al. "Cosmological Parameter Estimation Using 21 cm Radiation from the Epoch of Reionization". In: Astrophys. J. 653 (Dec. 2006), pp. 815–834. DOI: `10.1086/505167`. eprint: `astro-ph/0512263` (cit. on p. 18).

[62] T.-C. Chang et al. "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. 18).

[63] 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. 18).

[64] P. Bull et al. "Late-time Cosmology with 21 cm Intensity Mapping Experiments". In: Astrophys. J. 803, 21 (Apr. 2015), p. 21. DOI: `10.1088/0004-637X/803/1/21`. arXiv: `1405.1452` (cit. on pp. 18, 28, 39, 61, 62).

[65] M. Santos et al. "Cosmology from a SKA HI intensity mapping survey". In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 19 (Apr. 2015), p. 19. arXiv: `1501.03989` (cit. on pp. 18, 19, 62).

[66] F. Villaescusa-Navarro, P. Bull, and M. Viel. "Weighing Neutrinos with Cosmic Neutral Hydrogen". In: Astrophys. J. 814, 146 (Dec. 2015), p. 146. DOI: `10.1088/0004-637X/814/2/146`. arXiv: `1507.05102` (cit. on pp. 18, 61, 73).

[67] R. A. C. Croft et al. “Large-scale clustering of Lyman α emission intensity from SDSS/BOSS”. In: Mon. Not. R. Astron. Soc. 457 (Apr. 2016), pp. 3541–3572. DOI: `10.1093/mnras/stw204`. arXiv: `1504.04088` (cit. on p. 18).

[68] L. B. Newburgh et al. “Calibrating CHIME: a new radio interferometer to probe dark energy”. In: *Ground-based and Airborne Telescopes V*. Vol. 9145. Proceedings of the SPIE. July 2014, p. 91454V. DOI: `10.1117/12.2056962`. arXiv: `1406.2267 [astro-ph.IM]` (cit. on p. 19).

[69] L. B. Newburgh et al. “HIRAX: a probe of dark energy and radio transients”. In: *Ground-based and Airborne Telescopes VI*. Vol. 9906. Proceedings of the SPIE. Aug. 2016, p. 99065X. DOI: `10.1117/12.2234286`. arXiv: `1607.02059 [astro-ph.IM]` (cit. on p. 19).

[70] M.-A. Bigot-Sazy et al. “HI Intensity Mapping with FAST”. In: *Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015*. Ed. by L. Qain and D. Li. Vol. 502. Astronomical Society of the Pacific Conference Series. Feb. 2016, p. 41. arXiv: `1511.03006` (cit. on p. 19).

[71] C. R. Subrahmanya, P. K. Manoharan, and J. N. Chengalur. “The Ooty Wide Field Array”. In: *Journal of Astrophysics and Astronomy* 38, 10 (Mar. 2017), p. 10. DOI: `10.1007/s12036-017-9430-4`. arXiv: `1703.00621 [astro-ph.IM]` (cit. on p. 19).

[72] V. Ram Marthi et al. “Simulated foreground predictions for HI at $z = 3.35$ with the Ooty Wide Field Array: I. Instrument and the foregrounds”. In: *ArXiv e-prints* (July 2017). arXiv: `1707.05335 [astro-ph.IM]` (cit. on p. 19).

[73] A. K. Sarkar, S. Bharadwaj, and S. S. Ali. “Fisher Matrix-based Predictions for Measuring the $z = 3.35$ Binned 21-cm Power Spectrum using the Ooty Wide Field Array (OWFA)”. In: *Journal of Astrophysics and Astronomy* 38, 14 (Mar. 2017), p. 14. DOI: `10.1007/s12036-017-9432-2`. arXiv: `1703.00634` (cit. on p. 19).

[74] S. Chatterjee, S. Bharadwaj, and V. R. Marthi. “Simulating the $z = 3.35$ HI 21-cm Visibility Signal for the Ooty Wide Field Array (OWFA)”. In: *Journal of Astrophysics and Astronomy* 38, 15 (Mar. 2017), p. 15. DOI: `10.1007/s12036-017-9433-1`. arXiv: `1703.00628` (cit. on p. 19).

[75] A. K. Sarkar, S. Bharadwaj, and V. Ram Marthi. “An analytical method to simulate the HI 21-cm visibility signal for intensity mapping experiments”. In: *ArXiv e-prints* (Sept. 2017). arXiv: `1709.03984` (cit. on pp. 19, 61).

[76] E. Castorina and F. Villaescusa-Navarro. “On the spatial distribution of neutral hydrogen in the Universe: bias and shot-noise of the H i power spectrum”. In: Mon. Not. R. Astron. Soc. 471 (Oct. 2017), pp. 1788–1796. DOI: `10.1093/mnras/stx1599`. arXiv: `1609.05157` (cit. on pp. 19, 20, 21, 30, 59, 78).

[77] M. A. Zwaan et al. “The HIPASS catalogue: Ω_{HI} and environmental effects on the HI mass function of galaxies”. In: Mon. Not. R. Astron. Soc. 359 (May 2005), pp. L30–L34. DOI: `10.1111/j.1745-3933.2005.00029.x`. eprint: `astro-ph/0502257` (cit. on p. 19).

[78] A. M. Martin et al. “The Clustering Characteristics of H I-selected Galaxies from the 40% ALFALFA Survey”. In: Astrophys. J. 750, 38 (May 2012), p. 38. DOI: `10.1088/0004-637X/750/1/38`. arXiv: `1202.6005 [astro-ph.CO]` (cit. on pp. 19, 84, 86, 87, 88).

[79] M. A. Zwaan et al. “Reconciling the local galaxy population with damped Lyman α cross-sections and metal abundances”. In: Mon. Not. R. Astron. Soc. 364 (Dec. 2005), pp. 1467–1487. DOI: `10.1111/j.1365-2966.2005.09698.x`. eprint: `astro-ph/0510127` (cit. on p. 19).

[80] P. Noterdaeme et al. “Column density distribution and cosmological mass density of neutral gas: Sloan Digital Sky Survey-III Data Release 9”. In: Astron. Astrophys. 547, L1 (Nov. 2012), p. L1. DOI: `10.1051/0004-6361/201220259`. arXiv: `1210.1213` (cit. on p. 19).

[81] T. Zafar et al. “The ESO UVES advanced data products quasar sample. II. Cosmological evolution of the neutral gas mass density”. In: Astron. Astrophys. 556, A141 (Aug. 2013), A141. DOI: `10.1051/0004-6361/201321154`. arXiv: `1307.0602 [astro-ph.CO]` (cit. on pp. 19, 59).

[82] N. H. M. Crighton et al. “The neutral hydrogen cosmological mass density at z = 5”. In: Mon. Not. R. Astron. Soc. 452 (Sept. 2015), pp. 217–234. DOI: `10.1093/mnras/stv1182`. arXiv: `1506.02037` (cit. on pp. 19, 59, 60).

[83] S. M. Rao, D. A. Turnshek, and D. B. Nestor. “Damped Lyα Systems at $z < 1.65$: The Expanded Sloan Digital Sky Survey Hubble Space Telescope Sample”. In: Astrophys. J. 636 (Jan. 2006), pp. 610–630. DOI: `10.1086/498132`. eprint: `astro-ph/0509469` (cit. on p. 19).

[84] I. Pérez-Ràfols et al. “The SDSS-DR12 large-scale cross-correlation of damped Lyman alpha systems with the Lyman alpha forest”. In: Mon. Not. R. Astron. Soc. 473 (Jan. 2018), pp. 3019–3038. DOI: `10.1093/mnras/stx2525`. arXiv: `1709.00889` (cit. on p. 19).

[85] R. C. Kennicutt Jr. “The Global Schmidt Law in Star-forming Galaxies”. In: Astrophys. J. 498 (May 1998), pp. 541–552. DOI: `10.1086/305588`. eprint: `astro-ph/9712213` (cit. on p. 19).

[86] M. A. Zwaan and J. X. Prochaska. “Where Is the Molecular Hydrogen in Damped Lyα Absorbers?” In: Astrophys. J. 643 (June 2006), pp. 675–679. DOI: `10.1086/503191`. eprint: `astro-ph/0601655` (cit. on p. 19).

[87] S. M. Rao, D. A. Turnshek, and D. B. Nestor. “Damped Lya Systems at z<1.65: The Expanded Sloan Digital Sky Survey Hubble Space Telescope Sample”. In: Astrophys. J. 636 (Jan. 2006), pp. 610–630. DOI: `10.1086/498132`. eprint: `astro-ph/0509469` (cit. on p. 19).

[88] P. Lah et al. “The HI content of star-forming galaxies at z = 0.24”. In: Mon. Not. R. Astron. Soc. 376 (Apr. 2007), pp. 1357–1366. DOI: `10.1111/j.1365-2966.2007.11540.x`. eprint: `astro-ph/0701668` (cit. on p. 19).

[89] A. Songaila and L. L. Cowie. “The Evolution of Lyman Limit Absorption Systems to Redshift Six”. In: Astrophys. J. 721 (Oct. 2010), pp. 1448–1466. DOI: `10.1088/0004-637X/721/2/1448`. arXiv: `1007.3262` (cit. on p. 19).

[90] A. M. Martin et al. “The Arecibo Legacy Fast ALFA Survey. X. The H I Mass Function and Ω_H I from the 40% ALFALFA Survey”. In: Astrophys. J. 723 (Nov. 2010), pp. 1359–1374. DOI: `10.1088/0004-637X/723/2/1359`. arXiv: `1008.5107` (cit. on pp. 19, 91).

[91] P. Noterdaeme et al. “Column density distribution and cosmological mass density of neutral gas: Sloan Digital Sky Survey-III Data Release 9”. In: Astron. Astrophys. 547, L1 (Nov. 2012), p. L1. DOI: `10.1051/0004-6361/201220259`. arXiv: `1210.1213 [astro-ph.CO]` (cit. on pp. 19, 59).

[92] R. Braun. “Cosmological Evolution of Atomic Gas and Implications for 21 cm H I Absorption”. In: Astrophys. J. 749, 87 (Apr. 2012), p. 87. DOI: `10.1088/0004-637X/749/1/87`. arXiv: `1202.1840` (cit. on p. 19).

[93] J. Rhee et al. “Neutral atomic hydrogen (H I) gas evolution in field galaxies at z 0.1 and 0.2”. In: Mon. Not. R. Astron. Soc. 435 (Nov. 2013), pp. 2693–2706. DOI: `10.1093/mnras/stt1481`. arXiv: `1308.1462` (cit. on p. 19).

[94] J. Delhaize et al. “Detection of H I in distant galaxies using spectral stacking”. In: Mon. Not. R. Astron. Soc. 433 (Aug. 2013), pp. 1398–1410. DOI: `10.1093/mnras/stt810`. arXiv: `1305.1968 [astro-ph.CO]` (cit. on p. 19).

[95] S. Basilakos et al. “Large-scale structure in the HI Parkes All-Sky Survey: filling the voids with HI galaxies?” In: Mon. Not. R. Astron. Soc. 378 (June 2007), pp. 301–308. DOI: `10.1111/j.1365-2966.2007.11781.x`. eprint: `astro-ph/0703713` (cit. on p. 19).

[96] H. Guo et al. “Constraining the H i - Halo Mass Relation from Galaxy Clustering”. In: Astrophys. J. 846, 61 (Sept. 2017), p. 61. DOI: `10.3847/1538-4357/aa85e7`. arXiv: `1707.01999` (cit. on pp. 19, 84, 86, 88, 98).

[97] R. A. Battye et al. “H I intensity mapping: a single dish approach”. In: Mon. Not. R. Astron. Soc. 434 (Sept. 2013), pp. 1239–1256. DOI: `10.1093/mnras/stt1082`. arXiv: `1209.0343 [astro-ph.CO]` (cit. on p. 19).

[98] J. A. Peacock and R. E. Smith. “Halo occupation numbers and galaxy bias”. In: Mon. Not. R. Astron. Soc. 318 (Nov. 2000), pp. 1144–1156. DOI: `10.1046/j.1365-8711.2000.03779.x`. eprint: `astro-ph/0005010` (cit. on p. 20).

[99] U. Seljak. “Analytic model for galaxy and dark matter clustering”. In: Mon. Not. R. Astron. Soc. 318 (Oct. 2000), pp. 203–213. DOI: `10.1046/j.1365-8711.2000.03715.x`. eprint: `astro-ph/0001493` (cit. on p. 20).

[100] R. Scoccimarro et al. “How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering”. In: Astrophys. J. 546 (Jan. 2001), pp. 20–34. DOI: `10.1086/318261`. eprint: `astro-ph/0006319` (cit. on p. 20).

[101] R. E. Smith et al. “Stable clustering, the halo model and non-linear cosmological power spectra”. In: Mon. Not. R. Astron. Soc. 341 (June 2003), pp. 1311–1332. DOI: `10.1046/j.1365-8711.2003.06503.x`. eprint: `astro-ph/0207664` (cit. on pp. 20, 62).

[102] A. Cooray and R. Sheth. “Halo models of large scale structure”. In: Phys. Rep. 372 (Dec. 2002), pp. 1–129. DOI: `10.1016/S0370-1573(02)00276-4`. eprint: `astro-ph/0206508` (cit. on p. 20).

[103] H. Padmanabhan, A. Refregier, and A. Amara. “A halo model for cosmological neutral hydrogen : abundances and clustering”. In: Mon. Not. R. Astron. Soc. 469 (Aug. 2017), pp. 2323–2334. DOI: `10.1093/mnras/stx979`. arXiv: `1611.06235` (cit. on pp. 20, 24, 81, 82).

[104] J. L. Tinker et al. “The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests”. In: Astrophys. J. 724 (Dec. 2010), pp. 878–886. DOI: `10.1088/0004-637X/724/2/878`. arXiv: `1001.3162` (cit. on pp. 20, 82).

[105] H. Padmanabhan and A. Refregier. “Constraining a halo model for cosmological neutral hydrogen”. In: Mon. Not. R. Astron. Soc. 464 (Feb. 2017), pp. 4008–4017. DOI: `10.1093/mnras/stw2706`. arXiv: `1607.01021` (cit. on pp. 20, 60).

[106] V. Springel et al. “First results from the IllustrisTNG simulations: matter and galaxy clustering”. In: *ArXiv e-prints, 1707.03397* (July 2017). arXiv: `1707.03397` (cit. on p. 21).

[107] A. Pillepich et al. “First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies”. In: *ArXiv e-prints, 1707.03406* (July 2017). arXiv: `1707.03406` (cit. on p. 21).

[108] D. Nelson et al. “First results from the IllustrisTNG simulations: the galaxy colour bimodality”. In: Mon. Not. R. Astron. Soc. 475 (Mar. 2018), pp. 624–647. DOI: `10.1093/mnras/stx3040`. arXiv: `1707.03395` (cit. on pp. 21, 22).

[109] J. P. Naiman et al. “First results from the IllustrisTNG simulations: A tale of two elements – chemical evolution of magnesium and europium”. In: *ArXiv e-prints, 1707.03401* (July 2017). arXiv: `1707.03401` (cit. on p. 21).

[110] F. Marinacci et al. “First results from the IllustrisTNG simulations: radio haloes and magnetic fields”. In: *ArXiv e-prints, 1707.03396* (July 2017). arXiv: `1707.03396` (cit. on pp. 21, 22).

[111] V. Springel. “E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh”. In: Mon. Not. R. Astron. Soc. 401 (Jan. 2010), pp. 791–851. DOI: `10.1111/j.1365-2966.2009.15715.x`. arXiv: `0901.4107` (cit. on p. 21).

[112] M. Davis et al. “The evolution of large-scale structure in a universe dominated by cold dark matter”. In: Astrophys. J. 292 (May 1985), pp. 371–394. DOI: `10.1086/163168` (cit. on pp. 21, 31).

[113] A. H. Maller and J. S. Bullock. “Multiphase galaxy formation: high-velocity clouds and the missing baryon problem”. In: Mon. Not. R. Astron. Soc. 355 (Dec. 2004), pp. 694–712. DOI: `10.1111/j.1365-2966.2004.08349.x`. eprint: `astro-ph/0406632` (cit. on pp. 22, 81).

[114] L. A. Barnes and M. G. Haehnelt. “The bias of DLAs at z
2.3: evidence for very strong stellar feedback in shallow potential wells”. In: Mon. Not. R. Astron. Soc. 440 (May 2014), pp. 2313–2321. DOI: `10.1093/mnras/stu445`. arXiv: `1403.1873` (cit. on pp. 22, 81).

[115] J. Wang et al. “An observational and theoretical view of the radial distribution of H I gas in galaxies”. In: Mon. Not. R. Astron. Soc. 441 (July 2014), pp. 2159–2172. DOI: `10.1093/mnras/stu649`. arXiv: `1401.8164` (cit. on p. 24).

[116] D. Obreschkow et al. “A Virtual Sky with Extragalactic H I and CO Lines for the Square Kilometre Array and the Atacama Large Millimeter/Submillimeter Array”. In: Astrophys. J. 703 (Oct. 2009), pp. 1890–1903. DOI: `10.1088/0004-637X/703/2/1890`. arXiv: `0908.0983` (cit. on p. 24).

[117] M. G. Santos, A. Cooray, and L. Knox. “Multifrequency Analysis of 21 Centimeter Fluctuations from the Era of Reionization”. In: Astrophys. J. 625 (June 2005), pp. 575–587. DOI: `10.1086/429857`. eprint: `astro-ph/0408515` (cit. on p. 24).

[118] L. Wolz et al. “The effect of foreground subtraction on cosmological measurements from intensity mapping”. In: Mon. Not. R. Astron. Soc. 441 (July 2014), pp. 3271–3283. DOI: `10.1093/mnras/stu792`. arXiv: `1310.8144` (cit. on p. 24).

[119] J. R. Shaw et al. "Coaxing cosmic 21 cm fluctuations from the polarized sky using m -mode analysis". In: Phys. Rev. D 91.8, 083514 (Apr. 2015), p. 083514. DOI: `10.1103/PhysRevD.91.083514`. arXiv: `1401.2095` (cit. on pp. 24, 59, 63).

[120] D. Alonso et al. "Blind foreground subtraction for intensity mapping experiments". In: Mon. Not. R. Astron. Soc. 447 (Feb. 2015), pp. 400–416. DOI: `10.1093/mnras/stu2474`. arXiv: `1409.8667` (cit. on p. 24).

[121] L. Wolz et al. "Erasing the Milky Way: new cleaning technique applied to GBT intensity mapping data". In: *ArXiv e-prints* (Oct. 2015). arXiv: `1510.05453` (cit. on p. 24).

[122] D. Alonso, P. G. Ferreira, and M. G. Santos. "Fast simulations for intensity mapping experiments". In: Mon. Not. R. Astron. Soc. 444 (Nov. 2014), pp. 3183–3197. DOI: `10.1093/mnras/stu1666`. arXiv: `1405.1751` (cit. on pp. 24, 26).

[123] M. Schmittfull et al. "Eulerian BAO reconstructions and N -point statistics". In: Phys. Rev. D 92.12, 123522 (Dec. 2015), p. 123522. DOI: `10.1103/PhysRevD.92.123522`. arXiv: `1508.06972` (cit. on p. 27).

[124] H.-J. Seo and C. M. Hirata. "The foreground wedge and 21-cm BAO surveys". In: Mon. Not. R. Astron. Soc. 456 (Mar. 2016), pp. 3142–3156. DOI: `10.1093/mnras/stv2806`. arXiv: `1508.06503` (cit. on pp. 27, 29, 33, 35, 58, 63).

[125] F. Villaescusa-Navarro, D. Alonso, and M. Viel. "Baryonic acoustic oscillations from 21 cm intensity mapping: the Square Kilometre Array case". In: Mon. Not. R. Astron. Soc. 466 (Apr. 2017), pp. 2736–2751. DOI: `10.1093/mnras/stw3224`. arXiv: `1609.00019` (cit. on pp. 28, 29, 30, 50).

[126] K. Bandura et al. "Canadian Hydrogen Intensity Mapping Experiment (CHIME) pathfinder". In: *Ground-based and Airborne Telescopes V*. Vol. 9145. SPIE proceedings. July 2014, p. 914522. DOI: `10.1117/12.2054950`. arXiv: `1406.2288 [astro-ph.IM]` (cit. on p. 28).

[127] J. D. Cohn et al. "Combining galaxy and 21-cm surveys". In: Mon. Not. R. Astron. Soc. 457 (Apr. 2016), pp. 2068–2077. DOI: `10.1093/mnras/stw108`. arXiv: `1511.07377` (cit. on pp. 28, 35).

[128] F. Villaescusa-Navarro et al. "Neutral hydrogen in galaxy clusters: impact of AGN feedback and implications for intensity mapping". In: Mon. Not. R. Astron. Soc. 456 (Mar. 2016), pp. 3553–3570. DOI: `10.1093/mnras/stv2904`. arXiv: `1510.04277` (cit. on pp. 30, 60, 101).

[129] R. Scoccimarro and R. K. Sheth. "PTHALOS: a fast method for generating mock galaxy distributions". In: Mon. Not. R. Astron. Soc. 329 (Jan. 2002), pp. 629–640. DOI: `10.1046/j.1365-8711.2002.04999.x`. eprint: `astro-ph/0106120` (cit. on p. 30).

[130] F.-S. Kitaura and S. Heß. "Cosmological structure formation with augmented Lagrangian perturbation theory". In: Mon. Not. R. Astron. Soc. 435 (Aug. 2013), pp. L78–L82. DOI: `10.1093/mnrasl/slt101`. arXiv: `1212.3514` (cit. on p. 30).

[131] F.-S. Kitaura, G. Yepes, and F. Prada. "Modelling baryon acoustic oscillations with perturbation theory and stochastic halo biasing". In: Mon. Not. R. Astron. Soc. 439 (Mar. 2014), pp. L21–L25. DOI: `10.1093/mnrasl/slt172`. arXiv: `1307.3285` (cit. on p. 30).

[132] S. Tassev, M. Zaldarriaga, and D. J. Eisenstein. "Solving large scale structure in ten easy steps with COLA". In: J. Cosmol. Astropart. Phys. 6, 036 (June 2013), p. 036. DOI: `10.1088/1475-7516/2013/06/036`. arXiv: `1301.0322 [astro-ph.CO]` (cit. on pp. 30, 31).

[133] S. Tassev et al. "sCOLA: The N-body COLA Method Extended to the Spatial Domain". In: *ArXiv e-prints* (Feb. 2015). arXiv: `1502.07751` (cit. on p. 30).

[134] C. Howlett, M. Manera, and W. J. Percival. "L-PICOLA: A parallel code for fast dark matter simulation". In: *Astronomy and Computing* 12 (Sept. 2015), pp. 109–126. DOI: `10.1016/j.ascom.2015.07.003`. arXiv: `1506.03737` (cit. on pp. 30, 31).

[135] C.-H. Chuang et al. "EZmocks: extending the Zel'dovich approximation to generate mock galaxy catalogues with accurate clustering statistics". In: Mon. Not. R. Astron. Soc. 446 (Jan. 2015), pp. 2621–2628. DOI: `10.1093/mnras/stu2301`. arXiv: `1409.1124` (cit. on p. 30).

[136] Y. Feng et al. "FastPM: a new scheme for fast simulations of dark matter and haloes". In: Mon. Not. R. Astron. Soc. 463 (Dec. 2016), pp. 2273–2286. DOI: `10.1093/mnras/stw2123`. arXiv: `1603.00476` (cit. on p. 30).

[137] P. Monaco, T. Theuns, and G. Taffoni. "The pinocchio algorithm: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field". In: Mon. Not. R. Astron. Soc. 331 (Apr. 2002), pp. 587–608. DOI: `10.1046/j.1365-8711.2002.05162.x`. eprint: `astro-ph/0109323` (cit. on p. 30).

[138] G. Taffoni, P. Monaco, and T. Theuns. "PINOCCHIO and the hierarchical build-up of dark matter haloes". In: Mon. Not. R. Astron. Soc. 333 (July 2002), pp. 623–632. DOI: `10.1046/j.1365-8711.2002.05441.x`. eprint: `astro-ph/0109324` (cit. on p. 30).

[139] P. Monaco et al. "Predicting the Number, Spatial Distribution, and Merging History of Dark Matter Halos". In: Astrophys. J. 564 (Jan. 2002), pp. 8–14. DOI: `10.1086/324182`. eprint: `astro-ph/0109322` (cit. on p. 30).

[140] P. Monaco et al. "An accurate tool for the fast generation of dark matter halo catalogues". In: Mon. Not. R. Astron. Soc. 433 (Aug. 2013), pp. 2389–2402. DOI: `10.1093/mnras/stt907`. arXiv: `1305.1505` (cit. on p. 31).

[141] C.-H. Chuang et al. "nIFTy cosmology: Galaxy/halo mock catalogue comparison project on clustering statistics". In: Mon. Not. R. Astron. Soc. 452 (Sept. 2015), pp. 686–700. DOI: `10.1093/mnras/stv1289`. arXiv: `1412.7729` (cit. on p. 31).

[142] L. Anderson et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample". In: Mon. Not. R. Astron. Soc. 427 (Dec. 2012), pp. 3435–3467. DOI: `10.1111/j.1365-2966.2012.22066.x`. arXiv: `1203.6594` (cit. on p. 32).

[143] Nikhil Padmanabhan et al. "A 2 per cent distance to $z = 0.35$ by reconstructing baryon acoustic oscillations – II. Fitting techniques". In: *Mon. Not. R. Astron. Soc.* 427.3 (2012), pp. 2146–2167. ISSN: 00358711. DOI: `10.1111/j.1365-2966.2012.21573.x`. arXiv: `1202.0092`. URL: `http://mnras.oxfordjournals.org/cgi/doi/10.1111/j.1365-2966.2012.21573.x` (cit. on pp. 32, 36).

[144] L. Anderson et al. “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample”. In: Mon. Not. R. Astron. Soc. 427 (Dec. 2012), pp. 3435–3467. DOI: `10.1111/j.1365-2966.2012.22066.x`. arXiv: `1203.6594` (cit. on pp. 32, 36, 54).

[145] H.-J. Seo et al. “High-precision Predictions for the Acoustic Scale in the Nonlinear Regime”. In: Astrophys. J. 720 (Sept. 2010), pp. 1650–1667. DOI: `10.1088/0004-637X/720/2/1650`. arXiv: `0910.5005 [astro-ph.CO]` (cit. on p. 32).

[146] K. T. Mehta et al. “Galaxy Bias and Its Effects on the Baryon Acoustic Oscillation Measurements”. In: Astrophys. J. 734, 94 (June 2011), p. 94. DOI: `10.1088/0004-637X/734/2/94`. arXiv: `1104.1178 [astro-ph.CO]` (cit. on p. 32).

[147] Y. B. Zel’dovich. “Gravitational instability: An approximate theory for large density perturbations.” In: Astron. Astrophys. 5 (Mar. 1970), pp. 84–89 (cit. on p. 32).

[148] H.-J. Seo et al. “Modeling the reconstructed BAO in Fourier space”. In: Mon. Not. R. Astron. Soc. 460 (Aug. 2016), pp. 2453–2471. DOI: `10.1093/mnras/stw1138`. arXiv: `1511.00663` (cit. on pp. 32, 35).

[149] M. White. “Shot noise and reconstruction of the acoustic peak”. In: *ArXiv e-prints* (Apr. 2010). arXiv: `1004.0250 [astro-ph.CO]` (cit. on pp. 33, 35, 41).

[150] S. Anselmi, G. D. Starkman, and R. K. Sheth. “Beating non-linearities: improving the baryon acoustic oscillations with the linear point”. In: Mon. Not. R. Astron. Soc. 455 (Jan. 2016), pp. 2474–2483. DOI: `10.1093/mnras/stv2436`. arXiv: `1508.01170` (cit. on p. 34).

[151] N. Padmanabhan et al. “A 2 per cent distance to z = 0.35 by reconstructing baryon acoustic oscillations - I. Methods and application to the Sloan Digital Sky Survey”. In: Mon. Not. R. Astron. Soc. 427 (Dec. 2012), pp. 2132–2145. DOI: `10.1111/j.1365-2966.2012.21888.x`. arXiv: `1202.0090` (cit. on pp. 35, 54).

[152] A. Burden et al. “Efficient reconstruction of linear baryon acoustic oscillations in galaxy surveys”. In: Mon. Not. R. Astron. Soc. 445 (Dec. 2014), pp. 3152–3168. DOI: `10.1093/mnras/stu1965`. arXiv: `1408.1348` (cit. on p. 35).

[153] M. Vargas-Magaña et al. “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Effect of smoothing of density field on reconstruction and anisotropic BAO analysis”. In: *ArXiv e-prints* (Sept. 2015). arXiv: `1509.06384` (cit. on p. 35).

[154] N. Padmanabhan and M. White. “Constraining anisotropic baryon oscillations”. In: Phys. Rev. D 77.12, 123540 (June 2008), p. 123540. DOI: `10.1103/PhysRevD.77.123540`. arXiv: `0804.0799` (cit. on pp. 36, 37).

[155] X. Xu et al. “Measuring D_A and H at z=0.35 from the SDSS DR7 LRGs using baryon acoustic oscillations”. In: Mon. Not. R. Astron. Soc. 431 (May 2013), pp. 2834–2860. DOI: `10.1093/mnras/stt379`. arXiv: `1206.6732` (cit. on pp. 36, 37, 40).

[156] D. J. Eisenstein and W. Hu. “Baryonic Features in the Matter Transfer Function”. In: Astrophys. J. 496 (Mar. 1998), pp. 605–614. DOI: `10.1086/305424`. eprint: `astro-ph/9709112` (cit. on p. 38).

[157] D. Foreman-Mackey et al. “emcee: The MCMC Hammer”. In: 125 (Mar. 2013), p. 306. DOI: `10.1086/670067`. arXiv: `1202.3665 [astro-ph.IM]` (cit. on pp. 41, 98).

[158] J. Hartlap, P. Simon, and P. Schneider. “Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix”. In: Astron. Astrophys. 464 (Mar. 2007), pp. 399–404. DOI: `10.1051/0004-6361:20066170`. eprint: `astro-ph/0608064` (cit. on pp. 42, 89).

[159] W. J. Percival et al. “The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: including covariance matrix errors”. In: Mon. Not. R. Astron. Soc. 439 (Apr. 2014), pp. 2531–2541. DOI: `10.1093/mnras/stu112`. arXiv: `1312.4841` (cit. on p. 42).

[160] DESI Collaboration et al. “The DESI Experiment Part I: Science,Targeting, and Survey Design”. In: *ArXiv e-prints* (Oct. 2016). arXiv: `1611.00036 [astro-ph.IM]` (cit. on pp. 57, 67, 69, 71, 75, 103).

[161] R. Laureijs et al. “Euclid Definition Study Report”. In: *ArXiv e-prints* (Oct. 2011). arXiv: `1110.3193 [astro-ph.CO]` (cit. on pp. 57, 72, 103).

[162] Z. Ivezic et al. “LSST: from Science Drivers to Reference Design and Anticipated Data Products”. In: *ArXiv e-prints* (May 2008). arXiv: `0805.2366` (cit. on pp. 57, 103).

[163] D. Spergel et al. “Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA Final Report”. In: *ArXiv e-prints* (May 2013). arXiv: `1305.5422 [astro-ph.IM]` (cit. on pp. 57, 70, 103).

[164] G. J. Hill et al. “The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX): Description and Early Pilot Survey Results”. In: *Panoramic Views of Galaxy Formation and Evolution*. Ed. by T. Kodama, T. Yamada, and K. Aoki. Vol. 399. Astronomical Society of the Pacific Conference Series. Oct. 2008, p. 115. arXiv: `0806.0183` (cit. on p. 57).

[165] S. Dodelson et al. “Cosmic Visions Dark Energy: Science”. In: *ArXiv e-prints* (Apr. 2016). arXiv: `1604.07626` (cit. on p. 57).

[166] M. Doran and G. Robbers. “Early dark energy cosmologies”. In: J. Cosmol. Astropart. Phys. 6, 026 (June 2006), p. 026. DOI: `10.1088/1475-7516/2006/06/026`. eprint: `astro-ph/0601544` (cit. on p. 57).

[167] L. Verde et al. “Early cosmology constrained”. In: J. Cosmol. Astropart. Phys. 4, 023 (Apr. 2017), p. 023. DOI: `10.1088/1475-7516/2017/04/023`. arXiv: `1611.00376` (cit. on p. 57).

[168] 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. 57).

[169] I. P. Carucci et al. “Warm dark matter signatures on the 21cm power spectrum: intensity mapping forecasts for SKA”. In: J. Cosmol. Astropart. Phys. 7, 047 (July 2015), p. 047. DOI: `10.1088/1475-7516/2015/07/047`. arXiv: `1502.06961` (cit. on p. 57).

[170] M. Tegmark. “Measuring Cosmological Parameters with Galaxy Surveys”. In: *Physical Review Letters* 79 (Nov. 1997), pp. 3806–3809. DOI: `10.1103/PhysRevLett.79.3806`. eprint: `astro-ph/9706198` (cit. on pp. 58, 65).

[171] Antony Lewis. “CAMB Notes”. `https://cosmologist.info/notes/CAMB.pdf` (cit. on p. 58).

[172] S. Bird, R. Garnett, and S. Ho. “Statistical properties of damped Lyman-alpha systems from Sloan Digital Sky Survey DR12”. In: Mon. Not. R. Astron. Soc. 466 (Apr. 2017), pp. 2111–2122. DOI: `10.1093/mnras/stw3246`. arXiv: `1610.01165` (cit. on p. 59).

[173] G. Dalton et al. “Project overview and update on WEAVE: the next generation wide-field spectroscopy facility for the William Herschel Telescope”. In: *Ground-based and Airborne Instrumentation for Astronomy V*. Vol. 9147. Proceedings of the SPIE. July 2014, p. 91470L. DOI: `10.1117/12.2055132`. arXiv: `1412.0843 [astro-ph.IM]` (cit. on p. 59).

[174] J. R. Shaw et al. “All-sky Interferometry with Spherical Harmonic Transit Telescopes”. In: Astrophys. J. 781, 57 (Feb. 2014), p. 57. DOI: `10.1088/0004-637X/781/2/57`. arXiv: `1302.0327 [astro-ph.CO]` (cit. on pp. 59, 63).

[175] I. Pérez-Ràfols et al. “The SDSS-DR12 large-scale cross-correlation of Damped Lyman Alpha Systems with the Lyman Alpha Forest”. In: *ArXiv e-prints* (Sept. 2017). arXiv: `1709.00889` (cit. on p. 60).

[176] 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: J. Cosmol. Astropart. Phys. 11, 059 (Nov. 2012), p. 059. DOI: `10.1088/1475-7516/2012/11/059`. arXiv: `1209.4596` (cit. on p. 60).

[177] D. Sarkar, S. Bharadwaj, and S. Anathpindika. “Modelling the post-reionization neutral hydrogen (H I) bias”. In: Mon. Not. R. Astron. Soc. 460 (Aug. 2016), pp. 4310–4319. DOI: `10.1093/mnras/stw1111`. arXiv: `1605.02963` (cit. on p. 60).

[178] A. Pénin, O. Umeh, and M. G. Santos. “A scale-dependent bias on linear scales: the case for H I intensity mapping at z = 1”. In: Mon. Not. R. Astron. Soc. 473 (Feb. 2018), pp. 4297–4305. DOI: `10.1093/mnras/stx2635`. arXiv: `1706.08763` (cit. on p. 60).

[179] A. Datta, J. D. Bowman, and C. L. Carilli. “Bright Source Subtraction Requirements for Redshifted 21 cm Measurements”. In: Astrophys. J. 724 (Nov. 2010), pp. 526–538. DOI: `10.1088/0004-637X/724/1/526`. arXiv: `1005.4071` (cit. on p. 63).

[180] M. F. Morales et al. “Four Fundamental Foreground Power Spectrum Shapes for 21 cm Cosmology Observations”. In: Astrophys. J. 752, 137 (June 2012), p. 137. DOI: `10.1088/0004-637X/752/2/137`. arXiv: `1202.3830 [astro-ph.IM]` (cit. on p. 63).

[181] A. Liu, A. R. Parsons, and C. M. Trott. “Epoch of reionization window. I. Mathematical formalism”. In: Phys. Rev. D 90.2, 023018 (July 2014), p. 023018. DOI: `10.1103/PhysRevD.90.023018`. arXiv: `1404.2596` (cit. on p. 63).

[182] A. Liu, A. R. Parsons, and C. M. Trott. “Epoch of reionization window. II. Statistical methods for foreground wedge reduction”. In: Phys. Rev. D 90.2, 023019 (July 2014), p. 023019. DOI: `10.1103/PhysRevD.90.023019`. arXiv: `1404.4372` (cit. on p. 63).

[183] A. R. Parsons et al. “A Per-baseline, Delay-spectrum Technique for Accessing the 21 cm Cosmic Reionization Signature”. In: Astrophys. J. 756, 165 (Sept. 2012), p. 165. DOI: `10.1088/0004-637X/756/2/165`. arXiv: `1204.4749 [astro-ph.IM]` (cit. on p. 63).

[184] J. C. Pober. “The impact of foregrounds on redshift space distortion measurements with the highly redshifted 21-cm line”. In: Mon. Not. R. Astron. Soc. 447 (Feb. 2015), pp. 1705–1712. DOI: `10.1093/mnras/stu2575`. arXiv: `1411.2050` (cit. on p. 63).

[185] A. Pourtsidou, D. Bacon, and R. Crittenden. “H i and cosmological constraints from intensity mapping, optical and CMB surveys”. In: Mon. Not. R. Astron. Soc. 470 (Oct. 2017), pp. 4251–4260. DOI: `10.1093/mnras/stx1479`. arXiv: `1610.04189` (cit. on p. 64).

[186] C. Alcock and B. Paczynski. “An evolution free test for non-zero cosmological constant”. In: 281 (Oct. 1979), p. 358. DOI: `10.1038/281358a0` (cit. on p. 66).

[187] W. E. Ballinger, J. A. Peacock, and A. F. Heavens. “Measuring the cosmological constant with redshift surveys”. In: Mon. Not. R. Astron. Soc. 282 (Oct. 1996), p. 877. DOI: `10.1093/mnras/282.3.877`. eprint: `astro-ph/9605017` (cit. on p. 67).

[188] F. Beutler et al. “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: testing gravity with redshift space distortions using the power spectrum multipoles”. In: Mon. Not. R. Astron. Soc. 443 (Sept. 2014), pp. 1065–1089. DOI: `10.1093/mnras/stu1051`. arXiv: `1312.4611` (cit. on p. 67).

[189] H.-J. Seo and D. J. Eisenstein. “Probing Dark Energy with Baryonic Acoustic Oscillations from Future Large Galaxy Redshift Surveys”. In: Astrophys. J. 598 (Dec. 2003), pp. 720–740. DOI: `10.1086/379122`. eprint: `astro-ph/0307460` (cit. on p. 67).

[190] B. Audren et al. “Neutrino masses and cosmological parameters from a Euclid-like survey: Markov Chain Monte Carlo forecasts including theoretical errors”. In: J. Cosmol. Astropart. Phys. 1, 026 (Jan. 2013), p. 026. DOI: `10.1088/1475-7516/2013/01/026`. arXiv: `1210.2194` (cit. on pp. 67, 76).

[191] T. Baldauf et al. “LSS constraints with controlled theoretical uncertainties”. In: *ArXiv e-prints* (Feb. 2016). arXiv: `1602.00674` (cit. on pp. 67, 76, 105).

[192] A. Font-Ribera et al. “DESI and other Dark Energy experiments in the era of neutrino mass measurements”. In: J. Cosmol. Astropart. Phys. 5, 023 (May 2014), p. 023. DOI: `10.1088/1475-7516/2014/05/023`. arXiv: `1308.4164` (cit. on pp. 67, 71, 72, 75, 77).

[193] L. Amendola et al. “Cosmology and Fundamental Physics with the Euclid Satellite”. In: *Living Reviews in Relativity* 16, 6 (Sept. 2013), p. 6. DOI: `10.12942/lrr-2013-6`. arXiv: `1206.1225` (cit. on pp. 67, 69, 73).

[194] P. Bull. “Extending Cosmological Tests of General Relativity with the Square Kilometre Array”. In: Astrophys. J. 817, 26 (Jan. 2016), p. 26. DOI: `10.3847/0004-637X/817/1/26`. arXiv: `1509.07562` (cit. on p. 70).

[195] K. N. Abazajian et al. “CMB-S4 Science Book, First Edition”. In: *ArXiv e-prints* (Oct. 2016). arXiv: `1610.02743` (cit. on p. 72).

[196] J. R. Pritchard and E. Pierpaoli. “Constraining massive neutrinos using cosmological 21cm observations”. In: Phys. Rev. D 78.6, 065009 (Sept. 2008), p. 065009. DOI: `10.1103/PhysRevD.78.065009`. arXiv: `0805.1920` (cit. on p. 73).

[197] K. Kohri et al. “Precise measurements of primordial power spectrum with 21 cm fluctuations”. In: J. Cosmol. Astropart. Phys. 10, 065 (Oct. 2013), p. 065. DOI: `10.1088/1475-7516/2013/10/065`. arXiv: `1303.1688` (cit. on p. 73).

[198] Y. Oyama, K. Kohri, and M. Hazumi. “Constraints on the neutrino parameters by future cosmological 21 cm line and precise CMB polarization observations”. In: J. Cosmol. Astropart. Phys. 2, 008 (Feb. 2016), p. 008. DOI: `10.1088/1475-7516/2016/02/008`. arXiv: `1510.03806` (cit. on p. 73).

[199] F. Villaescusa-Navarro et al. “Cosmology with massive neutrinos I: towards a realistic modeling of the relation between matter, haloes and galaxies”. In: J. Cosmol. Astropart. Phys. 3, 011 (Mar. 2014), p. 011. DOI: `10.1088/1475-7516/2014/03/011`. arXiv: `1311.0866` (cit. on p. 73).

[200] E. Castorina et al. “Cosmology with massive neutrinos II: on the universality of the halo mass function and bias”. In: J. Cosmol. Astropart. Phys. 2, 049 (Feb. 2014), p. 049. DOI: `10.1088/1475-7516/2014/02/049`. arXiv: `1311.1212 [astro-ph.CO]` (cit. on p. 73).

[201] E. Castorina et al. “DEMNUni: the clustering of large-scale structures in the presence of massive neutrinos”. In: J. Cosmol. Astropart. Phys. 7, 043 (July 2015), p. 043. DOI: `10.1088/1475-7516/2015/07/043`. arXiv: `1505.07148` (cit. on p. 73).

[202] F. Villaescusa-Navarro et al. “The imprint of neutrinos on clustering in redshift-space”. In: *ArXiv e-prints* (Aug. 2017). arXiv: `1708.01154` (cit. on p. 73).

[203] C. D. Leonard, P. Bull, and R. Allison. “Spatial curvature endgame: Reaching the limit of curvature determination”. In: Phys. Rev. D 94.2, 023502 (July 2016), p. 023502. DOI: `10.1103/PhysRevD.94.023502`. arXiv: `1604.01410` (cit. on p. 75).

[204] The Planck Collaboration. “The Scientific Programme of Planck”. In: *ArXiv Astrophysics e-prints* (Apr. 2006). eprint: `astro-ph/0604069` (cit. on p. 75).

[205] T. Sprenger et al. “Cosmology in the era of Euclid and the Square Kilometre Array”. In: *ArXiv e-prints* (Jan. 2018). arXiv: `1801.08331` (cit. on pp. 75, 76).

[206] A. Liu et al. “Eliminating the optical depth nuisance from the CMB with 21 cm cosmology”. In: Phys. Rev. D 93.4, 043013 (Feb. 2016), p. 043013. DOI: `10.1103/PhysRevD.93.043013`. arXiv: `1509.08463` (cit. on p. 75).

[207] O. Umeh. “Imprint of non-linear effects on HI intensity mapping on large scales”. In: J. Cosmol. Astropart. Phys. 6, 005 (June 2017), p. 005. DOI: `10.1088/1475-7516/2017/06/005`. arXiv: `1611.04963` (cit. on p. 76).

[208] Daniel Baumann et al. “Cosmological Non-Linearities as an Effective Fluid”. In: *JCAP* 1207 (2012), p. 051. DOI: `10.1088/1475-7516/2012/07/051`. arXiv: `1004.2488 [astro-ph.CO]` (cit. on p. 76).

[209] John Joseph M. Carrasco, Mark P. Hertzberg, and Leonardo Senatore. “The Effective Field Theory of Cosmological Large Scale Structures”. In: *JHEP* 09 (2012), p. 082. DOI: `10.1007/JHEP09(2012)082`. arXiv: `1206.2926 [astro-ph.CO]` (cit. on p. 76).

[210] M. Raveri et al. “CosmicFish Validation Notes V1.0”. In: *ArXiv e-prints* (July 2016). arXiv: `1607.01005` (cit. on p. 77).

[211] G. Mangano et al. “Relic neutrino decoupling including flavour oscillations”. In: *Nuclear Physics B* 729 (Nov. 2005), pp. 221–234. DOI: `10.1016/j.nuclphysb.2005.09.041`. eprint: `hep-ph/0506164` (cit. on p. 77).

[212] Christopher Brust, David E. Kaplan, and Matthew T. Walters. “New Light Species and the CMB”. In: *JHEP* 12 (2013), p. 058. DOI: `10.1007/JHEP12(2013)058`. arXiv: `1303.5379 [hep-ph]` (cit. on p. 77).

[213] Zackaria Chacko et al. “Hidden dark matter sector, dark radiation, and the CMB”. In: *Phys. Rev.* D92 (2015), p. 055033. DOI: `10.1103/PhysRevD.92.055033`. arXiv: `1505.04192 [hep-ph]` (cit. on p. 77).

[214] Sergei Bashinsky and Uros Seljak. “Neutrino perturbations in CMB anisotropy and matter clustering”. In: *Phys. Rev.* D69 (2004), p. 083002. DOI: `10.1103/PhysRevD.69.083002`. arXiv: `astro-ph/0310198 [astro-ph]` (cit. on p. 77).

[215] Daniel Baumann et al. “Phases of New Physics in the CMB”. In: *JCAP* 1601 (2016), p. 007. DOI: `10.1088/1475-7516/2016/01/007`. arXiv: `1508.06342 [astro-ph.CO]` (cit. on p. 77).

[216] D. Baumann, D. Green, and B. Wallisch. “Searching for Light Relics with Large-Scale Structure”. In: *ArXiv e-prints* (Dec. 2017). arXiv: `1712.08067` (cit. on p. 77).

[217] B. Follin et al. “First Detection of the Acoustic Oscillation Phase Shift Expected from the Cosmic Neutrino Background”. In: *Physical Review Letters* 115.9, 091301 (Aug. 2015), p. 091301. DOI: `10.1103/PhysRevLett.115.091301`. arXiv: `1503.07863` (cit. on p. 77).

[218] D. Baumann, D. Green, and M. Zaldarriaga. “Phases of New Physics in the BAO Spectrum”. In: *ArXiv e-prints* (Mar. 2017). arXiv: `1703.00894` (cit. on p. 77).

[219] J. S. Bullock et al. “Profiles of dark haloes: evolution, scatter and environment”. In: Mon. Not. R. Astron. Soc. 321 (Mar. 2001), pp. 559–575. DOI: `10.1046/j.1365-8711.2001.04068.x`. eprint: `astro-ph/9908159` (cit. on p. 81).

[220] A. V. Macciò et al. “Concentration, spin and shape of dark matter haloes: scatter and the dependence on mass and environment”. In: Mon. Not. R. Astron. Soc. 378 (June 2007), pp. 55–71. DOI: `10.1111/j.1365-2966.2007.11720.x`. eprint: `astro-ph/0608157` (cit. on p. 81).

[221] A. J. S. Hamilton. “Uncorrelated modes of the non-linear power spectrum”. In: Mon. Not. R. Astron. Soc. 312 (Feb. 2000), pp. 257–284. DOI: `10.1046/j.1365-8711.2000.03071.x`. eprint: `astro-ph/9905191` (cit. on p. 82).

[222] M. G. Jones et al. “The ALFALFA HI mass function: A dichotomy in the low-mass slope and a locally suppressed ‘knee’ mass”. In: Mon. Not. R. Astron. Soc. (Feb. 2018). DOI: `10.1093/mnras/sty521`. arXiv: `1802.00053` (cit. on pp. 84, 85, 86, 91, 92, 93, 94, 95, 97, 98).

[223] R. Giovanelli et al. “The Arecibo Legacy Fast ALFA Survey. I. Science Goals, Survey Design, and Strategy”. In: Astron. J. 130 (Dec. 2005), pp. 2598–2612. DOI: `10.1086/497431`. eprint: `astro-ph/0508301` (cit. on p. 84).

[224] M. P. Haynes et al. “The Arecibo Legacy Fast ALFA Survey: The α.40 H I Source Catalog, Its Characteristics and Their Impact on the Derivation of the H I Mass Function”. In: Astron. J. 142, 170 (Nov. 2011), p. 170. DOI: `10.1088/0004-6256/142/5/170`. arXiv: `1109.0027` (cit. on pp. 84, 89).

[225] E. Papastergis et al. “The Clustering of ALFALFA Galaxies: Dependence on H I Mass, Relationship with Optical Samples, and Clues of Host Halo Properties”. In: Astrophys. J. 776, 43 (Oct. 2013), p. 43. DOI: `10.1088/0004-637X/776/1/43`. arXiv: `1308.2661` (cit. on pp. 84, 86).

[226] M. P. Haynes et al. “The Arecibo Legacy Fast ALFA Survey: The ALFALFA Extragalactic H I Source Catalog”. In: Astrophys. J. 861, 49 (July 2018), p. 49. DOI: `10.3847/1538-4357/aac956`. arXiv: `1805.11499` (cit. on p. 84).

[227] I. Yoon and J. L. Rosenberg. “The Effect of Halo Mass on the H I Content of Galaxies in Groups and Clusters”. In: Astrophys. J. 812, 4 (Oct. 2015), p. 4. DOI: `10.1088/0004-637X/812/1/4`. arXiv: `1509.00497` (cit. on p. 84).

[228] X. Yang et al. “Galaxy Groups in the SDSS DR4. I. The Catalog and Basic Properties”. In: Astrophys. J. 671 (Dec. 2007), pp. 153–170. DOI: `10.1086/522027`. arXiv: `0707.4640` (cit. on p. 85).

[229] X. Yang et al. “A halo-based galaxy group finder: calibration and application to the 2dFGRS”. In: Mon. Not. R. Astron. Soc. 356 (Feb. 2005), pp. 1293–1307. DOI: `10.1111/j.1365-2966.2005.08560.x`. eprint: `arXiv:astro-ph/0405234` (cit. on p. 85).

[230] D. N. Spergel et al. “Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology”. In: Astrophys. J. Suppl. 170 (June 2007), pp. 377–408. DOI: `10.1086/513700`. eprint: `astro-ph/0603449` (cit. on p. 85).

[231] S. D. Landy and A. S. Szalay. “Bias and variance of angular correlation functions”. In: Astrophys. J. 412 (July 1993), pp. 64–71. DOI: `10.1086/172900` (cit. on p. 86).

[232] H. A. Feldman, N. Kaiser, and J. A. Peacock. “Power-spectrum analysis of three-dimensional redshift surveys”. In: Astrophys. J. 426 (May 1994), pp. 23–37. DOI: `10.1086/174036`. eprint: `astro-ph/9304022` (cit. on p. 87).

[233] D. Alonso. “CUTE solutions for two-point correlation functions from large cosmological datasets”. In: *ArXiv e-prints* (Oct. 2012). arXiv: `1210.1833 [astro-ph.IM]` (cit. on p. 87).

[234] R. Takahashi et al. “Revising the Halofit Model for the Nonlinear Matter Power Spectrum”. In: Astrophys. J. 761, 152 (Dec. 2012), p. 152. DOI: `10.1088/0004-637X/761/2/152`. arXiv: `1208.2701` (cit. on p. 88).

[235] R. Lupton. *Statistics in theory and practice*. 1993 (cit. on p. 88).

[236] I. Zehavi et al. “Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data”. In: Astrophys. J. 571 (May 2002), pp. 172–190. DOI: `10.1086/339893`. eprint: `astro-ph/0106476` (cit. on p. 88).

[237] G. Efstathiou, R. S. Ellis, and B. A. Peterson. “Analysis of a complete galaxy redshift survey. II - The field-galaxy luminosity function”. In: Mon. Not. R. Astron. Soc. 232 (May 1988), pp. 431–461. DOI: `10.1093/mnras/232.2.431` (cit. on p. 90).

[238] S. Cole et al. “The 2dF galaxy redshift survey: near-infrared galaxy luminosity functions”. In: Mon. Not. R. Astron. Soc. 326 (Sept. 2001), pp. 255–273. DOI: `10.1046/j.1365-8711.2001.04591.x`. eprint: `astro-ph/0012429` (cit. on p. 90).

[239] M. G. Jones et al. “Environmental dependence of the H I mass function in the ALFALFA 70% catalogue”. In: Mon. Not. R. Astron. Soc. 457 (Apr. 2016), pp. 4393–4405. DOI: `10.1093/mnras/stw263`. arXiv: `1510.07050` (cit. on p. 90).

[240] R. Braun et al. “A Wide-Field High-Resolution H I Mosaic of Messier 31. I. Opaque Atomic Gas and Star Formation Rate Density”. In: Astrophys. J. 695 (Apr. 2009), pp. 937–953. DOI: `10.1088/0004-637X/695/2/937`. arXiv: `0901.4154 [astro-ph.CO]` (cit. on p. 98).

[241] C. D. P. Lagos et al. “Which galaxies dominate the neutral gas content of the Universe?” In: Mon. Not. R. Astron. Soc. 440 (May 2014), pp. 920–941. DOI: `10.1093/mnras/stu266`. arXiv: `1310.4178` (cit. on p. 101).

[242] A. Zoldan et al. “H I-selected galaxies in hierarchical models of galaxy formation and evolution”. In: Mon. Not. R. Astron. Soc. 465 (Feb. 2017), pp. 2236–2253. DOI: `10.1093/mnras/stw2901`. arXiv: `1610.02042` (cit. on p. 101).

[243] A. G. Riess et al. “A 2.4% Determination of the Local Value of the Hubble Constant”. In: Astrophys. J. 826, 56 (July 2016), p. 56. DOI: `10.3847/0004-637X/826/1/56`. arXiv: `1604.01424` (cit. on p. 103).

[244] J. L. Bernal, L. Verde, and A. G. Riess. “The trouble with H_0”. In: J. Cosmol. Astropart. Phys. 10, 019 (Oct. 2016), p. 019. DOI: `10.1088/1475-7516/2016/10/019`. arXiv: `1607.05617` (cit. on p. 103).

[245] B. P. Abbott et al. “Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A”. In: Astrophys. J. Lett. 848, L13 (Oct. 2017), p. L13. DOI: `10.3847/2041-8213/aa920c`. arXiv: `1710.05834 [astro-ph.HE]` (cit. on p. 103).

[246] Z. Ding et al. “Theoretical systematics of Future Baryon Acoustic Oscillation Surveys”. In: Mon. Not. R. Astron. Soc. 479 (Sept. 2018), pp. 1021–1054. DOI: `10.1093/mnras/sty1413`. arXiv: `1708.01297` (cit. on p. 105).

[247] J. A. Blazek, J. E. McEwen, and C. M. Hirata. “Streaming Velocities and the Baryon Acoustic Oscillation Scale”. In: *Physical Review Letters* 116.12, 121303 (Mar. 2016), p. 121303. DOI: `10.1103/PhysRevLett.116.121303`. arXiv: `1510.03554` (cit. on p. 105).

[248] F. Schmidt and F. Beutler. “Imprints of reionization in galaxy clustering”. In: Phys. Rev. D 96.8, 083533 (Oct. 2017), p. 083533. DOI: `10.1103/PhysRevD.96.083533`. arXiv: `1705.07843` (cit. on p. 106).