The cross-correlation between 21cm intensity mapping maps and the Lyα forest in the post-reionization era

Isabella P. Carucci,[a,b] **Francisco Villaescusa-Navarro,**[c,b,d] **Matteo Viel**[d,b]

[a]SISSA- International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, Italy
[b]INFN sez. Trieste, Via Valerio 2, 34127 Trieste, Italy
[c]Center for Computational Astrophysics, 160 5th Avenue, New York, NY, 10010, USA
[d]INAF - Osservatorio Astronomico di Trieste, Via Tiepolo 11, 34143, Trieste, Italy

E-mail: ipcarucci@sissa.it, fvillaescusa@simonsfoundation.org, viel@oats.inaf.it

Abstract. We investigate the cross-correlation signal between 21cm intensity mapping maps and the Lyα forest in the fully non-linear regime using state-of-the-art hydrodynamic simulations. The cross-correlation signal between the Lyα forest and 21cm maps can provide a coherent and comprehensive picture of the neutral hydrogen (HI) content of our Universe in the post-reionization era, probing both its mass content and volume distribution. We compute the auto-power spectra of both fields together with their cross-power spectrum at $z = 2.4$ and find that on large scales the fields are completely anti-correlated. This anti-correlation arises because regions with high (low) 21cm emission, such as those with a large (low) concentration of damped Lyα systems, will show up as regions with low (high) transmitted flux. We find that on scales smaller than $k \simeq 0.2 \ h\mathrm{Mpc}^{-1}$ the cross-correlation coefficient departs from -1, at a scale where non-linearities show up. We use the anisotropy of the power spectra in redshift-space to determine the values of the bias and of the redshift-space distortion parameters of both fields. We find that the errors on the value of the cosmological and astrophysical parameters could decrease by 30% when adding data from the cross-power spectrum, in a conservative analysis. Our results point out that linear theory is capable of reproducing the shape and amplitude of the cross-power up to rather non-linear scales. Finally, we find that the 21cm-Lyα cross-power spectrum can be detected by combining data from a BOSS-like survey together with 21cm intensity mapping observations by SKA1-MID with a S/N ratio higher than 3 in $k \in [0.06, 1] \ h\mathrm{Mpc}^{-1}$. We emphasize that while the shape and amplitude of the 21cm auto-power spectrum can be severely affected by residual foreground contamination, cross-power spectra will be less sensitive to that and therefore can be used to identify systematics in the 21cm maps.

Contents

1   Introduction

The discovery of the existence of both dark matter and dark energy has revolutionized our understanding of the fundamental constituents of the Universe. The standard model of cosmology has been extremely successful in providing the theoretical framework that is capable of explaining a very diverse set of cosmological observations ranging from the anisotropies in the cosmic microwave background (CMB) to the spatial distribution of galaxies at late times.

The nature of dark energy, assumed to be a cosmological constant in the standard ΛCDM model, remains a mystery. Its presence and properties leaves however a signature on the growth and spatial distribution of matter perturbations. Thus, cosmological observations of those can be used to improve our understanding on the physical nature of dark energy.

While the spatial distribution of matter in the Universe is not directly observable, its statistical properties can be examined through its tracers such as galaxies or cosmic neutral hydrogen (HI). The underlying idea is that the clustering properties of the tracers will, up to a bias factor, resemble those of the matter perturbations on large linear scales.

Neutral hydrogen can be detected in the Universe either in emission or in absorption. For the latter, observations of the Lyα forest in quasar spectra represents a powerful way to trace high-redshift ($z \sim [2-4]$) highly ionized low-density gas clouds where HI is found in absorption, while for the former the 21cm line of neutral hydrogen can be used to trace HI in emission. 21cm intensity mapping is a technique [1–7] that consists in detecting the integrated 21cm emission from cosmic neutral hydrogen from unresolved galaxies or HI clouds by means of low angular resolution radio observations. The very large volumes that can be surveyed together with the intrinsic spectroscopic nature of these observations make this technique a potential game changer in the epoch of precision cosmology.

The total radiation a radio-telescope collects is the sum of the HI cosmological signal, system noise and foregrounds. Unfortunately, the amplitude of the foregrounds (mainly synchrotron radiation) is several orders of magnitude above the one of the cosmological signal. The potential to investigate cosmological problems depends dramatically on the precision with which we will be able to remove the foregrounds from the 21cm maps [8]. While *well behaved* foregrounds can be robustly subtracted, the presence of some foregrounds such as polarized synchrotron radiation may not be removed and can significantly bias the inferred shape and amplitude of the 21cm power spectrum [9, 10].

A way to avoid residual foreground contamination is to carry out cross-correlations: since the spatial location and amplitude of the residual foregrounds is not correlated with the cosmological location of other tracers, cross-correlations should retrieve the underlying cosmological signal [10–16]. Furthermore, cross-correlations are very useful because they provide extra information that can be used to tighten the constraints on bias parameters and therefore to break degeneracies with other parameters.

The goal of this paper is to investigate the signal arising from the cross-correlation between the Lyα forest and 21cm intensity mapping maps. We notice that this analysis has already been carried out at linear order in [14], while here we investigate this cross-correlation in the fully non-linear regime by means of hydrodynamic simulations. In this paper we study the degree and scales where the two tracers are correlated and how well linear theory is able to reproduce the results.

We carry out our analysis making use of state-of-the-art hydrodynamic simulations and perform convergence tests using simulations with different resolutions. We also forecast the precision with which the future Square Kilometre Array[1] will be able to detect the 21cm auto-power spectrum together with the 21cm-Lyα cross-power spectrum at $z \simeq 2.4$.

This paper is organized as follows. We start in section 2 by describing the numerical simulations which we have analyzed. Then, in sections 3-4 we describe how we have modeled the observables we are interested in: the 21cm radiation and the Lyα forest flux. In section 5 we show the results of the cross-correlation procedure, performing a detectability study with present and future surveys and quantifying the bias parameters of the two fields. We summarize and draw our conclusions in section 6. Appendices A and B focus on the modelling of the two fields: on the way in which we distribute neutral hydrogen in halos and on the way we generate mock Lyα forest spectra. In appendix C we derive the analytical formulas we use to model errors in our analysis.

2   Simulations and halo catalogues

We rely on two high-resolution hydrodynamic simulations, that we label 80-2048 and 160-1024: the first number indicates the box size in h^{-1}Mpc while the second one represents the cube root number

---

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

of gas (and dark matter) particles. Both of them are part of the Sherwood simulation suite of Bolton et al. [17]. They are run using the TreePM+SPH code GADGET-III, an updated and extended version of GADGET-II [18]. The values of the cosmological parameters are in agreement with recent Planck data [19]: $(\Omega_{\rm m}, \Omega_{\Lambda}, \Omega_{\rm b}, h, n_s, \sigma_8)$ have the following values (0.308, 0.692, 0.0482, 0.678, 0.961, 0.829). The 80-2048 (160-1024) simulation follows the evolution of 2×2048^3 (2×1024^3) cold dark matter plus baryon particles within a periodic box of linear comoving size of 80 (160) h^{-1} Mpc from $z = 99$ to $z = 2.4$. In table 1 we show the mass resolution and softening length for the two different runs. Here we briefly summarize the main physical ingredients of the hydrodynamics.

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

Table 1: Summary of the resolution parameters of the two different simulations.

Star formation processes are not properly followed in the simulations: gas particles with temperature $T < 10^5$ K and an overdensity > 1000 are converted to collisionless particles, resulting in a significant increase in computation speed at the expense of removing cold, dense gas from the model. The simulations do not include metal line cooling. The photoionisation and photo-heating of the hydrogen and helium gas is calculated using the spatially uniform Haardt & Madau (2012) ionising background model, where HI reionization happens at $z \sim 12$. Moreover, the gas is assumed to be optically thin and in ionisation equilibrium. We stress that 80-2048 and 160-1024 share the same thermal history, since for both of them the cooling routine and the UV background have been modified following the reference model of [20]. These simulations are tailored to give converged properties for low density intergalactic medium statistics as probed by the Lyα forest. For a detailed description of the simulations we refer the reader to the paper [17].

We analyze snapshots at $z = 2.4$ for both simulations, for which halos are identified using the Friend-of-Friends (FoF) [21] algorithm with a linking parameter length of $b = 0.2$. The redshift choice of $z = 2.4$ is convenient for our purposes since it is within the redshift range for which the highest number of quasars is observed (hence most Lyα forest data available, for example see [22]) and for which 21cm observations are planned to be carried out (e.g. by SKA[2] [23] or CHIME[3]).

We made use of two simulations in order to carry out convergence tests. In this paper we are interested in studying the shape and amplitude of the cross-signal on linear and mildly non-linear scales. Thus, the 160-1024 simulation would be the best suited for this analysis as its volume is eight times higher than the one of 80-2048. On the other hand, the mass and spatial resolution of the 80-2048 simulation is much higher than the one of the 160-1024, so we can investigate whether the ingredients we need to carry out our analysis are converged already in the 160-1024 simulation and therefore focus our analysis on that run.

As we discuss in detail in section 3 we model the 21cm signal using the halo information from the simulation catalogues and therefore we do not attempt to account for HI self-shielding and presence of molecular hydrogen for each gas particle in our simulations. In this sense, an N-body simulation would have been enough for our purposes, with the advantage to be less computational expensive to run and to reach larger scales more easily. On the other hand, a precise modelling of the

---

[2]https://www.skatelescope.org
[3]http://www.mcgillcosmology.ca/chime
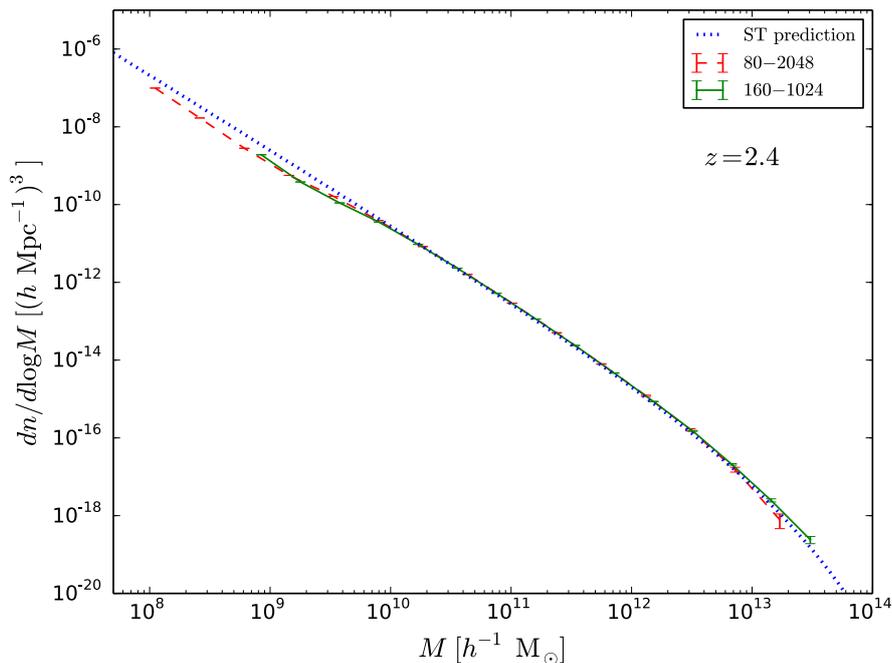
Figure 1: Halo mass functions of the 80-2048 simulation (red dashed line) and of the 160-1024 simulation (green solid line) compared with the Sheth-Tormen prediction [24] (blue dotted line). The error bars are the uncertainties in each of the mass bins used for the computation assuming that the number of halos follows a Poissonian distribution.

Lyα forest requires high-resolution hydrodynamic simulations as we describe in detail in section 4. In this paper we decided to use state-of-the-art hydrodynamic simulations that accurately model the properties of the Lyα forest at the expenses of neglecting the contribution of HI outside halos to the global 21cm signal (which was found negligible in [25]) and avoiding modelling the intrinsic scatter in the $M_{\rm HI}(M, z)$ function, that represents the average HI mass that a dark matter halo of mass M hosts at redshift z. However, we notice that we are also testing a method to model the Lyα signal in pure dark matter simulations (see appendix B) and this will be the subject of a future study.

As already stated, the halo information is crucial for this work since our 21cm maps are built from it. In Fig. 1 we plot the halo mass function for the 80-2048 and 160-1024 simulations in red and green, respectively, while the dotted blue line shows the Sheth & Tormen (ST) prediction [24] for the corresponding redshift $z = 2.4$. We find that the halo abundances from the simulations agree well with each other in the common mass range and also with the ST prediction. The higher mass resolution of the simulation 80-2048 allows us to sample the low mass end of the halo mass function while the larger box size of the 160-1024 simulation is better suited to explore the high-mass end. Thus, the combination of both simulations enables us to sample the halo mass function on a wide range of masses as shown in Fig. 1 which is important for our modelling of the 21cm signal.

3 Modelling the neutral hydrogen distribution

In this section we illustrate how we model the 21cm signal. First, in subsection 3.1 we describe the model for spatially distributing neutral hydrogen (HI), responsible for the 21cm emission. Next in

[3.2](#), we explain the method we use to transform the spatial distribution of HI into 21cm maps, and we show a comparison of the clustering properties of the two simulations.

3.1 The halo-based model

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

In order to avoid the above problem we can model the spatial distribution of HI assuming that all HI is confined within dark matter halos. This is a reasonable assumption since the UV background will prevent the formation of large HI clouds unless they are self-shielded. Besides, this statement has been verified through hydrodynamic simulations in [25]. Under these conditions it is possible to develop a HI halo model whose main ingredient is the function $M_{\mathrm{HI}}(M, z)$, that represents the average HI mass that a dark matter halo of mass M hosts at redshift z. Therefore, instead of computing the hydrogen phase fractions of each gas particle, we can estimate the spatial distribution of HI by modelling the $M_{\mathrm{HI}}(M, z)$ function.

We emphasize that by modelling the $M_{\mathrm{HI}}(M, z)$ function we are implicitly neglecting the intrinsic scatter expected in it. However, the scatter will not affect the clustering properties of the HI since it only depends on the average value. On the other hand, if the scatter correlates with environment cross-correlations may be affected by it. We will investigate this issue in a subsequent paper. Here we assume that the scatter in the $M_{\mathrm{HI}}(M, z)$ does not correlate with environment and therefore we can avoid modelling it. To reinforce this point we notice that such halo-based models have been already extensively investigated at similar redshifts against other methods such as post-processing pseudo radiative-transfer calculations, and the results have been proven to be robust against the model used: e.g. [10, 25, 26].

We model the $M_{\mathrm{HI}}(M, z)$ as

$$M_{\mathrm{HI}}(M) = \begin{cases} f\, M^{\alpha} & \text{if } M_{\min} \leq M \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

where we set $\alpha = 3/4$. This functional form arises from the result of high-resolution and zoom-in hydrodynamic simulations [27, 28] and it has been shown that it is capable of reproducing the abundance and clustering of the Damped Lyman-α systems (DLAs) [29]. $M_{\min}$ is a lower mass cut-off that considers that a minimum hydrogen density (clustered in a minimum halo potential well) is necessary to have hydrogen self-shielding and prevent the gas to be totally ionised. Here we assume that the mass parameter $M_{\min}$ corresponds to a dark matter halo with circular velocitiy $v_{\mathrm{circ}} = 25$ km s^{-1}, calculated using the virial relation:

$$M = 10^{10} \mathrm{M}_{\odot} \left(\frac{v_{\mathrm{circ}}}{60\,\mathrm{km\,s^{-1}}} \right)^{3} \left(\frac{1+z}{4} \right)^{-1.5}. \tag{3.2}$$
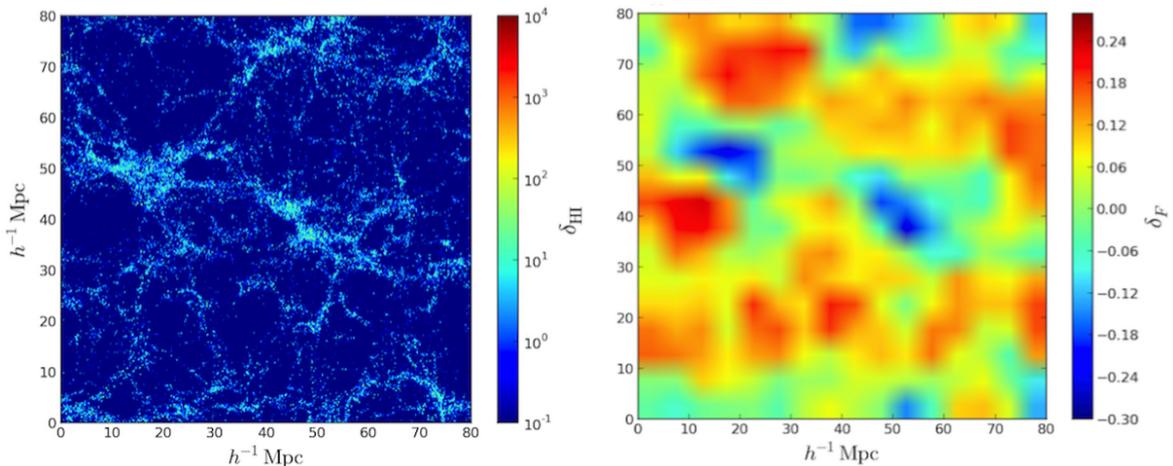
Figure 2: Spatial distribution of the density contrast of HI in redshift-space (left) and of the density contrast of the Lyα forest flux (right) at $z = 2.4$. We show the whole 80-2048 simulation box, taking a slice of $10\,h^{-1}\,\mathrm{Mpc}$ width.

We choose the value of the free parameter f by requiring that our model reproduces the measured value of Ω_{HI}:

$$\Omega_{\mathrm{HI}}(z) = \frac{1}{\rho_c^0} \int_0^\infty n(M) M_{\mathrm{HI}}(M, z) dM \ , \tag{3.3}$$

where $n(M)$ is the halo mass function and ρ_c^0 is the present day critical density of the universe. Since we have a particular realization of the halo field we impose the above condition in our simulations as

$$f \sum_{i=0}^n M^\alpha \, \Theta(M_i - M_{\mathrm{min}}) = \Omega_{\mathrm{HI}} L^3 \rho_c^0 \ , \tag{3.4}$$

where L is the simulation box size, $\Theta(x)$ is the Heaviside step function and the index i runs over all the dark matter halos of the simulation.

Since the value of Ω_{HI} is poorly constrained by observations at redshift $z \simeq 2.4$ (see [30] for a recent report) we set $\Omega_{\mathrm{HI}} = 10^{-3}$ in both simulations and notice that our conclusions are not affected by this choice as it only controls the amplitude of the 21cm power spectrum, not the HI clustering. We also notice that the value of f derived for the two simulations agree well with each other, pointing out that the mass function of both simulations overlap and M_{min} is resolved in both.

While the clustering of HI on large-scales is fully determined by the function $M_{\mathrm{HI}}(M, z)$, on smaller scales it depends on the way the HI is distributed within halos. Here for simplicity we avoid modelling the density profile of the HI inside halos (see e.g. [31]), and place it all in the halo center. In the appendix A we show that, for the scales we are interested in, this approximation does not bias our results.

By looking at the HI spatial distribution in Fig. 2, it is evident how HI tracks the halo positions in the simulation box, thus revealing the web-like structure of the matter in the Universe. We notice that since we are not modelling the HI 1-halo term, the presence of fingers-of-God in Fig. 2 is highly suppressed.

Since our model consists in assigning HI in a deterministic manner to dark matter halos of a given mass, we can compute the linear HI bias that we should recover in our simulations as

$$b_{\mathrm{HI}}(z) = \frac{\int_0^\infty b(M,z)n(M,z)M_{\mathrm{HI}}(M,z)dM}{\int_0^\infty n(M,z)M_{\mathrm{HI}}(M,z)dM}, \tag{3.5}$$

where $b(M,z)$ is the halo bias. By using our HI model together with the halo mass function and halo bias from Sheth & Tormen [32] and Sheth, Tormen & Mo [33], we get a value for the HI bias equal to $b_{\mathrm{HI}}(z=2.4)=1.45$, in disagreement with the constraints on the bias of the DLAs [34]; however, the purpose of this work is to be able to retrieve the input HI model values when fitting the results of the simulations. We also emphasize that a higher value of the HI bias will turn out to make the HI bias more scale-dependent [25, 35], and therefore shifting to large scales the onset of non-linearities, a situation we want to avoid given the relatively small volume of our simulations.

3.2 The 21cm signal

Radio telescopes are sensitive to the 21cm radiation emitted by HI. Therefore, the quantity directly measured in these observations is not the HI power spectrum, but the 21cm power spectrum, which is nothing but the power spectrum of the spatial distribution of neutral hydrogen in redshift-space, times an overall normalization factor. In this subsection we go through the steps for converting the HI distribution to the observed signal, the so called brightness temperature contrast: the radiation temperature against the CMB one at the redshift of observation.

After obtaining the HI distribution in comoving real-space as described in the previous subsection 3.1, we first move to redshift-space by displacing the position $\mathbf{x}$ of the particles (or of the halos) to $\mathbf{s}$ as

$$\mathbf{s} = \mathbf{x} + \frac{1+z}{H(z)}\mathbf{v}_{\mathrm{los}}(\mathbf{x}), \tag{3.6}$$

with z being the redshift of observation, $\mathbf{v}_{\mathrm{los}}$ the line of sight component of the peculiar velocity and $H(z)$ the Hubble parameter. Then, we compute the brightness temperature fluctuations using [36]:

$$\delta T_b(\mathbf{s}) = \overline{\delta T_b}(z)\left[\frac{\rho_{\mathrm{HI}}(\mathbf{s})}{\bar{\rho}_{\mathrm{HI}}}\right], \tag{3.7}$$

where $\rho_{\mathrm{HI}}(\mathbf{s})$ is the HI density in the redshift-space, $\bar{\rho}_{\mathrm{HI}}$ the HI mean density and

$$\overline{\delta T_b}(z) = \Omega_{\mathrm{HI}}\left(\frac{23.88h^2}{0.02}\frac{100}{76}\right)\sqrt{\frac{0.15}{\Omega_{\mathrm{m}}h^2}\frac{(1+z)}{10}}\ \mathrm{mK}. \tag{3.8}$$

Finally, the 21cm power spectrum is defined as $P_{21\mathrm{cm}}(k) = \langle \delta T_b(\mathbf{k})\delta T_b^*(\mathbf{k})\rangle$. Notice that the amplitude of the signal depends on the total amount of cosmic neutral hydrogen, expressed by the density parameter Ω_{HI} in Eq. 3.8.

As we discussed before, in our formalism we do not model the HI density profile. In the appendix A we show the differences arising in the 21cm power spectrum by the model used to describe the HI density profile.

We do a direct comparison between the two simulation 21cm power spectra in Fig. 3, upper panel, with the residuals $P_{21\mathrm{cm}}^{160-1024}(k)$ over $P_{21\mathrm{cm}}^{80-2048}(k)$ in green in the lower panel. We find that the shape of both power spectra agree on all scales, with a $\sim 20\%$ offset in amplitude among the two: HI is more clustered in the big box simulation. The reason for this amplitude difference is due to the fact that the larger box is able to capture the large-scale modes, which are not present in the low volume box. The presence of those modes enhances the clustering on lower scales, producing the effect we find here.
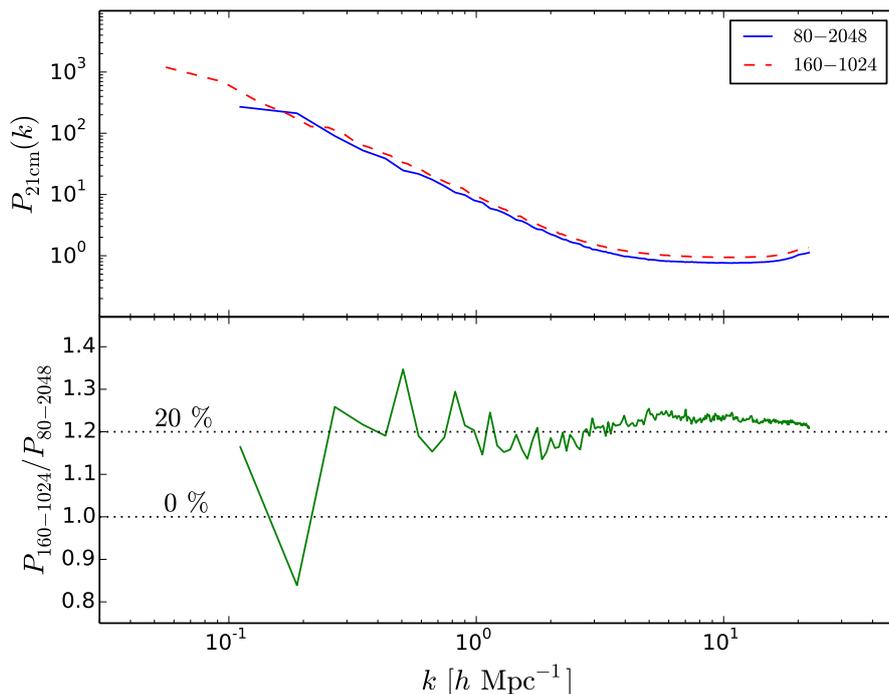
Figure 3: The upper panel displays the 21cm power spectrum from the 80-2048 (solid blue) and 160-1048 (dashed red) simulations at $z = 2.4$ while the lower panel shows their ratio.

4 Modelling the Lyα forest

As we discussed above, we use state-of-the-art hydrodynamic simulations to model the gas properties in low-density environments. We extract mock Lyα absorption spectra skewers from the simulations. We have three different catalogues, containing 2500, 1600 and 900 spectra that are obtained from regular grids of 50^2, 40^2 and 30^2 points along the x, y and z directions, respectively (in appendix B we discuss how regularizing the spectra position on a grid does not bias our results). Each spectrum contains 2048 pixels, evenly distributed along the simulation box length.

The gas density, weighted temperature, the neutral fraction and gas peculiar velocities are extracted following the SPH interpolation scheme described in [37]; for the Lyα optical depth τ along each line of sight we make use of the Voigt profile approximation, as in [38]. Once τ is determined in every pixel, we define the absorption flux as $F = e^{-\tau}$: this is the field we consider. For computing the power spectrum $P_{\mathrm{Ly}\alpha}(k)$ we compute the flux contrast as

$$\delta_F(\mathbf{x}) = \frac{F(\mathbf{x}) - \langle F \rangle}{\langle F \rangle} \tag{4.1}$$

where $\langle F \rangle$ is the flux mean (see appendix B for more details). For a more detailed description of the 160-1024 absorption spectra and comparison with observational data in terms of Lyα forest statistics we refer the reader to [17].

4.1 The Lyα forest auto-power spectrum

We first consider the one-dimensional Lyα power spectrum $P_{\mathrm{Ly}\alpha}^{\mathrm{1D}}(k)$. For each spectrum we compute the flux contrast δ_F in every pixel along the line of sight and the flux power spectrum. In Fig. 4
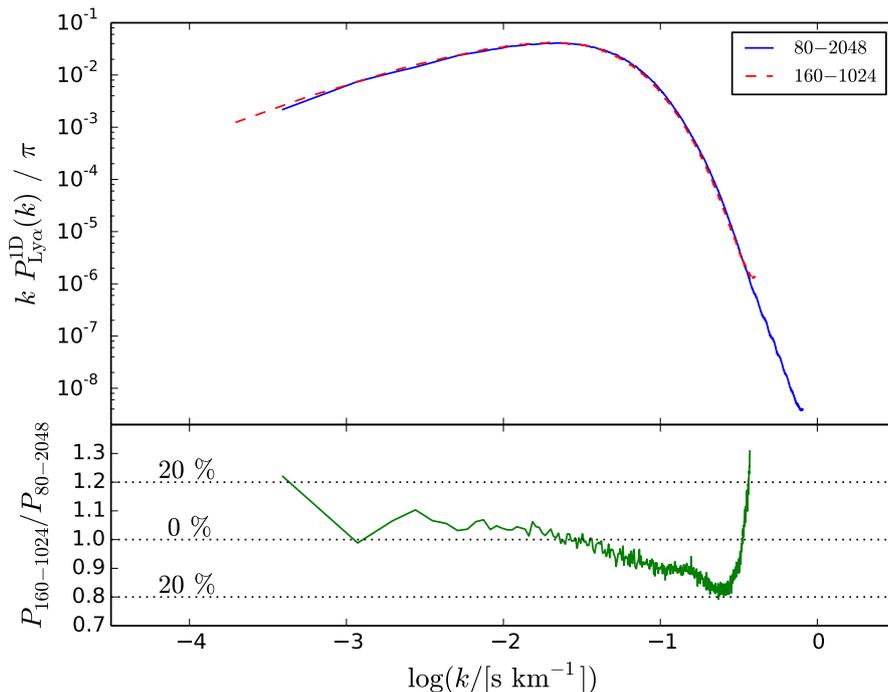
Figure 4: *Upper panel:* The dimensionless 1D Lyα flux power spectrum for the 80-2048 (blue solid) and the 160-1024 (red dashed) simulations. *Lower panel:* Ratio of the two power spectra; we highlight the 20 % difference region.

we show in the upper panel the two dimensionless power spectra, $k\,P_{\mathrm{Ly}\alpha}^{\mathrm{1D}}(k)/\pi$; with solid blue line for the 80-2048 simulation and in dashed red for the 160-1024. The shown power spectra have been computed by averaging the measured power spectrum from each individual skewer from the catalogue (5000 spectra in total: 2500 l.o.s. along x plus 1600 along y plus 900 along z). We find that our results are basically converged against resolution, as can be seen from the bottom panel, where we show their ratio, which stays within the 20% difference region highlighted within dotted lines.

Next we compute the 3D Lyα power spectrum $P_{\mathrm{Ly}\alpha}(k)$, shown in Fig. 5. There is good overlap among the six power spectra (giving that these catalogues have not been normalised, see appendix B), especially on large scales, with the 80-2048 spectra slightly flattening at $k \simeq 0.2h\,\mathrm{Mpc}^{-1}$. At smaller scales the 160-1024 spectra display an increase of power compared to the 80-2048 spectra because of the lack of small scale information, and the same effect is visible also within each simulation $P_{\mathrm{Ly}\alpha}(k)$: as the number of spectra per catalogue decreases (i.e. employing a smaller number of skewers in the box) we see an increase of power due to a sub-sampling of the field.

Given the good agreement on both the 1D and 3D Lyα power spectra among the two simulations (on large-scales for the 3D power spectrum and on all scales in the 1D case) we conclude that the spatial distribution of gas in the Lyα forest is converged against resolution in the simulation 160-1048. We will thus focus our analysis on that simulation, since its larger box size allow us to explore larger scales than the 80-2048 simulation.
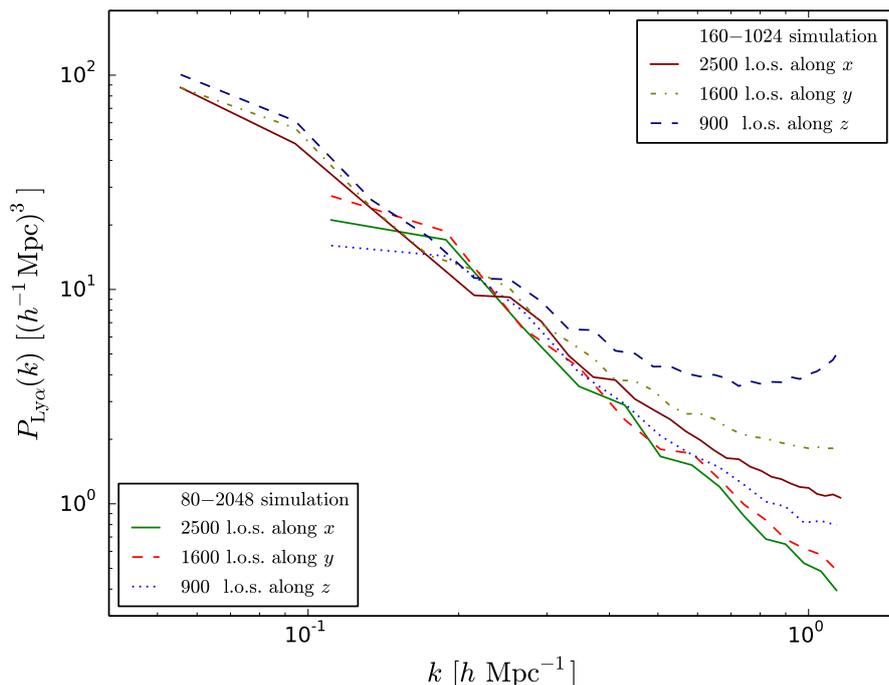
Figure 5: The 3D Lyα flux power spectrum computed for the 3 different catalogues (with spectra along different directions) for each simulation.

5 Results

In this section, we start by computing the cross-correlation of the 21cm and Lyα forest signals in 5.1. Next, in subsection 5.2 we check whether the detection of the cross-signal is possible in terms of realistic uncertainties on the measurements, whereas in 5.3 we infer the respective bias parameters by analysing the power spectra within a linear theory framework.

In the analysis of this section, we use the 160-1024 simulation since we are interested in investigating the shape and amplitude of the cross-power spectrum on linear and mildly non-linear scales, which are better probed by that simulation.

5.1 The 21cm-Lyα cross-correlation

In Fig. 6 we show with a dotted green line the 21cm power spectrum $P_{21cm}(k)$, in dashed blue the 3D Lyα flux power spectrum $P_{Ly\alpha}(k)$ and in solid magenta the absolute value of the 21cm-Lyα cross-power spectrum $P_X(k)$[4].

We have plotted the absolute value of the cross-power spectrum because $P_X(k)$ is negative, i.e. the fields are anti-correlated. Qualitatively, this results agrees with the picture of having the HI responsible for the 21cm radiation in dense environments as galaxies in halos, whereas the Lyα forest arises mainly from low-density, highly ionized, gas clouds in the intergalactic medium mostly residing in the filaments of the cosmic web. We will thus expect underdensities in the Lyα flux in places where 21cm overdensities are located (e.g. halos); this explains why the two fields are anti-

---

[4]We emphasize that we are neglecting instrumental effects such as beam size or system noise as well as the impact of galactic and extragalactic foregrounds.
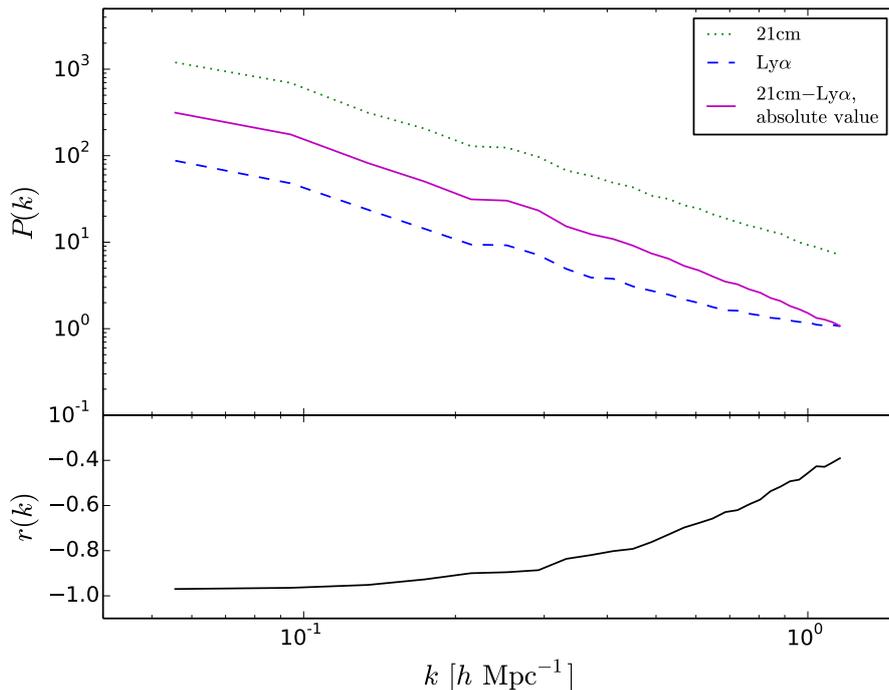
Figure 6: The upper panel displays the power spectrum of the 21cm signal (dotted green), of the Lyα forest flux (dashed blue) and of their cross-correlation in absolute value (solid magenta). The bottom panel shows the cross-correlation coefficient, $r(k)$, among the two fields.

correlated. We notice that this result agrees with recent observations by Mukae et al. [39], where they analyze HI-rich galaxies at $z \sim 2 - 3$ together with Lyα forest spectra finding anti-correlation between the two fields.

In the bottom panel of Fig. 6 we plot the cross-correlation coefficient, defined as

$$r(k) = \frac{P_{\mathrm{X}}(k)}{\sqrt{P_{21\mathrm{cm}}(k)P_{\mathrm{Ly}\alpha}(k)}} \; . \tag{5.1}$$

As expected from the arguments given above, we find the sign of the cross-correlation to be negative on all the scales, indicating that the fields are anti-correlated. On the largest scales probed by our simulation the value of the cross-correlation is close to -1, while for scales $k \geq 0.2 \, h\mathrm{Mpc}^{-1}$ the value of the cross-correlation increases. We can naively associate this scale to non-linearities, as linear theory predicts a scale-independent cross-correlation coefficient. Thus, since in this paper we are interested in extract cosmological information from linear scales[5], we limit our analysis to modes with wavenumbers $k < 0.2 \, h\mathrm{Mpc}^{-1}$.

In a nutshell: we are looking at signals coming from different regions in the sky. By considering together 21cm radiation and Lyα flux, we are looking at HI both in emission and in absorption, thus probing its cosmological amount and its spatial distribution at the same time. By visually inspecting the two fields in Fig. 2, we indeed notice how HI-poor regions (left panel, dark blue) correspond to high Lyα forest transmitted flux regions (right panel, yellow and red).

---

[5]We notice that perturbation theory is a powerful method to model mildly non-linear scales.

To quantify the improvement gained by looking for HI information in cross-correlation, in next section we make predictions on the accuracy with which one should be able to make such measurements, taking care of the system noise only, ignoring effects arising from calibration, astrophysical foregrounds, variations in the ionosphere, radio-frequency interference, etc, which are beyond the scope of this paper.

5.2 Error estimation and forecasts for SKA

In this section we describe the way we have computed the Gaussian errors for the measurements of the different power spectra of the previous section, and use the formalism to

In this section we estimate the signal-to-noise ratio of a measurement of the auto- and cross-power spectrum of the Lyα forest and the 21cm fields performed on the volume where that overlaps observations from SKA1-MID [23] and a BOSS-like survey [40] for the Lyα flux field. We notice that in this work we just focus on the cosmological signal, i.e. we neglect contributions from residual foreground contamination.

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

$$\left(\frac{S}{N}\right)_a^2(\mathbf{k}) = N_k \frac{P_a^2(\mathbf{k})}{\sigma^2[P_a(\mathbf{k})]}, \tag{5.2}$$

where N_k is the number of modes in each given bin centred at $(k, \cos(\theta))$, where θ is the angle between $\mathbf{k}$ and the line of sight. Here for concreteness and clearness we focus on the errors of the monopoles, but notice that the extensions to quadrupoles and monopoles-quadrupoles covariances is straightforward. For monopoles we average the amplitude of the power spectrum of modes with $k \in [k, k + dk]$, thus, the S/N ratio becomes:

$$\left(\frac{S}{N}\right)_a^2 = \frac{2\pi k^2 \mathrm{d}k V_{\mathrm{survey}}}{(2\pi)^3} \int_0^{\frac{\pi}{2}} \frac{P_a^2(k,\theta)\sin(\theta)\mathrm{d}\theta}{\sigma^2[P_a(k,\theta)]}, \tag{5.3}$$

where V_{survey} represents the survey volume. In our case we take it to be the volume of the region where both Lyα observations and the 21cm survey overlap. In particular, for a 21cm detection experiment we can write $V_{\mathrm{survey}} = D^2 \Delta D(\lambda^2/A)$, where D is the comoving distance to the redshift of observation, ΔD the comoving distance associated with the bandwidth of the instrument[6] $B = 32\,\mathrm{MHz}$, λ the wavelength of observation (corresponding 21cm line at the redshift of detection) and $A = 15\,\mathrm{m}$ the collecting area of a single antenna.

We can cast the error on the 21cm detection via interferometric observations as[7]:

$$\sigma^2[P_{21\mathrm{cm}}(k,\theta)] = \left[P_{21\mathrm{cm}}(k,\theta) + \frac{T_{\mathrm{sys}}^2}{2Bt_0} \frac{D^2 \Delta D}{n(k_\perp)} \left(\frac{\lambda^2}{A_e}\right)^2\right]^2, \tag{5.4}$$

where T_{sys} is the system temperature of the radio telescope which is the sum of the temperature of the sky at this redshift $T_{\mathrm{sky}} \simeq 60(300\mathrm{MHz}/\nu_{\mathrm{HI}}(z))^{2.55}$, with $\nu_{\mathrm{HI}}(z) = 1420/(1+z)$ MHz, the temperature receiver $T_{\mathrm{rcvr}} = 0.1T_{\mathrm{sky}} + T_{\mathrm{inst}}$, with $T_{\mathrm{inst}} = 28\mathrm{K}$ the SKA1-MID instrument temperature and. $n(k_\perp)$ is the number density of the interferometer baselines sensitive to the transverse mode $k_\perp$ which depends on the spatial distribution of the antennae that we calculate using the SKA1-MID baseline

---

[6]We choose this bandwidth to match the common volume of both surveys.

[7]We notice that we are neglecting the contribution to the error from shot-noise. In [29] it was shown that for the relevant scales we are interested here this term is subdominant.

density distribution in [10]. A_e is the effective collective area of a single antennae: for SKA1-MID $A_e = 140\,\mathrm{m}^2$. The parameter that we can tune is the total observation time t_0 that we conservatively set to 100 hours.

We express the error on the 3D Lyα flux power spectrum as a combination of the noise term with an aliasing term, due to the sparse sampling of the Lyα field made by the discrete lines of sight [41]:

$$\sigma^2[P_{\mathrm{Ly}\alpha}(k,\theta)] = \left[P_{\mathrm{Ly}\alpha}(k,\theta) + P_{\mathrm{Ly}\alpha}^{\mathrm{1D}}(k\cos\theta)n_{\mathrm{eff}}^{-1}\right]^2. \tag{5.5}$$

The aliasing term contributes in the line of sight direction with the 1D flux power spectrum $P_{\mathrm{Ly}\alpha}^{\mathrm{1D}}(k\cos\theta)$ multiplied by a noise-weighted density of lines of sight per unit area n_{eff}. From [22] we know that for a redshift bin to which $z = 2.4$ belongs the lines of sight density for BOSS is $n_{\mathrm{eff}} \simeq 15\,\mathrm{deg}^{-2}$. We choose to use both this value and its double ($30\,\mathrm{deg}^{-2}$) to estimate errors for a BOSS-like and a next generation BOSS-like survey like DESI [8]. Using the conversion table in [42], we obtained the values $n_{\mathrm{eff}} = 0.003$ and $0.006\,(h^{-1}\,\mathrm{Mpc})^{-2}$, respectively.

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

$$\sigma^2[P_{\mathrm{X}}(k,\theta)] = \frac{1}{2}\left(P_{\mathrm{X}}^2(k,\theta) + \sigma[P_{21\mathrm{cm}}(k,\theta)]\sigma[P_{\mathrm{Ly}\alpha}(k,\theta)]\right). \tag{5.6}$$

On small scales, the error budget of the 21cm power spectrum is dominated by the system noise and, by looking at the formula 5.4, we know it scales with the observing time as $1/t_0$, whereas for the cross-correlation power spectrum (see equation 5.6) goes as $1/\sqrt{t_0}$, so if we choose a more optimistic survey observing time of $t_0 = 1000$ hours instead of 100 hours, $\sigma[P_{21\mathrm{cm}}(k)]$ would improve by a factor 10 and $\sigma[P_{\mathrm{X}}(k)]$ would be ~ 3 times smaller. Anyways, at the scales we are looking at ($k < 0.2h\,\mathrm{Mpc}^{-1}$) this is no longer valid: observing for longer does not beat cosmic variance. Hence, we keep $t_0 = 100$ hours for our analysis and we do not explore other observing time possibilities.

We use the above formalism to compute the errors on the multipoles of the different power spectra, and their covariances, for the measurements we carried out in the previous section. We notice that since the scales we are interested when performing the fit are almost linear ($k < 0.2\,h\mathrm{Mpc}^{-1}$), the error budget is dominated by cosmic variance. Thus, in order to speed calculations up we have neglected the contribution of system noise to the 21cm power spectrum when computing the errors on the power spectra of the previous section. This is the reason of the low amplitude of the errors on the 21cm power spectra in Fig. 9, but notice that our results do not change if we include them.

However, we account for the contribution of system noise when estimating the S/N ratio for our forecasts in this section. We summarize the errors estimation in Fig. 7, where we plot their magnitude together with the amplitude of the different auto- and cross-power spectra.

In the upper panel of Fig. 7 we plot the dimensionless power spectrum $\Delta_a^2(k) = k^3 P_a(k)/(2\pi^2)$ with solid lines: green for the 21cm radiation, blue for the Lyα flux and magenta for the absolute value of their cross-signal, together with the errors $\sigma[P_a(k)]$ in dashed lines using $n_{\mathrm{eff}} = 0.003\,(h^{-1}\,\mathrm{Mpc})^{-2}$ and dotted lines for $n_{\mathrm{eff}} = 0.006\,(h^{-1}\,\mathrm{Mpc})^{-2}$. Clearly, the error on the 21cm power spectrum does not depend on the value of n_{eff}, and the improvement on the $P_{21\mathrm{cm}-\mathrm{Ly}\alpha}(k)$ is evident but smaller than that on $P_{\mathrm{Ly}\alpha}(k)$. Nonetheless, we can notice that the aliasing term in $\sigma^2[P_{\mathrm{Ly}\alpha}(k)]$ (and consequently in $\sigma^2[P_{\mathrm{X}}(k)]$) dominates the variance budget: therefore the way to do better would be to conduct more sensitive Lyα spectra surveys, i.e. by increasing the surveyed volume and/or increasing the quasars number density.

In the bottom panel of Fig. 7 we plot the signal to noise ratio S/N for $P_{21\mathrm{cm}}(k)$ and $P_{21\mathrm{cm}-\mathrm{Ly}\alpha}(k)$. We find that S/N ratio for both power spectra peak around $k \simeq 0.25\,h\mathrm{Mpc}^{-1}$. On larger scales cos-

---

[8] http://desi.lbl.gov/
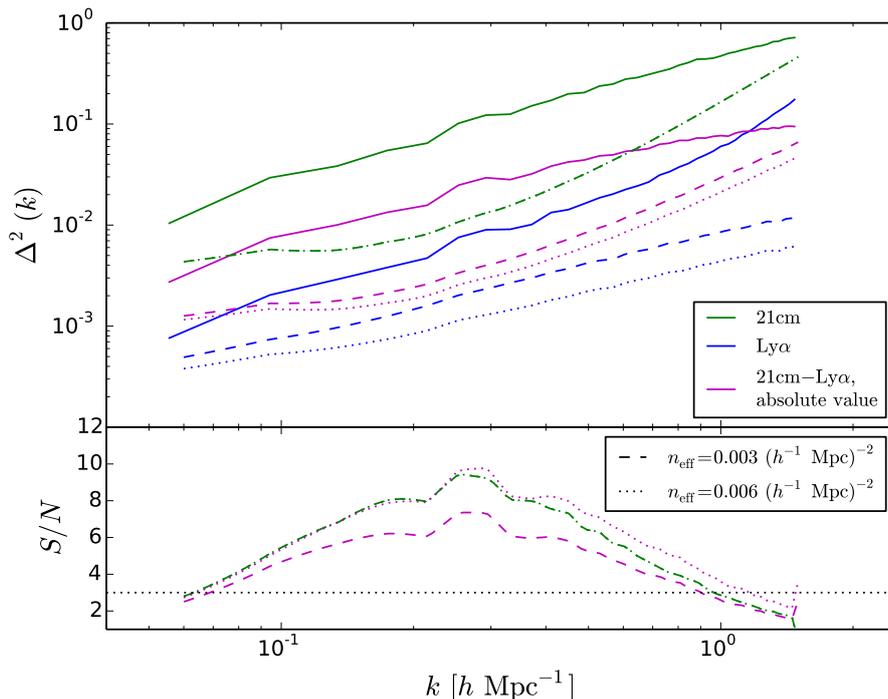
Figure 7: *Upper panel:* The dimensionless power spectrum $\Delta^2(k) = k^3 P(k)/2\pi^2$ in solid lines for the 21cm radiation (green), the Lyα forest flux (blue) and for their cross-correlation in absolute value (magenta). The non-solid lines are the estimated errors on the power spectra for a SKA1-MID (21cm intensity mapping) and BOSS like (Lyα flux) surveys, the latter with an effective density of lines of sight $n_{\rm eff} = 0.003 (h^{-1}\,{\rm Mpc})^{-2}$ in dashed lines or $n_{\rm eff} = 0.006 (h^{-1}\,{\rm Mpc})^{-2}$ in dotted lines. See section 5.2 for details. *Lower panel:* Signal-to-noise ratio defined as the ratio between the power spectrum and its error. The horizontal dotted line marks $S/N = 3$.

mic variance dominates the total error, while on smaller scales the noise from the instrument starts dominating.

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

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

5.3 Linear theory comparison and bias parameters estimation

In this section we introduce the theoretical model for the auto- and cross-power spectra and use it to blindly estimate the value of the bias parameters of both the Lyα forest and the 21cm fields. We

also investigate how much information we gain by using the cross-power spectrum together with the auto-power spectrum measurements.

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

$$P_{21cm}(k, \mu) = A^2 \Omega_{HI}^2 b_{HI}^2 \left(1 + \beta_{HI} \mu^2\right)^2 P_m(k), \tag{5.7}$$

$$P_{Ly\alpha}(k, \mu) = b_F^2 \left(1 + \beta_F \mu^2\right)^2 P_m(k), \tag{5.8}$$

$$P_X(k, \mu) = A \Omega_{HI} b_{HI} \left(1 + \beta_{HI} \mu^2\right) b_F \left(1 + \beta_F \mu^2\right) P_m(k), \tag{5.9}$$

where the normalisation factor A is computed from Eq. 3.8 defining $\overline{\delta T}_b = A \Omega_{HI}$, with $\Omega_{HI} = 10^{-3}$ in our modelling, b_{HI} and β_{HI} are the linear bias and the redshift-space distortion parameter for the 21cm field, b_F and β_F those for the Lyα forest, $P_m(k)$ is the linear matter power spectrum and μ is the cosine of the angle between the Fourier mode vector **k** and the line of sight.

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

Method 1: auto-power spectrum multipoles. This method consists in determining the value of the two bias parameters of each observable through fitting the monopole and quadrupole of each respective field in redshift-space. The power spectrum multipoles can be computed from the 2D power spectrum as:

$$P_l(k) = \frac{2l + 1}{2} \int_{-1}^{1} P(k, \mu) L_l(\mu) d\mu, \tag{5.10}$$

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

$$P_{a,0}(k) = b_a^2 \left(1 + \frac{2}{3} \beta_a + \frac{1}{5} \beta_a^2\right) P_m(k), \tag{5.11}$$

$$P_{a,2}(k) = b_a^2 \beta_a \left(\frac{4}{3} + \frac{4}{7} \beta_a\right) P_m(k), \tag{5.12}$$

where a stands either for 21cm or for Lyα (and here we omitted the normalisation factors for the 21cm case, see above Eq. 5.7). If we were interested in the value of β_a only, we could extract it from the ratio between quadrupole and monopole, which depends only on this parameter; to constrain also b_a, we need to assume a cosmological model through $P_m(k)$. The 21cm redshift-space distortion parameters contain information on the growth rate, f, since $f = \beta_{HI} b_{HI}$; the latter relation does not hold for the Lyα case, due to its intrinsically non-linear correspondence to the underlying matter density field [22] and this is the reason why adding information from the cross-power spectrum cannot directly improve the measurement of the linear growth rate, but it tightens the constraints on the (b_{HI}, β_{HI}) parameters as we later show.

We do the best fit to Eqs. 5.11-5.12 via a Monte Carlo Markov Chain (MCMC) on the two free parameters (b_a and β_a) making use of the *emcee* package [43]. Errors on the monopoles and quadrupoles of the auto-power spectra are computed assuming the modes follow a Gaussian distribution (see appendix C), which is a good description on large, linear scales, and just taking into account the contribution from cosmic variance. We consider a cosmological volume equal to the one probed by our simulations and using only modes with $k < 0.2 \, h\mathrm{Mpc}^{-1}$. We also account for the correlation between the monopoles and quadrupoles in the covariance matrix. We notice that when performing the fit we are neglecting the correlation between multipoles of the two fields through the cross-power spectrum. The best-fit values together with their 1σ errors are shown in the upper row of table 2.

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

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

The recovered values for the HI, $b_{\mathrm{HI}} = 1.520^{+0.058}_{-0.060}$ and $\beta_{\mathrm{HI}} = 0.720^{+0.14}_{-0.13}$ are in agreement with the input ones, $b_{\mathrm{HI}} = 1.45$, $\beta_{\mathrm{HI}} = 0.67$, at $\sim 1\sigma$. Notice that we have not inputed any model for the Lyα forest, whose properties are directly extracted from the output of the simulations. We find $\beta_F = 1.480^{+0.21}_{-0.20}$, $b_F = -0.144^{+0.007}_{-0.007}$, while from observations it has been measured $\beta_{\mathrm{F}} = 1.39 \pm 0.1$ and $b_{\mathrm{F}}(1 + \beta_{\mathrm{F}}) = -0.374 \pm 0.007$ [44], thus, in perfect agreement. We notice that we obtain a good normalized χ^2 of $13.6/12$, showing the model is a good description of the data.

Method 2: cross $P(k)$ multipoles. In this second method we exploit also the information contained in the cross-correlation power spectrum $P_{21\mathrm{cm}-\mathrm{Ly}\alpha}(\mathbf{k})$. Following Eqs. 5.10 and 5.9, we can write the cross-power spectrum multipoles as:

$$P_0(k) = A\Omega_{\mathrm{HI}}b_{\mathrm{HI}}b_F \left(1 + \frac{1}{3}(\beta_F + \beta_{\mathrm{HI}}) + \frac{1}{5}\beta_F\beta_{\mathrm{HI}}\right) P_{\mathrm{m}}(k), \tag{5.13}$$

$$P_2(k) = A\Omega_{\mathrm{HI}}b_{\mathrm{HI}}b_F \left(\frac{2}{3}(\beta_F + \beta_{\mathrm{HI}}) + \frac{4}{7}\beta_F\beta_{\mathrm{HI}}\right) P_{\mathrm{m}}(k). \tag{5.14}$$

where the normalisation factor A is such that $\overline{\delta T_b} = A\Omega_{\mathrm{HI}}$ from Eq. 3.8 and the total amount of HI is set to $\Omega_{\mathrm{HI}} = 10^{-3}$ in our modelling. We fit simultaneously the monopoles and quadrupoles of the auto- and cross-power spectra, again employing the *emcee* package [43] to perform Monte Carlo Markov Chain (MCMC) on the four free parameters ($b_F, \beta_F, b_{\mathrm{HI}}, \beta_{\mathrm{HI}}$) employing only power spectra measurements for $k < 0.2\ h\mathrm{Mpc}^{-1}$. As in the case of the auto-power spectra, we estimate the errors on the multipoles of the cross-power spectra assuming the modes follow a Gaussian distribution and accounting for the correlation between monopoles and quadrupoles among the different auto- and cross-power spectra (see appendix C). The results are shown in the bottom row of table 2 and the degeneracies among parameters are displayed in Fig. 8.

The best-fit values for the monopoles and quadrupoles of the auto- and cross-power spectra from the joint fitting are shown in Fig. 9. For this method we obtain a normalized χ^2 of $30.2/20$, showing that we may be underestimating some errors. We obtain values of the parameters equal to $b_{\mathrm{HI}} = 1.472^{+0.043}_{-0.044}$, $\beta_{\mathrm{HI}} = 0.762^{+0.10}_{-0.10}$, $b_F = -0.139^{+0.005}_{-0.005}$, $\beta_F = 1.580^{+0.16}_{-0.15}$. The derived values for the HI are in perfect agreement with the input ones, and the combination $b_F(1 + \beta_F)$ also reproduces the observational constraints [44]. We also notice that the values derived with this method are in agreement with the ones obtained by fitting only the auto-power spectra.

We show in Fig. 10 a more detailed comparison between the derived values of the 21cm from the two methods and the input ones. The dotted line in that plot represents the function $\beta_{\mathrm{HI}} = (f/b_{\mathrm{HI}})$, where $f(z) \simeq [\Omega_{\mathrm{m}}(z)]^{0.545}$ from linear theory; for the cosmological set-up employed in this simulation at redshift $z = 2.4$, $f = 0.97$. There is agreement among the two $(b_{\mathrm{HI}}, \beta_{\mathrm{HI}})$ points, and both are compatible with the Kaiser approximation (dotted line) within 1σ.

It is important to notice that by adding the information embedded into the cross-power spectra, together with the assumed perfectly correlation on linear scales, the errors on the 21cm and Lyα forest
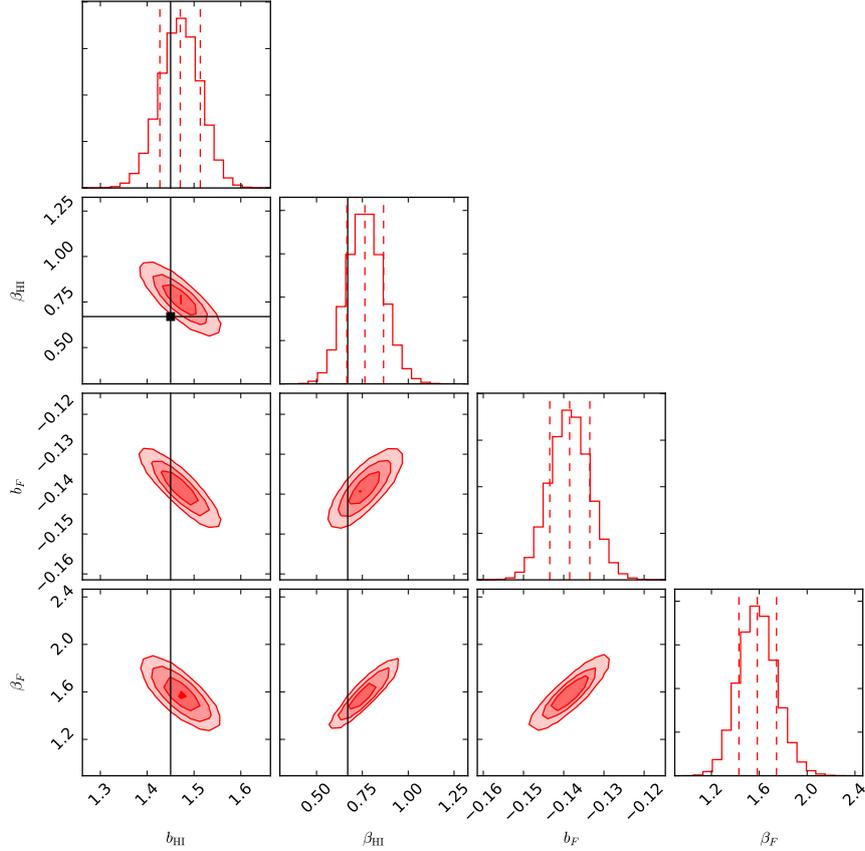
Figure 8: Constraints and degeneracies on the bias parameters of the 21cm and Lyα forest obtained by performing a joint fit to the auto- and cross-power spectra of the two fields. The blue lines indicate the expected value for the 21cm, that we know by construction.

parameters decrease by $\sim 30\%$. Therefore cross-power spectra have to be seen as a powerful way to look for systematics in the 21cm field but also as a way to add extra information that can shrink the error on the model parameters.

We also notice that although we are only using power spectra measurements with $k < 0.2 \, h\mathrm{Mpc}^{-1}$ to fit the results of the simulations, our linear model for the cross-power spectrum, used with the best-fit value parameters from the fit, is capable of reproducing the amplitude and shape of the cross-power spectrum multipoles to a remarkable accuracy down to the smallest scales we probe: see Fig. 9. This could be just a coincidence, it may arises because we are not properly modeling the 1-halo term of the HI distribution or may be something more deep. Investigating this issue is however beyond the scope of this paper.

We want to stress again that here our task was to retrieve the input b_{HI} value through our simulated data as a consistency check and to show the error improvement by using the information coming from the cross-power spectrum. We do not attempt to constraint the absolute b_{HI} since it is tuned in our model by construction.

One possible systematic in our analysis is the impact of the HI density profile we have used to compute 21cm power spectra. In the appendix A we show that on the scales where we carry out
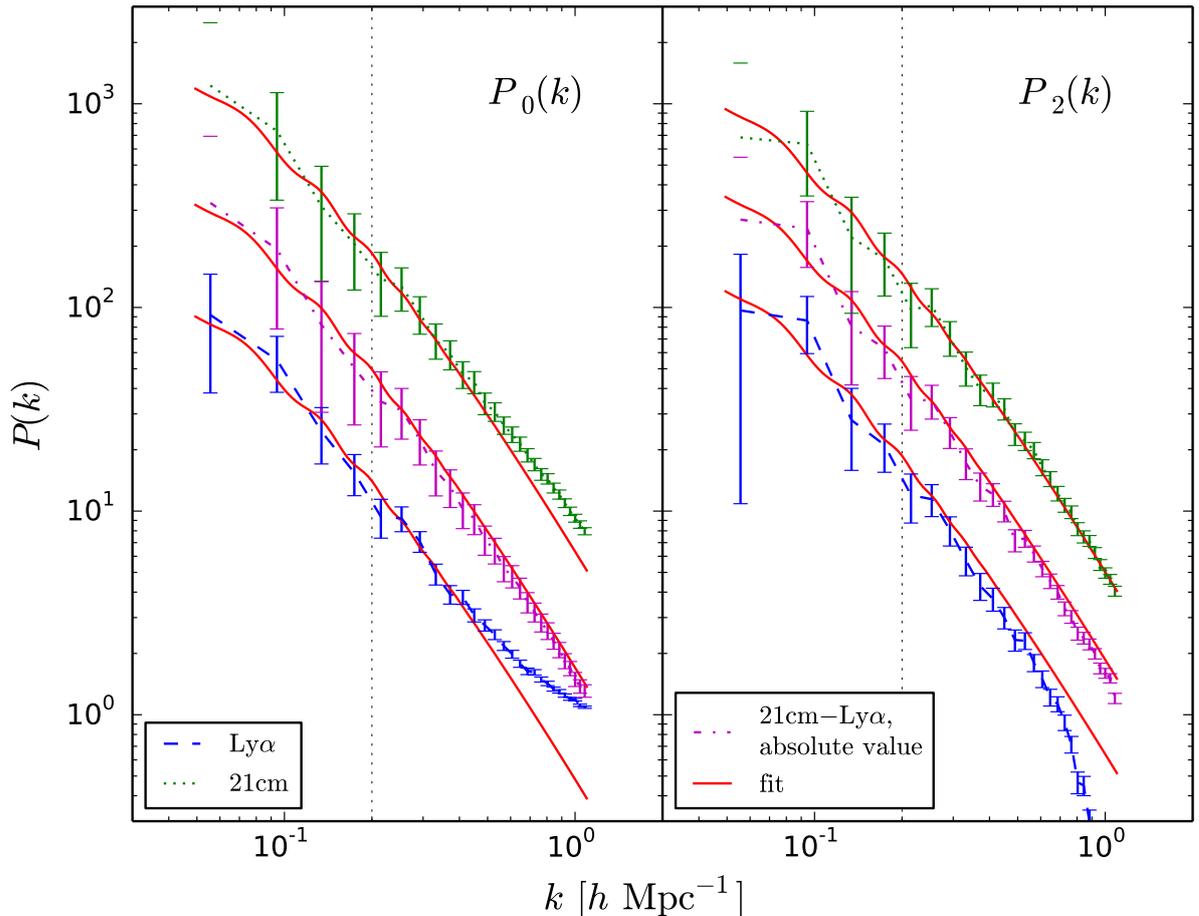
Figure 9: The red solid lines are the result of fitting the $P(k)$ monopoles (left panel) and the quadrupoles (right), the green dotted lines refer to the 21cm, the blue dashed to Lyα forest flux and the magenta dashed-dotted to their cross-correlation in absolute value, following the analysis described in section 5.3. The black dotted vertical line marks $k = 0.2h\,\mathrm{Mpc}^{-1}$, the mode up to which we perform the analysis.

the fit, the 1-halo term does not have a noticeable effect. We also notice that using a model with a different value of Ω_{HI} will only shift the amplitude of the 21cm power spectrum, but not its shape, i.e. the values of $(b_{\mathrm{HI}}, \beta_{\mathrm{HI}})$ will not be affected.

Since the amplitude of the 21cm power is proportional to $\Omega_{\mathrm{HI}}b_{\mathrm{HI}}$, 21cm intensity mapping surveys are only sensitive to that product: e.g. $\Omega_{\mathrm{HI}}b_{\mathrm{HI}} = 0.62\times10^{-3}$ at $z \sim 0.8$ [45]; for comparison, we remind the reader that in our modelling we set $\Omega_{\mathrm{HI}} = 10^{-3}$ as an input parameter. We stress that the degeneracy $\Omega_{\mathrm{HI}} - b_{\mathrm{HI}}$ can be broken by adding information from other surveys, such as the HI column density distribution function from the Lyα forest, that is directly sensitive to Ω_{HI}. From the 21cm signal alone (after reionization) the Ω_{HI} measurement is expected to be hard if not impossible, as result of the drastically dropping of the brightness temperature T_b (the strength of the signal), making it less easy to distinguish from foregrounds.

The effect of shifting the 21cm power spectrum amplitude can be produced also by allowing the
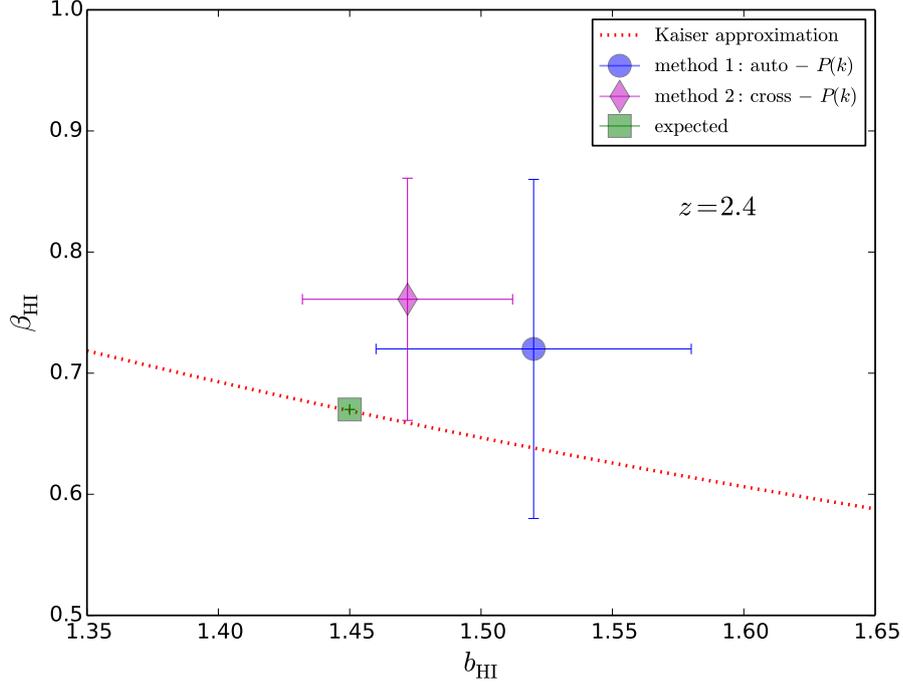
Figure 10: Values of the bias parameters b_{HI} and β_{HI} derived with the two procedures described in section 5.3 and summarised in table 2. The green square indicates the expected values, that we know by construction. The red dotted line is the Kaiser approximation: $\beta_{\mathrm{HI}} \times b_{\mathrm{HI}} = f(\Omega_{\mathrm{m}}) \simeq \Omega_{\mathrm{m}}^{0.55}(z)$.

dark matter particle to have mass of the order of keV (thermal relic) instead of the *cold* approximation [26], and this mass scale is, for example, that of the sterile neutrino, one of today most favourable dark matter candidates [46]. Hence, to extract information about Ω_{HI} by looking at the observed $P_{21\mathrm{cm}}(k)$, one should also consider the nature of the dark matter particle and disentangle the two responses on the $P_{21\mathrm{cm}}(k)$.

5.3.1 Degeneracies between the bias b_{HI}, Ω_{HI} and the growth factor f: a fisher matrix analysis

Given the results shown in Sec. 5.3, we can infer that the Kaiser approximation from linear theory $f = \beta_{\mathrm{HI}} b_{\mathrm{HI}}$ agrees well with our 21cm-Lyα modelling (Fig. 10) over a compelling range of k (Fig. 9) and thus we can rewrite the linear order 21cm power spectrum of Eq. 5.7 replacing $\beta_{\mathrm{HI}} = f/b_{\mathrm{HI}}$; its monopole and quadrupole become:

$$P_{21\mathrm{cm},0}(k) = A^2 \Omega_{\mathrm{HI}}^2 \left(b_{\mathrm{HI}}^2 + \frac{2}{3} f b_{\mathrm{HI}} + \frac{1}{5} f^2 \right) P_{\mathrm{m}}(k), \tag{5.15}$$

$$P_{21\mathrm{cm},2}(k) = A^2 \Omega_{\mathrm{HI}}^2 f \left(\frac{4}{3} b_{\mathrm{HI}} + \frac{4}{7} f \right) P_{\mathrm{m}}(k). \tag{5.16}$$

Analogously, the cross-power spectrum multipoles of Eqs. 5.13-5.14 become:

$$P_0(k) = A \Omega_{\mathrm{HI}} b_F \left(b_{\mathrm{HI}} + \frac{1}{3}(f + b_{\mathrm{HI}}\beta_{\mathrm{F}}) + \frac{1}{5} f \beta_{\mathrm{F}} \right) P_{\mathrm{m}}(k), \tag{5.17}$$

$$P_2(k) = A \Omega_{\mathrm{HI}} b_F \left(\frac{2}{3}(\beta_F b_{\mathrm{HI}} + f) + \frac{4}{7} \beta_F f \right) P_{\mathrm{m}}(k). \tag{5.18}$$

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

Table 3: Value of the bias and β parameters and of the cosmological growth factor f derived by carrying out fit to the results of the simulations using the auto-power spectrum multipoles alone (upper row) and making a joint fit to all auto- and cross-power spectra of the 2 fields (bottom row).

We thus perform a new MCMC analysis as described in Sec. 5.3, for determining the parameters (b_F, β_F, $b_{\rm HI}$, f) employing only power spectra measurements for $k < 0.2\ h{\rm Mpc}^{-1}$, first fitting together the monopoles and quadrupoles of the auto-power spectra, and then adding also the cross-power spectrum. Cosmic variance errors (no system noise) are estimated again as in appendix C. The results are shown in table 3.

There is agreement between the (b_F, β_F, $b_{\rm HI}$, f) values determined with the two fits and adding information coming from the cross-power spectrum again shrinks the associated errors. Both fits prefer an unreasonable growth factor value greater than unity, $f \simeq 1.1 \pm 0.1$, although being in agreement with linear theory $f = f(\Omega_{\rm m}) \simeq \Omega_{\rm m}^{0.55}(z) = 0.97$ within 1σ. This can be seen already in Fig. 10 where both points were above the dotted line corresponding to the linear relation $\beta_{\rm HI} \times b_{\rm HI} = f(\Omega_{\rm m})$. We remind that we do not add any prior on the physical value of f.

To better understand the degree of degeneracy of parameters involved in the 21cm characterization ($\Omega_{\rm HI}$, $b_{\rm HI}$ and f), we perform a Fisher matrix analysis using monopoles and quadrupoles of the 21cm auto-power spectrum and the of cross-power spectrum with the Lyα flux. This is a good exercise especially to check the effect of the uncertainty of $\Omega_{\rm HI}$, that is tuned by construction in the previous analysis and in the simulated 21cm field.

The Fisher matrix analysis quantifies the amount of information that the 21 cm power spectrum as observable carries about the three parameters $\Omega_{\rm HI}$ - $b_{\rm HI}$ - f. Practically, we use as prior the values found with the MCMC fit as in table 3 (i.e. $\Omega_{\rm HI} = 10^{-3}$, $b_{\rm HI} = 1.5$ and $f = 1.1$), and we calculate analytically how much the 21cm power spectrum varies by varying the values of those parameters. We make use of Eqs. 5.11-5.12-5.13-5.14 as templates and of the expressions shown in the appendix C for building the covariance matrix, i.e. the Gaussian uncertainties linked to these parameters. A quick and clear reference for this kind of analysis is in [47].

In Fig. 11 we show the results. In the top left panel we show the uncertainty contours for $\Omega_{\rm HI}$ - $b_{\rm HI}$ (fixing f), top right the uncertainty contours for f - $b_{\rm HI}$ (fixing $\Omega_{\rm HI}$), bottom left the uncertainty contours for $\Omega_{\rm HI}$ - f (fixing $b_{\rm HI}$) and in last panel on bottom left the uncertainty contours for the product $\Omega_{\rm HI} b_{\rm HI}$ - f. The orange ellipses refer to the uncertainty using the auto-power spectrum information only, the blue ellipses using information coming from both auto- and cross- power spectra: in the second case the uncertainties are always reduced, i.e. the cross-power spectrum shrinks our constraints in any case. Especially for the $\Omega_{\rm HI}$ - $b_{\rm HI}$ correlation: fixing the growth factor helps reducing the degeneracy and adding the cross- information shrinks the errors by $\sim 50\%$.

Concerning the ability of a 21cm - Lyα joint analysis to constrain the cosmological growth factor f: we will need an independent measurement of $\Omega_{\rm HI}$ in order to reach sufficient precision. Anyway, we point out that in the redshift range probed by 21cm intensity mapping surveys and Lyα flux experiments ($z \sim 2 - 3$) we have no other precise f measurement.
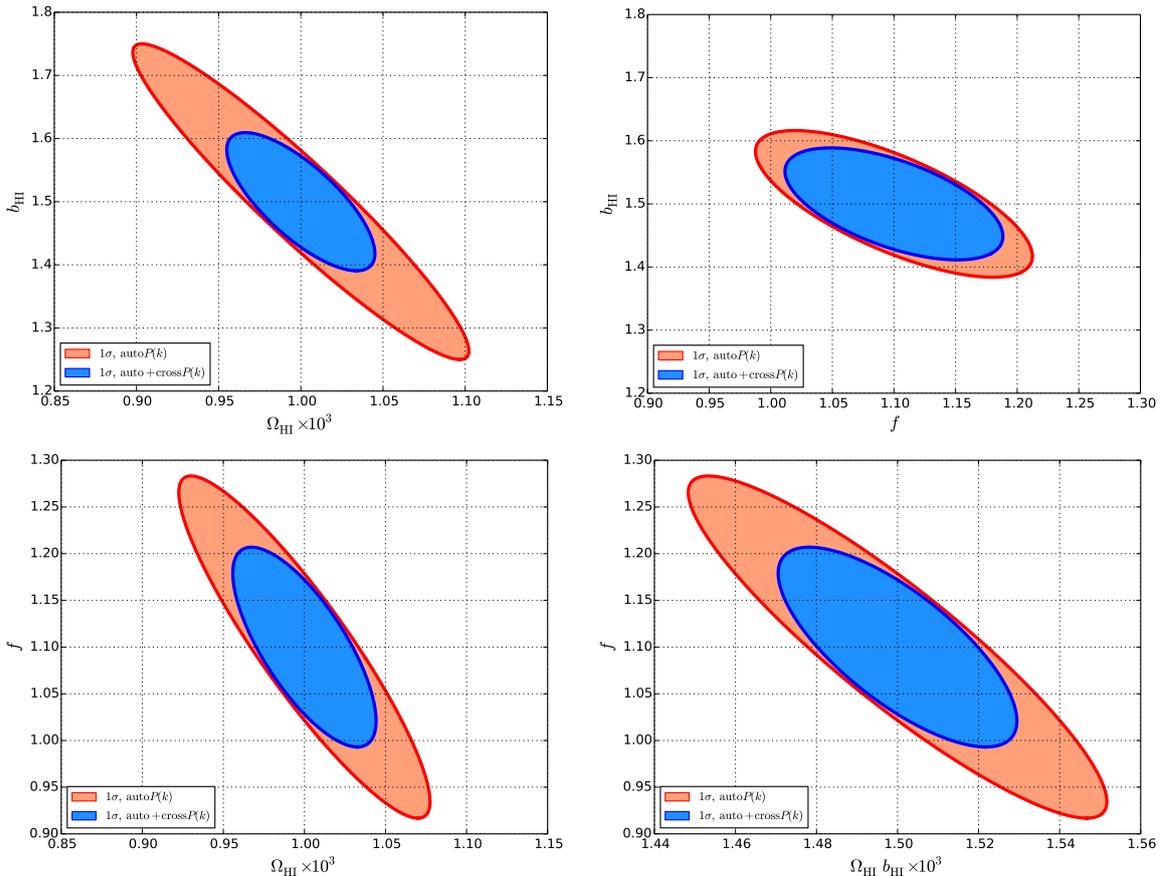
Figure 11: 1σ contours of the values of the bias b_{HI}, the density parameter Ω_{HI} and the growth factor f determined using either only the 21cm power spectrum (orange areas) or adding the cross 21cm-Lyα power spectrum (blue areas). The Fisher matrix analysis is performed using the theoretical templates of the power spectra multipoles and errors described in the results Sec. 5 and in the appendix C, using the $k < 0.2h\,\mathrm{Mpc}^{-1}$ information.

6   Summary and conclusions

The spatial distribution of matter in the Universe embeds a huge amount of information on the fraction that each component contributes to the total energy content of the Universe, on the nature of gravity and of the initial conditions of the Universe, on the geometry of the Universe and so on. Unfortunately, the spatial distribution of matter is not directly observable, but can be mapped through tracers of it such as galaxies or cosmic neutral hydrogen.

The 21cm intensity mapping technique consists in carrying out low angular resolution radio-observations with the goal of measuring 21cm flux perturbations from large patches of the sky where the galaxies that host the HI are not resolved. It is expected that this technique will play a major role in cosmology, given the spectroscopic nature of these observations and the large volumes it can sample. Unfortunately, the cosmological signal is buried by the galactic and extragalactic foregrounds, whose amplitude is several order of magnitude larger. Therefore, the cosmological information that can be extracted from these surveys critically depends on the precision with which the foregrounds can be cleaned from the 21cm maps.

While *well behaved* foregrounds can be robustly removed, the presence of some foregrounds such as polarized synchrotron radiation may not be completely removed and can ruin the inferred shape and amplitude of the 21cm power spectrum [9, 10]. A way to get rid of this problem is through cross-correlations [10–14, 16], since the foregrounds should not be correlated with the cosmological signal from the different tracers.

In this paper we have studied the cross-correlation between 21cm maps and the Lyα forest at redshift $z = 2.4$. We performed this study by means of state-of-the-art hydrodynamic simulations: the Sherwood suite [17]. While the properties of the Lyα forest are obtained directly from the simulation, we use a simple deterministic model to model the abundance of HI. We compute auto- and cross-power spectra for the two different fields in redshift space.

We find that on large, linear scales, the Lyα forest is completely anti-correlated ($r = -1$) with the 21cm field. This happens because regions with large HI densities exhibit large 21cm emission, but those regions are dominated by damped Lyα systems (DLAs), i.e. the mean transmitted flux is low. The fact that the Lyα forest traces low-density, highly ionized gas, while the 21cm signal arises from high density regions where hydrogen is self-shielded and therefore mainly neutral is the origin of this anti-correlation.

We have computed the cross-correlation coefficient between the two fields and we find it to be -1 until $k \simeq 0.2 \ h\mathrm{Mpc}^{-1}$, while on smaller scales its value increases, showing that on those scales the two fields are not perfectly anti-correlated. We thus determine that a joint fit to the different auto- and cross-power spectra using linear theory as template should be performed until $k = 0.2 \ h\mathrm{Mpc}^{-1}$. This is important for fisher matrix analysis, like the one in [14], used to forecast the improvement on the cosmological parameters from future surveys.

We compute errors on the monopoles of the Lyα forest and 21cm auto-power spectra and on the monopole of the cross-power spectrum for a cosmological volume equal to the one jointly sampled by a BOSS like survey and by SKA1-MID by means of interferometry observations. We find that the cross-power spectrum can be measured with a high S/N ratio to very small scales: S/N> 3 for $k \in [0.06 - 1] \ h\mathrm{Mpc}^{-1}$.

We have exploited the anisotropy of the power spectra in redshift-space to determine the values of the bias and redshift-space distortion parameter, β, for both fields by computing the monopole and quadrupole for each auto- and cross-power spectrum. We have fit simultaneously the monopole and quadrupole of each field using the linear theory prediction up to the above $k_{\max} = 0.2 \ h\mathrm{Mpc}^{-1}$ value. The errors on the multipoles are computed assuming the fields are Gaussian and the covariance is built taken into account the correlation between the monopole and quadrupole. We find that we are able to retrieve the input value for the HI distribution of b_{HI} together with $\beta_{\mathrm{HI}} = f/b_{\mathrm{HI}}$. In terms of the Ly$\alpha$ forest we obtain values of β_{F} and b_{F} that reproduce the observational constraints: $\beta_{\mathrm{F}} = 1.39 \pm 0.1$ and $b_{\mathrm{F}}(1 + \beta_{\mathrm{F}}) = -0.374 \pm 0.007$ [44].

We also perform a joint fitting of the multipoles and quadrupoles of the Lyα and 21cm auto-power spectra and their cross-power spectrum. In this case the values we obtain for the bias parameters are in agreement with those derived from the auto-power spectra, but their errors are smaller by $\sim 30\%$.

We find that the multipoles of the 21cm-Lyα cross-power spectrum can be surprisingly well reproduced by linear theory down to the smallest scales we probe in our analysis: $k \simeq 1 \ h\mathrm{Mpc}^{-1}$. It is beyond the scope of this paper to investigate whether this is just a coincidence or the result of some non-trivial physical reason.

Finally, we performed a Fisher matrix analysis for inspecting the degeneracy of the parameters characterizing the 21cm power spectrum, again showing how adding the information from the cross-correlation helps reducing the uncertainties.

In summary: while residual foreground contamination can bias the shape and amplitude of the 21cm auto-power spectrum, and therefore the estimated value of the cosmological and astrophysical parameters, the cross-power spectrum will be less affected by this problem and therefore represents a powerful method to verify the presence of systematics in the 21cm maps. If the 21cm maps are free of large systematics, data from auto- and cross-power spectra can be combined to break degeneracies and to tighten the value of the parameters. Determining the scale where linear theory holds is important since using an incorrect theoretical template will bias the inferred values of the model parameters.

Acknowledgements

## A   21cm power spectrum: dependence on $\rho_{\mathrm{HI}}(r)$

Here, we investigate the dependence of the HI and 21cm power spectra on the model used for the HI density profile within halos: $\rho_{\mathrm{HI}}(r|M,z)$. We study this using an hydrodynamic simulation that we coin 60-512. As the simulations used in this work, the 60-512 has been run using the TreePM+SPH code Gadget-3, with cosmological parameters in agreement with recent Planck data [19]: $(\Omega_{\mathrm{m}}, \Omega_{\Lambda}, \Omega_{\mathrm{b}}, h, n_s, \sigma_8)$ have the following values: (0.3175, 0.6825, 0.049, 0.671, 0.9624, 0.834).

The 60-512 simulation follows the evolution of 512^3 cold dark matter particles and 512^3 baryon particles within a periodic box of linear comoving size of $60 \, h^{-1}$ Mpc from $z = 99$ down to $z = 3$. Star formation is modeled using the effective multi-phase model of Springel & Hernquist [48]. The code also simulates radiative cooling by hydrogen and helium and heating by an uniform Ultraviolet (UV) background. 60-512 has mass resolutions of $m_{\mathrm{CDM}} = 1.2 \times 10^8 \, h^{-1}\mathrm{M}_\odot$ and $m_{\mathrm{baryon}} = 2.2 \times 10^7 \, h^{-1}\mathrm{M}_\odot$ and softening length with value $l_{\mathrm{soft}} = 2.9 \, h^{-1}\mathrm{kpc}$.

We have performed our tests using this simulation and not 80-2048 or 160-1048 because these are affected by an unrealistic star formation rate that make the gas reservoir of those simulations unreliable. In our analysis we have modeled the HI density profile as:

$$\rho_{\mathrm{HI}}(r|M,z) = M_{\mathrm{HI}}(M,z)\delta(\mathbf{r}) \qquad (\mathrm{A.1})$$

where $\delta(\mathbf{x})$ is the Dirac delta. Thus, we are collapsing the HI density distribution into a single point located in the halo center. This is of course a very crude approximation, but given the fact that in our simulations the spatial distribution of gas is unreliable this is the most simple choice that does not involve ad-hoc assumptions. In order to check the dependence of our results on $\rho_{\mathrm{HI}}(r|M,z)$ we have also distributed the HI within halos evenly among all gas particles belonging to it, i.e. we have used $\rho_{\mathrm{HI}}(r|M,z) = \rho_{\mathrm{g}}(r|M,z)$, where $\rho_{\mathrm{g}}(r|M,z)$ represents the density profile of gas within a dark matter halo of mass M at redshift z.
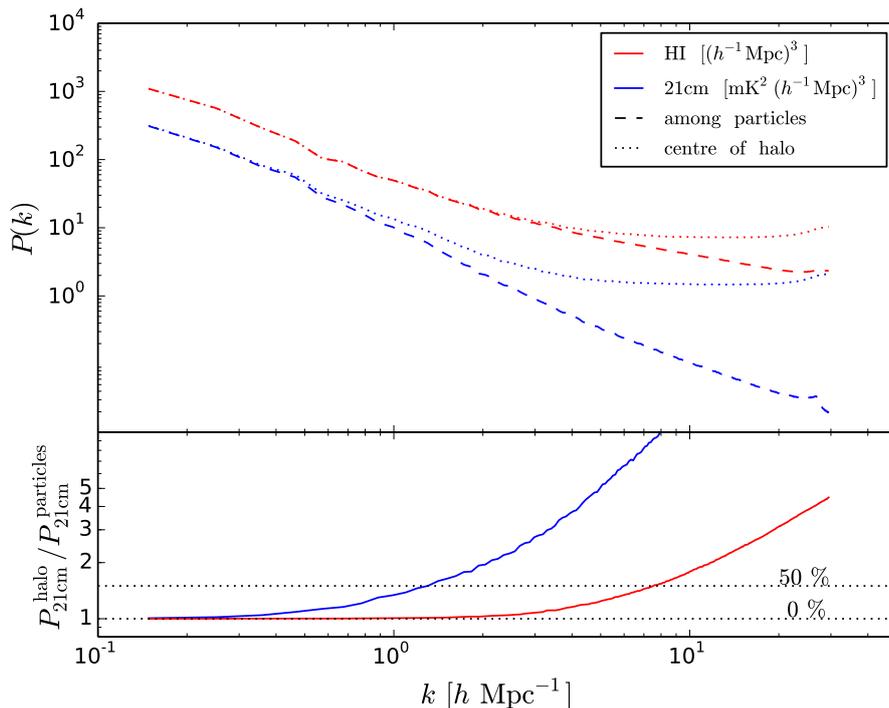
Figure 12: *Upper panel:* The HI power spectrum in real space $P_{\mathrm{HI}}(k)$ (in red) and the 21cm $P_{21\mathrm{cm}}(k)$ (in blue), using two ways for distributing the HI: at the centre of each halo (dotted lines) or spread among all gas particles belonging to the halo (dashed lines). *Lower panel:* Ratio between the two $P(k)$ computed with the HI at the centre of halos over the one with HI among all halo particles. The dotted horizontal lines mark when a 50% difference among the two $P(k)$'s is reached and when there is no difference among the two. For a description of the two ways of implementing the halo-based model see subsection 3.1. All power spectra here refer to the 60-512 simulation.

In Fig. 12 we show the HI and 21cm power spectra that we obtain using the two different HI density profiles. As expected, the HI density profile only affects the 1-halo term and therefore differences among different models for the $\rho_{\mathrm{HI}}(r|M, z)$ only show up on relatively small scales. On the other hand, dispersion velocities within halos propagate into large scales through redshift-space distortions; we find that the 21cm power spectra start deviating from each other on scales $k \lesssim 0.03\ h\mathrm{Mpc}^{-1}$, and differ by a $\sim 50\%$ at $k \sim 1.5\ h\mathrm{Mpc}^{-1}$. These differences are thus expected since with our fiducial collapsed HI density profile we are not modeling the finger-of-God because no halo substructure is used to place the neutral hydrogen.

In this paper we are interested on the amplitude and shape of the auto- and cross-power spectra on large-scales, where the different models for the HI density profile produce almost identical results. Thus, we conclude that our findings are robust against the simplified model we use to distribute HI within halos.

B   On some standard practices in generating mock Lyα forest spectra

In this appendix we discuss some of the problems we have faced when computing the 3D power spectrum of the Lyα forest. As in appendix A, here we also make use of the 60-512 simulation. We
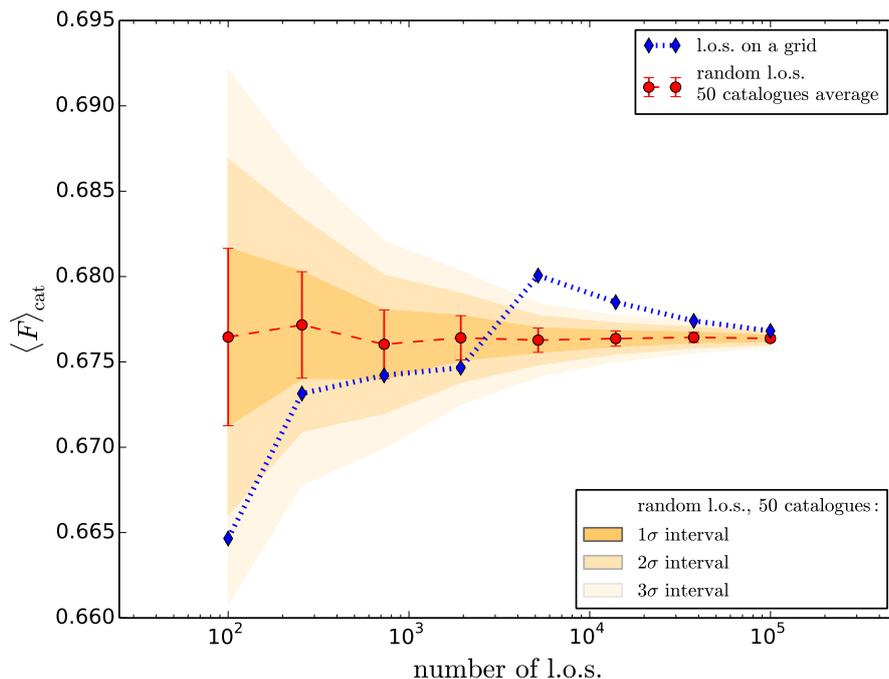
Figure 13: Mean flux $\langle F \rangle$ of the Lyα absorption spectra catalogues, plotted versus the number of lines of sight (l.o.s.) in each catalogue. 8 catalogues have the l.o.s. placed on a regular grid (blue diamonds). Each red point represents the average of the mean flux from 50 independent catalogues with same number of random l.o.s.. The error bars display the 1σ variation from the catalogues, while the shaded orange contours show the 1σ, 2σ and 3σ deviations. All absorption spectra are made of 256 pixels skewers along the x direction from the 60-512 simulation.

extract mock Lyα absorption spectra skewers from it as described in section 4, i.e. in the same way as the spectra analysed in the whole paper. All absorption spectra used in this appendix contain 256 pixels each and are taken along the x direction of the 60-512 simulation box.

B.1 Normalising the spectra with τ_{eff}

The amplitude of the 3D Lyα power spectrum, $P_{\mathrm{Ly}\alpha}(k)$, depends on the actual observed cosmological mean transmitted flux $\langle F \rangle_{\mathrm{obs}}$, measured in e.g. [49–51]. When the absorption spectra are recovered artificially by piercing the simulation box with skewers, the mean value $\langle F \rangle_{\mathrm{cat}}$ of the catalogue of spectra can vary quite a lot due primarily to the number of skewers that are drawn, as already seen earlier in Fig. 5, i.e. the more we sample the box, the more flux we get and $\langle F \rangle_{\mathrm{cat}}$ increases. Of course in a survey we have a finite amount of l.o.s. ($nlos$). The question is: how should the mean flux be computed to get sensible results for the 3D Lyα power spectrum?

A possibility is to take a real survey l.o.s. density, for example we know that BOSS roughly detects 15 quasars' spectra per deg^2 [22], but this would make a very low number of skewers in a typical hydrodynamical simulation box and let arise other computational problems (see next subsection B.2).

It is also possible to take the $nlos$ necessary to have $\langle F \rangle_{\mathrm{cat}} = \langle F \rangle_{\mathrm{obs}}$, i.e. to match the simulation flux mean with the measured one, so it would be naively expected that the simulation $P_{\mathrm{Ly}\alpha}(k)$ converges with the measured one. But this is not completely true as we now discuss.
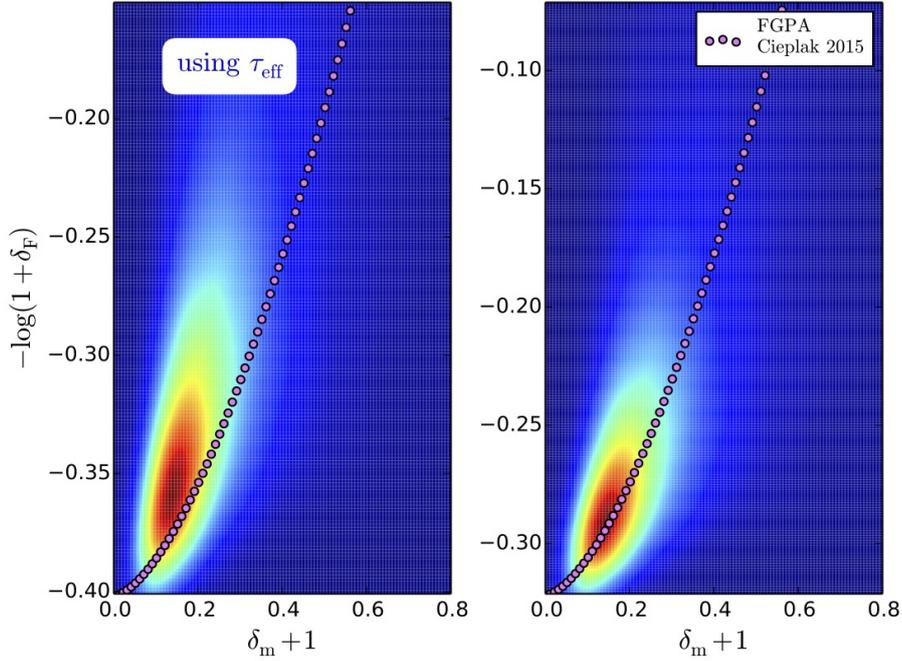
Figure 14: Density of points (red color for high density values, blue for small) for the scatter of $(\delta_{\mathrm{m}}+1)$ - $(-\log(1+\delta_{\mathrm{F}}))$ calculated for the 60-512 simulation pierced by 100 random lines of sight with 256 pixels each, as described in the subsection B.1. The magenta dotted line is the fluctuating Gunn-Peterson approximation (FGPA) relation of equation B.2 with the parameters in [52]. *Left panel:* The spectra have been normalised by using an effective optical depth τ_{eff}. *Right panel:* The spectra have not been normalised. By eye it is clear that the right panel (unnormalised spectra) has a much better agreement with FGPA.

In Fig. 13 we show $\langle F \rangle_{\mathrm{cat}}$ coming from catalogues with the following $nlos$: 10^2, 16^2, 27^2, 44^2, 72^2, 118^2, 194^2 and 316^2. The skewers have been placed either on a $\sqrt{nlos} \times \sqrt{nlos}$ regular grid (blue diamonds, each point corresponds to one catalogue) or randomly (red dot, each point correspond to the average among 50 different catalogues). In case of random l.o.s., we also plot the 1σ error bar on the average $\langle F \rangle_{\mathrm{cat}}$ and show with shaded area the 2σ and 3σ intervals too. The scatter is big for the catalogues with smaller $nlos$, nevertheless their average $\langle F \rangle_{\mathrm{cat}}$ agrees with the what we obtain with a huge amount of l.o.s. We thus conclude that the catalogue mean flux is not very sensitive to the value of the surface density with which the Lyα field is sampled through skewers, i.e. the quantity $\langle F \rangle_{\mathrm{cat}}$ is a property of the specific realization itself, marginally dependent on the choice for $nlos$.

Throughout the literature (e.g. [17, 37, 51, 53, 54]) the general practice consisted in artificially change the amplitude of the UV background strength to obtain the desired mean transmitted flux:

$$\langle e^{-B\tau_i} \rangle = \langle e^{-\tau_{\mathrm{eff}}} \rangle = \langle F \rangle_{\mathrm{obs}}, \qquad (\text{B.1})$$

where τ_i is the optical depth of a single pixel, B is a scalar representing the variation in the amplitude of the UV background and τ_{eff} is called effective optical depth (e.g. [51]).

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

To clarify this point, we make use of the fluctuating Gunn-Peterson approximation (FGPA, [55]) that relates the matter and transmitted flux fields through:

$$F = e^{-A(1+\delta_{\mathrm{m}})^\alpha}, \tag{B.2}$$

where δ_{m} is the matter density contrast and for the other parameters we adopt the values in [52]: $A = 0.3((1+z)/(1+2.4))^{4.5}$ and $\alpha = 1.6$. The FGPA, although neglecting some small scales physics, has been first derived analytically and has had great success in explaining statistical observed properties of the Lyα forest. We check whether FGPA holds for our mock spectra. We generate two catalogues composed by the same 100 random l.o.s. (i.e. BOSS resolution in the $60^2 \, (h^{-1}\mathrm{Mpc})^2$ simulation box) with 256 pixels each, but one of them is normalised to $\langle F \rangle_{\mathrm{obs}}$ using τ_{eff}. We interpolate the flux F pixels and matter particles back onto a 256^3 grid to obtained two 3D density maps. In Fig. 14 in color code we display the density of points in the Lyα flux $F -$ matter scatter plot and we plot with magenta dots the FGPA relation. The left panel refers to the normalised catalogue, the right panel to the unnormalised. We find that the flux renormalization shifts the whole flux field (see the different y scale on the two panels) and modifies the FGPA relation.

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

None of the spectra catalogues of this work has been normalized using τ_{eff}, at the expenses of having the tiny (and understood) large scale discrepancies in $P_{\mathrm{Ly}\alpha}(k)$ (see Fig. 5); still, the statistical properties of the Lyα forest flux have been remarkably recovered (see bias and redshift space distortion parameter determination in section 5.3).

B.2 Placing lines of sight on a regular grid

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

There are several solutions to this problem: 1) compute the 3D Lyα power spectrum masking out the regions not sampled by the mock spectra and 2) working in configuration-space (i.e. compute correlation functions), as in [56, 57]. Unfortunately, these solutions are computationally expensive and in the case of 1) technically hard to implement.

The standard practice is to place the skewers into a regular grid, this way the whole box is sampled uniformly, although real quasars are not on a grid. We check what is the effect of regularising the spectra position. In Fig. 15 we plot with a red dashed line the average of 50 Lyα power spectra, each corresponding to a catalogue with $256^2 = 65536$ l.o.s. randomly placed, and with a dotted blue line the $P_{\mathrm{Ly}\alpha}(k)$ of a catalogue of the same number of l.o.s. but placed on a 256×256 regular grid in the $y - z$ plane. The two power spectra converge at large scales, the sub-sampling effect discussed above becomes relevant on small scales.

We thus conclude that our method to compute the 3D Lyα power spectrum is robust, on large-scales, against the different ways of sampling the Lyα field.

C Gaussian errors derivation

Here we derive the equations governing the Gaussian errors of the multipoles of auto- and cross-power spectra.
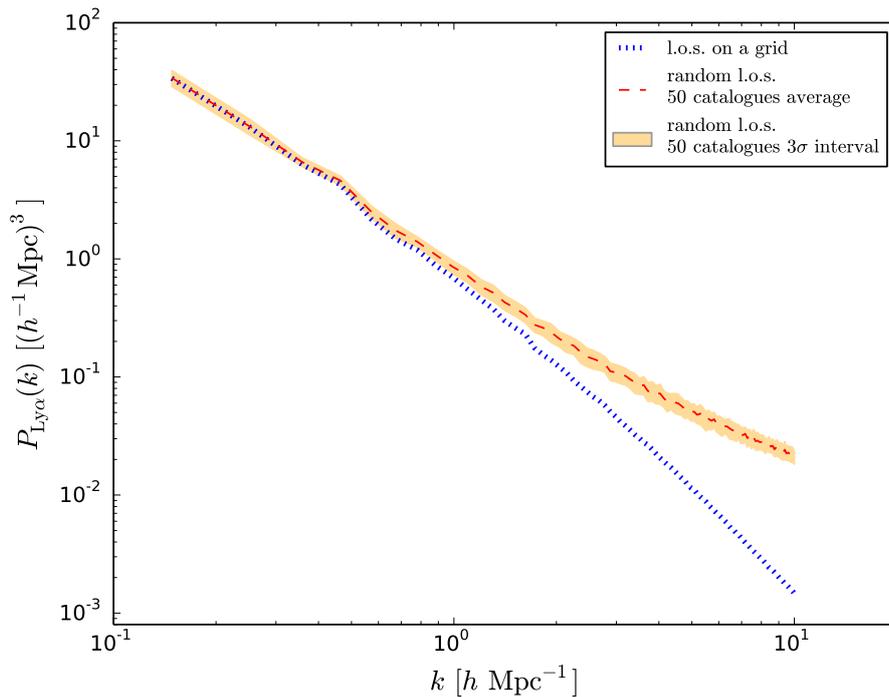
Figure 15: 3D Lyα flux power spectra $P_{\mathrm{Ly}\alpha}(k)$ at $z = 3$ obtained by placing the skewers into a regular grid (dotted blue) or distributing them randomly (dashed red). For the regular grid we employ 256^2 spectra while for the randomly distributed skewers we create 50 independent catalogues, each of them containing 256^2 spectra, and we show the mean of them and the 3σ variation.

Fitting simultaneously monopoles and quadrupoles of the auto-power spectra. If we have measurements of the monopoles and quadrupoles of the auto-power spectra of the two fields that occupy the same volume and we want to fit them simultaneously we need a theoretical model and an estimation of the errors. The theoretical model is given by the Legendre expansion of Eq. 5.10, that translates to Eqs. 5.11-5.12 for the two fields we are considering (dubbed with α and β subscripts in

what follows). We derive the associated covariance:

$$\sigma^2\left(P_{\ell_1,\alpha}(k_i), P_{\ell_2,\beta}(k_j)\right) = \left\langle \left(\hat{P}_{\ell_1,\alpha}(k_i) - P_{\ell_1,\alpha}(k_i)\right)\left(\hat{P}_{\ell_2,\beta}(k_j) - P_{\ell_2,\beta}(k_j)\right)\right\rangle$$

$$= \langle \hat{P}_{\ell_1,\alpha}(k_i)\hat{P}_{\ell_2,\beta}(k_j)\rangle - P_{\ell_1,\alpha}(k_i)P_{\ell_2,\beta}(k_j)$$

$$= \frac{(2\ell_1+2)(2\ell_2+1)}{N_k^2}\sum_{l=1}^{N_k}\sum_{m=1}^{N_k}\langle \delta_\alpha(k_{i,l})\delta_\alpha^*(k_{i,l})\delta_\beta(k_{j,m})\delta_\beta^*(k_{j,m})\rangle L_{\ell_1}(\mu_l)L_{\ell_2}(\mu_m)$$

$$-P_{\ell_1,\alpha}(k_i)P_{\ell_2,\beta}(k_j)$$

$$= \frac{(2\ell_1+2)(2\ell_2+1)}{N_k^2}\sum_{l=1}^{N_k}\sum_{m=1}^{N_k}\langle \delta_\alpha(k_{i,l})\delta_\beta^*(k_{j,m})\rangle\langle \delta_\alpha^*(k_{i,l})\delta_\beta(k_{j,m})\rangle L_{\ell_1}(\mu_l)L_{\ell_2}(\mu_m)$$

$$= \frac{(2\ell_1+2)(2\ell_2+1)}{N_k^2}\sum_{l=1}^{N_k}\sum_{m=1}^{N_k}P_{\alpha\beta}^2(k_{i,l})L_{\ell_1}(\mu_l)L_{\ell_2}(\mu_m)\delta_{l,m}\delta_{k_i,k_j}$$

$$= \frac{(2\ell_1+2)(2\ell_2+1)}{N_k^2}\sum_{l=1}^{N_k}P_{\alpha\beta}^2(k_{i,j})L_{\ell_1}(\mu_l)L_{\ell_2}(\mu_l)\delta_{k_i,k_j} \tag{C.1}$$

where N_k is the number of independent modes in the k-interval $[k, k+dk]$, i.e. $N_k = \frac{1}{2}\frac{4\pi k^2 dk}{k_F^3}$ where $k_F = 2\pi/L$ is the value of the fundamental frequency, with L being the size of the cubic volume and dk is the k-bin size, usually chosen as $dk = k_F$. We notice that in the previous expression we have taken into account that the imaginary part of the cross-power spectrum is 0[9], i.e.

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

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

$$\sigma^2(P_{\ell_1,\alpha}(k_1), P_{\ell_2,\beta}(k_2)) = \delta_{k_1,k_2}\frac{(2\ell_1+2)(2\ell_2+1)}{2N_k}\int_{-1}^1 P_{\alpha\beta}^2(k,\mu)L_{\ell_1}(\mu)L_{\ell_2}(\mu)d\mu \tag{C.3}$$

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

$$\sigma^2(P_{0,\alpha}(k_1), P_{0,\beta}(k_2)) = \gamma(k)\left[\frac{1}{9}\beta_\alpha^2\beta_\beta^2 + \frac{2}{7}(\beta_\alpha^2\beta_\beta + \beta_\alpha\beta_\beta^2) + \right.$$

$$\left. +\frac{1}{5}(\beta_\alpha^2 + 4\beta_\alpha\beta_\beta + \beta_\beta^2) + \frac{2}{3}(\beta_\alpha + \beta_\beta) + 1\right]\delta_{k_1,k_2}$$

$$\sigma^2(P_{2,\alpha}(k_1), P_{2,\beta}(k_2)) = 5\gamma(k)\left[\frac{415}{1287}\beta_\alpha^2\beta_\beta^2 + \frac{170}{231}(\beta_\alpha^2\beta_\beta + \beta_\alpha\beta_\beta^2) + \right.$$

$$\left. +\frac{3}{7}(\beta_\alpha^2 + 4\beta_\alpha\beta_\beta + \beta_\beta^2) + \frac{22}{21}(\beta_\alpha + \beta_\alpha) + 1\right]\delta_{k_1,k_2}$$

$$\sigma^2(P_{0,\alpha}(k_1), P_{2,\beta}(k_2)) = \tfrac{4}{693}\gamma(k)\left[70\beta_\alpha^2\beta_\beta^2 + 165(\beta_\alpha^2\beta_\beta + \beta_\alpha\beta_\beta^2) + \right.$$

$$\left. +99(\beta_\alpha^2 + 4\beta_\alpha\beta_\beta + \beta_\beta^2) + 231(\beta_\alpha + \beta_\beta)\right]\delta_{k_1,k_2} \tag{C.4}$$

---

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

where

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

Fitting simultaneously auto- and cross-power spectra. If we have measurements of monopoles and quadrupoles of the two auto-power spectra and measurements of monopole and quadrupole of their cross-power spectrum and we want to fit all six functions together, we need again a theoretical model and an estimation of the errors. The theoretical model is given again by the expansion in Legendre polinomial as in Eqs. 5.11-5.12 and 5.13-5.14. It follows that the covariance will be given by

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

and taken into account that $P_\alpha(k,\mu) = b_\alpha^2 (1 + \beta_\alpha \mu^2)^2 P_{\rm m}(k)$ and $P_{12}(k,\mu) = b_1 b_2 (1 + \beta_1 \mu^2)(1 + \beta_2 \mu^2) P_{\rm m}(k)$ we get

$$\sigma^2(P_{0,1}(k_1), P_{0,12}(k_2)) = \gamma(k) \left[\frac{1}{9}\beta_1^3 \beta_2 + \frac{1}{7}(\beta_1^3 + 3\beta_1^2 \beta_2) + \frac{1}{5}(3\beta_1^2 + 3\beta_1 \beta_2) + \frac{1}{3}(3\beta_1 + \beta_2) + 1 \right] \delta_{k_1,k_2}$$

$$\sigma^2(P_{2,1}(k_1), P_{2,12}(k_2)) = 5\gamma(k) \left[\frac{415}{1287}\beta_1^3 \beta_2 + \frac{85}{231}(\beta_1^3 + 3\beta_1^2 \beta_2) + \frac{9}{7}(\beta_1^2 + \beta_1 \beta_2) + \frac{11}{21}(3\beta_1 + \beta_2) + 1 \right] \delta_{k_1,k_2}$$

$$\sigma^2(P_{0,1}(k_1), P_{2,12}(k_2)) = \frac{2}{693}\gamma(k) \left[140\beta_1^3 \beta_2 + 165(\beta_1^3 + 3\beta_1^2 \beta_2) + 594(\beta_1^2 + \beta_1 \beta_2) + 231(3\beta_1 + \beta_2) \right] \delta_{k_1,k_2}$$

where

$$\gamma(k) = \frac{b_1^3 b_2 P_m^2(k)}{N_k} \,. \tag{C.7}$$

References

[1] S. Bharadwaj, B. B. Nath, and S. K. Sethi, *Using HI to Probe Large Scale Structures at z ~ 3*, Journal of Astrophysics and Astronomy **22** (Mar., 2001) 21, [astro-ph/0003200].

[2] S. Bharadwaj and S. K. Sethi, *HI Fluctuations at Large Redshifts: I–Visibility correlation*, Journal of Astrophysics and Astronomy **22** (Dec., 2001) 293–307, [astro-ph/0203269].

[3] R. A. Battye, R. D. Davies, and J. Weller, *Neutral hydrogen surveys for high redshift galaxy clusters and proto-clusters*, Mon.Not.Roy.Astron.Soc. **355** (2004) 1339–1347, [astro-ph/0401340].

[4] M. McQuinn, O. Zahn, M. Zaldarriaga, L. Hernquist, and S. R. Furlanetto, *Cosmological Parameter Estimation Using 21 cm Radiation from the Epoch of Reionization*, Astrophysical Journal **653** (Dec., 2006) 815–834, [astro-ph/0512263].

[5] T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald, *Baryon Acoustic Oscillation Intensity Mapping of Dark Energy*, Physical Review Letters **100** (Mar., 2008) 091303, [arXiv:0709.3672].

[6] 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*, Physical Review Letters **100** (Apr., 2008) 161301, [arXiv:0801.1677].

[7] P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos, *Late-time Cosmology with 21 cm Intensity Mapping Experiments*, Astrophysical Journal **803** (Apr., 2015) 21, [arXiv:1405.1452].

[8] L. Wolz, F. B. Abdalla, D. Alonso, C. Blake, P. Bull, T. C. Chang, P. Ferreira, C. Y. Kuo, M. Santos, and J. R. Shaw, *Foreground Subtraction in Intensity Mapping with the SKA*, Advancing Astrophysics with the Square Kilometre Array (AASKA14) (Apr., 2015) 35, [arXiv:1501.0382].

[9] D. Alonso, P. Bull, P. G. Ferreira, and M. G. Santos, *Blind foreground subtraction for intensity mapping experiments*, *Montly Notices of the Royal Astronomical Society* **447** (Feb., 2015) 400–416, [arXiv:1409.8667].

[10] F. Villaescusa-Navarro, M. Viel, D. Alonso, K. K. Datta, P. Bull, and M. G. Santos, *Cross-correlating 21cm intensity maps with Lyman Break Galaxies in the post-reionization era*, *Journal of Cosmology and Astroparticle Physics* **3** (Mar., 2015) 034, [arXiv:1410.7393].

[11] E. Sobacchi, A. Mesinger, and B. Greig, *Cross-correlation of the cosmic 21-cm signal and Lyman α emitters during reionization*, *Montly Notices of the Royal Astronomical Society* **459** (July, 2016) 2741–2750, [arXiv:1602.0483].

[12] L. Wolz, C. Tonini, C. Blake, and J. S. B. Wyithe, *Intensity mapping cross-correlations: connecting the largest scales to galaxy evolution*, *Montly Notices of the Royal Astronomical Society* **458** (May, 2016) 3399–3410, [arXiv:1512.0418].

[13] A. Lidz, O. Zahn, S. R. Furlanetto, M. McQuinn, L. Hernquist, and M. Zaldarriaga, *Probing Reionization with the 21 cm Galaxy Cross-Power Spectrum*, *Astrophysical Journal* **690** (Jan., 2009) 252–266, [arXiv:0806.1055].

[14] T. Guha Sarkar, S. Bharadwaj, T. R. Choudhury, and K. K. Datta, *Cross-correlation of the H I 21-cm signal and Lyα forest: a probe of cosmology*, *Montly Notices of the Royal Astronomical Society* **410** (Jan., 2011) 1130–1134, [arXiv:1002.1368].

[15] T. Guha Sarkar and K. K. Datta, *On using large scale correlation of the Ly-α forest and redshifted 21-cm signal to probe HI distribution during the post reionization era*, *Journal of Cosmology and Astroparticle Physics* **8** (Aug., 2015) 001, [arXiv:1501.0230].

[16] T. Guha Sarkar, K. K. Datta, A. K. Pal, T. R. Choudhury, and S. Bharadwaj, *The redshifted HI 21 cm signal from the post-reionization epoch: Cross-correlations with other cosmological probes*, *ArXiv e-prints* (Oct., 2016) [arXiv:1610.0818].

[17] J. S. Bolton, E. Puchwein, D. Sijacki, M. G. Haehnelt, T.-S. Kim, A. Meiksin, J. A. Regan, and M. Viel, *The Sherwood simulation suite: overview and data comparisons with the Lyman-alpha forest at redshifts 2 $<z<$ 5*, *ArXiv e-prints* (May, 2016) [arXiv:1605.0346].

[18] V. Springel, *The cosmological simulation code GADGET-2*, *Montly Notices of the Royal Astronomical Society* **364** (Dec., 2005) 1105–1134, [astro-ph/].

[19] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al., *Planck 2015 results. XIII. Cosmological parameters*, *ArXiv e-prints* (Feb., 2015) [arXiv:1502.0158].

[20] M. Viel, G. D. Becker, J. S. Bolton, and M. G. Haehnelt, *Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data*, *Physical Review D* **88** (Aug., 2013) 043502, [arXiv:1306.2314].

[21] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, *The evolution of large-scale structure in a universe dominated by cold dark matter*, *Astrophysical Journal* **292** (May, 1985) 371–394.

[22] A. Slosar, A. Font-Ribera, M. M. Pieri, J. Rich, J.-M. Le Goff, É. Aubourg, J. Brinkmann, N. Busca, B. Carithers, R. Charlassier, M. Cortês, R. Croft, K. S. Dawson, D. Eisenstein, J.-C. Hamilton, S. Ho, K.-G. Lee, R. Lupton, P. McDonald, B. Medolin, D. Muna, J. Miralda-Escudé, A. D. Myers, R. C. Nichol, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, Y. Piškur, E. Rollinde, N. P. Ross, D. J. Schlegel, D. P. Schneider, E. Sheldon, B. A. Weaver, D. H. Weinberg, C. Yeche, and D. G. York, *The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data*, *Journal of Cosmology and Astroparticle Physics* **9** (Sept., 2011) 001, [arXiv:1104.5244].

[23] R. Braun, T. Bourke, J. A. Green, E. Keane, and J. Wagg, *Advancing Astrophysics with the Square Kilometre Array*, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)* (Apr., 2015) 174.

[24] R. K. Sheth and G. Tormen, *An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier*, *Montly Notices of the Royal Astronomical Society* **329** (Jan., 2002) 61–75, [astro-ph/0105113].

[25] F. Villaescusa-Navarro, M. Viel, K. K. Datta, and T. R. Choudhury, *Modeling the neutral hydrogen distribution in the post-reionization Universe: intensity mapping*, *Journal of Cosmology and Astroparticle Physics* **9** (Sept., 2014) 050, [arXiv:1405.6713].

[26] I. P. Carucci, F. Villaescusa-Navarro, M. Viel, and A. Lapi, *Warm dark matter signatures on the 21cm power spectrum: intensity mapping forecasts for SKA*, *Journal of Cosmology and Astroparticle Physics* **7** (July, 2015) 047, [arXiv:1502.0696].

[27] F. Villaescusa-Navarro, P. Bull, and M. Viel, *Weighing Neutrinos with Cosmic Neutral Hydrogen*, *Astrophysical Journal* **814** (Dec., 2015) 146, [arXiv:1507.0510].

[28] F. Villaescusa-Navarro, S. Planelles, S. Borgani, M. Viel, E. Rasia, G. Murante, K. Dolag, L. K. Steinborn, V. Biffi, A. M. Beck, and C. Ragone-Figueroa, *Neutral hydrogen in galaxy clusters: impact of AGN feedback and implications for intensity mapping*, *Montly Notices of the Royal Astronomical Society* **456** (Mar., 2016) 3553–3570, [arXiv:1510.0427].

[29] E. Castorina and F. Villaescusa-Navarro, *On the spatial distribution of neutral hydrogen in the Universe: bias and shot-noise of the HI Power Spectrum*, *ArXiv e-prints* (Sept., 2016) [arXiv:1609.0515].

[30] R. Sánchez-Ramírez, S. L. Ellison, J. X. Prochaska, T. A. M. Berg, S. López, V. D'Odorico, G. D. Becker, L. Christensen, G. Cupani, K. D. Denney, I. Pâris, G. Worseck, and J. Gorosabel, *The evolution of neutral gas in damped Lyman α systems from the XQ-100 survey*, *Montly Notices of the Royal Astronomical Society* **456** (Mar., 2016) 4488–4505, [arXiv:1511.0500].

[31] H. Padmanabhan, A. Refregier, and A. Amara, *A halo model for cosmological neutral hydrogen : abundances and clustering*, *ArXiv e-prints* (Nov., 2016) [arXiv:1611.0623].

[32] R. K. Sheth and G. Tormen, *Large-scale bias and the peak background split*, *Montly Notices of the Royal Astronomical Society* **308** (Sept., 1999) 119–126, [astro-ph/9901122].

[33] R. K. Sheth, H. J. Mo, and G. Tormen, *Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes*, *Monthly Notices of the Royal Astronomical Society* **323** (May, 2001) 1–12, [astro-ph/9907024].

[34] A. Font-Ribera, J. Miralda-Escudé, E. Arnau, B. Carithers, K.-G. Lee, P. Noterdaeme, I. Pâris, P. Petitjean, J. Rich, E. Rollinde, N. P. Ross, D. P. Schneider, M. White, and D. G. York, *The large-scale cross-correlation of Damped Lyman alpha systems with the Lyman alpha forest: first measurements from BOSS*, *Journal of Cosmology and Astroparticle Physics* **11** (Nov., 2012) 059, [arXiv:1209.4596].

[35] M. Santos, P. Bull, D. Alonso, S. Camera, P. Ferreira, G. Bernardi, R. Maartens, M. Viel, F. Villaescusa-Navarro, F. B. Abdalla, M. Jarvis, R. B. Metcalf, A. Pourtsidou, and L. Wolz, *Cosmology from a SKA HI intensity mapping survey*, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)* (Apr., 2015) 19, [arXiv:1501.0398].

[36] Y. Mao, P. R. Shapiro, G. Mellema, I. T. Iliev, J. Koda, and K. Ahn, *Redshift-space distortion of the 21-cm background from the epoch of reionization - I. Methodology re-examined*, *Montly Notices of the Royal Astronomical Society* **422** (May, 2012) 926–954, [arXiv:1104.2094].

[37] T. Theuns, A. Leonard, G. Efstathiou, F. R. Pearce, and P. A. Thomas, *P^3M-SPH simulations of the Lyalpha forest*, *Montly Notices of the Royal Astronomical Society* **301** (Dec., 1998) 478–502, [astro-ph/9805119].

[38] T. Tepper-García, *Voigt profile fitting to quasar absorption lines: an analytic approximation to the Voigt-Hjerting function*, *Montly Notices of the Royal Astronomical Society* **369** (July, 2006) 2025–2035, [astro-ph/0602124].

[39] S. Mukae, M. Ouchi, K. Kakiichi, N. Suzuki, Y. Ono, Z. Cai, A. K. Inoue, Y.-K. Chiang, T. Shibuya,

and Y. Matsuda, *Cosmic Galaxy-IGM HI Relation at $z \sim 2 - 3$ Probed in the COSMOS/UltraVISTA* 1.6 *deg^2 Field*, *ArXiv e-prints* (May, 2016) [`arXiv:1605.0037`].

[40] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, A. Beifiori, A. A. Berlind, V. Bhardwaj, D. Bizyaev, C. H. Blake, M. R. Blanton, M. Blomqvist, A. S. Bolton, A. Borde, J. Bovy, W. N. Brandt, H. Brewington, J. Brinkmann, P. J. Brown, J. R. Brownstein, K. Bundy, N. G. Busca, W. Carithers, A. R. Carnero, M. A. Carr, Y. Chen, J. Comparat, N. Connolly, F. Cope, R. A. C. Croft, A. J. Cuesta, L. N. da Costa, J. R. A. Davenport, T. Delubac, R. de Putter, S. Dhital, A. Ealet, G. L. Ebelke, D. J. Eisenstein, S. Escoffier, X. Fan, N. Filiz Ak, H. Finley, A. Font-Ribera, R. Génova-Santos, J. E. Gunn, H. Guo, D. Haggard, P. B. Hall, J.-C. Hamilton, B. Harris, D. W. Harris, S. Ho, D. W. Hogg, D. Holder, K. Honscheid, J. Huehnerhoff, B. Jordan, W. P. Jordan, G. Kauffmann, E. A. Kazin, D. Kirkby, M. A. Klaene, J.-P. Kneib, J.-M. Le Goff, K.-G. Lee, D. C. Long, C. P. Loomis, B. Lundgren, R. H. Lupton, M. A. G. Maia, M. Makler, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, D. Margala, K. L. Masters, C. K. McBride, P. McDonald, I. D. McGreer, R. G. McMahon, O. Mena, J. Miralda-Escudé, A. D. Montero-Dorta, F. Montesano, D. Muna, A. D. Myers, T. Naugle, R. C. Nichol, P. Noterdaeme, S. E. Nuza, M. D. Olmstead, A. Oravetz, D. J. Oravetz, R. Owen, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. K. Parejko, I. Pâris, W. J. Percival, I. Pérez-Fournon, I. Pérez-Ràfols, P. Petitjean, R. Pfaffenberger, J. Pforr, M. M. Pieri, F. Prada, A. M. Price-Whelan, M. J. Raddick, R. Rebolo, J. Rich, G. T. Richards, C. M. Rockosi, N. A. Roe, A. J. Ross, N. P. Ross, G. Rossi, J. A. Rubiño-Martin, L. Samushia, A. G. Sánchez, C. Sayres, S. J. Schmidt, D. P. Schneider, C. G. Scóccola, H.-J. Seo, A. Shelden, E. Sheldon, Y. Shen, Y. Shu, A. Slosar, S. A. Smee, S. A. Snedden, F. Stauffer, O. Steele, M. A. Strauss, A. Streblyanska, N. Suzuki, M. E. C. Swanson, T. Tal, M. Tanaka, D. Thomas, J. L. Tinker, R. Tojeiro, C. A. Tremonti, M. Vargas Magaña, L. Verde, M. Viel, D. A. Wake, M. Watson, B. A. Weaver, D. H. Weinberg, B. J. Weiner, A. A. West, M. White, W. M. Wood-Vasey, C. Yeche, I. Zehavi, G.-B. Zhao, and Z. Zheng, *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *The Astronomical Journal* **145** (Jan., 2013) 10, [`arXiv:1208.0022`].

[41] A. Font-Ribera, D. Kirkby, N. Busca, J. Miralda-Escudé, N. P. Ross, A. Slosar, J. Rich, É. Aubourg, S. Bailey, V. Bhardwaj, J. Bautista, F. Beutler, D. Bizyaev, M. Blomqvist, H. Brewington, J. Brinkmann, J. R. Brownstein, B. Carithers, K. S. Dawson, T. Delubac, G. Ebelke, D. J. Eisenstein, J. Ge, K. Kinemuchi, K.-G. Lee, V. Malanushenko, E. Malanushenko, M. Marchante, D. Margala, D. Muna, A. D. Myers, P. Noterdaeme, D. Oravetz, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, M. M. Pieri, G. Rossi, D. P. Schneider, A. Simmons, M. Viel, C. Yeche, and D. G. York, *Quasar-Lyman α forest cross-correlation from BOSS DR11: Baryon Acoustic Oscillations*, *Journal of Cosmology and Astroparticle Physics* **5** (May, 2014) 027, [`arXiv:1311.1767`].

[42] M. McQuinn and M. White, *On estimating Lyα forest correlations between multiple sightlines*, *Montly Notices of the Royal Astronomical Society* **415** (Aug., 2011) 2257–2269, [`arXiv:1102.1752`].

[43] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *emcee: The MCMC Hammer*, *Publications of the Astronomical Society of the Pacific* **125** (Mar., 2013) 306–312, [`arXiv:1202.3665`].

[44] M. Blomqvist, D. Kirkby, J. E. Bautista, A. Arinyo-i-Prats, N. G. Busca, J. Miralda-Escudé, A. Slosar, A. Font-Ribera, D. Margala, D. P. Schneider, and J. A. Vazquez, *Broadband distortion modeling in Lyman-α forest BAO fitting*, *Journal of Cosmology and Astroparticle Physics* **11** (Nov., 2015) 034, [`arXiv:1504.0665`].

[45] E. R. Switzer, K. W. Masui, K. Bandura, L.-M. Calin, T.-C. Chang, X.-L. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek, *Determination of $z \sim 0.8$ neutral hydrogen fluctuations using the 21 cm intensity mapping autocorrelation*, *Montly Notices of the Royal Astronomical Society* **434** (July, 2013) L46–L50, [`arXiv:1304.3712`].

[46] R. Adhikari, M. Agostini, N. A. Ky, T. Araki, M. Archidiacono, M. Bahr, J. Behrens, F. Bezrukov, P. S. Bhupal Dev, D. Borah, A. Boyarsky, A. de Gouvea, C. A. d. S. Pires, H. J. de Vega, A. G. Dias, P. Di Bari, Z. Djurcic, K. Dolde, H. Dorrer, M. Durero, O. Dragoun, M. Drewes, C. E. Düllmann,

K. Eberhardt, S. Eliseev, C. Enss, N. W. Evans, A. Faessler, P. Filianin, V. Fischer, A. Fleischmann, J. A. Formaggio, J. Franse, F. M. Fraenkle, C. S. Frenk, G. Fuller, L. Gastaldo, A. Garzilli, C. Giunti, F. Glück, M. C. Goodman, M. C. Gonzalez-Garcia, D. Gorbunov, J. Hamann, S. Hannestad, J. Heeck, V. Hannen, S. H. Hansen, C. Hassel, F. Hofmann, T. Houdy, A. Huber, D. Iakubovskyi, A. Ianni, A. Ibarra, R. Jacobsson, T. Jeltema, S. Kempf, T. Kieck, M. Korzeczek, V. Kornoukhov, T. Lachenmaier, M. Laine, P. Langacker, T. Lasserre, J. Lesgourgues, D. Lhuillier, Y. F. Li, W. Liao, A. W. Long, M. Maltoni, G. Mangano, N. E. Mavromatos, N. Menci, A. Merle, S. Mertens, A. Mirizzi, B. Monreal, A. Nozik, A. Neronov, V. Niro, Y. Novikov, L. Oberauer, E. Otten, N. Palanque-Delabrouille, M. Pallavicini, V. S. Pantuev, E. Papastergis, S. Parke, S. Pastor, A. Patwardhan, A. Pilaftsis, D. C. Radford, P. C.-O. Ranitzsch, O. Rest, D. J. Robinson, P. S. Rodrigues da Silva, O. Ruchayskiy, N. G. Sanchez, M. Sasaki, N. Saviano, A. Schneider, F. Schneider, T. Schwetz, S. Schönert, F. Shankar, N. Steinbrink, L. Strigari, F. Suekane, B. Suerfu, R. Takahashi, N. T. H. Van, I. Tkachev, M. Totzauer, Y. Tsai, C. G. Tully, K. Valerius, J. Valle, D. Venos, M. Viel, M. Y. Wang, C. Weinheimer, K. Wendt, L. Winslow, J. Wolf, M. Wurm, Z. Xing, S. Zhou, and K. Zuber, *A White Paper on keV Sterile Neutrino Dark Matter*, *ArXiv e-prints* (Feb., 2016) [`arXiv:1602.0481`].

[47] D. Coe, *Fisher Matrices and Confidence Ellipses: A Quick-Start Guide and Software*, *ArXiv e-prints* (June, 2009) [`arXiv:0906.4123`].

[48] V. Springel and L. Hernquist, *Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation*, *Montly Notices of the Royal Astronomical Society* **339** (Feb., 2003) 289–311, [`astro-ph/0206393`].

[49] T.-S. Kim, J. S. Bolton, M. Viel, M. G. Haehnelt, and R. F. Carswell, *An improved measurement of the flux distribution of the Lyα forest in QSO absorption spectra: the effect of continuum fitting, metal contamination and noise properties*, *Montly Notices of the Royal Astronomical Society* **382** (Dec., 2007) 1657–1674, [`arXiv:0711.1862`].

[50] C.-A. Faucher-Giguère, J. X. Prochaska, A. Lidz, L. Hernquist, and M. Zaldarriaga, *A Direct Precision Measurement of the Intergalactic Lyα Opacity at* $2 < z < 4.2$, *Astrophysical Journal* **681** (July, 2008) 831–855, [`arXiv:0709.2382`].

[51] G. D. Becker, P. C. Hewett, G. Worseck, and J. X. Prochaska, *A refined measurement of the mean transmitted flux in the Lyα forest over* $2 < z < 5$ *using composite quasar spectra*, *Montly Notices of the Royal Astronomical Society* **430** (Apr., 2013) 2067–2081, [`arXiv:1208.2584`].

[52] A. M. Cieplak and A. Slosar, *Towards physics responsible for large-scale Lyman-α forest bias parameters*, *Journal of Cosmology and Astroparticle Physics* **3** (Mar., 2016) 016, [`arXiv:1509.0787`].

[53] M. White, A. Pope, J. Carlson, K. Heitmann, S. Habib, P. Fasel, D. Daniel, and Z. Lukic, *Particle Mesh Simulations of the Lyα Forest and the Signature of Baryon Acoustic Oscillations in the Intergalactic Medium*, *Astrophysical Journal* **713** (Apr., 2010) 383–393, [`arXiv:0911.5341`].

[54] Z. Lukić, C. W. Stark, P. Nugent, M. White, A. A. Meiksin, and A. Almgren, *The Lyman α forest in optically thin hydrodynamical simulations*, *Montly Notices of the Royal Astronomical Society* **446** (Feb., 2015) 3697–3724, [`arXiv:1406.6361`].

[55] R. A. C. Croft, D. H. Weinberg, N. Katz, and L. Hernquist, *Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyα Forest*, *Astrophysical Journal* **495** (Mar., 1998) 44–62, [`astro-ph/9708018`].

[56] A. Slosar, S. Ho, M. White, and T. Louis, *The acoustic peak in the Lyman alpha forest*, *Journal of Cosmology and Astroparticle Physics* **10** (Oct., 2009) 019, [`arXiv:0906.2414`].

[57] T. Delubac, J. E. Bautista, N. G. Busca, J. Rich, D. Kirkby, S. Bailey, A. Font-Ribera, A. Slosar, K.-G. Lee, M. M. Pieri, J.-C. Hamilton, É. Aubourg, M. Blomqvist, J. Bovy, J. Brinkmann, W. Carithers, K. S. Dawson, D. J. Eisenstein, S. G. A. Gontcho, J.-P. Kneib, J.-M. Le Goff, D. Margala, J. Miralda-Escudé, A. D. Myers, R. C. Nichol, P. Noterdaeme, R. O'Connell, M. D. Olmstead, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, N. P. Ross, G. Rossi, D. J. Schlegel, D. P. Schneider, D. H. Weinberg, C. Yèche,

and D. G. York, *Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars*, *Astronomy and Astrophysics* **574** (Feb., 2015) A59, [`arXiv:1404.1801`].