. SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

SISSA Digital Library

Optimal galaxy survey for detecting the dipole in the cross-correlation with 21 cm Intensity Mapping

Original
Optimal galaxy survey for detecting the dipole in the cross-correlation with 21 cm Intensity Mapping / Lepori, F., Dio, E.D., Villa, E., Viel, M.. - In: JOURNAL OF COSMOLOGY AND ASTROPARTICLE PHYSICS. - ISSN 1475-7516. - 2018:5(2018), pp. 1-35. [10.1088/1475-7516/2018/05/043]

Availability:
This version is available at: 20.500.11767/85526 since: 2018-12-18T10:32:38Z

Publisher:

Published
DOI:10.1088/1475-7516/2018/05/043

Terms of use:

Testo definito dall'ateneo relativo alle clausole di concessione d'uso

Publisher copyright
IOP- Institute of Physics

This version is available for education and non-commercial purposes.

note finali coverpage

(Article begins on next page)

13 October 2026

Prepared for submission to JCAP

Optimal galaxy survey for detecting the dipole in the cross-correlation with 21 cm Intensity Mapping

Francesca Lepori,[a,b] **Enea Di Dio,**[c,a,b] **Eleonora Villa,**[a,b] **Matteo Viel**[a,b,c]

[a]SISSA- International School for Advanced Studies,
Via Bonomea 265, 34136 Trieste, Italy
[b]INFN, Sezione di Trieste,
Via Valerio 2, I-34127 Trieste, Italy
[c]INAF - Osservatorio Astronomico di Trieste,
Via G. B. Tiepolo 11, I-34143 Trieste, Italy

E-mail: flepori@sissa.it, enea.didio@oats.inaf.it, evilla@sissa.it, viel@sissa.it

Abstract. We investigate the future perspectives of the detection of the relativistic dipole by cross-correlating the 21 cm emission in Intensity Mapping (IM) and galaxy surveys at low redshift. We model the neutral hydrogen (HI) and the galaxy population by means of the halo model to relate the parameters that affect the dipole signal such as the biases of the two tracers and the Poissonian noise. We investigate the behavior of the signal-to-noise as a function of the galaxy and magnification biases, for two fixed models of the neutral hydrogen. In both cases we found that the signal-to-noise does not grow by increasing the difference between the biases of the two tracers, due to the larger shot-noise yields by highly biased tracers. We also study and provide an optimal luminosity-threshold galaxy catalogue to enhance the signal-to-noise ratio of the relativistic dipole. Interestingly, we show that the maximum magnitude provided by the survey does not lead to the maximum signal-to-noise for detecting relativistic effects and we predict the optimal value for the limiting magnitude. Our work suggests that an optimal analysis could increase the signal-to-noise ratio up to a factor five compared to a standard one.

Contents

1 Introduction **1**

2 Cross-correlation odd multipoles **3**
- 2.1 Galaxy number counts 3
- 2.2 21 cm brightness temperature fluctuation 5
- 2.3 21 cm - Galaxies cross-correlation 5
- 2.4 Contaminations to the relativistic dipole 9
- 2.5 Covariance for the cross-correlation dipole 11

3 Halo Model approach **12**
- 3.1 HI model 13
- 3.2 Galaxy model 16

4 Signal-to-noise analysis **18**

5 HOD approach for modelling Luminosity-threshold galaxy catalogues **21**

6 Conclusions **23**

A Flat-sky versus full-sky dipole **24**

B Contribution to the wide-angle correction to the covariance **28**

1 Introduction

The state-of-the-art analysis of the Cosmic Microwave Background (CMB) anisotropies [1] and of the Large Scale Structure (LSS) of the universe [2] offers robust observational evidence in favor of the standard cosmological model, known as ΛCDM. However, the ΛCDM model presents some theoretical troubles, above all the cosmological constant problem [3] and the nature of the dark components in the matter sector [4, 5]. These issues keep raising an increasing interest in the scientific community and strongly necessitate further investigation of the foundational principles from which the standard model was built [6]. A pivotal role in testing the laws of gravity at cosmological scales will be played by the forthcoming LSS surveys aiming to map with stunning precision the large scale structure of the universe. The huge amount of incoming data requires an equally powerful advance from the theoretical and modelling side. One of the issues related to the correct modelling of the LSS observables is the fact that we do not observe directly the dark matter density field, but we observe some biased tracers at a given observed redshift and direction in the sky. A fully relativistic computation of what is really measured in a galaxy survey was performed at linear order [7–14] and at second order in perturbation theory [15–24]. This formalism has been further extended to other observables such as the lensing convergence [25], the HI brightness temperature [26] and the Lyman-α transmitted flux [27]. Since we measure galaxy positions by collecting photons which have travelled in a clumpy universe, the LSS observables are also affected by other contributions, which were commonly neglected, beyond the standard local overdensity

and Redshift Space Distortion (RSD) [28]. These terms, which include gravitational lensing convergence, gravitational redshift, Doppler and integrated Sachs-Wolfe effect, may bias the estimation of some cosmological parameters [29–34], in particular for deviations from ΛCDM [35]. More interestingly, a measurement of these effects offers the opportunity to extract further cosmological information, allowing to test the Equivalence Principle or the theory of gravity at the largest observable scales. Even if all these corrective terms are subdominant on sub-horizon scales with respect to the local density and RSD, it has been shown that it is possible to isolate some of them, by correlating two different tracers[1] [27, 37–46]. This can be achieved using the fact that some terms carry an odd symmetry with respect to the line of sight, which can be exploited with different tracers. In fact, in the same way as RSD introduces an anisotropy that sources other even multipoles in the correlation function of a LSS tracer apart from the monopole, gravitational redshift and Doppler effects break the symmetry with respect to the exchange of two tracers along the line of sight. Therefore, they induce non-vanishing odd multipoles in the cross-correlation function or an imaginary part in the Fourier power spectrum. Measuring the relativistic effects is a new opportunity for testing the consistency of general relativity [47, 48] and may offer an alternative method to measure the peculiar velocity field of the sources [49, 50]. The first measurement of the dipole for the cross-correlation of two galaxies populations [43] was a measurement of the dipole induced by the so-called large-angle effect. Nevertheless, at the lowest order in the distant observer approximation, the large-angle effect can be written as a combination of the monopole and the quadrupole and therefore it does not provide new information, at least in the regime where this approximation is reliable. In this work we will investigate the possibility to detect the relativistic Doppler corrections by cross-correlating galaxies and 21 cm Intensity Mapping (IM). The prospect of measuring the relativistic dipole by cross-correlating these two tracers was addressed in [49], where a signal-to-noise analysis is presented for specific IM experiments and galaxy survey. Here we present a survey-independent analysis, with the aim of understanding which types of galaxies are more suitable for detecting the relativistic dipole and how the signal-to-noise can be optimized by properly choosing the luminosity threshold of the galaxy catalogue. Since the Doppler corrections are more relevant at low redshift, this is the regime we are focused on.

Cross-correlation observational studies between galaxies and IM at low redshift ($z < 1$) have already been performed in recent years with the goal of detecting the diffuse neutral hydrogen (HI) [51, 52] and also at high redshift exploiting the cross-correlation between IM and Lyman-break galaxies [53] and the Lyman-α forest [54] in the post-reionization era. Furthermore, higher redshift investigations of the cross correlations are also of primary importance during or around HI reionization (e.g. [55]).

The paper is structured as follows. In section 2 we revise the relativistic formalism for galaxy surveys and IM experiments. Furthermore, we summarize the formalism that will be used in the following sections to compute the multipoles of the cross-correlation for the two tracers and its covariance, and we discuss two possible contaminations to the relativistic dipole: the lensing dipole and the wide-angle effect. More details on the formalism and the approximation we adopt in this work can be found in the Appendices A and B. In section 3 we use the halo model to describe in a fairly general way the HI and galaxy distribution properties, and we show how the biases of the two tracers can be related to their shot-noise. In section

[1] It has also been claimed [36] that futuristic surveys may be able to detect them through a single tracer analysis.

4 we present a signal-to-noise analysis for the relativistic dipole. In particular, we investigate the behavior of the signal-to-noise as a function of the galaxy bias and magnification bias, for two HI models. In section 5 we model luminosity-threshold galaxy catalogues with an halo occupation distribution model and we studied the signal-to-noise for the relativistic Doppler dipole as a function of the limiting magnitude. Finally, in section 6 we sum up the results of our work and we draw the conclusions.

We stress that the rationale of this work is to provide a first quantitative investigation of the optimal strategy to detect relativistic effects by exploiting the cross-correlation signal between IM and galaxies, in doing that we will learn that the low redshift regime is important to have a high value of the signal-to-noise-ratio of the relativistic effects. This work is also motivated by the fact that there are indeed wide area low redshift surveys planned or under way that could provide the necessary data for the galaxy populations (e.g. the planned EMU [56]) or the WISE data set [57]) to be interfaced with IM data provided by radio telescopes like LOFAR [58], Murchison Wide-field Array [59], GMRT [60], the Ooty Radio Telescope[60], CHIME [61], ASKAP [62], MeerKAT [63] and SKA [64].

Throughout all the paper we assume a spatially flat ΛCDM cosmology with parameters $h = 0.67556$, $\Omega_{\rm cdm}h^2 = 0.12038$, $\Omega_{\rm b}h^2 = 0.022032$. The primordial amplitude and spectral index are $A_{\rm s} = 2.215 \times 10^{-9}$ and $n_{\rm s} = 0.9619$, respectively. The matter power spectrum was computed with the Cosmic Linear Anisotropy Solving System (CLASS) code [65, 66], with pivot scale $k = 0.05\,{\rm Mpc}^{-1}$.

2 Cross-correlation odd multipoles

In this section we will report the expression for the observable quantities in a galaxy redshift survey and in a 21 cm intensity mapping experiment and the corresponding cross-correlation. We consider a perturbed Friedmann-Lemaître-Robertson-Walker (FLRW) metric and we work in longitudinal gauge[2]

$$ds^2 = a(\eta)^2 \left(-\left(1+2\Psi\right) d\eta^2 + \left(1-2\Phi\right) d\mathbf{x}^2\right) , \tag{2.1}$$

where η denotes the conformal time, $a(\eta)$ is the scale factor, and the metric perturbations, Ψ and Φ, are the Bardeen potentials. We also remark that the equations summarized in this section do not assume General Relativity (GR).

2.1 Galaxy number counts

In a galaxy clustering experiment we measure the number of galaxies $N\left(\mathbf{n}, z\right)$ in terms of an angular direction $\mathbf{n}$,[3] and a measured redshift z. We can then define the galaxy number counts as

$$\Delta_{\rm gal}(\mathbf{n}, z) = \frac{N\left(\mathbf{n}, z\right) - \langle N\rangle\left(z\right)}{\langle N\rangle\left(z\right)}, \tag{2.2}$$

where $\langle..\rangle$ denotes the angular average at fixed observed redshift z. The galaxy number counts were computed to first order in perturbation theory [7–10], by accounting that galaxies are

[2] Since we consider only observable quantities, the gauge choice will not affect any result.

[3] $\mathbf{n}$ denotes here the unit vector (direction) in which the photons propagate, while the angular position in the celestial sphere is $-\mathbf{n}$.

a biased tracer of the underline dark matter field, and they can schematically be expressed[4] as [67]

$$\begin{aligned}\Delta_{\rm gal}(\mathbf{n},z,m^*) = {} & b_{\rm gal}(z,m^*)D + \frac{1}{\mathcal{H}(z)}\partial_r(\mathbf{V}\cdot\mathbf{n}) \\ & + (5s(m^*,z)-2)\int_0^{r(z)}\frac{r(z)-r}{2r(z)r}\Delta_\Omega(\Phi+\Psi)dr \\ & + \left(\frac{\mathcal{H}'}{\mathcal{H}^2} + \frac{2-5s(m^*,z)}{r\mathcal{H}} + 5s(m^*,z) - f^{\rm gal}_{\rm evo}(m^*,z)\right)(\mathbf{V}\cdot\mathbf{n}) \\ & + (f^{\rm gal}_{\rm evo}-3)\mathcal{H}V + (5s-2)\Phi + \Psi + \frac{1}{\mathcal{H}}\Phi' + \frac{2-5s}{r(z)}\int_0^{r(z)} dr(\Phi+\Psi) \\ & + \left(\frac{\mathcal{H}'}{\mathcal{H}^2} + \frac{2-5s}{r(z)\mathcal{H}} + 5s - f^{\rm gal}_{\rm evo}\right)\left(\Psi + \int_0^{r(z)} dr(\Phi'+\Psi')\right), \end{aligned} \tag{2.4}$$

where $b_{\rm gal}(z,m^*)$ is the galaxy bias of the sources whose magnitudes are smaller than the magnitude limit of the survey m^* [5], assumed to be linear and local; D is the dark matter density fluctuation in synchronous gauge; $\mathcal{H}=a'/a$ is the conformal Hubble factor; $\mathbf{V}$ is the peculiar velocity in longitudinal gauge; V the velocity potential defined by $\mathbf{V}=-\nabla V$; η_o is the present time and $r(z)=\eta_o-\eta$ is the conformal distance at redshift z; Δ_Ω represents the angular laplacian operator and a prime denotes a partial derivative with respect to the conformal time.

The bias factors in the expression above, $s(m^*,z)$ and $f_{\rm evo}(m^*,z)$, are the magnification and the evolution biases of the galaxy catalogue, respectively. The magnification bias s is the slope of the cumulative luminosity function [67, 68]

$$s(m^*,z) = \frac{\partial\log_{10}\bar{N}(z,m<m^*)}{\partial m^*}, \tag{2.5}$$

and $\bar{N}(z,m<m^*)$ denotes the cumulative luminosity function

$$\bar{N}(z,m<m^*) = \int_{\ln L*}^{\infty}\phi(\eta(z),\ln L)d\ln L, \tag{2.6}$$

where $\phi(\eta(z),\ln L)$ is the luminosity function and L^* is the luminosity threshold of the survey that is related to the flux threshold by $L^*=4\pi(1+z)^2r^2(z)F^*$. The evolution bias describes the departure from a sample of sources conserved in a comoving region

$$f^{\rm gal}_{\rm evo}(m^*,z) = \frac{\partial\ln\bar{N}(z,m<m^*)}{\mathcal{H}\partial\eta} = -(1+z)\frac{\partial\ln\bar{N}(z,m<m^*)}{\partial z}. \tag{2.7}$$

Modelling the magnification and the evolution biases of a galaxy catalogue requires a prior knowledge on the luminosity function of the targeted galaxy population.

[4] The expression for the number counts in (2.4) assumes that galaxies follow geodesics

$$\partial_r\Psi = \mathbf{V}'\cdot\mathbf{n} + \mathcal{H}\mathbf{V}\cdot\mathbf{n}. \tag{2.3}$$

[5] The magnitude limit of the survey is related to the flux limit through $m^* = -\frac{5}{2}\log_{10}\left[\frac{F^*}{F_0}\right]$, where F_0 is a reference value for the flux.

2.2 21 cm brightness temperature fluctuation

Intensity mapping [69] is a novel technique which aims to map the Large Scale Structure of the universe by measuring the collective emission of many galaxies without resolving individual sources. 21 cm IM experiments target the emission line of neutral atomic hydrogen. The observable quantity is the flux density, i.e. the integral of the specific intensity over the solid angle of the telescope beam, which can be related to the HI brightness temperature in the Rayleigh-Jeans regime (see the Appendix in [70] for a more detailed discussion). The observed fluctuation in the 21 cm brightness temperature has been computed in linear theory in Ref. [26], including all the relativistic corrections. The full expression is mathematically equivalent to the expression in (2.4), with the magnification bias value set to $s = 2/5$, such that the lensing contribution vanishes. The fact that the observable in an intensity mapping survey is not affected by gravitational lensing to linear order is due to surface brightness conservation. Indeed the change in the solid angle $d\Omega$ is exactly compensated by the change in the observed flux. The full expression for the observed fluctuation in the 21 cm brightness temperature is

$$\Delta_{21\mathrm{cm}}(\mathbf{n}, z) = b_{\mathrm{HI}}(z) D + \frac{1}{\mathcal{H}(z)} \partial_r (\mathbf{V}\cdot\mathbf{n}) + \left(\frac{\mathcal{H}'}{\mathcal{H}^2} + 2 - f^{\mathrm{HI}}_{\mathrm{evo}}(z) \right) (\mathbf{V}\cdot\mathbf{n}) + (f^{\mathrm{HI}}_{\mathrm{evo}} - 3)\mathcal{H} V$$
$$+ \Psi + \frac{1}{\mathcal{H}}\Phi' + \left(\frac{\mathcal{H}'}{\mathcal{H}^2} + 2 - f^{\mathrm{HI}}_{\mathrm{evo}} \right) \left(\Psi + \int_0^{r(z)} dr (\Phi' + \Psi') \right), \tag{2.8}$$

where $b_{\mathrm{HI}}(z)$ and $f^{\mathrm{HI}}_{\mathrm{evo}}(z)$ are the bias and the evolution bias of the neutral hydrogen, respectively.

The evolution bias for the HI can be defined similarly to the galaxy evolution bias in (2.6). Taking into account that we observe all the HI emissions from a patch of the sky, it depends on the redshift evolution of the HI comoving density $\bar{\rho}_{\mathrm{HI}}$ (see the Appendix in [68])

$$f^{\mathrm{HI}}_{\mathrm{evo}} = \frac{\partial \ln \bar{\rho}_{\mathrm{HI}}(z)}{\mathcal{H}\partial\eta} = -(1+z)\frac{\partial \ln \bar{\rho}_{\mathrm{HI}}(z)}{\partial z}. \tag{2.9}$$

2.3 21 cm - Galaxies cross-correlation

Cross-correlation studies are promising techniques to study relativistic effects. In fact, in the past years it has been pointed out that relativistic effects source odd multiples of the correlation function or Fourier space power spectrum, when two different tracers are cross-correlated [27, 37–43, 45].

The two-point cross-correlation between the HI temperature and galaxy number count fluctuations, in terms of the observed coordinates, is defined as

$$\xi^{\mathrm{HI,gal}}(z_1, z_2, \theta) = \langle \Delta_{21\mathrm{cm}}(\mathbf{n}_1, z_1) \Delta_{\mathrm{gal}}(\mathbf{n}_2, z_2) \rangle \,, \qquad \cos\theta \equiv \mathbf{n_1}\cdot\mathbf{n_2}, \tag{2.10}$$

where the $\langle ... \rangle$ denotes the ensemble average, replaced in observation by the average over observed directions at a fixed observed redshift. In figure 1 we represent a scheme of the observed coordinates for the system under investigation.

The observed coordinates z_1, z_2 and θ can be converted, by assuming a cosmology, into comoving distances. We denote with r_1 and r_2 the comoving distance at z_1 and z_2, respectively, while $-\mathbf{n}_1$ and $-\mathbf{n}_2$ are the unit vectors pointing in the direction of the two tracers (HI and galaxies, respectively). Due to angular statistical isotropy, the cross-correlation can

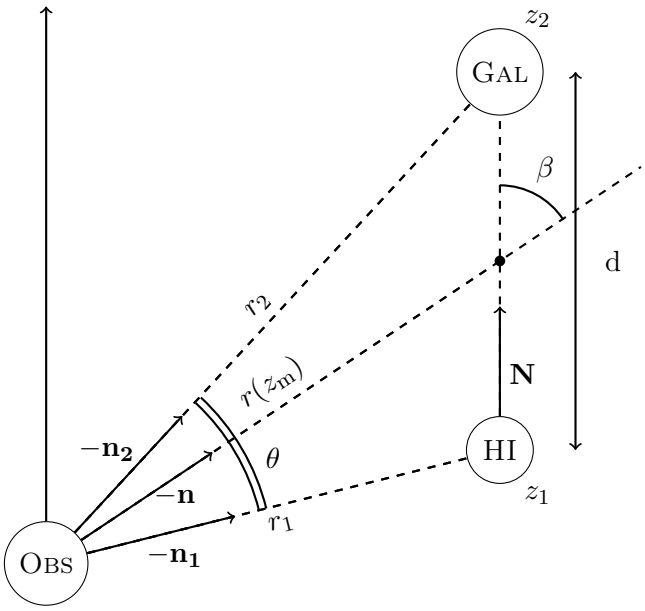


Figure 1: Illustration of the position of the two tracers under investigation with respect to the observer, in terms of the observed coordinates $\mathbf{z}_1$, $\mathbf{z}_2$ and θ and the coordinate system adopted in this work.

be the written in terms of three coordinates. In this work, we will adopt the following coordinate system (see figure 1 for a schematic representation): the comoving separation r at the mean redshift $z_m = (z_1 + z_2)/2$, the separation between the two sources d and the angle β (or $\mu = \cos\beta$), where $\cos\beta = -\mathbf{n} \cdot \mathbf{N}$, while $-\mathbf{n}$ and $\mathbf{N}$ are the unit vectors pointing in the direction of the mean redshift and the distance between the two tracers, respectively, as defined in figure 1.

The cross-correlation function can be then expanded into Legendre polynomials L_ℓ,

$$\xi^{\rm HI,gal}(r, d, \mu) = \sum_\ell \xi_\ell(r, d) L_\ell(\mu), \tag{2.11}$$

and the coefficients of this expansion, the multipoles of the correlation function, are defined as

$$\xi_\ell(r, d) = \frac{2\ell + 1}{2} \int_{-1}^{1} \xi^{\rm HI,gal}(r, d, \mu) L_\ell(\mu) d\mu. \tag{2.12}$$

This definition does not include an optimal weight based on the galaxy number density, that is generally included in order to reduce the shot-noise [71]. We do not include this effect for the following reasons. First, the weight is survey dependent, therefore a correct modelling of this effect will require to make some assumptions on the survey specifics. Furthermore, this weight is expected to reduce the noise, therefore our analysis without any assumption on the weight can be considered conservative.

In a similar way we can define[6] the Fourier cross power spectrum $P^{\rm HI,gal}(z_{\rm m}, \mathbf{k})$ as

$$\langle \Delta_{21\rm cm}(\mathbf{k}_1, z_1) \Delta_{\rm gal}(\mathbf{k}_2, z_2) \rangle = (2\pi)^3 P^{\rm HI,gal}(z_{\rm m}, \mathbf{k}_1) \delta_D(\mathbf{k}_1 + \mathbf{k}_2), \tag{2.13}$$

[6]We neglect the redshift evolution in the following definition. As shown in Ref. [39] the redshift evolution corrections are subdominant compared to the wide-angle dipole contamination of the standard terms. We

where $\Delta_{21\rm cm}(\mathbf{k}_1, z_1)$ and $\Delta_{\rm gal}(\mathbf{k}_2, z_2)$ denotes the Fourier transform of $\Delta_{21\rm cm}(\mathbf{n}_1, z_1)$ and $\Delta_{\rm gal}(\mathbf{n}_2, z_2)$, respectively, and δ_D is the Dirac delta. The multipoles of the power spectrum can be computed similarly to the correlation function multipoles, and they are proportional to the linear matter power spectrum at the mean redshift, $P(k, z_{\rm m})$,

$$\langle D(\mathbf{k}_1, z_{\rm m}) D(\mathbf{k}_2, z_{\rm m})\rangle = (2\pi)^3 P(k_1, z_{\rm m})\delta_D(\mathbf{k}_1 + \mathbf{k}_2). \tag{2.14}$$

In the distant observer limit, i.e. $d \ll r$, the angular position $-\mathbf{n}$ is assumed to be fixed for the two observed sources $\mathbf{n}_1 = \mathbf{n}_2 = \mathbf{n}$ and the full-sky correlation function can be simplified. Indeed, the full expression of the correlation function (Eq. (A.1) in Appendix A) can be written as power series expansion in d/r, and by taking only the lowest order (i.e. assuming $d \ll r$) the multipoles of the correlation function can be expressed in terms of multipoles of the power spectrum [49],

$$\xi_\ell(d, r(z_{\rm m})) = (-i)^\ell \int \frac{k^2 dk}{2\pi^2} P_\ell(k, z_{\rm m}) j_\ell(k\, d), \tag{2.15}$$

where

$$P_\ell(k, z_{\rm m}) = \frac{2\ell + 1}{2} \int_{-1}^{1} P^{\rm HI,gal}(k, z_{\rm m}, \mu) L_\ell(\mu) d\mu \tag{2.16}$$

are the coefficients of the expansion of the angle dependent power spectrum in Legendre polynomials, at fixed redshift z, and j_ℓ are the spherical Bessel function of order ℓ. Following the same strategy adopted in [49], we consider a local expansion of the multipoles of the power spectrum in power of $(\mathcal{H}/k)$ and we include the leading terms with respect to the expansion parameter $(\mathcal{H}/k)$. The even multipoles are dominated by the Newtonian contribution (i.e. the first line in (2.4) and the first two terms in (2.8)), which are simply proportional to the matter power spectrum, whilst the odd multipoles are suppressed by a factor $(\mathcal{H}/k)$, which is provided by the correlation of Doppler contribution with density and redshift space distortions (i.e. the third line in Eq. (2.4) and the third term in Eq. (2.8)). For the cross-correlation of galaxies and 21 cm brightness temperature, the leading contributions to the power spectrum multipoles are

$$\begin{aligned}
P_0(k) =& \left[b_{\rm HI} b_{gal} + \frac{f}{3}(b_{\rm HI} + b_{\rm gal}) + \frac{f^2}{5} \right] P(k)\,, \\
P_1(k) =& (-i) \left[\Big(b_{\rm gal} C_{\rm HI} - b_{\rm HI} C_{\rm gal} \Big) f + \frac{3}{5} \Big(C_{\rm HI} - C_{\rm gal} \Big) f^2 \right] \frac{\mathcal{H}}{k} P(k)\,, \\
P_2(k) =& \left[\frac{2}{3} f (b_{\rm HI} + b_{\rm gal}) + \frac{4}{7} f^2 \right] P(k)\,, \\
P_3(k) =& \, i \frac{2}{5} \Big(C_{\rm gal} - C_{\rm HI} \Big) f^2 \frac{\mathcal{H}}{k} P(k)\,, \\
P_4(k) =& \frac{8}{35} f^2 P(k),
\end{aligned} \tag{2.17}$$

also do not consider the integrated terms here. While time-delay and ISW effects are negligible, lensing magnification may contaminate the measurement of the Doppler dipole. We therefore study the contamination of magnification lensing in Sec. 2.4.

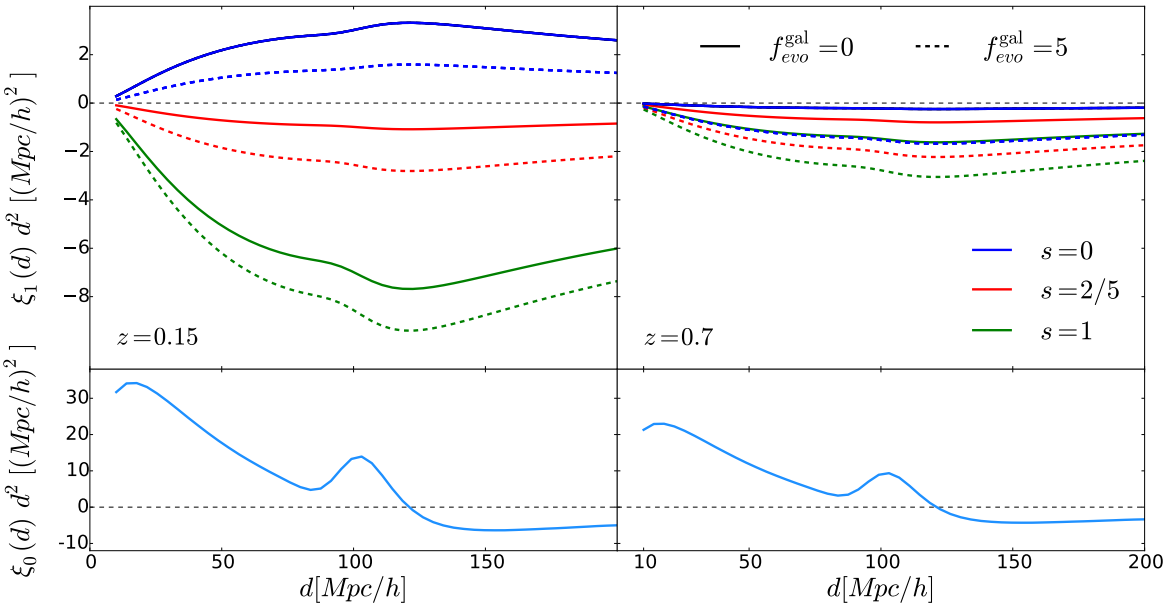


Figure 2: Cross-correlation dipole (top panels) and monopole (bottom panels), computed from Eq. (2.15), at $z = 0.15$ (left panel) and $z = 0.7$ (right panel). In the top panels different colors denotes different values for the magnification bias of the galaxy catalogue, whilst different line-styles refer to two different values of the galaxy evolution bias. The clustering biases are set to the values $b_{\rm gal} = 1$ and $b_{\rm HI} = 0.6$, the evolution bias for the HI is set to be $f_{\rm evo}^{\rm HI} = -1.5$.

where $f = d\ln D/d\ln a$ is the growth factor. In ΛCDM the growth factor is given by $f(z) = \Omega_m(z)^{4/7}$ [72, 73]. In Eq. (2.17) the mean redshift is assumed to be fixed and, for the sake of simplicity, it is omitted from the notation. The coefficients $C_{\rm gal}$ and $C_{\rm HI}$ are defined as

$$C_{\rm gal} = \left(\frac{\mathcal{H}'}{\mathcal{H}^2} + \frac{2-5s}{r\mathcal{H}} + 5s - f_{\rm evo}^{\rm gal}\right),$$
$$C_{\rm HI} = \left(\frac{\mathcal{H}'}{\mathcal{H}^2} + 2 - f_{\rm evo}^{\rm HI}(z)\right). \tag{2.18}$$

In figure 2 we compare the cross-correlation dipole (top panels) and monopole (bottom panels) at two different mean redshifts $z_{\rm m} = 0.15$ (left panel) and $z_{\rm m} = 0.7$ (right panel). In this plot, we set the values of the clustering biases to be $b_{\rm gal} = 1$ and $b_{\rm HI} = 0.6$. The evolution bias for the HI is assumed to be $f_{\rm evo}^{\rm HI} = -1.5$, while we denote with different colors different values of the magnification bias s and with different line-styles two different values of the galaxy evolution bias. Interestingly, we remark that the sign of the dipole depends strongly on the magnification bias factor and, therefore, it can not be omitted in the analysis.

We see that both the dipole and the monopole signals decrease at larger redshift and that the monopole is significantly larger, in amplitude, than the dipole. Furthermore, at $z_{\rm m} = 0.7$ the terms depending on the galaxy evolution bias become dominant in the dipole. Since modelling the evolution properties of a galaxy population is not an easy task, the cosmological information we can extrapolate from the dipole at large redshift can be contaminated and limited from a prior knowledge about the evolution bias. The decrease of the dipole at larger redshift depends on two elements: the time evolution of the linear matter power spectrum

(which affects the monopole as well) and the terms in the coefficients (2.18) proportional to $(r\mathcal{H})^{-1}$. Therefore, in order to detect relativistic effects we expect an ideal galaxy catalogue in the low redshift regime, and for this reason we will set the mean redshift of observation for our analysis to be $z_{\rm m} = 0.15$.

2.4 Contaminations to the relativistic dipole

The dipole of the cross-correlation that we discussed in the previous section is sourced by the Doppler corrections to the galaxies number counts and to the observed brightness temperature of the 21 cm emission. This is usually considered the main contribution to the dipole of the cross-correlation between two tracers, in fact the Doppler corrections depend on the projection of the peculiar velocity along the line of sight and therefore they are intrinsically anisotropic.

Nevertheless, a measurement of the dipole would be contaminated by other sources of anisotropy [39]. In this section we will discuss two possible contaminations: the dipole induced by gravitational lensing and the wide-angle effects. The latter have been extensively studied in Refs. [12, 14, 74–81]. Beside these two contaminants, there are further corrections induced by the redshift evolution of the bias and the growth factors. They are generally subdominant with respect to the wide-angle correction [39], therefore they will be neglected.

The gravitational lensing asymmetry [39, 50] comes from the cross-correlation of the HI density and the gravitational lensing term in the galaxy number counts

$$\xi^{\rm lens}(r,d,\beta) \equiv \left\langle \left(b_{\rm HI}(z_1) D\left(\mathbf{n}_1,z_1\right)\right)\left(\frac{5s-2}{2}\int_0^{r_2}\frac{r_2-r'}{r_2 r'}\Delta_\Omega(\Phi+\Psi)\left(\mathbf{n}_2,z'\right)dr'\right)\right\rangle, \quad (2.19)$$

and it emerges from the fact that galaxies behind an HI overdensity with respect to the observer will be lensed, while the HI temperature fluctuations are not lensed, to linear order, by the galaxies in front. The lensing correlation function defined above has been computed in [39] and further studied in [50]. In the Limber approximation and to the lower order in d/r, it reads

$$\xi^{\rm lens}(r,d,\beta) = (1+z_{\rm m})\frac{3\Omega_m\pi}{4} b_{\rm HI}(5s-2)\, d\,\mathcal{H}_0\,\cos(\beta)\,\Theta(r_2-r_1)\mu_{\rm lens}(\beta), \quad (2.20)$$

where Θ is the Heaviside function, D_1 is the linear growth factor and the function $\mu_{\rm lens}$ is

$$\mu_{\rm lens}(\beta) = \int_0^\infty \frac{k_\perp dk_\perp}{2\pi^2}\mathcal{H}_0 P(k_\perp) J_0(k_\perp\, d\,\sin(\beta)), \quad (2.21)$$

being J_0 is the order-0 Bessel function. The lensing dipole can therefore be computed similarly to the relativistic dipole

$$\xi_1^{\rm lens}(r,d) = \frac{3}{2}\int_{-1}^{1}\xi^{\rm lens}(r,d,\mu)L_1(\mu)d\mu. \quad (2.22)$$

The lensing dipole is proportional to the radial distance between the two tracers and the redshift. Therefore, we expect it to be relevant in the high redshift regime, in particular because the Doppler dipole decreases with redshift. In figure 3 we show the amplitude of the lensing dipole compared to the dipole sourced by the Doppler terms, at redshift $z_{\rm m} = 0.15$. We see that the lensing dipole is always few orders of magnitude smaller then the Doppler dipole, in this redshift regime. Therefore, in the rest of the paper it will be neglected.

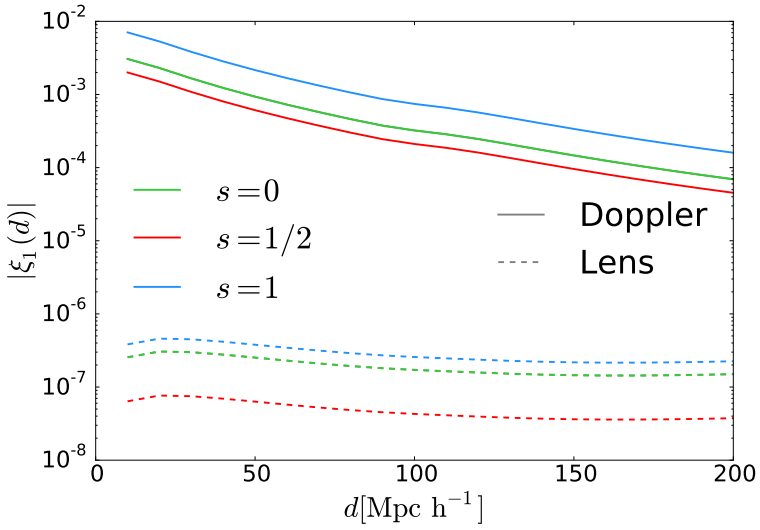


Figure 3: Comparison between the amplitude of lensing dipole (dashed lines) and the Doppler dipole (continuous lines) at $z_{\rm m} = 0.15$. The clustering bias of the two tracers are fixed to the values $b_{\rm HI} = 0.6$ and $b_{\rm gal} = 1$. Different colors denote different values of the magnification bias. The HI evolution bias is $f_{\rm HI}^{\rm evo} = -1.5$, while we fix the galaxy evolution bias to be zero.

The second correction we will discuss here is the wide-angle effect. In the limit $d \ll r$, we can expand the full-sky correlation function in power of d/r. Therefore the Taylor expansion of all the functions of r_1 and r_2 around r in the leading terms, namely density and redshift space distortions, induces a non-vanishing dipole suppressed by d/r. Considering that the largest contributions to the relativistic dipole at low redshift is of the order $d/r \sim 1/(kr) \sim \mathcal{H}/k$, we need to account for the wide-angle contamination in the dipole. The leak from the monopole to the dipole due to the wide-angle contribution, at fixed redshift and at the lowest order in d/r, is described by [39, 49]

$$\xi_1^{\rm WA}(d) = \frac{2f}{5}(b_{\rm gal} - b_{\rm HI})\frac{d}{r}\int \frac{k^2 dk}{2\pi^2} P(k) j_2(k\,d). \tag{2.23}$$

This correction can be written as a combination of the quadrupole of the autocorrelation of two tracers. Therefore, we correct the dipole estimator $\hat{\xi}_1$ for the bias due to the wide-angle effect [49]

$$\hat{\xi}_1(d,r) \to \hat{\xi}_1(d,r) - \frac{3}{10}(\hat{\xi}_2^{gal} - \hat{\xi}_2^{\rm HI})\frac{d}{r}. \tag{2.24}$$

In figure 4 we show how the signal changes when the wide-angle correction in Eq. (2.24) is applied to the estimator. The magenta dotted line represent the wide-angle contribution to the dipole, computed from Eq. (2.23). Its magnitude is comparable to the one of the relativistic dipole, thus it is clearly a not negligible contribution. Furthermore we see that correcting for the wide-angle effect does not necessarily reduces the signal: if its sign agrees with the one of the dipole, the signal is boosted. This leads also to an extra contribution to the covariance of the estimator [49]. In the next section we will show how to compute the full covariance of the estimator, the explicit contribution to the covariance of the wide-angle correction can be found in Appendix B.

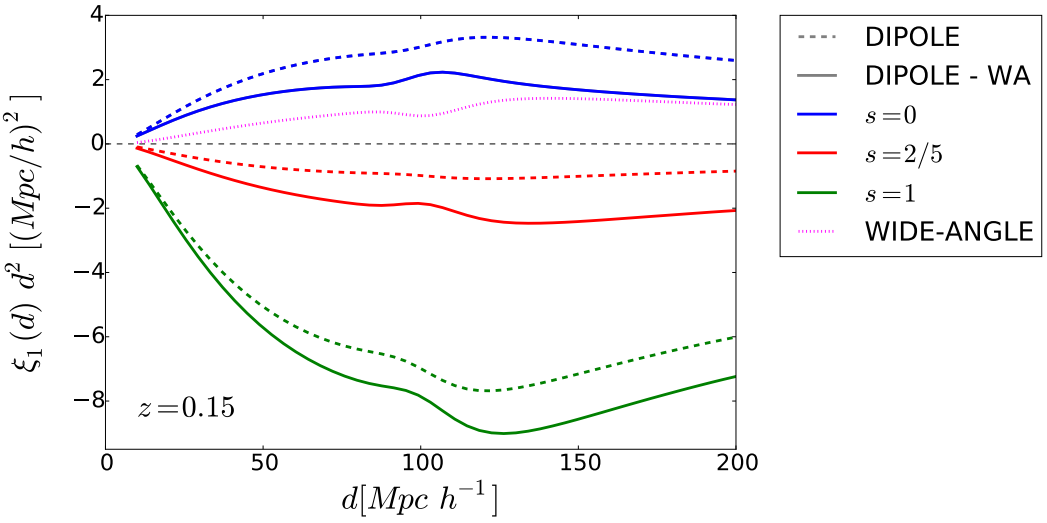


Figure 4: Cross-correlation dipole with (continuous line) and without (dashed line) the wide-angle correction in (2.24). Different colors denotes different values of the galaxy magnification bias. The clustering biases are set to be $b_{\rm gal} = 1$ and $b_{\rm HI} = 0.6$. The evolution bias is set $f_{\rm HI}^{\rm evo} = -1.5$ for the neutral hydrogen, while for galaxies it is set to be zero. The magenta dotted line represents the wide-angle correction in (2.23).

2.5 Covariance for the cross-correlation dipole

The full covariance matrix for the multipoles of the 2-point correlation function (2PCF) is presented in [82] for the single tracer case and in [49] for the multiple tracers case. Here we will apply the generic expression in [49] (Eq. 17) to the dipole.

$$\begin{aligned}
\mathrm{COV}(d_1,d_2) = &-\frac{9}{V}\int\frac{k^2dk}{2\pi^2}j_1(k\,d_1)j_1(k\,d_2)\left(\frac{1}{5}P_1^2(k)+\frac{8}{35}P_1(k)P_3(k)+\frac{23}{315}P_3^2(k)\right)\\
&+\frac{9}{V}\int\frac{k^2dk}{2\pi^2}j_1(k\,d_1)j_1(k\,d_2)\left[\frac{1}{3}\left(N_{\rm HI}P_0^{\rm gal}(k)+\frac{1}{n_{\rm gal}}P_0^{\rm HI}(k)\right)\right.\\
&\left.+\frac{2}{15}\left(N_{\rm HI}P_2^{\rm gal}(k)+\frac{1}{n_{\rm gal}}P_2^{\rm HI}(k)\right)\right]\\
&+\delta_{d_1,d_2}\frac{3}{4\pi V d_1 d_2 L_p}N_{\rm HI}\frac{1}{n_{\rm gal}}\\
&+\frac{9}{V}P_{\rm N}\int\frac{k^2dk}{2\pi^2}j_1(k\,d_1)j_1(k\,d_2)\left(\frac{1}{3}P_0^{\rm gal}(k)+\frac{2}{15}P_2^{\rm gal}(k)\right)\\
&+\delta_{d_1,d_2}\frac{3}{4\pi V d_1 d_2 L_p}P_{\rm N}\frac{1}{n_{\rm gal}},
\end{aligned}\tag{2.25}$$

where V is the overlapping volume of the galaxy and the intensity mapping surveys; $P_\ell^{\rm gal}$ and $P_\ell^{\rm HI}$ denotes the galaxy and the HI power spectrum multipoles, respectively; $n_{\rm gal}$ is the comoving number density of galaxies; $N_{\rm HI}$ and $P_{\rm N}$ are the shot-noise and the interferometer

noise for the HI, respectively; $L_{\rm p}$ denotes the resolution of the IM survey and $\delta_{x,y}$ is the Kronecker delta.

The first line represents the purely cosmic variance contribution to the covariance, the second and the third lines are the cosmic variance - Poisson noise terms, the fourth line is the purely Poissonian contribution, while the last two lines are the interferometer noise - cosmic variance term and the interferometer noise - galaxy Poisson noise term, respectively. The terms, which do not involve the integral of the power spectrum, have been integrated by using the orthogonality relation of the spherical Bessel functions

$$\int_0^\infty dk k^2 j_1(kd_1) j_1(kd_2) = \frac{\pi}{2d_1 d_2}\delta_D(d_1 - d_2), \tag{2.26}$$

where in the discrete limit [49] $\delta_D(d_1 - d_2) \to \delta_{d_1,d_2}/L_p$. The other integrals have been solved numerically, and a smooth cutoff is applied in all the integrals to model the finite resolution of the interferometer. To be more precise, all the integrands are multiplied by a top-hat filter in Fourier space $W^4(kR)$, defined as

$$W(kR) = \frac{3[\sin(kR) - kR\cos(kR)]}{(kR)^3}, \tag{2.27}$$

where the scale R is set to be the size of the pixel for the IM, $L_{\rm p}$.

Figure 5 shows the different terms contributing to the diagonal covariance entries, at redshift $z_{\rm m} = 0.15$. The bias values for the two tracers are the same adopted in figure 4. Furthermore, we set the shot-noise and the interferometer noise for the HI to $N_{\rm HI} = 100\,({\rm Mpc/h})^3$ and $P_{\rm N} = 100\,({\rm Mpc/h})^3$, respectively. The comoving galaxy number density is assumed to be $n_{\rm gal} = 10^{-3}\,({\rm h/Mpc})^3$ and the volume of the survey is computed assuming that the sky coverage of the cross-correlation is $f_{\rm sky} = 0.2$ and that the redshift bin is $z \in [0.05, 0.25]$. The size of the pixel is chosen to be $L_{\rm p} = 2\,{\rm Mpc/h}$. At small scales, the dominant components of the covariance are the terms in which the cosmic variance is cross-correlated with the noise. On large scales, the covariance introduced by the wide-angle correction becomes more important and on scales $\sim 190\,{\rm Mpc}/h$ or larger is the dominant contributor (see the Appendix B for details on how to estimate the covariance introduced by the wide-angle correction). The magnification bias enters only in the computation of the purely cosmic variance term, but since this term result to be subdominant it does not affect significantly the full covariance (at least its diagonal components).

Both the interferometer noise $P_{\rm N}$ and the size of the pixel $L_{\rm p}$ depend on the specifics of the IM survey (see [49, 70]). In order to stay as general as possible, we choose fiducial values ($P_{\rm N} = 100\,({\rm Mpc}/h)^3$ and $L_{\rm p} = 2\,{\rm Mpc}/h$, respectively). They are approximately the values of a survey similar to the Canadian Hydrogen Intensity Mapping Experiment (CHIME) [61, 83], extrapolated to low redshifts ($z \sim 0.1 - 0.2$).

3 Halo Model approach

As shown in section 2, the signal and the covariance for the dipole depend on many different parameters (clustering and evolution biases, galaxy magnification bias, shot-noise of the two tracers, sky fraction of the cross-correlation and redshift range). In this section we aim to model the properties of the two tracers in order to find a relation between these parameters, which are not all independent. This will allow us to find the optimal specification to look for

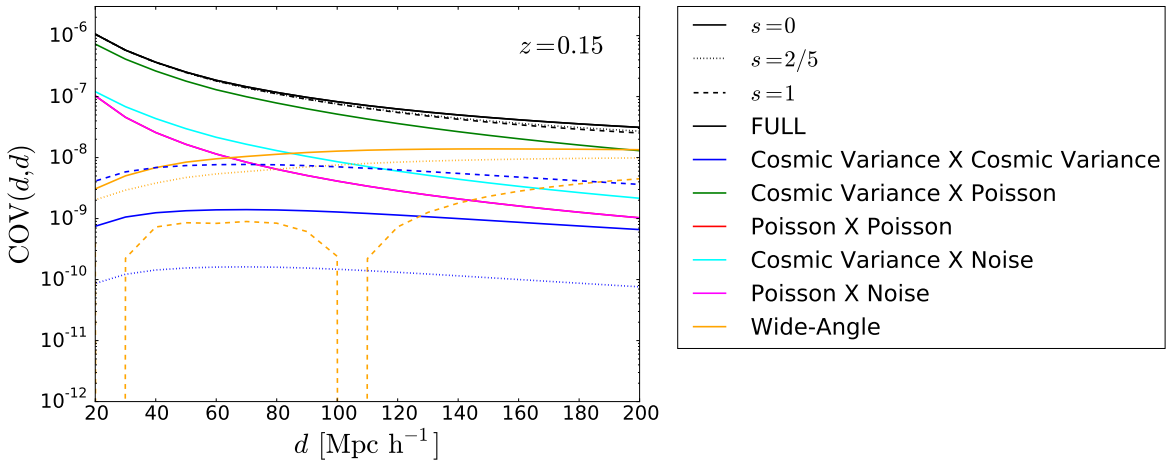


Figure 5: Diagonal entries of the covariance, computed from (2.25). Different colors denote different terms contributing to the covariance. The full covariance is the black lines. Different line-styles represent different values of the magnification bias, which affects only the cosmic variance X cosmic variance contribution (blue line).

relativistic effects detection in LSS. Moreover, by determining the relation between all these different parameters we can parameterize better their uncertainties if we need to marginalize over them. In order to model a generic galaxy population and the neutral hydrogen distribution we will adopt the so-called halo model (see [84] for a general review). The halo model was first proposed for modelling the galaxies properties [85–87], but it has been more recently successfully applied to neutral hydrogen [88–91]. With this approach we are able to model the relation between the tracers clustering bias and their shot-noises.

3.1 HI model

We adopt for the neutral hydrogen the model based on [89]. As supported by numerical simulations [88], we can safely assume that the contribution of neutral hydrogen outside the dark matter halos is negligible. Hence, the neutral hydrogen comoving density $\bar{\rho}_{\rm HI}$ at a given redshift z can be computed as

$$\bar{\rho}_{\rm HI}(z) = \int_0^\infty n(M,z) M_{\rm HI}(M,z) dM, \tag{3.1}$$

where $n(M,z)$ is the halo mass function at redshift z, i.e. the comoving number density of halos with masses in the range between M and $M+dM$, $M_{\rm HI}(M,z)$ is the average HI mass in a halo of mass M at redshift z. The halo mass function can be expressed as

$$n(M,z) = -\frac{\bar{\rho}_m}{M^2} f(\sigma) \frac{d\ln\sigma}{d\ln M}, \tag{3.2}$$

where $\bar{\rho}_m = \Omega_m(z)\,\rho_c^0$, ρ_c^0 is the critical density at $z=0$, σ is the root mean square of the variance of the linear density field, smoothed on the scale $R(M)$, the radius enclosing an amount of mass equal to M

$$\sigma^2(R,z) = \frac{1}{2\pi^2}\int_0^\infty k^2 P(k,z) W^2(kR) dk, \qquad R = \left(\frac{3M}{4\pi\bar{\rho}_m}\right)^{\frac{1}{3}}. \tag{3.3}$$

The smoothing function W is the Fourier transform of a top-hat filter (the same functional form of (2.27)). The function $f(\sigma)$ is generally calibrated from N-body simulation. In this work we use a Tinker mass function [92]. Within this framework, the shot-noise and the HI bias can be written as [70, 89]

$$N_{\rm HI}(z) = \left(\frac{1}{\bar{\rho}_{\rm HI}(z)}\right)^2 \int_0^\infty n(M,z) M_{\rm HI}^2(z) dM, \tag{3.4}$$

$$b_{\rm HI}(z) = \frac{1}{\bar{\rho}_{\rm HI}(z)} \int_0^\infty n(M,z) b(M,z) M_{\rm HI}(z) dM, \tag{3.5}$$

where $b(M,z)$ is the halo bias, calibrated on the N-body simulation from [93].

We model the average HI mass within an halo of mass M as redshift independent [89]

$$M_{\rm HI}(M,z) = {\rm C}\ (1-Y_{\rm p}) \frac{\Omega_{\rm b}}{\Omega_{\rm m}} \exp\Big[-(M_{\rm min}/M)\Big] M^\alpha, \tag{3.6}$$

where $Y_{\rm p} = 0.24$ is the Helium fraction (note that $(1-Y_{\rm p})\frac{\Omega_{\rm b}}{\Omega_{\rm m}}$ is the HI mass fraction), $M_{\rm min}$, α and C are the free parameters of the model. Eq. (3.6) entails that the mass of cosmic hydrogen within a halo scales as a power law of the total virial mass of the halo, at large halo masses. The efficiency of this scaling is regulated by the exponent α (larger values of α correspond to larger amount of hydrogen within a halo of fixed mass). At low halo masses, we expect instead an exponential suppression, due to different physical processes, such as photoionization from the UV background or galactic winds. The parameter $M_{\rm min}$ regulates the range of halo masses for which this suppression is effective (larger values of $M_{\rm min}$ imply a larger range of masses for which the suppression is relevant). The parameter C is an overall normalization constant, that needs to be fixed by matching the theoretical abundance of neutral hydrogen, predicted by (3.1), with what is measured by HI galaxy survey at a given redshift. As we can notice from Eqs. (3.4) and (3.5), both the shot-noise and the HI bias do not depend on this normalization factor, therefore we can fix C = 1 without loss of generality.

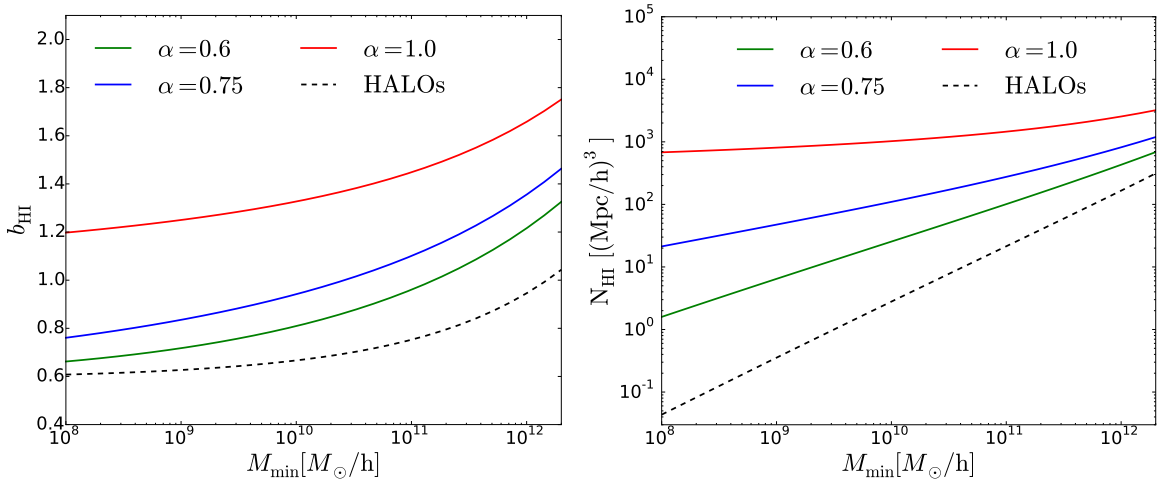
In figure 6 we show the HI bias and HI shot-noise, respectively, as a function of the cutoff parameter $M_{\rm min}$, for different values of α. We can see that if the neutral hydrogen is concentrated in halos with large mass (this corresponds to larger values for $M_{\rm min}$ and α), both the shot-noise and the HI bias have larger values with respect to the case in which we find considerable amount of neutral hydrogen also in small halos. The halo bias (in black dashed lines) is computed from Eq. (15) in [94]

$$b_{\rm halos}(z) = \frac{\int_0^\infty n(M,z) b(M,z) dM}{\int_0^\infty n(M,z) dM}, \tag{3.7}$$

while the shot-noise for the halo population is computed as

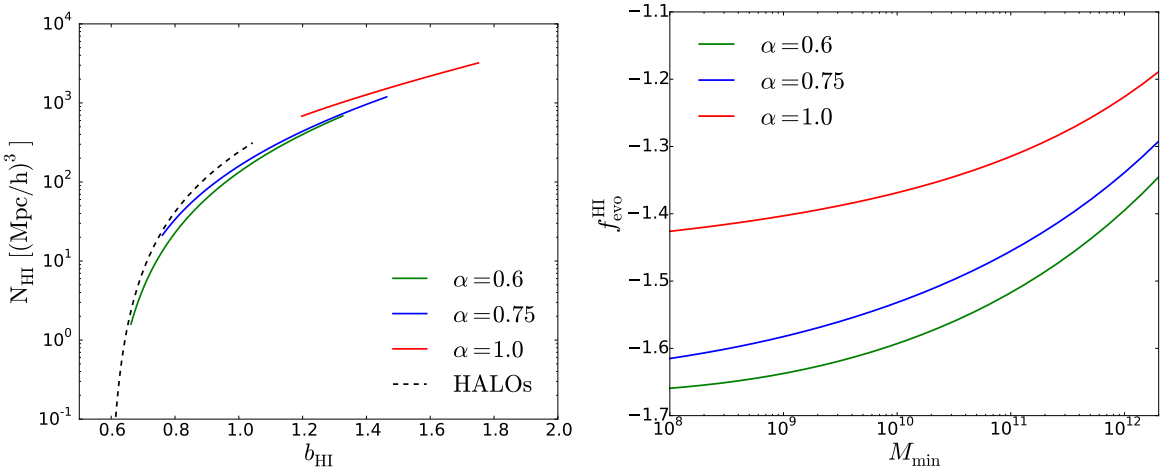
$$N_{\rm halos} = \frac{1}{\int_0^\infty n(M,z) dM}. \tag{3.8}$$

From figure 6 we can find a relation between the bias of HI, $b_{\rm HI}$, and the shot-noise, $N_{\rm HI}$. In figure (7a) we show how the HI bias and the HI shot-noise are related to each others, for different values of the parameter α, by considering $M_{\rm min}$ as the parametrisation of the different curves. As expected, in all models higher biased HI models correspond to higher shot-noise. Furthermore, models with an higher efficiency in accreting neutral hydrogen within



(a) HI bias **(b)** HI shot-noise

Figure 6: HI bias (left panel) and HI shot-noise (right panel) as a function of the cutoff parameter $M_{\rm min}$, at redshift $z = 0.15$. Different colors denote different values of the exponent α in (3.6). The black dashed lines represent the halo bias and the shot-noise of the halo population.



(a) Shot noise VS bias - HI **(b)** Evolution bias - HI

Figure 7: Left panel: Shot-noise of the HI as a function of the HI bias. Right panel: Evolution bias as a function of the cutoff parameter $M_{\rm min}$. The notation is the same as figure 6.

halos (larger values of α) correspond to higher values of bias and shot-noise. According to hydrodynamics simulations [95, 96] and semi-analytic models [97], values of $\alpha < 0.9$ are more realistic.

In figure 7b we show the dependence of the evolution bias, computed from (2.9), for different models. The evolution bias for the HI does not vary significantly within the parameter space, because it is described by the redshift evolution of the halo mass function. Its negative value indicates that the density of HI drops at low redshift.

We note that the model outlined in this section is more focussed on the high redshift regime and needs to be better tested and improved in the low redshift Universe in order to

capture the complex astrophysical effects [98, 99] that could impact on the HI distribution inside galaxies.

3.2 Galaxy model

As discussed in the previous section for neutral hydrogen, we can use the halo model to relate the bias parameters and the shot-noise of galaxy distributions. In this framework, we assume that all the galaxies are found within dark matter halos. The relevant quantity we need to model is their comoving number density, which can be computed from the halo mass function as

$$n_{\rm g}(z) = \int_0^\infty n(M,z) N_{\rm av}(M,z) dM, \tag{3.9}$$

where $N_{\rm av}$ is the average number of galaxies for an halo of mass M. The galaxy bias can be modeled as

$$b_{\rm g}(z) = \frac{1}{n_{\rm g}(z)} \int_0^\infty n(M,z) b(M,z) N_{\rm av}(M,z) dM, \tag{3.10}$$

while the shot-noise is simply given by the inverse of the number density, i.e.

$$N_{\rm gal} \equiv \frac{1}{n_{\rm g}} = \frac{1}{\int_0^\infty n(M,z) N_{\rm av}(M,z)\, dM}. \tag{3.11}$$

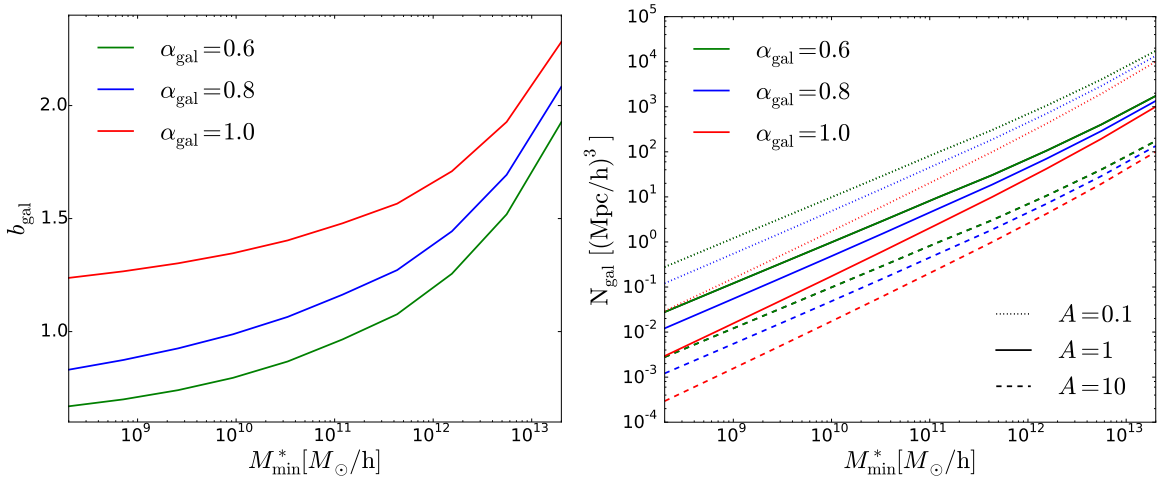
We model the average number of galaxies within a halo of mass M as

$$N_{\rm av}(M) = \begin{cases} 0 & \text{if } M \leq M^*_{\rm min} \\ A\left(\frac{M}{M^*_{\rm min}}\right)^{\alpha_{\rm gal}} & \text{if } M > M^*_{\rm min} \end{cases} \tag{3.12}$$

This model is similar to the model employed in [84] to model red and blue galaxies, which provides a good fit to the number of subhalos expected from numerical simulations [100]. The parameter $M^*_{\rm min}$ represents a threshold mass, below which a halo cannot host a galaxy, because their potential wells are shallower with respect to more massive halos. Therefore, some physical processes such as supernova feedback, can be efficient enough to expel a huge percent of baryons from the halo and therefore suppress the star formation within the halo itself. For halos with masses larger than this threshold, we assume that the average number of galaxies per halo increases with the halo mass, following a power law $N_{\rm av}(M) \propto M^{\alpha_{\rm gal}}$, so that $\alpha_{\rm gal}$ represents its slope. The parameter A represents a normalization constant. From Eq. (3.10) we see that, as we found for the HI, the galaxy bias is not affected by the value of the normalization, while the shot-noise, Eq. (3.11), strongly depends on A. In the literature more flexible and physically motivated models have been proposed (see e.g. [101, 102]). Nevertheless, the simple model described above captures the features that we need for the purpose of our analysis with the minimal number of parameters.

In figure (8) we show the galaxy bias (left panel) and galaxy shot-noise (right panel) as a function of the threshold mass $M^*_{\rm min}$. The bias increases exponentially with the threshold mass, while the shot-noise increase as a power law whose slope depends the slope of the model (3.12).

In figure (9) we summarize the information of figure (8) by displaying the relation between galaxy bias and shot-noise, for different models. Similarly to HI, highly biased galaxy populations (which are more massive) show a larger shot-noise, compared to the one with lower bias.



(a) Galaxy bias **(b)** Galaxy shot-noise

Figure 8: Galaxy bias (left panel) and galaxy shot-noise (right panel) as a function of the cutoff parameter $M^*_{\rm min}$, at redshift $z = 0.15$. Different colors denote different values of the exponent $\alpha_{\rm gal}$ in (3.6). In the right panel, different line-styles denote different values of the normalization constant A.

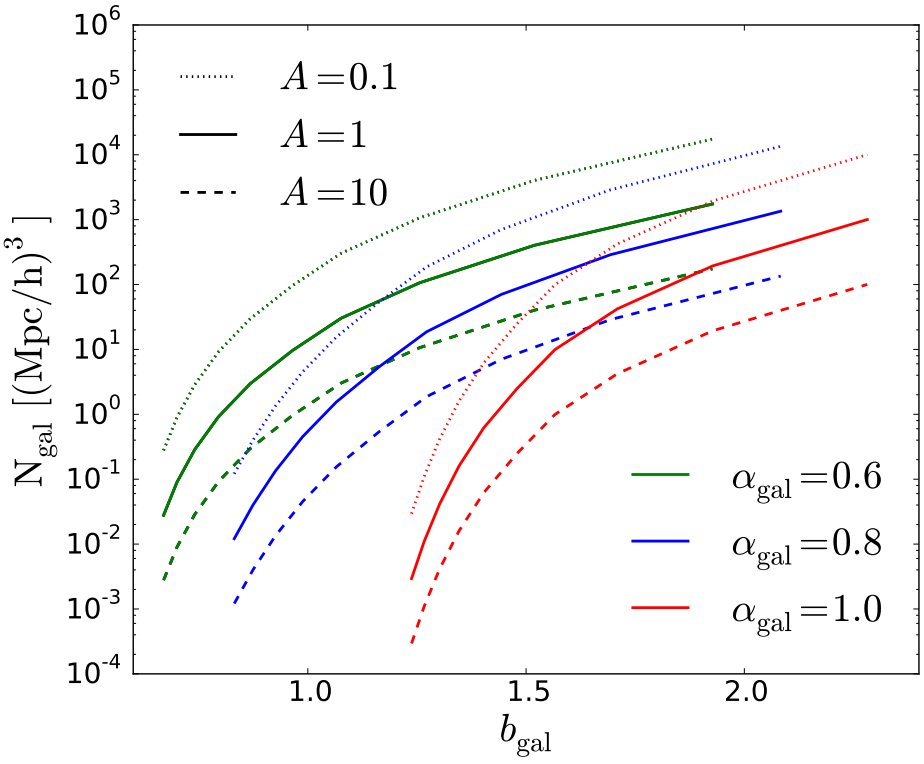


Figure 9: Shot-noise of the galaxies as a function of the galaxy bias, for different values of the slope $\alpha_{\rm gal}$ and different values of the normalization constant A.

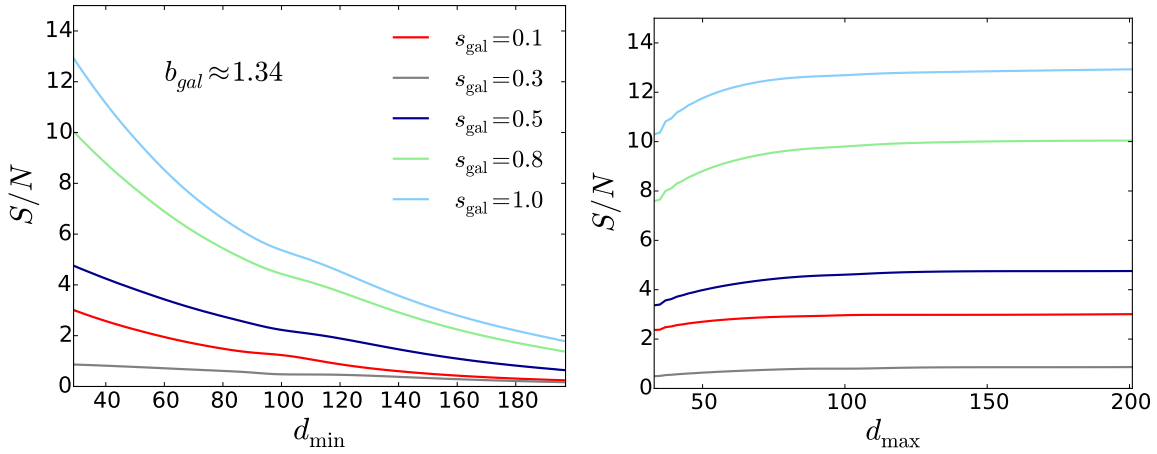


Figure 10: Signal-to-noise as a function of $d_{\rm min}$ and $d_{\rm max}$. The HI model is the model 1 in Table 1. The galaxy bias is fixed to be $b_{\rm gal} \approx 1.34$ here, and the interferometer noise is set to be $P_{\rm N} = 100\,({\rm Mpc}/h)^3$. Different colors denotes different values for the magnification bias.

4 Signal-to-noise analysis

In this section we will present a signal-to-noise analysis for the dipole of the HI-galaxies cross-correlation. The signal-to-noise for the dipole is defined as [49]

$$\left(\frac{S}{N}\right)^2 = \sum_{d_1,d_2=d_{\rm min}+L_{\rm p}/2}^{d_{\rm max}-L_{\rm p}/2} \xi_1(d_1){\rm COV}^{-1}(d_1,d_2)\xi_1(d_2)\,. \tag{4.1}$$

The minimum distance $d_{\rm min}$ is chosen to be the non-linear scale. We set $d_{\rm min} = 30\,{\rm Mpc}/h$, which correspond to $k_{\rm NL} \approx 0.2\,h/{\rm Mpc}$. We also study the dependence of the signal-to-noise ratio on both $d_{\rm min}$ and $d_{\rm max}$. The maximum distance is set to be the distance at which the dipole estimator, computed to lowest order in the d/r expansion from (2.24), coincides with the full-sky quantity up to 3%. Since we are interested in relatively low redshift measurements, we provide a comparison between the wide-angle correction to linear order in d/r and the exact full-sky contributions. The details of this comparison are shown in the Appendix A. Our results show that for $d \leq 200\,{\rm Mpc}/h$ the discrepancy between the two quantities is smaller than the 3% threshold, therefore we set $d_{\rm max} = 200\,{\rm Mpc}/h$. Although more numerically expensive, it is possible to consider the full-sky expression if further accuracy is required. Nevertheless, as shown in figure 10, the main information is encoded on scales where the expansion to linear order in d/r is accurate enough.

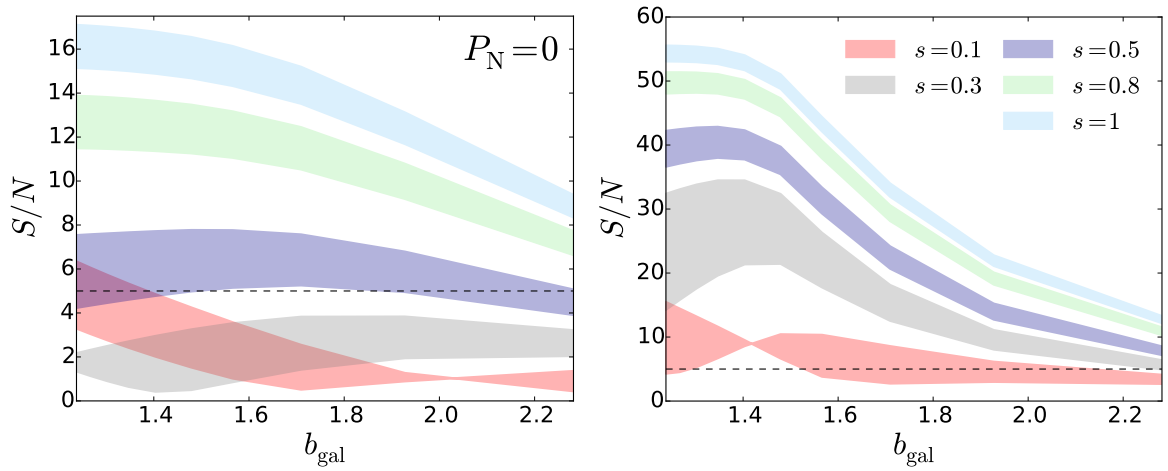
The cross-correlation volume V is computed as

$$V = \frac{4}{3}\pi\Big(r^3(z_{\rm max}) - r^3(z_{\rm min})\Big)f_{\rm sky}, \tag{4.2}$$

where $r(z)$ is the comoving distance at redshift z, $z_{\rm max}$ and $z_{\rm min}$ are the maximum and minimum observed redshift, respectively, and $f_{\rm sky}$ is the fraction of the sky available for the observation. We consider a redshift range between $z_{\rm min} = 0.05$ and $z_{\rm max} = 0.25$ and a fraction of sky $f_{\rm sky} = 0.2$, which is the sky coverage of a survey similar to the Baryon Oscillation Spectroscopic Survey (BOSS) [2]. We assumed the interferometer employed for observing the neutral hydrogen in intensity mapping can resolve pixels of size $L_{\rm p} = 2\,{\rm Mpc}/h$.

	MODEL 1 (CONSERVATIVE)	MODEL 2 (OPTIMISTIC)
α	0.75	0.6
$M_{\rm min}$	$\approx 1.7 \times 10^{10} M_\odot/{\rm h}$	$10^8 M_\odot/{\rm h}$
$b_{\rm HI}$	≈ 0.99	≈ 0.67
$N_{\rm HI}$	$\approx 143{\rm Mpc}^3/{\rm h}^3$	$\approx 2{\rm Mpc}^3/{\rm h}^3$
$f_{\rm evo}^{\rm HI}$	≈ -4.33	≈ -4.40

Table 1: Values of the parameters for the two HI models we consider for our analysis. Model 1 has larger bias and larger shot-noise with respect to Model 2.



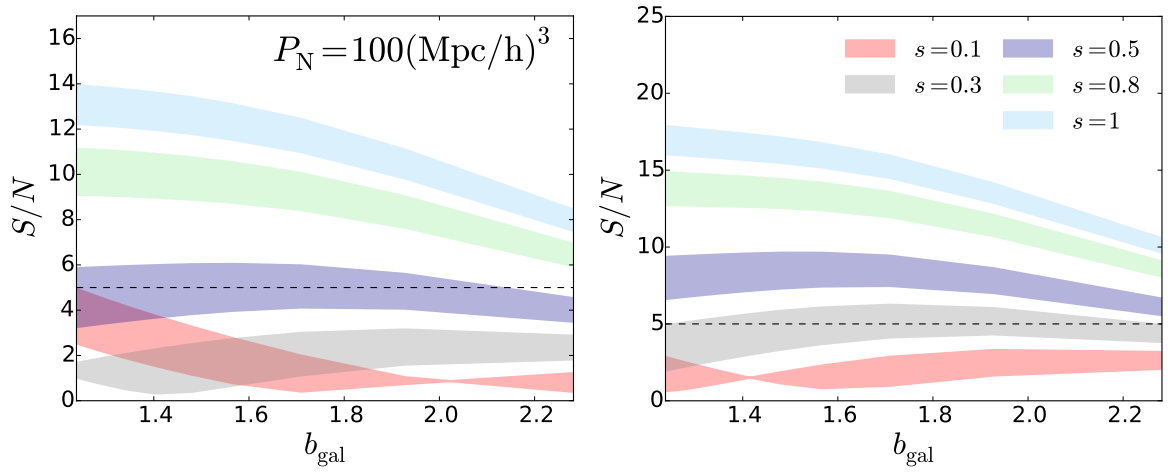
(a) MODEL 1 (CONSERVATIVE) **(b)** MODEL 2 (OPTIMISTIC)

Figure 11: Signal-to-noise for the cross-correlation dipole as a function of the galaxy bias. The slope of the model for the galaxy population is fixed at the value $\alpha_{\rm gal} = 1$. Different colors denote different values of the magnification bias. The corresponding shaded region represents the signal-to-noise for $f_{\rm evo}^{gal} \in [-2, 2]$. The interferometer noise is neglected here.

We will consider two models for the neutral hydrogen, based on the formalism presented in section 3.1. The first model (conservative model) is similar to the one employed in [103], with bias close to unity at low redshift and relatively high shot-noise. The second model (optimistic model) is similar to the one adopted for the forecast in [70], with smaller bias and lower shot-noise. The specific values of the parameters of the two models are highlighted in Table 1.

Concerning the galaxy model, the parameter A is a normalization constant and does not qualitatively affect the relations between shot-noise and bias. Therefore, in the rest of the paper we will fix its value to be $A = 1$, which correspond to the minimum value of the average number of galaxies per halos whose masses are above the threshold value $M^*_{\rm min}$. Furthermore, we fix the value of $\alpha_{\rm gal} = 1$, which corresponds to the most biased galaxies.

Figure 11 shows the dipole signal-to-noise at $z = 0.15$, as a function of the galaxy bias, for the two models described above. The signal-to-noise is computed for different values of the magnification bias, which is treated here as a free parameter. We find that the signal-to-noise is generally optimized for the largest value of the magnification bias (which in this case is $s_{\rm gal} = 1$). Let us remark that the signal-to-noise ratio does not grow monotonically with the



(a) MODEL 1 (CONSERVATIVE) **(b)** MODEL 2 (OPTIMISTIC)

Figure 12: Signal-to-noise for the cross-correlation dipole as a function of the galaxy bias, with an interferometer noise $P_N = 100\,(\mathrm{Mpc/h})^3$. The same conventions of figure 11 are adopted. The black dashed line highlight the $S/N = 5$ threshold.

magnification bias, but it has a minimum between the lines $s = 0.1$ and $s = 0.3$ of Fig. (11). Indeed as shown in Fig. (4), the largest pre-factor of the dipole in flat-sky is $(2 - 5s)/(r\mathcal{H})$. Then, considering wide-angle effects and the impact of the other bias factors (in particular the large negative value of the evolution bias of neutral hydrogen), the dipole vanishes for lower values of magnification bias.

In the optimistic model for the HI (figure 11b), the signal-to-noise is > 5 for all the values of the magnification bias. We see that it increases with the bias up to $b_{\rm gal} \approx 1.4$, then it declines for more biased galaxies. This is due to the fact that the dipole signal increases for larger difference between the biases of the two tracers, but highly biased tracers are more massive and we observe fewer of them. Therefore for the most biased galaxies the growth of the shot-noise dominates over growth of the signal and the S/N results to be suppressed.

Similar comments are valid for the more conservative HI model. In this case we need $s_{\rm gal} \geq 0.5$ in order to have a $S/N > 5$. The maximum value of the signal-to-noise in the two models ranges between $(S/N)_{\rm max} \approx 17$, for the more conservative model, to $(S/N)_{\rm max} \approx 53$ in the optimistic model.

In figure 12 we show how the interferometer noise affects the signal-to-noise analysis for the two models. We assume the fiducial value $P_N = 100\,(\mathrm{Mpc}/h)^3$, which corresponds to the noise of a CHIME-like survey [61, 83] observing at the mean redshift considered in our work, computed from Refs. [49, 70]. For both models the signal-to-noise is suppressed and the effect results to be more prominent for the optimistic model. We see that the interferometer noise suppresses the maximum value of the signal-to-noise ratio roughly from 17 to 14 for the conservative model, while the maximum S/N decreases from 53 to 18 in the optimistic model. Indeed, in the optimistic model the HI shot-noise results to be much smaller than the reference interferometer noise, while in the conservative model the two quantities are comparable. Interestingly, we observe that, even when the interferometer noise is included in the analysis, the signal-to-noise results to be > 5 for $s_{\rm gal} \geq 0.5$ in both models we considered for the HI.

5 HOD approach for modelling Luminosity-threshold galaxy catalogues

The galaxy model described in section 3.2, and applied in the signal-to-noise analysis in section 4, is intuitive and easy to implement. Nevertheless, its limitations are many. In fact, galaxy surveys can observe sources with luminosity larger that a threshold value. Within the framework described in 3.2 , it is not possible to model the luminosity function of a sample of galaxies, and therefore both magnification and evolution bias were treated as free parameters. In this section we will partially trade the generality of this framework with a more realistic model for a galaxy catalogue.

In order to model a luminosity-threshold galaxy catalogue, we will assume the model based on the Halo Occupation Distribution (HOD) described in [104, 105]. The average number of galaxies, within a halo of mass M and with apparent magnitude below a certain threshold value m^* is modelled as a sum of two contributions: the contributions from the central galaxies and the one from the satellite galaxies [101, 106]

$$N_{\rm AV}(<m^*,M)=N_{\rm cent}(<m^*,M)+N_{\rm sat}(<m^*,M). \tag{5.1}$$

The central galaxies are modeled as a step function

$$N_{\rm cent}(<m^*,M)=\frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M-\log M_{\rm min}(m^*)}{\sigma(m^*)}\right)\right], \tag{5.2}$$

while the satellite galaxies are a product of the same step function and a power law

$$N_{\rm sat}(<m^*,M)=N_{\rm cent}(<m^*,M)\times\left(\frac{M-M_0(m^*)}{M_1(m^*)}\right)^{\alpha(m^*)}. \tag{5.3}$$

The model described above involves 5 parameters. $M_{\rm min}$ is the halo mass such that the average number of central galaxies with luminosity above the cut luminosity is $1/2$, σ regulates the efficiency at which the number of galaxies increases from the small to the large halos regime, M_0 is the cutoff mass scale for the satellite galaxies, M_1 is a normalization factor and α is the slope of the power-law that determine the number of galaxies in the highly massive halos regime. These parameters are not independent: they all depend on the threshold magnitude of the considered catalogue. In Ref. [105] the parameters have been computed for the SDSS galaxy catalogue by considering samples with different luminosity thresholds. In this section we will assume a galaxy sample that follows the same behavior. Nevertheless, the functional form in the equations (5.1), (5.2) and (5.3) can be in principle applied to other galaxy catalogues. Figure 13 (left panel) represents the average number of galaxies as a function of the halo mass, for galaxy samples for different values of the maximum absolute magnitude M_r [7].

By assuming this behaviour for $N_{\rm AV}$, we can use the same framework defined in section 3.2 to relate the shot-noise, the clustering bias and the magnification bias to the magnitude threshold, at fixed redshift. In fact, we can compute the comoving number density of galaxies as

$$\bar{n}_{\rm gal}(<M_{\rm r})=\int_0^\infty n(M,z)N_{\rm AV}(<M_{\rm r},M)dM. \tag{5.4}$$

[7] The absolute magnitude is related to the apparent magnitude m^* by $M_r=m^*-5\log_{10}\frac{d_L(z)}{d_L(z_{\rm ref})}-K(z)$, where $d_L(z)$ is the luminosity distance at a given redshift and $K(z)$ is the k-correction, which corrects the measured magnitude into the one that would be measured in the source's rest frame.

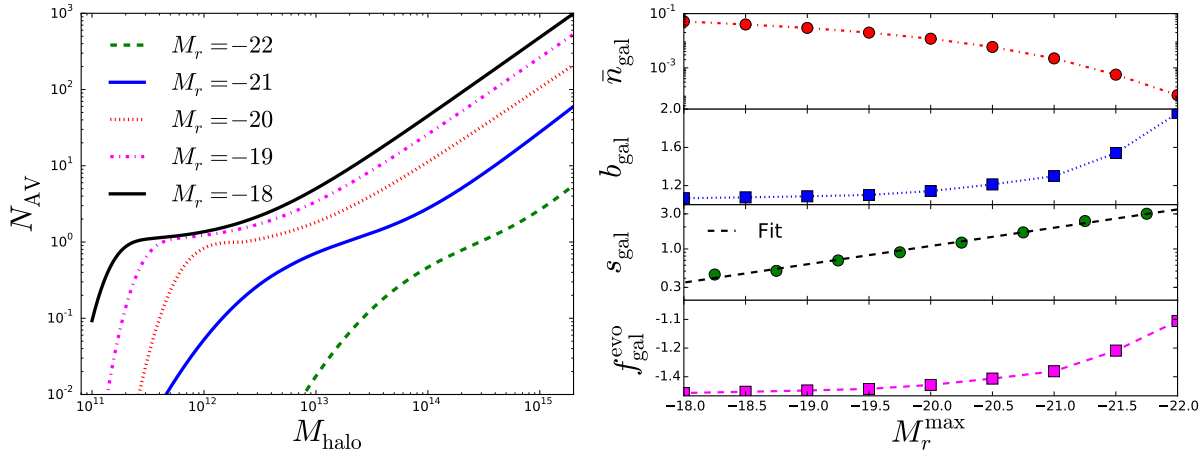


Figure 13: Left panel: Average number of galaxies within a halo as a function of the halo mass. Different lines denote galaxies samples with different values of the maximum absolute magnitude M_r. Right panel: Comoving number density (red), clustering bias (blue), magnification bias (green) and evolution bias (magenta) as a function of the maximum absolute magnitude M_r of the selected galaxy sample.

The shot-noise is the inverse of $\bar{n}_{\rm gal}(< M_{\rm r})$, while the galaxy and the magnification biases are

$$b_{\rm g}(M_{\rm r}) = \frac{1}{n_{\rm g}(z)} \int_0^\infty n(M,z) b(M,z) N_{\rm AV}(< m^*, M) dM, \tag{5.5}$$

$$s_{\rm gal}(M_{\rm r}) = \frac{\partial \log_{10}(\bar{n}_{\rm gal}(< M_{\rm r}))}{\partial M_{\rm r}}. \tag{5.6}$$

Note that since we fixed the redshift, the derivative with respect to the apparent magnitude coincides with the one respect the absolute magnitude $M_{\rm r}$. The evolution bias can be directly computed from (2.7). In figure 13 (right panel) we plot $\bar{n}_{\rm gal}(< M_{\rm r})$, $b_{\rm g}$ and $s_{\rm gal}$ as a function of the maximum magnitude of the sample. We see that samples with a smaller value of the magnitude threshold correspond to higher value of the biases and higher shot-noise (lower comoving density). Furthermore, we observe that the magnification bias increases for smaller values of the magnitude cut $M_{\rm r}$, reaching values up to ≈ 3. The evolution bias is also larger for smaller values of $M_{\rm r}$, but its value are relatively small in all the range of magnitude threshold.

Once we have the behaviour of the biases and of the shot-noise in terms of the magnitude cut, we can compute the correspondent signal to-noise for the dipole from (4.1). The result is plotted in figure 14, where different colors and line-styles refer to different values of the interferometer noise. We assumed the model 1 (the conservative one) in Table 1 for the HI, while all the galaxies properties shown in figure 13. We found that the signal-to-noise is optimized for a certain value of the limiting magnitude, which is not its maximum values (which corresponds to the minimum shot-noise), because the magnification bias, and therefore the dipole signal, has larger values if the maximum magnitude threshold is set to be smaller. Up to $M_{\rm r} \approx -21$ the signal increases faster than the noise, while for $M_{\rm r} < -21$ the shot-noise dominates over the signal growth rate. This result implies that, even if for a galaxy catalogue the expected signal-to-noise for the dipole is below the detection threshold (in figure 14 we

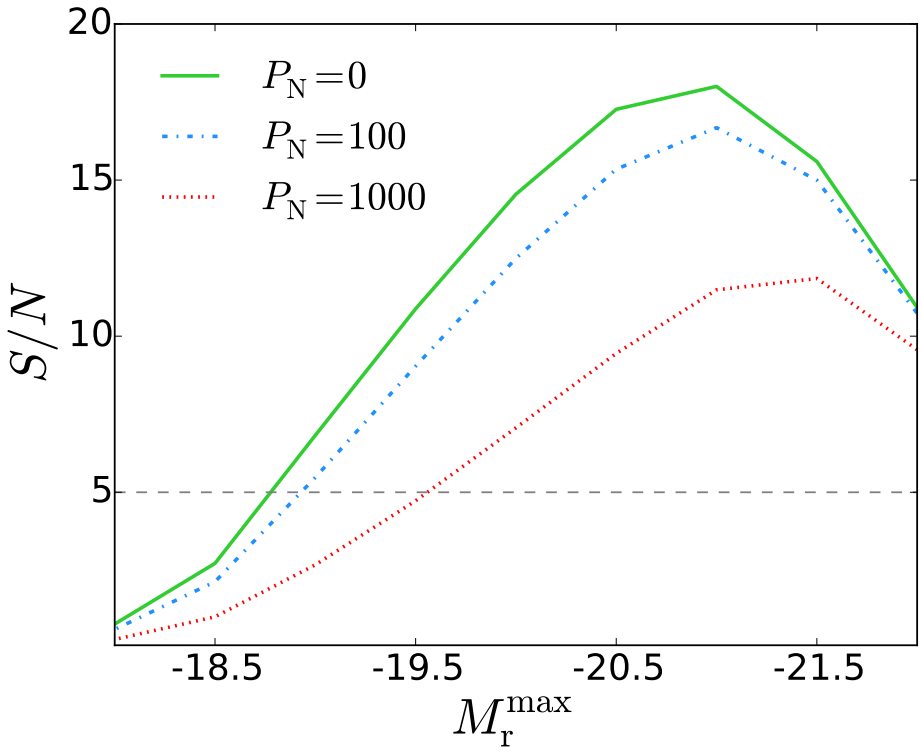


Figure 14: Signal-to-noise as a function of the magnitude threshold of the galaxy survey. The galaxies are described by the HOD model discussed in this section, while the model for the HI is the more conservative model described in the previous section (see Table 1). Different colors and line-styles denote different values of the interferometer noise (the unit is Mpc^3/h^3). The dashed horizontal line denotes a detection threshold of $S/N = 5$.

set $S/N \geq 5$ for a possible detection), it is possible to properly choose a smaller limiting magnitude for the sample and reject the galaxies with magnitude above this value, in order to amplify the S/N above the detection threshold.

6 Conclusions

In this work we studied the relativistic dipole in the cross-correlation function of HI intensity mapping and galaxies. We model the HI and galaxy parameters (clustering bias, magnification and evolution bias and shot-noise) that affects the signal in the general framework of the halo model.

We present a signal-to-noise analysis for the relativistic dipole sourced by the Doppler effect of the cross-correlation between the galaxy number counts and the HI brightness temperature. Our analysis aims to study the properties of a galaxy population that optimize the signal-to-noise ratio. We consider two HI models and in both cases we find that the signal-to-noise is suppressed for highly biased tracers, in despite that the dipole is proportional to the bias difference of the tracers. Therefore galaxies with lower bias are preferred for detecting the Doppler dipole in the cross-correlation of HI intensity mapping and galaxies. Furthermore the signal appears to be considerably larger for higher values of the magnification bias,

therefore galaxy surveys with steeper luminosity function are favored: we find that in order to have a $S/N > 5$ we need roughly $s \geq 0.5$.

Finally, we perform a similar signal-to-noise analysis to a luminosity threshold galaxy catalogue. We describe the number of galaxies with magnitude below a given threshold using a model based on the Halo Occupation Distribution. We study the relation between the parameters of the galaxy population and the limiting magnitude. In particular, we investigate the signal-to-noise of the dipole as a function the magnitude threshold. Our results, see figure 14, indicate that the maximum value for the limiting magnitude of the survey does not correspond to the higher signal-to-noise ratio. Indeed, we show how to select an optimal value for the magnitude threshold to maximize the signal-to-noise. This analysis depends on the specific Halo Occupation Distribution model that we employed, which was built to fit the SDSS data. Nevertheless, analogue methods can be applied to other galaxy catalogues and can be useful for selecting the optimal galaxy sample to measure the relativistic dipole.

We conclude with some words of caution. First of all both the modelling of the HI and the galaxy populations are based on suitable, physically motivated extensions of the halo model that are expected to be reliable. However, the modelling of HI inside galaxies in the low redshift Universe is crucial and it is not fully explored in this work how observational results of the HI content of galaxies impact on the halo model parameters. In this respect we notice that attempts of modelling the IM signal by incorporating these physical effects into the simple parametric model have been made [107]. Secondly, here we neglect any modelling of the foreground signal of the IM that we know dominates by several orders of magnitude. Foreground removal techniques are of primary importance in order to fully exploit the cross correlation signal and preliminary results are encouraging (e.g. [108]) also for the IM-galaxy cross-correlation [53]. The relatively high values of the signal-to-noise ratio of the effect under study is however suggesting that it could still be detected once these caveats are properly modelled.

Acknowledgments

We thank Francesco Montanari and Ruth Durrer for useful discussions. MV and ED are supported by the cosmoIGM starting grant. FL, EV, MV and ED are supported by the INFN PD51 grant INDARK.

Appendix

A Flat-sky versus full-sky dipole

In this Appendix we test the validity of the flat-sky approximation by comparing the flat-sky and the full-sky relativistic dipole at redshift $z_m = 0.15$. This comparison is crucial to test in which regime we can apply the flat-sky approximation and therefore which are the maximum scales that we can model in this framework. In particular, we test the dipole expansion in terms of d/r which, to lowest order, leads to the wide-angle correction, defined in Eq. (2.24), that we adopted in our analysis.

The full-sky correlation function for the cross-correlation of two populations of galaxies has been studied in detail in the literature, both in the newtonian approximation [74–76]

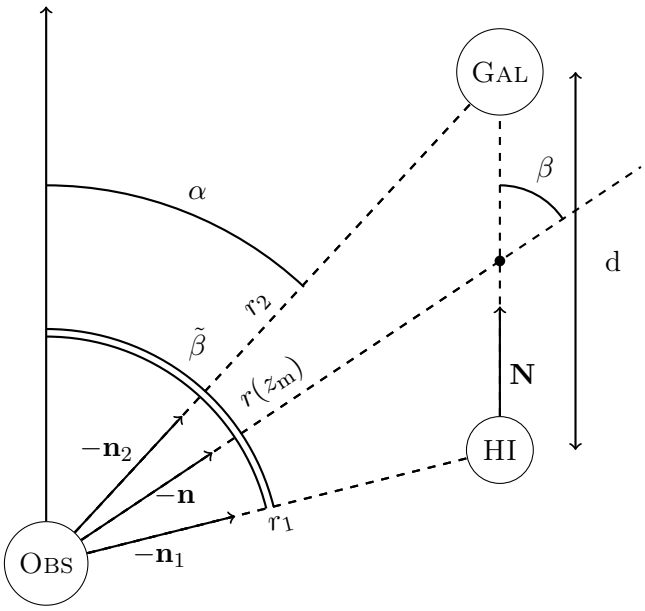


Figure 15: Coordinate system for the full-sky relativistic dipole.

and beyond [12, 14, 39, 79, 109–112]. Here we will use the same notation as [39], where the full-sky relativistic correlation function is computed for two galaxy populations, including the gravitational redshift and Doppler effects, which are the quantity relevant for this work (see the Appendix B in [39] for a detailed derivation of the relativistic part, i.e. the cross-correlation between the density plus redshift space distortion and the Doppler terms in the galaxy number counts). We will apply that expression to the cross-correlation between HI and galaxies. We will fix the values of the bias to $b_{\rm HI} = 0.6$ and $b_{\rm gal} = 1.0$, and we will compare three values of the galaxy magnification bias.

We will recast the expression for the full-sky correlation function in the way described below. The relativistic correlation function, in the coordinate system defined in section 2.3 (see also figure 15), can be written in terms of some coefficients $R_i = R_i(r_1, r_2, d)$

$$\begin{aligned}\xi^{\rm HI,gal}(r,d,\mu) = {} & R_1\cos(\alpha) + R_2\cos(\tilde{\beta}) + R_3\cos(\alpha)\cos(2\tilde{\beta}) + R_4\cos(\tilde{\beta})\cos(2\alpha) \\ & + R_5\sin(\alpha)\sin(2\tilde{\beta}) + R_6\sin(\tilde{\beta})\sin(2\alpha),\end{aligned} \tag{A.1}$$

where α is defined as $\cos(\alpha) = -\mathbf{n_2}\cdot\mathbf{N}$ and $\cos(\tilde{\beta}) = -\mathbf{n_1}\cdot\mathbf{N}$ (see figure 15), while the

coefficients R_i are

$$
\begin{aligned}
R_1 =& C_{\rm gal}(r_2)\frac{D_1(r_2)D_1(r_1)}{D_1^2(r)}f(r_2)\left[\left(b_{\rm HI}+\frac{2}{5}f(r_1)\right)\nu_1-\frac{1}{10}f(r_1)\nu_3\right],\\
R_2 =& -C_{\rm HI}(r_1)\frac{D_1(r_2)D_1(r_1)}{D_1^2(r)}f(r_1)\left[\left(b_{\rm gal}+\frac{2}{5}f(r_2)\right)\nu_1-\frac{1}{10}f(r_2)\nu_3\right],\\
R_3 =& C_{\rm gal}(r_2)\frac{D_1(r_2)D_1(r_1)}{D_1^2(r)}f(r_1)f(r_2)\frac{1}{5}\left[\nu_1-\frac{3}{2}\nu_3\right],\\
R_4 =& -C_{\rm HI}(r_1)\frac{D_1(r_2)D_1(r_1)}{D_1^2(r)}f(r_1)f(r_2)\frac{1}{5}\left[\nu_1-\frac{3}{2}\nu_3\right],\\
R_5 =& C_{\rm gal}(r_2)\frac{D_1(r_2)D_1(r_1)}{D_1^2(r)}f(r_1)f(r_2)\frac{1}{5}\left[\nu_1+\nu_3\right],\\
R_6 =& -C_{\rm HI}(r_1)\frac{D_1(r_2)D_1(r_1)}{D_1^2(r)}f(r_1)f(r_2)\frac{1}{5}\left[\nu_1+\nu_3\right],
\end{aligned}
\tag{A.2}
$$

where $C_{\rm gal}$ and $C_{\rm HI}$ are defined in (2.18), D_1 denotes the linear growth factor, and $\nu_{\ell=1,3}$ are defined through

$$
\nu_\ell(d)=\int\frac{k^2dk}{2\pi^2}\left(\frac{\mathcal{H}}{k}\right)P(k,z_m)j_\ell(k\,d). \tag{A.3}
$$

All the quantities involved in the computation of the correlation function in (A.1) can be expressed as a function of r, d and μ by using the following simple geometric relations

$$
\begin{aligned}
r_1 &= \frac{1}{2}\sqrt{d^2+4r^2-4d\mu r}\,, & r_2 &= \frac{1}{2}\sqrt{d^2+4r^2+4d\mu r}\,,\\
\cos(\tilde{\beta}) &= \frac{-d+2r\mu}{\sqrt{4r^2+d^2-4dr\mu}}\,, & \cos(\alpha) &= \frac{d+2r\mu}{\sqrt{4r^2+d^2+4dr\mu}}\,.
\end{aligned}
\tag{A.4}
$$

The dipole can then be computed as

$$
\xi_1^{\text{full sky}}=\frac{3}{2}\int_{-1}^{1}\xi^{\rm HI,gal}(r,d,\mu)L_1(\mu)d\mu. \tag{A.5}
$$

We see the full-sky correlation function depends in a not trivial way from the angular coordinate μ. Therefore, the angular integral in (A.5) cannot be performed analytically, as in the flat-sky approximation, but it needs to be solved numerically.

A similar computation can be done for the wide-angle corrections. The full sky expression can be written in terms of some coefficients $S_i = S_i(r_1, r_2, d)$ [39]

$$
\xi^{\rm WA}(r,d,\mu)=S_1+S_2\cos(2\tilde{\beta})+S_3\cos(2\alpha)+S_4\cos(2\alpha)\cos(2\tilde{\beta})+S_5\sin(2\tilde{\beta})\sin(2\alpha), \tag{A.6}
$$

$$
\begin{aligned}
S_1 =& \frac{D_1(r_2)D_1(r_1)}{D_1^2(r)} f(r_2)\left[\left(b_{\rm HI}b_{\rm gal} + \frac{b_{\rm HI}}{3}f(r) + \frac{b_{\rm gal}}{3}f(r) + \frac{2}{15}f^2(r)\right)\mu_0,\right.\\
&\left. - \frac{1}{3}\left(\frac{b_{\rm HI}}{2}f(r) + \frac{b_{\rm gal}}{2}f(r) + \frac{2}{7}f^2(r)\right)\mu_2 + \frac{3}{140}f^2(r)\mu_4\right],\\
S_2 =& -\frac{D_1(r_2)D_1(r_1)}{D_1^2(r)}\left[\left(\frac{b_{\rm gal}}{2}f(r) + \frac{3}{14}f^2(r)\right)\mu_2 - \frac{1}{28}f^2(r)\mu_4\right],\\
S_3 =& -\frac{D_1(r_2)D_1(r_1)}{D_1^2(r)}\left[\left(\frac{b_{\rm HI}}{2}f(r) + \frac{3}{14}f^2(r)\right)\mu_2 - \frac{1}{28}f^2(r)\mu_4\right],\\
S_4 =& \frac{D_1(r_2)D_1(r_1)}{D_1^2(r)} f(r)^2\frac{1}{5}\left[\mu_0 - \frac{1}{21}\mu_2 + \frac{19}{140}\mu_4\right],\\
S_5 =& \frac{D_1(r_2)D_1(r_1)}{D_1^2(r)} f(r)^2\frac{1}{5}\left[\mu_0 - \frac{1}{21}\mu_2 - \frac{4}{35}\mu_4\right],
\end{aligned}
\tag{A.7}
$$

where $\mu_{\ell=0,2,4}$ are defined by

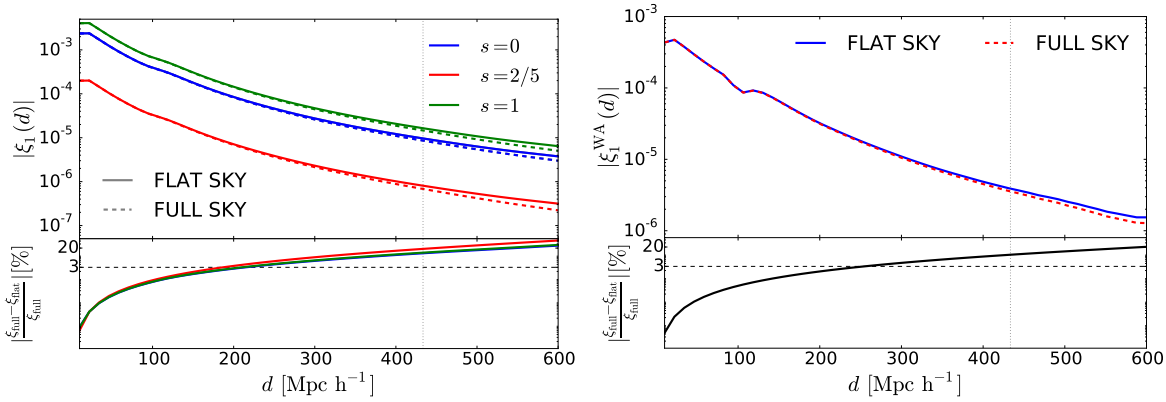
$$
\mu_\ell(d) = \int \frac{k^2 dk}{2\pi^2} P(k, z_m) j_\ell(k\, d). \tag{A.8}
$$

The dipole of the wide-angle effect, in full sky, can be computed as

$$
\xi_1^{\rm WA,\ full\ sky} = \frac{3}{2}\int_{-1}^{1} \xi^{\rm WA}(r, d, \mu) L_1(\mu) d\mu. \tag{A.9}
$$

In figure 16 we compare the full-sky dipole (dashed lines) and the corresponding dipole in the flat-sky approximation (continuous lines). Figure 16a represents the Doppler dipole, computed from (A.5) (full-sky) and (2.15) (flat-sky), while figure 16b shows the wide-angle dipole, computed from (A.9) (full-sky) and (2.23) (flat-sky). Different colors refer to different values of the magnification bias. In the top panel we plot the absolute value of the dipole. At small scale, the flat-sky approximation fairly reproduces the full-sky signal, while the full-sky dipoles significantly departs from the one in flat-sky at large scales. Note that the full-sky signal always results to be smaller, in absolute value, than the approximated one. Indeed two sources appear closer in the flat-sky limit and therefore they seems to be more correlated. In the bottom panel we plot the relative difference between the full-sky and the flat-sky dipole, in percentage. The black dashed line denotes a 3% difference between the two quantities, which we set as a threshold for the flat-sky approximation to be valid. We see that for scales larger than roughly 200Mpc/h the relative difference is beyond the threshold value and that it rapidly increases at larger scales. Nevertheless the amount of physical information at these scales is irrelevant as we see from figure 10.

In figure 17 we show the difference between the estimator defined in (2.24), which is unbiased from wide-angle effect, in full-sky and the one in flat-sky, at lower order in the d/r expansion. We see that at scales $d \leq 200\,{\rm Mpc/h}$ the approximated dipole fairly reproduces the full-sky quantity. Therefore, we set the maximum scale of our analysis to be $d_{\rm max} = 200\,{\rm Mpc/h}$.



(a) Doppler dipole. **(b)** Wide-angle dipole.

Figure 16: Comparison between the dipole of the cross-correlation HI-galaxies (right panels) and the wide-angle dipole (left panels) computed within the flat-sky approximation (continuous line) and the full-sky dipole signal (dashed line). Top panels: absolute value of the dipole. Bottom panels: relative difference between the full signal and the flat-sky approximation. The dashed line denotes a difference of relative difference of 3%. Different colors denote different values of the galaxy magnification bias. The dotted vertical lines denotes the distance $d = r$.

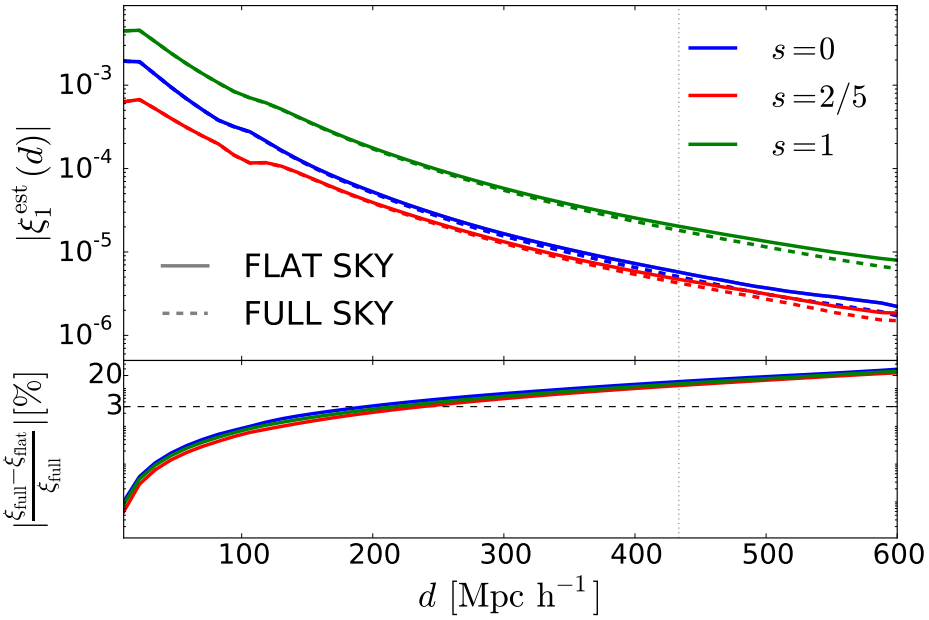


Figure 17: Comparison between the estimator of the cross-correlation HI-galaxies computed within the flat-sky approximation (continuous line) and the full-sky dipole signal (dashed line). Top panel: absolute value of the estimator. Bottom panel: relative difference between the full signal and the approximated signal. Different colors denote different values of the galaxy magnification bias.

B Contribution to the wide-angle correction to the covariance

The estimator we used in the signal-to-noise analysis, which is unbiased from wide-angle effects (at least at the leading order in the d/r expansion), involved the measurement of both the

dipole of the cross-correlation of galaxies and HI and the quadrupole of the autocorrelations of both tracers

$$\hat{\xi}_1(d,r) = \hat{\xi}_1(d,z_1) - \frac{3}{10}\left(\hat{\xi}_2^{gal}(d,z_1) - \hat{\xi}_2^{\rm HI}(d,z_1)\right)\frac{d}{r}. \tag{B.1}$$

We split the covariance for the estimator in Eq. (B.1) in two contributions [8],

$$\mathrm{COV}^{\rm est}(d_1,d_2) = \mathrm{COV}^{\rm dip}(d_1,d_2) + \mathrm{COV}^{\rm WA}(d_1,d_2), \tag{B.2}$$

where $\mathrm{COV}^{\rm dip}(d_1,d_2) = \langle\hat{\xi}_1(d_1)\hat{\xi}_1(d_2)\rangle$ is the dipole contribution to the estimator it is given by Eq. (2.25), while $\mathrm{COV}^{\rm WA}(d_1,d_2)$ is the contributions of the wide-angle correction,

$$\begin{aligned}\mathrm{COV}^{\rm WA}(d_1,d_2) = \left(\tfrac{3}{10r}\right)^2 d_1\, d_2 \left(\mathrm{COV}_{2,2}^{\rm gal,\ gal} + \mathrm{COV}_{2,2}^{\rm HI,\ HI} + 2\,\mathrm{COV}_{2,2}^{\rm gal,\ HI}\right)\\ -\tfrac{3}{10}\tfrac{d_1+d_2}{r}\left(\mathrm{COV}_{1,2}^{\rm Dip,\ gal} - \mathrm{COV}_{1,2}^{\rm Dip,\ HI}\right),\end{aligned} \tag{B.3}$$

where

$$\begin{aligned}\mathrm{COV}_{2,2}^{\rm gal,\ gal}(d_1,d_2) &= \left\langle\hat{\xi}_2^{gal}(d_1)\hat{\xi}_2^{gal}(d_2)\right\rangle - \left\langle\hat{\xi}_2^{gal}(d_1)\right\rangle\left\langle\hat{\xi}_2^{gal}(d_2)\right\rangle,\\ \mathrm{COV}_{2,2}^{\rm HI,\ HI}(d_1,d_2) &= \left\langle\hat{\xi}_2^{HI}(d_1)\hat{\xi}_2^{HI}(d_2)\right\rangle - \left\langle\hat{\xi}_2^{HI}(d_1)\right\rangle\left\langle\hat{\xi}_2^{HI}(d_2)\right\rangle,\\ \mathrm{COV}_{2,2}^{\rm gal,\ HI}(d_1,d_2) &= \left\langle\hat{\xi}_2^{gal}(d_1)\hat{\xi}_2^{HI}(d_2)\right\rangle - \left\langle\hat{\xi}_2^{gal}(d_1)\right\rangle\left\langle\hat{\xi}_2^{HI}(d_2)\right\rangle,\\ \mathrm{COV}_{1,2}^{\rm Dip,\ gal}(d_1,d_2) &= \left\langle\hat{\xi}_1(d_1)\hat{\xi}_2^{gal}(d_2)\right\rangle - \left\langle\hat{\xi}_1(d_1)\right\rangle\left\langle\hat{\xi}_2^{gal}(d_2)\right\rangle,\\ \mathrm{COV}_{1,2}^{\rm Dip,\ HI}(d_1,d_2) &= \left\langle\hat{\xi}_1(d_1)\hat{\xi}_2^{HI}(d_2)\right\rangle - \left\langle\hat{\xi}_1(d_1)\right\rangle\left\langle\hat{\xi}_2^{HI}(d_2)\right\rangle.\end{aligned} \tag{B.4}$$

The first line in (B.3) is the autocorrelation of the wide-angle contribution to the dipole, and the second one is the cross-correlation of the dipole with the wide-angle correction. Each term of the wide-angle contribution to the covariance can be computed similarly to the dipole contribution by applying the general formula in [49].

The cosmic variance contribution to the autocorrelation can be written as

$$\mathrm{COV}_{\rm CVCV}^{\rm WA,\ auto} = \left(\frac{3}{10r}\right)^2\frac{50 d_1 d_2}{V}\int\frac{k^2dk}{2\pi^2}j_2(kd_1)j_2(kd_2)\sum_{L_1,L_2}G_{22}^{L1L2}[P_{L1}^{\rm gal}P_{L2}^{\rm gal}+P_{L1}^{\rm HI}P_{L2}^{\rm HI}-2P_{L1}^{\rm gal}P_{L2}^{\rm HI}]; \tag{B.5}$$

the mixed cosmic variance – noise contribution to the autocorrelation, including both shot-noise and interferometer noise, is given by

$$\mathrm{COV}_{\rm CVNoise}^{\rm WA,\ auto} = \left(\frac{3}{10r}\right)^2\frac{50 d_1 d_2}{V}\int\frac{k^2dk}{2\pi^2}j_2(kd_1)j_2(kd_2)\sum_{L}G_{22}^{0L}[1/n_{\rm gal}P_L^{\rm gal}+(N_{\rm HI}+P_{\rm N})P_L^{\rm HI}]; \tag{B.6}$$

while the noise contribution to the covariance is

$$\mathrm{COV}_{\rm NoiseNoise}^{\rm WA,\ auto} = \left(\frac{3}{10r}\right)^2\frac{10}{4\pi L_p V}\delta_{d_1,d_2}\left[(N_{\rm HI}+P_{\rm N})^2+\frac{1}{n_{\rm gal}^2}\right]. \tag{B.7}$$

[8] We assume the redshift to be constant and we drop the redshift dependence from our notation

The coefficients G_{22}^{L1L2} in (B.5) and (B.6) are defined in terms of the Wigner $3j$ symbols

$$G_{22}^{L1L2} = \sum_L (2L+1) \begin{pmatrix} 2 & 2 & L \\ 0 & 0 & 0 \end{pmatrix}^2 \begin{pmatrix} L_1 & L_2 & L \\ 0 & 0 & 0 \end{pmatrix}^2. \tag{B.8}$$

The cosmic variance contribution to the cross-correlation term in (B.3) can be written as

$$\mathrm{COV}_{\mathrm{CVCV}}^{\mathrm{WA,\ cross}} = \left(\frac{3}{10}\frac{d_1+d_2}{r}\right)\frac{30}{V}\int \frac{k^2 dk}{2\pi^2} j_1(kd_1) j_2(kd_2) \sum_{L_1,L_2} G_{12}^{L1L2} P_{L1}[P_{L2}^{\mathrm{gal}} - P_{L2}^{\mathrm{HI}}]; \tag{B.9}$$

the cosmic variance x noise contribution to the cross-correlation is given by

$$\mathrm{COV}_{\mathrm{CVNoise}}^{\mathrm{WA,\ cross}} = \left(\frac{3}{10}\frac{d_1+d_2}{r}\right)\frac{30}{V}\int \frac{k^2 dk}{2\pi^2} j_1(kd_1) j_2(kd_2) \sum_L \left(\frac{1}{n_{\mathrm{gal}}} - N_{\mathrm{HI}} - P_{\mathrm{N}}\right) G_{12}^{0L} P_L. \tag{B.10}$$

The coefficients G_{12}^{L1L2} are defined similarly to the coefficients involved in the autocorrelation term,

$$G_{12}^{L1L2} = \sum_L (2L+1) \begin{pmatrix} 1 & 2 & L \\ 0 & 0 & 0 \end{pmatrix}^2 \begin{pmatrix} L_1 & L_2 & L \\ 0 & 0 & 0 \end{pmatrix}^2. \tag{B.11}$$

The purely noise contribution to the cross-correlation term vanishes.

References

[1] **Planck** Collaboration, P. A. R. Ade et al., *Planck 2015 results. XIII. Cosmological parameters*, arXiv:1502.01589.

[2] **BOSS** Collaboration, S. Alam et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, *Mon. Not. Roy. Astron. Soc.* **470** (2017), no. 3 2617–2652, [arXiv:1607.03155].

[3] S. Weinberg, *The Cosmological Constant Problem*, *Rev. Mod. Phys.* **61** (1989) 1–23.

[4] F. Zwicky, *On the Masses of Nebulae and of Clusters of Nebulae*, *Astrophysical Journal* **86** (Oct., 1937) 217.

[5] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, *A direct empirical proof of the existence of dark matter*, *Astrophys. J.* **648** (2006) L109–L113, [astro-ph/0608407].

[6] P. Bull et al., *Beyond ΛCDM: Problems, solutions, and the road ahead*, *Phys. Dark Univ.* **12** (2016) 56–99, [arXiv:1512.05356].

[7] J. Yoo, A. L. Fitzpatrick, and M. Zaldarriaga, *New perspective on galaxy clustering as a cosmological probe: General relativistic effects*, *Physical Review D* **80** (Oct., 2009) 083514, [arXiv:0907.0707].

[8] J. Yoo, *General relativistic description of the observed galaxy power spectrum: Do we understand what we measure?*, *Physical Review D* **82** (Oct., 2010) 083508, [arXiv:1009.3021].

[9] C. Bonvin and R. Durrer, *What galaxy surveys really measure*, *Phys. Rev.* **D84** (2011) 063505, [arXiv:1105.5280].

[10] A. Challinor and A. Lewis, *The linear power spectrum of observed source number counts*, *Phys. Rev.* **D84** (2011) 043516, [arXiv:1105.5292].

[11] F. Schmidt and D. Jeong, *Cosmic Rulers*, *Phys. Rev.* **D86** (2012) 083527, [arXiv:1204.3625].

[12] D. Bertacca, R. Maartens, A. Raccanelli, and C. Clarkson, *Beyond the plane-parallel and Newtonian approach: Wide-angle redshift distortions and convergence in general relativity*, *JCAP* **1210** (2012) 025, [arXiv:1205.5221].

[13] A. Raccanelli, D. Bertacca, R. Maartens, C. Clarkson, and O. Doré, *Lensing and time-delay contributions to galaxy correlations*, *Gen. Rel. Grav.* **48** (2016), no. 7 84, [arXiv:1311.6813].

[14] A. Raccanelli, D. Bertacca, O. Doré, and R. Maartens, *Large-scale 3D galaxy correlation function and non-Gaussianity*, *JCAP* **1408** (2014) 022, [arXiv:1306.6646].

[15] J. Yoo and M. Zaldarriaga, *Beyond the Linear-Order Relativistic Effect in Galaxy Clustering: Second-Order Gauge-Invariant Formalism*, *Phys. Rev.* **D90** (2014), no. 2 023513, [arXiv:1406.4140].

[16] D. Bertacca, R. Maartens, and C. Clarkson, *Observed galaxy number counts on the lightcone up to second order: I. Main result*, *JCAP* **1409** (2014), no. 09 037, [arXiv:1405.4403].

[17] D. Bertacca, *Observed galaxy number counts on the light cone up to second order: III. Magnification bias*, *Class. Quant. Grav.* **32** (2015), no. 19 195011, [arXiv:1409.2024].

[18] E. Di Dio, R. Durrer, G. Marozzi, and F. Montanari, *Galaxy number counts to second order and their bispectrum*, *JCAP* **1412** (2014) 017, [arXiv:1407.0376]. [Erratum: JCAP1506,no.06,E01(2015)].

[19] E. Di Dio, R. Durrer, G. Marozzi, and F. Montanari, *The bispectrum of relativistic galaxy number counts*, *JCAP* **1601** (2016) 016, [arXiv:1510.04202].

[20] J. T. Nielsen and R. Durrer, *Higher order relativistic galaxy number counts: dominating terms*, *JCAP* **1703** (2017), no. 03 010, [arXiv:1606.02113].

[21] O. Umeh, S. Jolicoeur, R. Maartens, and C. Clarkson, *A general relativistic signature in the galaxy bispectrum: the local effects of observing on the lightcone*, *JCAP* **1703** (2017), no. 03 034, [arXiv:1610.03351].

[22] E. Di Dio, *Lensing smoothing of BAO wiggles*, *JCAP* **1703** (2017), no. 03 016, [arXiv:1609.09044].

[23] S. Jolicoeur, O. Umeh, R. Maartens, and C. Clarkson, *Imprints of local lightcone projection effects on the galaxy bispectrum. Part II*, *JCAP* **1709** (2017), no. 09 040, [arXiv:1703.09630].

[24] D. Bertacca, A. Raccanelli, N. Bartolo, M. Liguori, S. Matarrese, and L. Verde, *Relativistic wide-angle galaxy bispectrum on the light-cone*, arXiv:1705.09306.

[25] C. Bonvin, *Effect of Peculiar Motion in Weak Lensing*, *Phys. Rev.* **D78** (2008) 123530, [arXiv:0810.0180].

[26] A. Hall, C. Bonvin, and A. Challinor, *Testing General Relativity with 21-cm intensity mapping*, *Phys. Rev.* **D87** (2013), no. 6 064026, [arXiv:1212.0728].

[27] V. Iršič, E. Di Dio, and M. Viel, *Relativistic effects in Lyman-α forest*, *JCAP* **1602** (2016), no. 02 051, [arXiv:1510.03436].

[28] N. Kaiser, *Clustering in real space and in redshift space*, *Monthly Notices of the Royal Astronomical Society* **227** (July, 1987).

[29] S. Camera, M. G. Santos, and R. Maartens, *Probing primordial non-Gaussianity with SKA galaxy redshift surveys: a fully relativistic analysis*, *Mon. Not. Roy. Astron. Soc.* **448** (2015), no. 2 1035–1043, [arXiv:1409.8286].

[30] S. Camera, R. Maartens, and M. G. Santos, *Einstein's legacy in galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **451** (2015), no. 1 L80–L84, [arXiv:1412.4781].

[31] A. Raccanelli, F. Montanari, D. Bertacca, O. Doré, and R. Durrer, *Cosmological Measurements with General Relativistic Galaxy Correlations*, *JCAP* **1605** (2016), no. 05 009, [arXiv:1505.06179].

[32] F. Montanari and R. Durrer, *Measuring the lensing potential with tomographic galaxy number counts*, *JCAP* **1510** (2015), no. 10 070, [arXiv:1506.01369].

[33] E. Di Dio, F. Montanari, A. Raccanelli, R. Durrer, M. Kamionkowski, and J. Lesgourgues, *Curvature constraints from Large Scale Structure*, *JCAP* **1606** (2016), no. 06 013, [arXiv:1603.09073].

[34] W. Cardona, R. Durrer, M. Kunz, and F. Montanari, *Lensing convergence and the neutrino mass scale in galaxy redshift surveys*, *Phys. Rev.* **D94** (2016), no. 4 043007, [arXiv:1603.06481].

[35] E. Villa, E. Di Dio, and F. Lepori, *Lensing convergence in galaxy clustering in ?CDM and beyond*, *JCAP* **1804** (2018), no. 04 033, [arXiv:1711.07466].

[36] A. Raccanelli, D. Bertacca, D. Jeong, M. C. Neyrinck, and A. S. Szalay, *Doppler term in the galaxy two-point correlation function: wide-angle, velocity, Doppler lensing and cosmic acceleration effects*, arXiv:1602.03186.

[37] P. McDonald, *Gravitational redshift and other redshift-space distortions of the imaginary part of the power spectrum*, *JCAP* **0911** (2009) 026, [arXiv:0907.5220].

[38] J. Yoo, N. Hamaus, U. Seljak, and M. Zaldarriaga, *Going beyond the redshift-space distortion formula: a full general relativistic account of the effects and their detectability in galaxy clustering*, *Phys. Rev.* **D86** (2012) 063514, [arXiv:1206.5809].

[39] C. Bonvin, L. Hui, and E. Gaztanaga, *Asymmetric galaxy correlation functions*, *Phys. Rev.* **D89** (2014), no. 8 083535, [arXiv:1309.1321].

[40] D. Alonso and P. G. Ferreira, *Constraining ultralarge-scale cosmology with multiple tracers in optical and radio surveys*, *Phys. Rev.* **D92** (2015), no. 6 063525, [arXiv:1507.03550].

[41] J. Fonseca, S. Camera, M. Santos, and R. Maartens, *Hunting down horizon-scale effects with multi-wavelength surveys*, *Astrophys. J.* **812** (2015), no. 2 L22, [arXiv:1507.04605].

[42] C. Bonvin, L. Hui, and E. Gaztanaga, *Optimising the measurement of relativistic distortions in large-scale structure*, *JCAP* **1608** (2016), no. 08 021, [arXiv:1512.03566].

[43] E. Gaztanaga, C. Bonvin, and L. Hui, *Measurement of the dipole in the cross-correlation function of galaxies*, *JCAP* **1701** (2017), no. 01 032, [arXiv:1512.03918].

[44] L. Dai, M. Kamionkowski, E. D. Kovetz, A. Raccanelli, and M. Shiraishi, *Antisymmetric galaxy cross-correlations as a cosmological probe*, *Phys. Rev.* **D93** (2016), no. 2 023507, [arXiv:1507.05618]. [Phys. Rev.D93,023507(2016)].

[45] M. Borzyszkowski, D. Bertacca, and C. Porciani, *LIGER: mock relativistic light-cones from Newtonian simulations*, *Mon. Not. Roy. Astron. Soc.* **471** (2017) 3899, [arXiv:1703.03407].

[46] L. R. Abramo and D. Bertacca, *Disantangling the effects of Doppler velocity and primordial non-Gaussianity in galaxy power spectra*, arXiv:1706.01834.

[47] L. Lombriser, J. Yoo, and K. Koyama, *Relativistic effects in galaxy clustering in a parametrized post-Friedmann universe*, *Phys. Rev.* **D87** (2013) 104019, [arXiv:1301.3132].

[48] C. Bonvin, *Isolating relativistic effects in large-scale structure*, *Class. Quant. Grav.* **31** (2014), no. 23 234002, [arXiv:1409.2224].

[49] A. Hall and C. Bonvin, *Measuring cosmic velocities with 21 cm intensity mapping and galaxy redshift survey cross-correlation dipoles*, *Phys. Rev.* **D95** (2017), no. 4 043530, [arXiv:1609.09252].

[50] C. Bonvin, S. Andrianomena, D. Bacon, C. Clarkson, R. Maartens, T. Moloi, and P. Bull, *Dipolar modulation in the size of galaxies: The effect of Doppler magnification*, *Mon. Not. Roy. Astron. Soc.* **472** (2017) 3936, [arXiv:1610.05946].

[51] T.-C. Chang, U.-L. Pen, K. Bandura, and J. B. Peterson, *An intensity map of hydrogen 21-cm emission at redshift $z \sim 0.8$*, *Nature* **466** (July, 2010) 463–465.

[52] K. W. Masui et al., *Measurement of 21 cm brightness fluctuations at $z \sim 0.8$ in cross-correlation*, *Astrophys. J.* **763** (2013) L20, [arXiv:1208.0331].

[53] 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*, *JCAP* **3** (Mar., 2015) 034, [arXiv:1410.7393].

[54] I. P. Carucci, F. Villaescusa-Navarro, and M. Viel, *The cross-correlation between 21 cm intensity mapping maps and the Lyα forest in the post-reionization era*, *JCAP* **4** (Apr., 2017) 001, [arXiv:1611.07527].

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

[56] R. P. Norris et al., *EMU: Evolutionary Map of the Universe*, *Publ. Astron. Soc. Austral.* **28** (2011) 215, [arXiv:1106.3219].

[57] M. Bilicki et al., *WISE x SuperCOSMOS photometric redshift catalog: 20 million galaxies over 3pi steradians*, arXiv:1607.01182. [Astrophys. J. Suppl.225,5(2016)].

[58] http://www.lofar.org.

[59] http://www.mwatelescope.org.

[60] http://gmrt.ncra.tifr.res.in.

[61] "The canadian hydrogen intensity mapping experiment." http://chime.phas.ubc.ca/.

[62] https://www.atnf.csiro.au/projects/askap/index.html.

[63] http://www.ska.ac.za/gallery/meerkat/.

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

[65] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, *ArXiv e-prints* (Apr., 2011) [arXiv:1104.2932].

[66] D. Blas, J. Lesgourgues, and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes*, *JCAP* **1107** (2011) 034, [arXiv:1104.2933].

[67] E. Di Dio, F. Montanari, J. Lesgourgues, and R. Durrer, *The CLASSgal code for Relativistic Cosmological Large Scale Structure*, *JCAP* **1311** (2013) 044, [arXiv:1307.1459].

[68] D. Alonso, P. Bull, P. G. Ferreira, R. Maartens, and M. Santos, *Ultra large-scale cosmology in next-generation experiments with single tracers*, *Astrophys. J.* **814** (2015), no. 2 145, [arXiv:1505.07596].

[69] J. B. Peterson et al., *21 cm Intensity Mapping*, arXiv:0902.3091.

[70] P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos, *Late-time cosmology with 21cm intensity mapping experiments*, *Astrophys. J.* **803** (2015), no. 1 21, [arXiv:1405.1452].

[71] **BOSS** Collaboration, A. J. Ross et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Analysis of potential systematics*, *Mon. Not. Roy. Astron. Soc.* **424** (2012) 564, [arXiv:1203.6499].

[72] O. Lahav, P. B. Lilje, J. R. Primack, and M. J. Rees, *Dynamical effects of the cosmological constant*, *Mon. Not. Roy. Astron. Soc.* **251** (1991) 128–136.

[73] S. M. Carroll, W. H. Press, and E. L. Turner, *The cosmological constant*, *Annual Review of Astronomy and Astrophysics* **30** (1992) 499–542.

[74] A. S. Szalay, T. Matsubara, and S. D. Landy, *Redshift space distortions of the correlation function in wide angle galaxy surveys*, *Astrophys. J.* **498** (1998) L1, [astro-ph/9712007].

[75] I. Szapudi, *Wide angle redshift distortions revisited*, *Astrophys. J.* **614** (2004) 51–55, [astro-ph/0404477].

[76] P. Papai and I. Szapudi, *Non-Perturbative Effects of Geometry in Wide-Angle Redshift Distortions*, *Mon. Not. Roy. Astron. Soc.* **389** (2008) 292, [arXiv:0802.2940].

[77] T. Matsubara, *The Correlation function in redshift space: General formula with wide angle effects and cosmological distortions*, *Astrophys. J.* **535** (2000) 1, [astro-ph/9908056].

[78] S. Bharadwaj, *Radial Redshift Space Distortions*, *Astrophys. J.* **516** (1999) 507–518, [astro-ph/9812274].

[79] A. Raccanelli, L. Samushia, and W. J. Percival, *Simulating Redshift-Space Distortions for Galaxy Pairs with Wide Angular Separation*, *Mon. Not. Roy. Astron. Soc.* **409** (2010) 1525, [arXiv:1006.1652].

[80] L. Samushia, W. J. Percival, and A. Raccanelli, *Interpreting large-scale redshift-space distortion measurements*, *Mon. Not. Roy. Astron. Soc.* **420** (2012) 2102–2119, [arXiv:1102.1014].

[81] A. Raccanelli, D. Bertacca, D. Pietrobon, F. Schmidt, L. Samushia, N. Bartolo, O. Dore, S. Matarrese, and W. J. Percival, *Testing Gravity Using Large-Scale Redshift-Space Distortions*, *Mon. Not. Roy. Astron. Soc.* **436** (2013) 89–100, [arXiv:1207.0500].

[82] J. N. Grieb, A. G. Sánchez, S. Salazar-Albornoz, and C. Dalla Vecchia, *Gaussian covariance matrices for anisotropic galaxy clustering measurements*, *Mon. Not. Roy. Astron. Soc.* **457** (2016), no. 2 1577–1592, [arXiv:1509.04293].

[83] K. Bandura et al., *Canadian Hydrogen Intensity Mapping Experiment (CHIME) Pathfinder*, *Proc. SPIE Int. Soc. Opt. Eng.* **9145** (2014) 22, [arXiv:1406.2288].

[84] A. Cooray and R. K. Sheth, *Halo models of large scale structure*, *Phys. Rept.* **372** (2002) 1–129, [astro-ph/0206508].

[85] U. Seljak, *Analytic model for galaxy and dark matter clustering*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 203, [astro-ph/0001493].

[86] J. A. Peacock and R. E. Smith, *Halo occupation numbers and galaxy bias*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 1144, [astro-ph/0005010].

[87] R. Scoccimarro, R. K. Sheth, L. Hui, and B. Jain, *How many galaxies fit in a halo? Constraints on galaxy formation efficiency from spatial clustering*, *Astrophys. J.* **546** (2001) 20–34, [astro-ph/0006319].

[88] F. Villaescusa-Navarro, M. Viel, K. K. Datta, and T. R. Choudhury, *Modeling the neutral hydrogen distribution in the post-reionization Universe: intensity mapping*, *JCAP* **1409** (2014), no. 09 050, [arXiv:1405.6713].

[89] 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*, *Mon. Not. Roy. Astron. Soc.* **471** (2017) 1788, [arXiv:1609.05157].

[90] H. Padmanabhan and A. Refregier, *Constraining a halo model for cosmological neutral hydrogen*, *Mon. Not. Roy. Astron. Soc.* **464** (2017), no. 4 4008–4017, [arXiv:1607.01021].

[91] H. Padmanabhan, A. Refregier, and A. Amara, *A halo model for cosmological neutral*

hydrogen: abundances and clustering, *Mon. Not. Roy. Astron. Soc.* **469** (2017) 2323, [arXiv:1611.06235].

[92] J. L. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. S. Warren, G. Yepes, S. Gottlober, and D. E. Holz, *Toward a halo mass function for precision cosmology: The Limits of universality*, *Astrophys. J.* **688** (2008) 709–728, [arXiv:0803.2706].

[93] J. L. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. S. Warren, G. Yepes, S. Gottlober, and D. E. Holz, *The large-scale bias of dark matter halos: Numerical calibration and model tests*, *The Astrophysical Journal* **724** (2010), no. 2 878.

[94] F. Marulli, C. Carbone, M. Viel, L. Moscardini, and A. Cimatti, *Effects of Massive Neutrinos on the Large-Scale Structure of the Universe*, *Mon. Not. Roy. Astron. Soc.* **418** (2011) 346, [arXiv:1103.0278].

[95] F. Villaescusa-Navarro, P. Bull, and M. Viel, *Weighing neutrinos with cosmic neutral hydrogen*, *Astrophys. J.* **814** (2015), no. 2 146, [arXiv:1507.05102].

[96] F. Villaescusa-Navarro et al., *Neutral hydrogen in galaxy clusters: impact of AGN feedback and implications for intensity mapping*, *Mon. Not. Roy. Astron. Soc.* **456** (2016), no. 4 3553–3570, [arXiv:1510.04277].

[97] H.-S. Kim, J. S. B. Wyithe, C. M. Baugh, C. d. P. Lagos, C. Power, and J. Park, *The spatial distribution of neutral hydrogen as traced by low hi mass galaxies*, *Monthly Notices of the Royal Astronomical Society* **465** (2017), no. 1 111, [arXiv:1603.02383].

[98] C. Evoli, P. Salucci, A. Lapi, and L. Danese, *The H I Content of Local Late-type Galaxies*, *ApJ* **743** (Dec., 2011) 45, [arXiv:1110.4199].

[99] C. Mancuso, A. Lapi, I. Prandoni, I. Obi, J. Gonzalez-Nuevo, F. Perrotta, A. Bressan, A. Celotti, and L. Danese, *Galaxy Evolution in the Radio Band: The Role of Star-forming Galaxies and Active Galactic Nuclei*, *ApJ* **842** (June, 2017) 95, [arXiv:1705.06539].

[100] M. J. White, L. Hernquist, and V. Springel, *The Halo model and numerical simulations*, *Astrophys. J.* **550** (2001) L129–L132, [astro-ph/0012518].

[101] Z. Zheng, A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Dave, C. S. Frenk, N. Katz, and C. G. Lacey, *Theoretical models of the halo occupation distribution: Separating central and satellite galaxies*, *Astrophys. J.* **633** (2005) 791–809, [astro-ph/0408564].

[102] X. Yang, H. J. Mo, F. C. v. d. Bosch, Y. Zhang, and J. Han, *Evolution of the Galaxy - Dark Matter Connection and the Assembly of Galaxies in Dark Matter Halos*, *Astrophys. J.* **752** (2012) 41, [arXiv:1110.1420].

[103] F. Villaescusa-Navarro, D. Alonso, and M. Viel, *Baryonic acoustic oscillations from 21cm intensity mapping: the Square Kilometre Array case*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 2736, [arXiv:1609.00019].

[104] Z. Zheng, A. L. Coil, and I. Zehavi, *Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering*, *Astrophys. J.* **667** (2007) 760–779, [astro-ph/0703457].

[105] **SDSS** Collaboration, I. Zehavi et al., *Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity*, *Astrophys. J.* **736** (2011) 59–88, [arXiv:1005.2413].

[106] A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottloeber, B. Allgood, and J. R. Primack, *The Dark side of the halo occupation distribution*, *Astrophys. J.* **609** (2004) 35–49, [astro-ph/0308519].

[107] 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*, *MNRAS* **456** (Mar., 2016) 3553–3570, [arXiv:1510.04277].

[108] D. Alonso, P. Bull, P. G. Ferreira, and M. G. Santos, *Blind foreground subtraction for intensity mapping experiments*, *MNRAS* **447** (Feb., 2015) 400–416, [arXiv:1409.8667].

[109] D. Jeong, F. Schmidt, and C. M. Hirata, *Large-scale clustering of galaxies in general relativity*, *Phys. Rev.* **D85** (2012) 023504, [arXiv:1107.5427].

[110] L. Samushia, W. J. Percival, and A. Raccanelli, *Interpreting large-scale redshift-space distortion measurements*, *Mon. Not. Roy. Astron. Soc.* **420** (Mar., 2012) 2102–2119, [arXiv:1102.1014].

[111] J. Yoo and U. Seljak, *Wide Angle Effects in Future Galaxy Surveys*, *Mon. Not. Roy. Astron. Soc.* **447** (2015), no. 2 1789–1805, [arXiv:1308.1093].

[112] V. Tansella, C. Bonvin, R. Durrer, B. Ghosh, and E. Sellentin, *The full-sky relativistic correlation function and power spectrum of galaxy number counts: I. Theoretical aspects*, arXiv:1708.00492.