Bias of Damped Lyman-α systems from their cross-correlation with CMB lensing

D. Alonso,[a,1] **J. Colosimo,**[b,c] **A. Font-Ribera**[d] **and A. Slosar**[b]

[a]Oxford Astrophysics, Keble Road, Oxford OX1 3RH, UK
[b]Physics Department, Brookhaven National Laboratory, Upton NY 11973, USA
[c]School of Physics, Georgia Tech, 837 State St NW, Atlanta, GA 30332, USA
[d]Department of Physics and Astronomy, University College London, 132 Hampstead Road, London, NW1 2PS, UK

E-mail: david.alonso@physics.ox.ac.uk, jcolosimo@gatech.edu, a.font@ucl.ac.uk, anze@bnl.gov

Abstract. We cross-correlate the positions of damped Lyman-α systems (DLAs) and their parent quasar catalog with a convergence map derived from the Planck cosmic microwave background (CMB) temperature data. We make consistent detections of the lensing signal of both samples in both Fourier and configuration space. By interpreting the excess signal present in the DLA catalog with respect to the parent quasar catalog as caused by the large scale structure traced by DLAs, we are able to infer the bias of these objects: $b_{\rm DLA} = 2.6\pm0.9$. These results are consistent with previous measurements made in cross-correlation with the Lyman-α forest, although the current noise in the lensing data and the low number density of DLAs limits the constraining power of this measurement. We discuss the robustness of the analysis with respect to a number different systematic effects and forecast prospects of carrying out this measurement with data from future experiments.

---

[1]Author list is alphabetized

Contents

1 Introduction

Damped Lyman-α systems (DLAs) are absorption features in the spectra of high redshift quasars, with a neutral hydrogen column density $N_{\mathrm{HI}} > 2 \times 10^{20} \mathrm{cm}^{-2}$ [1], with damped wings covering a large fraction of the spectra. DLAs are an important source of contamination in studies of the clustering of the Lyman-α forest [2–5], and it is important to model their clustering properties in order to use the Lyman-α forest as a cosmological tool.

On the other hand, DLAs are very interesting astrophysical objects themselves, since they contain large amounts of neutral hydrogen and are thought to play a key role in the formation of galaxies at high redshift. They are also considered to be the main source of emission in ongoing and future 21cm surveys [6–8], where most of the signal is expected to come from galaxies with a high density of neutral hydrogen. Understanding the distribution of neutral hydrogen as traced by DLAs will allow us to make better predictions for the amplitude of the 21cm signal, which is currently uncertain by a factor of a few [9–11].

A first measurement of the clustering strength of DLAs was presented in [12]. Using quasar spectra from the ninth data release (DR9, [13]) of the Sloan Digital Sky Survey (SDSS-III, [14]), in particular from the Baryon Oscillation Spectroscopic Survey (BOSS, [15]), the authors were able to measure at high significance the cross-correlation of DLAs with the Lyman-α forest absorption. Assuming a cosmological model, as well as bias parameters for the Lyman-α forest, they presented an estimate of the DLA bias of $b_{\mathrm{DLA}} = 2.17 \pm 0.20$. Recently, this result has been confirmed by [16], who used a similar technique on the twelfth data release (DR12, [17]) of SDSS-III to present an updated measurement of $b_{\mathrm{DLA}} = 1.99 \pm 0.11$.

These results came as a surprise, since they imply that DLAs are hosted by rare and massive halos, with typical masses upwards of $\sim 10^{11} \ h^{-1} M_{\odot}$. On the other hand, most hydrodynamical simulations predict that DLAs should have a lower clustering strength, since the main contribution to the total HI density in simulations comes from halos of smaller mass [18, 19], except when strong galactic winds are included [20–23]. Resolving these discrepancies might have significant repercussions for the astrophysics of small halos.
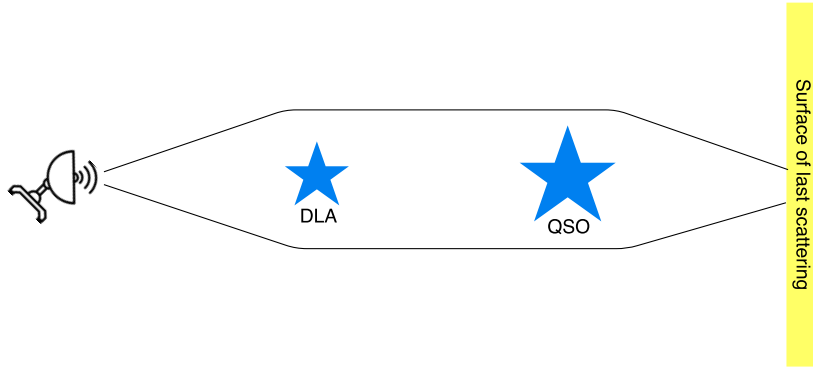
Figure 1: Figure outlining the effect. Photons emitted at the surface of last scattering are deflected by the intervening mass that correlates with positions of DLAs. However, since DLAs are found in quasar spectra, there is a quasar behind every DLA that also lenses the CMB photons. The latter contribution must be accounted for when measuring DLA bias.

In this paper we set out to measure the large-scale bias of DLAs with an alternative technique, by measuring their cross-correlation with CMB lensing data. Compared to cross-correlation with the other tracers in their vicinity, such as quasars, or the Lyman-α forest, this has an advantage that there is no bias factor associated with the CMB lensing map and that systematic effects should be very different from other methods of measuring the DLA bias. As we will discuss later, current data allows a measurement that is consistent with existing determinations, but with large error bars. However, we argue that future data will allow a decisive measurement to be made using this technique.

In Section 2 we describe the measurements and outline the relevant calculations for making theoretical predictions. In Section 3 we discuss the datasets used in the analysis and in Section 4 we present our results. We forecast the ability of future experiments to improve upon this measurements in Section 5 and we summarize our findings in Section 6.

2 Theory

In Figure 1 we show schematically the measurement that we are trying to perform. DLAs trace the overall structure of the Universe and are linearly biased with respect to the dark matter fluctuations δ_{DM}, leading to the following Fourier space relation valid on large scales (small wavenumbers $\mathbf{k}$):

$$b_{\mathrm{DLA}}(\mathbf{k}) = b_{\mathrm{DLA}}\delta_{\mathrm{DM}}(\mathbf{k}). \tag{2.1}$$

Since matter over-densities gravitationally lense the CMB photons, we expect projected DLA overdensities to do the same. Thus, for fixed cosmological parameters, we should be able to infer the bias parameter of the DLAs from the amplitude of the cross-correlation with a CMB lensing map. The linear biasing becomes exact in the large-scale limit, $\mathbf{k} \to 0$, and should be a very good approximation over all scales on which the data used in this paper is sensitive.

A subtlety in this measurement is the fact that DLAs are found in the spectra of distant quasars, and hence there is a quasar behind every DLA. These quasars in turn trace structures at higher redshift behind the DLAs that cause their own lensing of the CMB photons. Since the path-length of the forest is very large, these structures are, to a very good approximation, unrelated. I.e. the quasars are backlights that illuminate the DLAs in front of them, but are

sufficiently far away that they can be considered otherwise uncorrelated with the DLAs. Of course, the only way to subtract the effect of the quasar sample is to subtract an otherwise statistically equivalent sample of quasars but without DLAs in them. As we will later argue, it is sufficient to reweight the quasars based on their redshifts in order to obtain a sample with the same redshift distribution as that subsample which hosts DLAs.

We will now put this on a firmer mathematical footing. The paths of photons emitted by a source at comoving radial distance χ_s are perturbed by the intervening matter inhomogeneities through the gravitational lensing effect. In a flat Universe, the angular deflection in the trajectories of these photons arriving from the direction $\hat{\mathbf{n}}$ can be written as the gradient of an underlying lensing potential Φ_L, given by [24]:

$$\Phi_L(\hat{\mathbf{n}}) = \int_0^{\chi_s} d\chi \, \frac{\chi_s - \chi}{\chi_s \, \chi} \, \Phi(\chi \hat{\mathbf{n}}, \chi), \tag{2.2}$$

where $\Phi(\chi\hat{\mathbf{n}}, \chi)$ is the Weyl potential evaluated in the lightcone. The Weyl potential is $\Phi = \psi + \phi$, where ϕ and ψ are the two metric potentials in the Newtonian gauge. In the case of CMB lensing, χ_s is the distance to the surface of last scattering. Finally, the convergence κ is given by the Laplacian of Φ_L on the sphere $\kappa \equiv \nabla_{\hat{\mathbf{n}}}^2 \Phi_L / 2$.

On the other hand, assuming that DLAs and quasars are biased tracers of the underlying matter density perturbations, their angular overdensity is given by:

$$\delta^\alpha(\hat{\mathbf{n}}) = \int d\chi \, b_\alpha(\chi) \, p_\alpha(\chi) \, \delta_M(\chi \hat{\mathbf{n}}, \chi), \tag{2.3}$$

where α denotes either DLAs or QSOs, b_α is the linear bias factor of tracer α, $p_\alpha(\chi)$ is its normalized selection function and $\delta_M(\chi\hat{\mathbf{n}}, \chi)$ is the matter overdensity field in the lightcone. The equation above uses a number of approximations. First, we assume that a scale-independent, linear bias factor is enough to characterize the relation between the DLA/QSO distribution and the density field, which is known to be a good approximation on large, linear scales. The noise properties of the Planck lensing map, as well as the high shot noise of both source catalogs guarantee that only linear scales ($r \gtrsim 20 \, \mathrm{Mpc}/h$) contribute significantly to the total signal-to-noise of this measurement, and therefore this approximation should suffice. On the other hand, we have neglected the effect of lensing magnification. Although this effect is present on all scales, its contribution is subdominant, and therefore irrelevant given the large error bars of this measurement.

The angular cross-power spectrum between κ and δ^α, defined as the two-point correlator of the harmonic coefficients of both quantities, $C_\ell^{\kappa\alpha} \equiv \langle \kappa_{\ell m} \delta_{\ell m}^\alpha \rangle$, can be related to the matter power spectrum $P(k, z, z')$ as:

$$C_\ell^{\kappa\alpha} = \frac{2}{\pi} \int dk \, k^2 \int d\chi \int d\chi' P(k, z(\chi), z(\chi')) \, W_\ell^\alpha(\chi, k) \, W_\ell^\kappa(\chi', k), \tag{2.4}$$

where

$$W_\ell^\alpha(\chi, k) = p_\alpha(\chi) b_\alpha(\chi) j_\ell(k\chi), \quad W_\ell^\kappa(\chi, k) = \frac{3}{2} H_0^2 \Omega_M \frac{\ell(\ell+1)}{k^2 \, a(\chi)} \Theta(\chi; 0, \chi_s) \frac{\chi_s - \chi}{\chi \, \chi_s} j_\ell(k\chi). \tag{2.5}$$

Here $j_\ell(x)$ is the spherical Bessel function of order ℓ, $\Theta(x; x_1, x_2)$ is a top-hat window function in the interval (x_1, x_2), and $a(\chi)$ is the scale factor at which the radial comoving distance is χ.

To simplify the calculation of this observable we use Limber's approximation, which is equivalent to substituting $j_\ell(x) \sim \sqrt{\pi/2}\,\delta(\ell + 1/2 - x)/\sqrt{\ell + 1/2}$. This approximation is excellent for the wide radial kernels of the three tracers explored here. The expression for the angular power spectrum in this approximation becomes much simpler:

$$C_\ell^{\kappa\alpha} = b_\alpha \frac{3}{2} H_0^2 \Omega_M \frac{\ell(\ell+1)}{(\ell+1/2)^2} \int dk\, k\, \frac{p_\alpha(\chi_\ell)}{a(\chi_\ell)} \Theta(\chi_\ell; 0, \chi_s) \frac{\chi_s - \chi_\ell}{\chi_s} P(k, \chi_\ell). \qquad (2.6)$$

Here we have defined $\chi_\ell \equiv (\ell + 1/2)/k$. We have also assumed that the DLA/QSO bias is redshift-independent. While it would be very interesting to study the redshift dependence of this quantity, the noise in the current data does not allow for a detailed study beyond the overall normalization of their clustering amplitude.

Finally, we will also be interested in computing the two-point correlation function between κ and δ_α in configuration space. This quantity is defined as $\xi^{\kappa\alpha}(\theta) \equiv \langle \kappa(\hat{\mathbf{n}}_1) \delta_\alpha(\hat{\mathbf{n}}_2) \rangle$, where θ is the angle between $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$. In the small-angle limit, it is easy to prove that C_ℓ and $\xi(\theta)$ are related by a Hankel transform:

$$\xi^{\kappa\alpha}(\theta) = \frac{1}{2\pi} \int d\ell\, \ell\, C_\ell^{\kappa\alpha}\, J_0(\ell\theta), \qquad (2.7)$$

where $J_0(x)$ is the cylindrical Bessel function of order 0. At this point it is worth noting that our configuration-space analysis, described in Section 4, uses a filtered version of the convergence map. This can be easily taken into account by modifying the previous equation, substituting $C_\ell^{\kappa\alpha} \rightarrow W_\ell^\kappa C_\ell^{\kappa\alpha}$, where W_ℓ^κ is the filter applied to the convergence map in harmonic space.

All angular power spectra and correlation functions were computed using the Core Cosmology Library[1], using CLASS [25] to compute the matter power spectrum. In this analysis we fixed all cosmological parameters to their best-fit values found by [26].

3 Data

We use the CMB lensing convergence data provided by the Planck Collaboration [27]. We transformed the provided multipole coefficients $\kappa_{\ell m}$ into a HEALPix map [28] with resolution $N_{\mathrm{side}} = 2048$, and verified that this choice of resolution had a negligible impact on our results by re-running the analysis pipeline described in Section 4 for $N_{\mathrm{side}} = 1024$. When estimating both correlation functions and power spectra we mask this convergence map using the mask provided by Planck, covering around 67% of the sky. This mask was designed to remove regions dominated by galactic foregrounds as well as strong Sunyaev-Zel'dovich sources. Since these sources lie mostly on redshifts $z \lesssim 0.4$, this mask is completely uncorrelated with the quasar and DLA catalogs used here.

We use two different DLA catalogs, both created from quasar spectra from the 12th data release of the Sloan Digital Sky Survey-III (SDSS-III DR12 [17]). The first catalog, which we will refer to as N12, identified DLAs using an algorithm first described in [29], and has been used in the DR12 Lyman-α analyses of BOSS ([16, 30, 31]). In short, DLAs are identified by correlating the spectra with synthetic absorption profiles of increasing column density. The

---

[1]CCL is a public library to compute theoretical predictions for multiple cosmological observables. The library is maintained by the Dark Energy Science Collaboration of LSST, and can be found in https://github.com/LSSTDESC/CCL
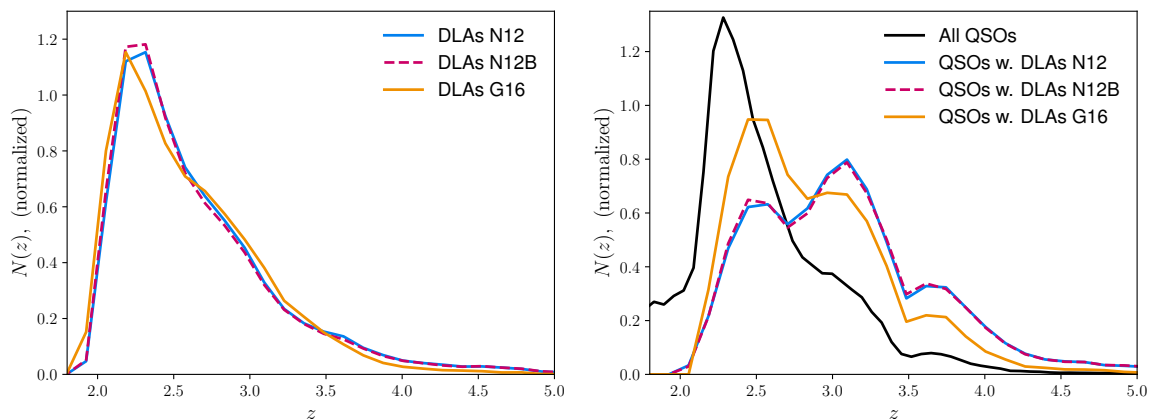
Figure 2: *Left:* redshift distribution of DLAs in the N12 (solid cyan), N12B (dashed dark magenta) and G16 (solid light orange) catalogs. *Right:* redshift distribution of the full DR12 quasar catalog (solid black) as well as that of the QSOs containing DLAs in their spectra in the N12 (solid cyan), N12B (dashed dark magenta) and G16 (solid light orange) DLA catalogs.
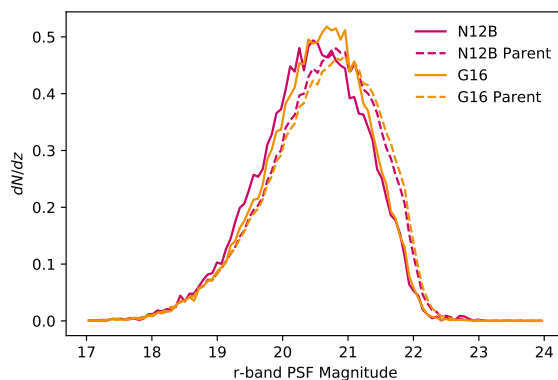

Figure 3: Distribution of r-magnitude for quasars where DLAs have been found (solid lines) and the parent catalogs (created by reweighting the full quasar catalog to match redshift distributions, dashed lines) for the N12 (red) and G16 (blue) catalogs.

total catalog consists of 34,050 systems, with a redshift distribution shown as a solid cyan line in the left panel of Figure 2. In Section 4 we also use a subset of this catalog, which we label N12B, where we exclude those systems identified in the spectra with lower signal to noise. The second catalog, which we will refer to as G16[2], used instead an algorithm presented in [32]. In this case the authors used a Gaussian process to model the quasar emission spectra, and used Bayesian model selection to assign a probability of hosting a DLA to each spectrum. We used G16 with an extra quality cut to ensure the purity of the catalog. In particular, we select only the DLA candidates that have a probability $p_{\mathrm{DLA}} > 0.9$, thus retaining a sample of 35,621 objects. This probability cut-off was chosen to generate a catalog of similar size to the N12 catalog. Since the distribution of probabilities is very bi-modal (most cluster around

---

[2]The catalog is publicly available at `http://tiny.cc/dla_catalog_gp_dr12q_lyb`

$p = 0$ or $p = 1$), the number of DLAs is surprisingly insensitive to the precise choice of probability cut. Note that during the training of the machine learning algorithm used in G16 the authors used information from the DLA catalog of [29] that used the same algorithm as N12, and therefore we expect the two catalog to be quite similar.

For both DLA catalogs, we also generate the parent quasar catalog. This is done by taking the full catalog of quasars where a DLA could have been found and reweighting it so that the redshift distribution of quasars where DLA have been found matches that of the reweighted full catalog. These parent catalogs allow us to measure the "null" signal, i.e the signal in absence of lensing by the DLAs.

The properties of the samples are summarized in Figure 2. We see that the redshift distribution of DLAs is very similar for both catalogs and peaks at around $z \sim 2.3$. The shape of this distribution is dominated by the total path length of the Lyman-α forest available as a function of redshift. Below $z \sim 1.8$, the forest goes into UV and is unobservable from the ground. At redshifts above $z > 2.5$, the number density of quasars decreases and they become fainter and thus noiser which both act to lower the effective number of DLAs. The right panel of the same figure shows the redshift distribution of the quasars in which these DLAs have been found. We see that the majority of quasars skew towards lower redshifts, but these have very short forests and thus not very many DLAs are found in them. However, the parent quasar redshift distributions for the N12 and G16 catalogs are surprisingly different given the similarity of the DLA distributions plotted in the left panel. This means, that although both have similar redshift distributions, the two samples actually detect *different* DLAs.

The method to generate synthetic parent QSO catalogs described above assumes that quasar redshift is the only property that determines the probability of detecting a DLA. We know that this must be an imperfect assumption at some level since, for instance, it is easier to identify DLAs in the spectra of bright quasars. If the clustering of quasars had a strong dependence on their luminosity, this could bias our result, since the clustering of quasars without DLAs would be on average lower than the clustering of quasars with DLAs in their spectra. In order to study this possibility, Figure 3 shows the r-band magnitude distribution for the quasars where DLAs have been found (solid lines) and for their synthetic parent catalogs generated using only redshift-dependent weights (dashed). We see that, although the synthetic catalogs are slightly fainter – this is expected as DLAs are preferentially identified in brighter quasars that have better signal-to-noise in the Lyman-α forest region – the effect is surprisingly small. This result, together with the independence of the quasar bias on luminosity, verified by different analyses using BOSS-like samples [33–35] gives us confidence that redshift-dependent weights are enough to mimic the angular clustering properties of the DLA parent sample.

4 Method & Results

4.1 Differential bias measurement

As mentioned in Section 3, the main complication of the measurement we are attempting here is the fact that the DLA catalogs cannot be directly interpreted as a homogeneous point process: the DLAs all lie along the lines of sight of illuminating quasars, which contribute to the total lensing signal. Thus, a stack of the CMB κ map around the positions of DLAs will contain contributions from the lensing signals of both the DLAs and the background quasars. The quasar contribution must therefore be estimated and subtracted in order to get at the pure DLA signal.

Now, let $\mathbf{d}_{\mathrm{D+Q}}$ be a measurement of the lensing signal on DLAs, including the contribution of the background quasars, and let $\mathbf{d}_{\mathrm{Q}}$ be the same measurement performed on the parent QSO sample. Here $\mathbf{d}$ stands for a set of measurements of either the correlation function or the power spectrum at N_d values of θ or ℓ. There are several ways of combining both data vectors to obtain a measurement of the bias of DLAs, b_{DLA}. Here we explore two possibilites:

1. **Direct subtraction.** As an estimator for the DLA-only signal we directly subtract the measured QSO contribution: $\mathbf{d}_{\mathrm{D}} = \mathbf{d}_{\mathrm{D+Q}} - \mathbf{d}_{\mathrm{Q}}$. We then model this estimator as $\mathbf{d}_{\mathrm{D}} = b_{\mathrm{DLA}}\mathbf{t}_{\mathrm{D}}$, where $\mathbf{t}_{\mathrm{D}}$ is a theoretical template for the DLA-only signal with unit bias computed as described in Section 2. This method has the advantage that no assumptions are made regarding the particular form of the QSO signal.

2. **Simultaneous modelling.** We simultaneously model the DLA+QSO and QSO-only measurements as

$$\mathbf{d}_{\mathrm{D+Q}} = b_{\mathrm{DLA}}\mathbf{t}_{\mathrm{D}} + b_{\mathrm{QSO}}\mathbf{t}_{\mathrm{Q}}, \quad \mathbf{d}_{\mathrm{Q}} = b_{\mathrm{QSO}}\mathbf{t}_{\mathrm{Q}}, \tag{4.1}$$

where $\mathbf{t}_{\mathrm{Q}}$ is the theoretical template for the QSO signal with unit bias, and b_{QSO} is the bias of this sample. This method has the advantage that, by modelling the QSO lensing signal, we reduce the impact of the noise of $\mathbf{d}_{\mathrm{Q}}$ on the final uncertainty on b_{DLA}.

In the following sections we present the measurement of $\mathbf{d}_{\mathrm{D+Q}}$ and $\mathbf{d}_{\mathrm{Q}}$ in both harmonic space and configuration space, as well as the resulting estimates of b_{DLA}.

4.2 Harmonic-space pipeline

This section describes the process used to estimate the power spectra $C_\ell^{\kappa\alpha}$, with α corresponding to the positions of DLAs or QSOs. The main complication to overcome in this case is the incomplete sky coverage of both the convergence map and the SDSS catalogs, which biases any direct estimate of either field's harmonic coefficients. In order to account for this effect, we use the so-called pseudo-C_ℓ (PCL) or MASTER approach [36], which we describe briefly here.

In short, we observe incomplete maps of the sky for a couple of fields f^1 and f^2: $\hat{f}^i(\hat{\mathbf{n}}) = w^i(\hat{\mathbf{n}})\, f^i(\hat{\mathbf{n}})$, where w^i is a generic weights map, which can be a binary mask defined by a given experiment's footprint. The harmonic coefficients of $\hat{f}^i$ are therefore a convolution of the harmonic coefficients of f^i and w^i: $\hat{f}^i_{\ell m} = \sum_{\ell'm'\ell''m''} D^{mm'm''}_{\ell\ell'\ell''} f^i_{\ell'm'} w^i_{\ell''m''}$. Thus, it is possible to prove that the standard power spectrum estimator in the full sky (essentially averaging over the redundant m) yields a biased estimate of the true power spectrum C_ℓ^{12}:

$$\left\langle \hat{C}_\ell^{12} \right\rangle \equiv \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \left\langle \mathrm{Re}\left(\hat{f}^1_{\ell m} \hat{f}^{2*}_{\ell m} \right) \right\rangle = \sum_{\ell'} W_{\ell\ell'} C_{\ell'}^{12}, \tag{4.2}$$

where $W_{\ell\ell'}$ is a mode-mixing matrix dependent on the properties of the weight maps. The PCL method is based on computing an analytical estimate of $W_{\ell\ell'}$ and inverting the previous equation to remove the bias associated to an incomplete sky coverage. We computed the PCL estimate of the κ-DLA and κ-QSO power spectra using the public code `NaMaster`[3]. The cross-power spectra were estimated in 20 linearly-spaced bandpowers of width $\Delta\ell = 50$ between $\ell = 2$ and $\ell = 1002$.

---

[3]https://github.com/damonge/NaMaster/

The PCL estimator requires a sky map and the associated weights map. Both are directly available in the case of the CMB convergence, as described in Section 3. For an object catalog, as is the case for the DLA and quasar samples, we must construct our own maps. The weights map should be built from the non-trivial completeness data associated with the DR12 quasar sample, since inaccurately accounting for this completeness introduces artificial inhomogeneities in the associated overdensity map. However, while these inhomogeneities would bias any estimate of the auto-correlation of the DLA/QSO two-point statistics, they are uncorrelated with κ, and therefore should not introduce any biases in cross-correlation (although they will contribute towards the noise). We verified this by running our pipeline using three different masks for the DR12 data:

1. A high-resolution ($N_{\mathrm{side}} = 2048$) sky mask built from the random catalogs associated with the BOSS CMASS galaxy sample. Essentially we count the number of random objects in each HEALPix pixel and mask out all empty pixels.

2. A low-resolution ($N_{\mathrm{side}} = 64$) mask built from the DR12 QSO catalog, again masking out all pixels with no QSOs in them.

3. An even lower resolution version of the previous mask with $N_{\mathrm{side}} = 32$.

The final results did not change significantly with the mask used, which demonstrates the robustness of the measurement with respect to this choice.

For both the DLA and QSO samples we build a map of their projected overdensity in each pixel p:

$$\delta_p \equiv \frac{N_p}{\bar{N}} - 1, \tag{4.3}$$

where N_p is the weighted number of objects in each pixel, $N_p \equiv \sum_{g \in p} w_g$, and $\bar{N}$ is the average of N_p over all unmasked pixels. As described in Section 4.1, QSOs were weighted to reproduce the properties of the DLA parent catalog, and unit weights were used for the DLAs themselves.

4.3 Configuration-space pipeline

We complement the harmonic-space measurement described in the previous section with a configuration-space pipeline based on the angular cross-correlation function $\xi^{\kappa\alpha}(\theta)$. Measurements in real space are arguably more robust in terms of systematics associated with survey geometry definition (e.g. defining the QSO weights map as discussed in the previous section). They also provide a more intuitive observable, since $\xi^{\kappa\alpha}(\theta)$ is essentially the average lensing convergence caused by the DLAs/QSOs as a function of angular separation. In any case, this parallel pipeline allows us to validate the measured value of b_{DLA} in harmonic space.

The estimator we use for $\xi^{\kappa\alpha}$ essentially consists of stacking the κ map on the position of the DLAs/QSOs:

$$\xi^{\kappa\alpha}(\theta) = \frac{\sum_g w_g \, \bar{\kappa}_g(\theta)}{\sum_g w_g}, \tag{4.4}$$

where g runs over all objects in either catalog and w_g is the weight associated with the g-th object. The mean convergence for each object $\bar{\kappa}$ is estimated as:

$$\bar{\kappa}_g(\theta) = \frac{\sum_p w_p^{\kappa} \kappa_p}{\sum_p w_p^{\kappa}}, \tag{4.5}$$

where the weights w_p^κ correspond to the mask associated with the convergence map provided by Planck, and p runs over all pixels separated from the position of the g-th object by an angle lying in the interval $[\theta - \Delta\theta/2, \theta + \Delta\theta/2]$ (here $\Delta\theta$ is the choice of bin width for our correlation function estimator). For our configuration-space measurement we estimated the correlation function in 16 bins of θ in the interval $0 < \theta < 3°$ (i.e. $\Delta\theta \sim 11'$).

In order to improve the signal-to-noise ratio S/N of this measurement, we pre-process the convergence map by Wiener-filtering it before computing the correlation function. This effectively down-weights noise-dominated modes optimally, reducing their impact in the final S/N. Since the noise-induced correlations associated with the quadratic estimator used by Planck to compute the lensing convergence map are diagonal in harmonic space, this down-weighting is automatically implemented in the analysis described in the previous section. However, we find (in agreement with [37]), that this procedure marginally improves the final uncertainties with respect to a straightforward measurement of the correlation function using the raw κ map. The Wiener filter applied to the harmonic coefficients $\kappa_{\ell m}$ is $W_\ell \equiv C_\ell^{\kappa\kappa}/(C_\ell^{\kappa\kappa} + N_\ell^{\kappa\kappa})$, where $C_\ell^{\kappa\kappa}$ is the convergence power spectrum for the best-fit Planck cosmological parameters, and $N_\ell^{\kappa\kappa}$ is the noise power spectrum of the Planck κ map. As mentioned in Section 2 this weighting is taken into account when modelling the theory prediction for the correlation function.

4.4 Measurement of b_{DLA}

As described in Section 4.1, both of our approaches to measure b_{DLA} model the data vector $\mathbf{d}$ as $\mathbf{d} = \hat{\mathsf{T}}\,\mathbf{b}$, where T is a matrix of theoretical templates and $\mathbf{b}$ is a vector containing all bias parameters. In the case of direct subtraction $\mathbf{d} \equiv \mathbf{d}_{\mathrm{D}}$, $\hat{\mathsf{T}} \equiv \mathbf{t}_{\mathrm{D}}$ and $\mathbf{b} = b_{\mathrm{DLA}}$, whereas for simultaneous modelling:

$$\mathbf{d} \equiv \begin{pmatrix} \mathbf{d}_{\mathrm{D+Q}} \\ \mathbf{d}_{\mathrm{Q}} \end{pmatrix}, \quad \hat{\mathsf{T}} \equiv \begin{pmatrix} \mathbf{t}_{\mathrm{D}} & \mathbf{t}_{\mathrm{Q}} \\ 0 & \mathbf{t}_{\mathrm{Q}} \end{pmatrix}, \quad \mathbf{b} \equiv \begin{pmatrix} b_{\mathrm{DLA}} \\ b_{\mathrm{QSO}} \end{pmatrix}. \tag{4.6}$$

Under the assumption that the data $\mathbf{d}$ is well described by Gaussian statistics, the maximum-likelihood estimate of $\mathbf{b}$ and its associated covariance are analytically given by

$$\bar{\mathbf{b}} = \left(\hat{\mathsf{T}}^T \hat{\mathsf{C}}^{-1} \hat{\mathsf{T}} \right)^{-1} \hat{\mathsf{T}}^T \mathsf{C}^{-1} \mathbf{d}, \quad \mathrm{Cov}\left(\bar{\mathbf{b}} \right) = \left(\hat{\mathsf{T}}^T \hat{\mathsf{C}}^{-1} \hat{\mathsf{T}} \right)^{-1}, \tag{4.7}$$

where $\hat{\mathsf{C}}$ is the covariance matrix of the data $\hat{\mathsf{C}} \equiv \left\langle (\mathbf{d} - \langle \mathbf{d} \rangle)(\mathbf{d} - \langle \mathbf{d} \rangle)^T \right\rangle$.

We estimated the fiducial data covariance matrix used in our analysis from a sample of 1000 simulated measurements of $\mathbf{d}$. Each of these were produced by first generating a random realization of the CMB lensing convergence map and correlating it with the positions of DLAs and QSOs in the original catalogs. The random κ maps were generated as Gaussian realizations of the best-fit power spectrum of the Planck map, including signal and noise. We verified our estimate of the covariance against the 100 random maps provided by Planck (which better account for the inhomogeneous and non-Gaussian nature of the Planck noise). We also compared the diagonal errors computed this way with analytical "$1/f_{\mathrm{sky}}$" estimates and with similar random realizations also involving randomly-positioned DLAs and QSOs. Our final results were not found to be sensitive to any of these variations. Figure 4 shows the covariance matrices used in the different variations of our pipeline (real-space, fourier-space, direct subtraction or joint modelling).
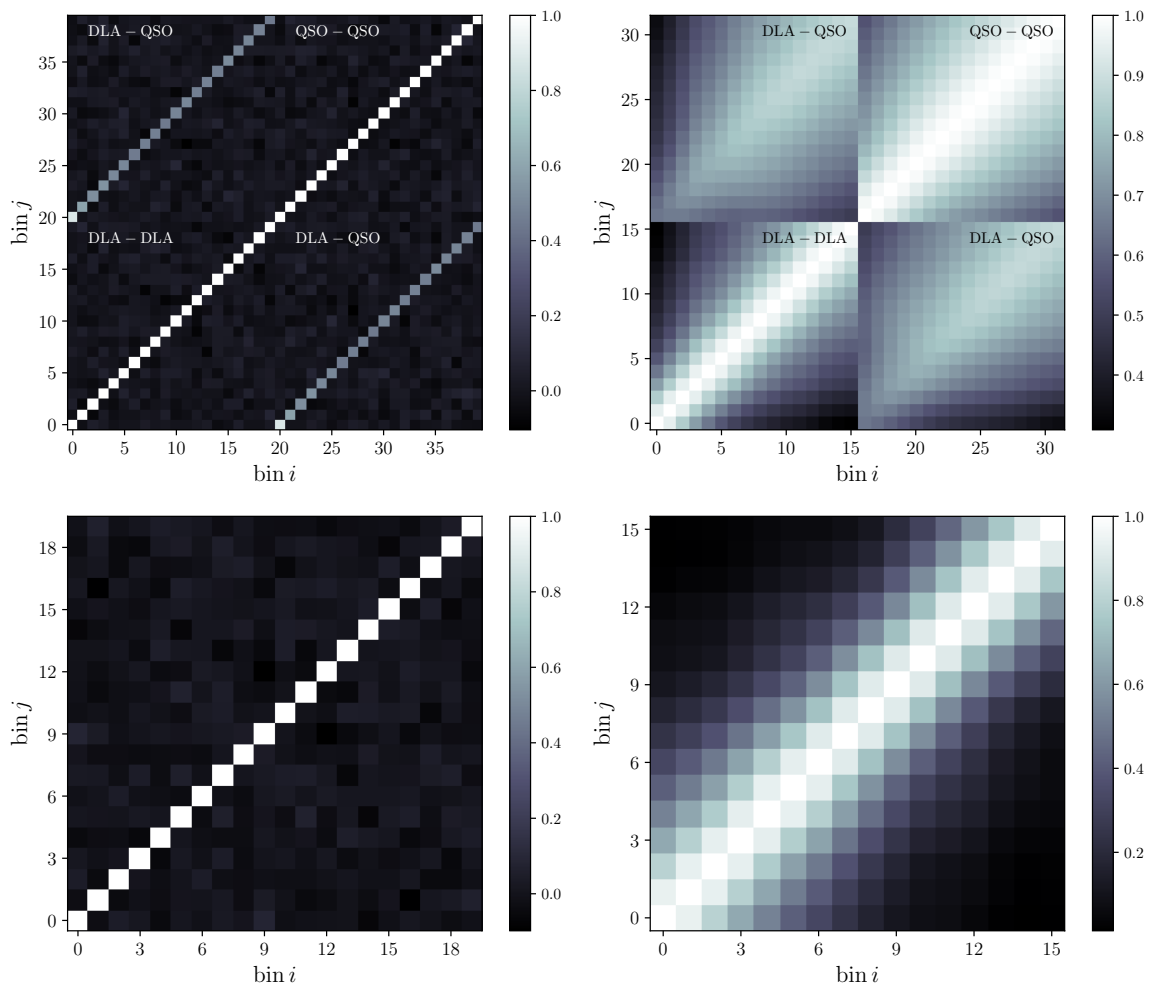
Figure 4: *Top:* data correlation matrix for the simultaneous-fitting pipeline. The four boxes in each plot correspond to the different combinations of pairs of correlations involving DLAs or QSOs. *Bottom:* data correlation matrix for the direct-subtraction pipeline. *Left:* results the C_ℓ-based analysis. *Right:* data correlation matrix for the $\xi(\theta)$-based analysis. The large-scale correlations are caused by the strong smoothing associated to the Wiener filter applied to the κ map before computing the correlation function.

Our measurements of the 2-point statistics and the best-fit theoretical estimates are shown in Figure 5. The best-fit values of b_{DLA} associated to these measurements are:

$$\begin{aligned}
\text{Direct subtraction:} \quad & b_{\mathrm{DLA}} = 2.55 \pm 0.94 \ (C_\ell), \quad b_{\mathrm{DLA}} = 2.69 \pm 0.92 \ (\xi(\theta)) \\
\text{Simultaneous model:} \quad & b_{\mathrm{DLA}} = 2.66 \pm 0.93 \ (C_\ell), \quad b_{\mathrm{DLA}} = 2.55 \pm 0.91 \ (\xi(\theta))
\end{aligned} \tag{4.8}$$

All our measurements are therefore highly compatible with each other, as well as with the latest measurement of $b_{\mathrm{DLA}} = 1.99 \pm 0.11$ from [16]. In all cases the model described in Section 2 is a good fit to the data, with reasonable χ^2 values and associated probabilities-to-exceed (PTE) above 31%.
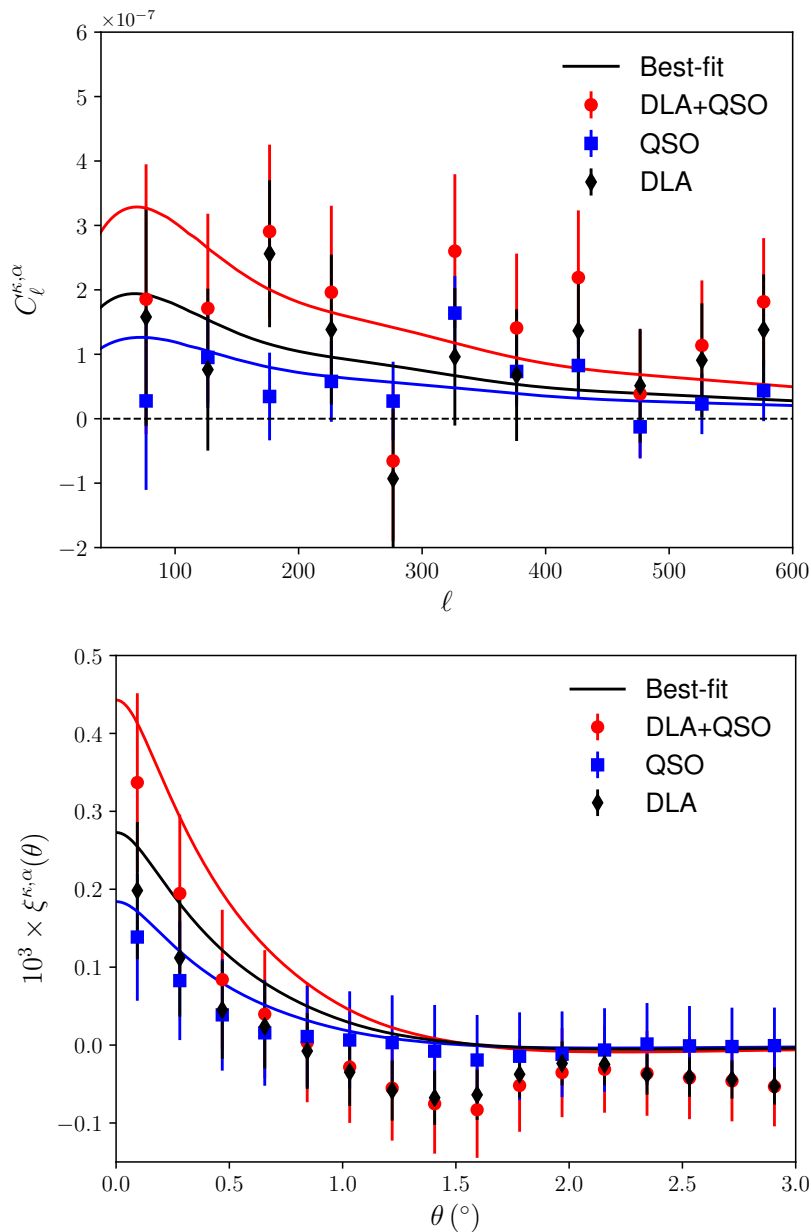
Figure 5: *Top:* measured power spectrum between the CMB lensing convergence κ and the raw DLA distribution (red circles), between κ and the parent QSO distribution (blue squares) and the difference of both (black diamonds), corresponding to the pure DLA lensing signal with the direct-subtraction pipeline. The solid lines show the best-fit theory predictions. *Bottom:* results for the two-point correlation function with the same color coding.

4.5 Consistency tests

In the previous sections we have shown that our measurement is robust against a number of effects, such as the choice of Fourier-space vs. configuration-space analysis, the modelling of the clustering of background QSOs, the definition of the DLA/QSO mask or the method used to estimate the covariance matrix. In this section we explore other possible systematic
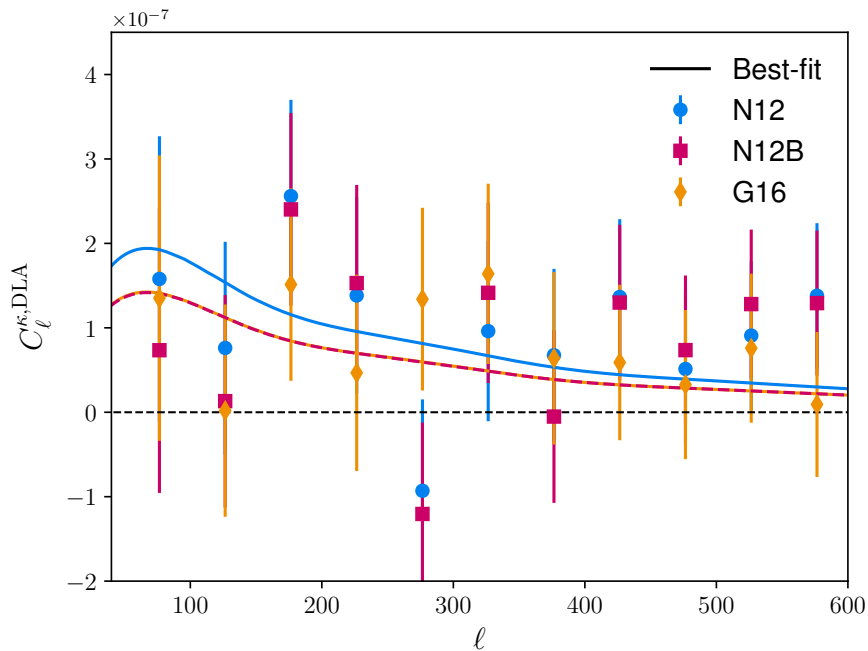
Figure 6: Cross-power spectrum between the CMB lensing convergence and three DLA catalogs: N12 (cyan circles), N12B (dark magenta squares) and G16 (light orange diamonds). In all cases, the contribution from the parent quasar catalog has been directly subtracted. The lines show the best-fit model for each measurement.

effects, especially associated with the properties of our fiducial DLA sample. The values of b_{DLA} quoted in this section will correspond to those measured using a C_ℓ-based analysis and simultaneously fitting b_{QSO}, although we find consistent values with all other analysis methods in all cases.

Our fiducial analysis makes use of the raw DLA sample, with no further quality cuts. Given the low signal-to-noise ratio of this measurement, the effects of spurious detections in the sample are likely to be negligible. However, a large number of fake DLAs in the catalog would systematically lower the measurement of b_{DLA} (since they would be uncorrelated with κ), and therefore it is worth verifying whether such a systematic shift is observed when using a cleaner sample. We do so by repeating our analysis on the so-called N12B catalog, defined by removing all DLAs on spectra with continuum-to-noise ratio CNR < 2, and containing $\sim 10\%$ fewer objects. The power spectrum measured on N12B for the direct-subtraction pipeline is shown as dark magenta squares in Fig. 6, together with our fiducial measurement on N12. We obtain a best-fit measurement of the DLA bias from N12B of $b_{\mathrm{DLA}}^{\mathrm{N12B}} = 2.00 \pm 1.01$. This corresponds to a $\sim 0.7\sigma$ downward shift with respect to our fiducial measurement, with a larger measurement error due to the smaller size of this sample. Since any contamination caused by spurious detections would have produced an upwards shift when removing them, we conclude that our fiducial measurement is robust with respect to this effect.

There exist different procedures to detect DLAs in quasar spectra, and we have also explored the robustness of our results with respect to the definition of DLA used. To do so, we have repeated our analysis on an alternative catalog, G16, produced by [32]. This sample was generated by modelling the quasar spectrum as a Gaussian process, and then determining

the detection probability of each DLA using a Bayesian model. We select a clean subsample of this catalog, given by all objects with probability $p > 0.9$, containing 35,621 objects. As shown in Section 3, although the redshift distribution of these DLAs is very similar to that of the N12 sample, the distributions of their parent QSO samples are markedly different (see Fig. 2). In practical terms, this implies that the raw DR12 QSO catalog must be re-weighted in order to reproduce the new redshift distribution. This modifies the measured quasar power spectrum and therefore the associated measurement of $b_{\rm DLA}$. The DLA-κ power spectrum measured by direct subtraction on G16 is shown as light orange diamonds in Fig. 6, and we obtain an associated best-fit value $b_{\rm DLA}^{\rm G16} = 1.92 \pm 0.69$.

This measurement is $\sim 1.1\sigma$ away from the value found for N12, it has a similar detection significance, and both measurements are compatible with the latest measurement of [16]. However, since both measurements are made on catalogs with a significant overlap, they are not completely uncorrelated. Therefore it is important to assess the degree of disagreement between both measurements more precisely. From the 1000 simulations used to estimate the data covariance matrix we can determine both measurements to be $\sim 50\%$ correlated, i.e.:

$$r \equiv \frac{\text{Cov(N12, G16)}}{\sqrt{\text{Var(N12)Var(G16)}}} = 0.501. \tag{4.9}$$

Now, let b_1 and b_2 be two measurements of $b_{\rm DLA}$ with covariance matrix C. Then, assuming Gaussian statistics, the difference $\Delta b \equiv b_1 - b_2$ should have a Gaussian distribution with standard deviation $\sigma(\Delta b) = \sqrt{C_{11} + C_{22} - 2C_{12}}$. Thus we can quantify the disagreement between both measurements in terms of the probability of a value Δb being larger than our measurement. Doing this, we obtain a probability PTE $= 0.19$, and therefore the measurements made in N12 and G16 are compatible.

Finally, we have also explored the impact of other variables in our analysis, including our choice of mask apodization in the C_ℓ-based pipeline and the binning scheme used both in θ and ℓ. None of these factors significantly impact our results. Our confidence in the robustness of our analysis pipeline is further strengthened by our ability to reproduce the measurement of $b_{\rm QSO}$ presented in [38].

5 Forecasts

Although the measurement of $b_{\rm DLA}$ presented here serves to validate previous estimates of this quantity made in cross-correlation with the Lyman-α forest $b_{\rm DLA} \sim 2$, the noise in the CMB lensing map, as well as the low density of detected DLAs prevent us from discriminating between the measured DLA bias and the lower value $b_{\rm DLA} \sim 1.2$ that is expected from hydrodynamical simulations without strong galactic wings (see Figure 14 in [12]). It is therefore relevant to explore the potential of future quasar surveys and CMB experiments to improve on this result. To do so, we have estimated the uncertainty $\sigma(b_{\rm DLA})$ for a number of future facilities using a Fisher-matrix formalism (see e.g. [39]). We have explored three different quasar samples:

1. The SDSS DR12 sample described in Section 3.

2. The sample achievable by the Dark Energy Spectroscopic Instrument (DESI, [40]). We define this as having a 3-times larger density of QSOs and DLAs covering 14,000 deg^2 of the northern celestial hemisphere.
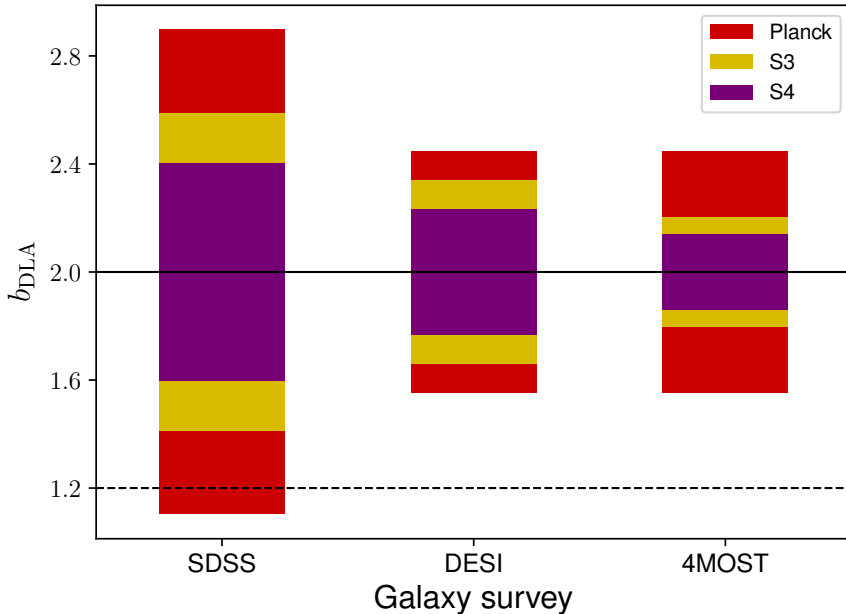
Figure 7: Forecast 1σ uncertainties on $b_{\rm DLA}$ achievable by current and future surveys and CMB experiments. The horizontal lines show the preferred values for this quantity measured in cross correlation with the Lyman-α forest (solid) and estimated from hydrodynamical simulations (dashed). Although current data is unable to discriminate between both cases, a combined analysis of DESI and Stage-4 CMB would make this possible.

3. A southern-hemisphere version of the DESI sample, which could be obtained by the 4MOST instrument [41]. We consider this sample in order to explore the benefit of having a full area overlap with southern-hemisphere, ground-based CMB experiments.

We studied the constraints on $b_{\rm DLA}$ achievable by each of these surveys in combination with a CMB lensing map produced by three different experiments:

1. The Planck satellite (as described in Section 3). Given the full-sky coverage provided by Planck, we assume full area overlap with the three previous samples.

2. A Stage-3 (S3) ground-based CMB experiment, such as Advanced ACT [42]. We characterize this experiment by a 1.4 arcmin FWHM Gaussian beam and a white noise level of 8.5μK arcmin spread over an area of 20,000 deg^2 in the southern hemisphere.

3. A Stage-4 (S4) ground-based experiment [43], also observing 20,000 deg^2 of the southern hemisphere, with a noise level of 1μK arcmin and a 3 arcmin beam.

We assumed a total area overlap of only 4,000 deg^2 between SDSS/DESI and S3/S4, and a full overlap (14,000 deg^2) with 4MOST.

The forecast values of $\sigma(b_{\rm DLA})$ for the different combinations of quasar sample and CMB experiment are shown in Figure 7. The forecast for SDSS $\times$ Planck agrees well with the errors obtained in this work, and shows that current sensitivities and number densities are insufficient to discriminate between $b_{\rm DLA} = 2$ and $b_{\rm DLA} = 1.2$. However, the combination of DESI and S4 would achieve $\sim 10\%$ errors on this parameter, making the distinction between these two

models possible at the $\gtrsim 3\sigma$ level. Optimal results would be obtained by maximizing area overlap, which a survey like 4MOST could achieve.

6   Conclusions

In this paper we have measured the bias of DLAs from the cross-correlation between DLAs found in Lyman-α forest spectra of quasars and the weak lensing map obtained from the Planck satellite. This is the first measurement of DLA properties based on cross-correlation with weak-lensing. We measure the DLA bias to be $b_{\rm DLA} = 2.6 \pm 0.9$. The formal statistical significance is at 2-3σ level, but since DLA clustering has been measured before using cross-correlations with other tracers, the correct way to look at our measurement is as a noisy measurement of the bias of DLAs. We have performed the measurement in both Fourier and configuration space, and performed a number of tests and data-splits. We find our result to be robust against a number of analysis choices. Measurements of the DLA bias based on cross-correlations with lensing fields do not rely on the bias model of any other tracer and derive their information on the very large scales where theoretical modeling is least prone to systematic uncertainties.

Our measurement is consistent with the best measurements derived from the cross-correlation between the Lyman-α forest fluctuations and the DLA density field ($b_{\rm DLA} = 1.99 \pm 0.11$), but does not add to the discussion of this surprisingly high value given our current understanding of astrophysics relevant to DLAs, since our result is also consistent with the considerably lower values favored by hydro simulations $b_{\rm DLA} \sim 1.2$.

We forecast the ability of future measurements to improve on this number and find that, with DESI or 4MOST in combination with future Stage-3 and Stage-4 CMB experiments, we will have statistical power to decisively differentiate between the high-bias scenario currently favored by data and low-bias scenario favored by simulations. With improvement in the lensing field, these measurements will also become statistically competitive with forest cross-correlation measurements. Resolving the DLA bias puzzle is not important just for the astrophysics of low-mass halos, but it also have repercussions for predicting future 21cm cosmology signals.

References

[1] A. M. Wolfe, D. A. Turnshek, H. E. Smith and R. D. Cohen, *Damped Lyman-alpha absorption by disk galaxies with large redshifts. I - The Lick survey*, *ApJS* **61** (June, 1986) 249–304.

[2] P. McDonald, U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, R. Cen et al., *The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey*, *ApJS* **163** (Mar., 2006) 80–109, [`arXiv:astro-ph/0405013`].

[3] A. Font-Ribera and J. Miralda-Escudé, *The effect of high column density systems on the measurement of the Lyman-α forest correlation function*, *JCAP* **7** (July, 2012) 028, [`1205.2018`].

[4] K. K. Rogers, S. Bird, H. V. Peiris, A. Pontzen, A. Font-Ribera and B. Leistedt, *Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers*, *ArXiv e-prints* (Nov., 2017) , [`1711.06275`].

[5] K. K. Rogers, S. Bird, H. V. Peiris, A. Pontzen, A. Font-Ribera and B. Leistedt, *Simulating the effect of high column density absorbers on the one-dimensional Lyman-alpha forest flux power spectrum*, *ArXiv e-prints* (June, 2017) , [`1706.08532`].

[6] F. B. Abdalla, P. Bull, S. Camera, A. Benoit-Lévy, B. Joachimi, D. Kirk et al., *Cosmology from HI galaxy surveys with the SKA*, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)* (Apr., 2015) 17, [`1501.04035`].

[7] K. Bandura, G. E. Addison, M. Amiri, J. R. Bond, D. Campbell-Wilson, L. Connor et al., *Canadian Hydrogen Intensity Mapping Experiment (CHIME) pathfinder*, in *Ground-based and Airborne Telescopes V*, vol. 9145 of *Proc. SPIEE*, p. 914522, July, 2014, `1406.2288`, DOI.

[8] L. B. Newburgh, K. Bandura, M. A. Bucher, T. C. Chang, H. C. Chiang, J. F. Cliche et al., *Hirax: A probe of dark energy and radio transients*, `arXiv:1607.02059v1`.

[9] A. M. Martin, R. Giovanelli, M. P. Haynes and L. Guzzo, *The Clustering Characteristics of H I-selected Galaxies from the 40% ALFALFA Survey*, *Astrophysical Journal* **750** (May, 2012) 38, [`1202.6005`].

[10] M. Santos, D. Alonso, P. Bull, M. B. Silva and S. Yahya, *HI galaxy simulations for the SKA: number counts and bias*, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)* (Apr., 2015) 21, [`1501.03990`].

[11] E. Castorina and F. Villaescusa-Navarro, *On the spatial distribution of neutral hydrogen in the Universe: bias and shot-noise of the H I power spectrum*, *MNRAS* **471** (Oct., 2017) 1788–1796, [`1609.05157`].

[12] A. Font-Ribera, J. Miralda-Escudé, E. Arnau, B. Carithers, K.-G. Lee, P. Noterdaeme et al., *The large-scale cross-correlation of Damped Lyman alpha systems with the Lyman alpha forest: first measurements from BOSS*, *JCAP* **11** (Nov., 2012) 59, [`1209.4596`].

[13] C. P. Ahn, R. Alexandroff, C. Allende Prieto, S. F. Anderson, T. Anderton, B. H. Andrews et al., *The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey*, *ApJS* **203** (Dec., 2012) 21, [`1207.7137`].

[14] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson et al., *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems*, *Astron. Journal* **142** (Sept., 2011) 72, [`1101.1529`].

[15] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey et al., *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *Astron. Journal* **145** (Jan., 2013) 10, [`1208.0022`].

[16] I. Pérez-Ràfols, A. Font-Ribera, J. Miralda-Escudé, M. Blomqvist, S. Bird, N. Busca et al., *The SDSS-DR12 large-scale cross-correlation of damped Lyman alpha systems with the Lyman alpha forest*, *MNRAS* **473** (Jan., 2018) 3019–3038, [`1709.00889`].

[17] S. Alam, F. D. Albareti, C. Allende Prieto, F. Anders, S. F. Anderson, T. Anderton et al., *The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III*, *ApJS* **219** (July, 2015) 12, [`1501.00963`].

[18] J. P. Gardner, N. Katz, L. Hernquist and D. H. Weinberg, *The Population of Damped Lyα and Lyman Limit Systems in the Cold Dark Matter Model*, *Astrophysical Journal* **484** (July, 1997) 31–39, [astro-ph/9609072].

[19] A. Pontzen, F. Governato, M. Pettini, C. M. Booth, G. Stinson, J. Wadsley et al., *Damped Lyman α systems in galaxy formation simulations*, *MNRAS* **390** (Nov., 2008) 1349–1371, [0804.4474].

[20] K. Nagamine, A. M. Wolfe, L. Hernquist and V. Springel, *Distribution of Damped Lyα Absorbers in a Λ Cold Dark Matter Universe*, *Astrophysical Journal* **660** (May, 2007) 945–958, [astro-ph/0510729].

[21] E. Tescari, M. Viel, L. Tornatore and S. Borgani, *Damped Lyman α systems in high-resolution hydrodynamical simulations*, *MNRAS* **397** (July, 2009) 411–430, [0904.3545].

[22] D. Erkal, N. Y. Gnedin and A. V. Kravtsov, *On the Origin of the High Column Density Turnover in the H I Column Density Distribution*, *Astrophysical Journal* **761** (Dec., 2012) 54, [1201.3653].

[23] S. Bird, M. Vogelsberger, M. Haehnelt, D. Sijacki, S. Genel, P. Torrey et al., *Damped Lyman α absorbers as a probe of stellar feedback*, *MNRAS* **445** (Dec., 2014) 2313–2324, [1405.3994].

[24] M. Bartelmann and P. Schneider, *Weak gravitational lensing*, *Physics Reports* **340** (Jan., 2001) 291–472, [astro-ph/9912508].

[25] D. Blas, J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes*, *JCAP* **7** (July, 2011) 034, [1104.2933].

[26] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont et al., *Planck 2015 results. XIII. Cosmological parameters*, *A&A* **594** (Sept., 2016) A13, [1502.01589].

[27] Planck Collaboration et al., *Planck 2015 results. XV. Gravitational lensing*, *A&A* **594** (Sept., 2016) A15, [1502.01591].

[28] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke et al., *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, *Astrophysical Journal* **622** (Apr., 2005) 759–771, [astro-ph/0409513].

[29] P. Noterdaeme, P. Petitjean, W. C. Carithers, I. Pâris, A. Font-Ribera, S. Bailey et al., *Column density distribution and cosmological mass density of neutral gas: Sloan Digital Sky Survey-III Data Release 9*, *A&A* **547** (Nov., 2012) L1, [1210.1213].

[30] J. E. Bautista, N. G. Busca, J. Guy, J. Rich, M. Blomqvist, H. du Mas des Bourboux et al., *Measurement of baryon acoustic oscillation correlations at z = 2.3 with SDSS DR12 Lyα-Forests*, *A&A* **603** (June, 2017) A12, [1702.00176].

[31] H. du Mas des Bourboux, J.-M. Le Goff, M. Blomqvist, N. G. Busca, J. Guy, J. Rich et al., *Baryon acoustic oscillations from the complete SDSS-III Lyα-quasar cross-correlation function at z = 2.4*, *ArXiv e-prints* (Aug., 2017) , [1708.02225].

[32] R. Garnett, S. Ho, S. Bird and J. Schneider, *Detecting damped Ly α absorbers with Gaussian processes*, *MNRAS* **472** (Dec., 2017) 1850–1865, [1605.04460].

[33] A. Font-Ribera, E. Arnau, J. Miralda-Escudé, E. Rollinde, J. Brinkmann, J. R. Brownstein et al., *The large-scale quasar-Lyman α forest cross-correlation from BOSS*, *JCAP* **5** (May, 2013) 018, [1303.1937].

[34] S. Eftekharzadeh, A. D. Myers, M. White, D. H. Weinberg, D. P. Schneider, Y. Shen et al., *Clustering of intermediate redshift quasars using the final SDSS III-BOSS sample*, *MNRAS* **453** (Nov., 2015) 2779–2798, [1507.08380].

[35] P. Laurent, S. Eftekharzadeh, J.-M. Le Goff, A. Myers, E. Burtin, M. White et al., *Clustering of quasars in SDSS-IV eBOSS: study of potential systematics and bias determination*, *JCAP* **7** (July, 2017) 017, [1705.04718].

[36] E. Hivon, K. M. Górski, C. B. Netterfield, B. P. Crill, S. Prunet and F. Hansen, *MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets*, *Astrophysical Journal* **567** (Mar., 2002) 2–17, [astro-ph/0105302].

[37] C. Doux, E. Schaan, E. Aubourg, K. Ganga, K.-G. Lee, D. N. Spergel et al., *First detection of cosmic microwave background lensing and Lyman-α forest bispectrum*, *Physical Review D* **94** (Nov., 2016) 103506, [1607.03625].

[38] C. Doux, M. Penna-Lima, S. D. P. Vitenti, J. Tréguer, E. Aubourg and K. Ganga, *Cosmological constraints from a joint analysis of cosmic microwave background and large-scale structure*, *ArXiv e-prints* (June, 2017) , [1706.04583].

[39] M. Tegmark, A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley and A. S. Szalay, *Measuring the Galaxy Power Spectrum with Future Redshift Surveys*, *Astrophysical Journal* **499** (May, 1998) 555–576, [astro-ph/9708020].

[40] M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein et al., *The DESI Experiment, a whitepaper for Snowmass 2013*, *ArXiv e-prints* (Aug., 2013) , [1308.0847].

[41] R. S. de Jong, S. C. Barden, O. Bellido-Tirado, J. G. Brynnel, S. Frey, D. Giannone et al., *4MOST: the 4-metre Multi-Object Spectroscopic Telescope project at preliminary design review*, in *Ground-based and Airborne Instrumentation for Astronomy VI*, vol. 9908 of *Proc. SPIEE*, p. 99081O, Aug., 2016, DOI.

[42] F. De Bernardis, J. R. Stevens, M. Hasselfield, D. Alonso, J. R. Bond, E. Calabrese et al., *Survey strategy optimization for the Atacama Cosmology Telescope*, in *Observatory Operations: Strategies, Processes, and Systems VI*, vol. 9910 of *Proc. SPIEE*, p. 991014, July, 2016, 1607.02120, DOI.

[43] K. N. Abazajian, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, K. S. Arnold et al., *CMB-S4 Science Book, First Edition*, *ArXiv e-prints* (Oct., 2016) , [1610.02743].