Small-scale Lyman alpha forest cosmology with PRIYA: Constraints from XQ100 and KODIAQ-SQUAD one-dimensional flux power spectra

Ming-Feng Ho[1,2,3]**, Mahdi Qezlou**[1,4]**, Simeon Bird**[1]**, Yanhui Yang**[1]**, Camille Avestruz**[2,3]**, M.A. Fernandez**[1,5]**, and Vid Iršič**[6]**,**

[1]Department of Physics and Astronomy, University of California Riverside, 900 University Ave, Riverside, CA 92521

[2]Department of Physics, University of Michigan, 450 Church St, Ann Arbor, MI 48109, USA

[3]Leinweber Institute for Theoretical Physics, 450 Church St, Ann Arbor, MI 48109, USA

[4]Department of Astronomy, University of Texas at Austin, 2515 Speedway Boulevard, Stop C1400, Austin, TX 78712, USA

[5]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA

[6]Center for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

E-mail: mfho@umich.edu

Abstract. We present a new cosmological analysis of the small-scale Lyman-α forest 1D flux power spectrum (P1D) using high-resolution quasar spectra from XQ100 and KODIAQ-SQUAD, interpreted through the PRIYA emulator. PRIYA is a suite of galaxy formation simulations spanning a range of cosmological and inhomogeneous He II reionization parameters, enabling few-percent-level predictions of the P1D. These datasets, probing down to $k \sim 6\,h\,\mathrm{Mpc}^{-1}$ at $z = 2 - 5$, offer access to non-linear scales inaccessible to large-volume surveys like eBOSS. We find that the XQ100 P1D yields constraints on the primordial power spectrum parameters (A_P, n_P) at pivot scale $k_0 = 0.78\,\mathrm{Mpc}^{-1}$ that are consistent with PRIYA results from eBOSS DR14 and Planck CMB, albeit with broader uncertainties. Notably, this is achieved without external IGM temperature data, showing that XQ100 alone provides stronger constraints on thermal history than eBOSS DR14. In contrast, the KODIAQ-SQUAD P1D favors a significantly higher A_P value, driven by the selection bias toward high-column density absorbers (HCDs). We also find that the P1D at $k > 0.045\,\mathrm{s/km}$ is more sensitive to Lyman limit system contamination and thermal history. When imposing a prior on (A_P, n_P), the reduced χ^2 remains unchanged and the inferred mean IGM temperature is unaffected, suggesting that cosmological and thermal parameters are largely sensitive to different scales. The XQ100 P1D therefore provides complementary information on thermal nuisance parameters, which can be jointly fit with eBOSS or DESI P1D measurements to improve cosmological constraints.

Contents

1 Introduction

The Lyman-α forest traces the distribution of neutral gas at low densities in the intergalactic medium (IGM) at redshifts $z = 2-5$ [1–9]. This neutral gas traces cosmic structure formation, making the Lyman-α forest an exceptionally powerful cosmological probe, sensitive to the distribution of dark matter. Lyman-α forest observations play a crucial role in cosmological spectroscopic surveys, such as the Sloan Digital Sky Survey (SDSS) [10–14] and the Dark Energy Spectroscopic Instrument (DESI) [15–21], enabling these surveys to detect Baryonic Acoustic Oscillations (BAO) and probe the growth of cosmic structure on scales, $k \sim 0.01-10$ h/Mpc [22–29], and at higher redshifts ($z = 2 - 5$) than most other large scale structure probes.

The Lyman-α forest not only serves as a probe of the cosmic web but also provides a unique window into the thermal and ionization history of the IGM [30–37]. This sensitivity to the IGM thermal history enables constraints on processes such as hydrogen and helium reionization, which left distinct imprints on the thermal evolution of the universe [38–41]. At the same time, the Lyman-α forest is sensitive to the free-streaming scale of warm dark matter (WDM), offering constraints on its particle mass [28, 42–47]. Dark matter free-streaming suppresses the formation of structure below a characteristic scale, which can be constrained by precision measurements of the Lyman-α flux power spectrum. In addition, analyses of the small-scale Lyman-α forest P1D have yielded constraints on other alternative dark matter models, such as fuzzy dark matter (see Ref. [48]). An important strength of the Lyman-α forest as a cosmological probe is its ability to access non-linear clustering statistics over nearly three orders of magnitude in scale: from the largest scales of ~ 100 Mpc probed by cross-sightline correlations (see, e.g., 3D Lyman-α power spectrum, P3D in Ref. [49–51]), to intermediate scales of ~ 1 Mpc accessible through the 1D power spectrum in systematics-limited surveys, and ~ 0.1 Mpc from small high-resolution datasets.

Large-scale spectroscopic surveys like the SDSS [14] and DESI [15] have measured the 1D flux power spectrum from the Lyman-α forest. The extended Baryon Oscillation Sky Survey (eBOSS), part of SDSS, mapped the flux power spectrum using 43,751 high signal to noise quasar sightlines [14]. Despite its scale, eBOSS remained limited by systematic uncertainties, particularly those arising from spectrograph resolution and calibration issues, that dominated its error budget. DESI addresses these limitations with a substantial improvement in spectral resolution, nearly doubling the resolving power of its predecessors. This enhancement allows DESI to probe smaller scales, down to $k \sim 0.035$ s km^{-1} [18, 52]. The full projected five-year dataset of 700,000 Lyman-α forest quasars from DESI will achieve percent level accuracy even on the smallest scales, substantially improving cosmological constraints and revealing fine details of the IGM thermal history.

A further source of complementary information comes from high-resolution quasar spectra surveys. These include the X-Shooter quasar catalog 100 (XQ100), a blind survey of 100 quasar spectra, and two catalogues assembled from the telescope archives: the Keck Observatory Database of Ionized Absorption toward Quasars (KODIAQ), and the UVES Spectral Quasar Absorption Database (SQUAD). These surveys provide information on the Lyman-α forest at the smallest physically accessible scales, $k \approx 0.001 - 0.2$ s km^{-1} [53–57][1], and are especially sensitive to the IGM thermal history. In this work, we present an analysis of the Lyman-α P1D from these high-resolution quasar spectroscopic surveys. We reanalyze the XQ100 P1D ($R \sim 5\,000$) [58] and, for the first time, perform joint inference of cosmological

---

[1]$k \approx 0.001 - 0.2$ s km^{-1} is probed by Ref. [57].

Simulation	Box Volume	N_{gas}	M_{gas} ($M_\odot$ h^{-1})
LF	$(120 \text{ Mpc } h^{-1})^3$	1536^3	$[5.29, 6.98] \times 10^6$
HF	$(120 \text{ Mpc } h^{-1})^3$	3072^3	$[6.73, 7.97] \times 10^5$

Table 1. Low-Fidelity (LF) and High-Fidelity (HF) simulation suite details. N_{gas} is the number of gas particles simulated, M_{gas} is the mass resolution of those particles, which depends on Ω_M.

parameters and IGM thermal history using the KODIAQ-SQUAD P1D ($R \sim 38\,000$) [55], which is the largest high-resolution dataset yet assembled, comprising ~ 700 quasar sightlines with $z_{QSO} = 2 - 5$.

High-resolution hydrodynamical simulations are required to capture the non-linear small-scale clustering of neutral hydrogen, but they are computationally expensive and face an inherent trade-off between resolution and volume. To enable cosmological inference, theoretical predictions of the flux power spectrum must achieve percent-level accuracy across a wide range of cosmological and astrophysical parameters. Currently, all available Lyman-α emulators built on hydrodynamical simulations, e.g., Refs. [59–62], are constrained by this resolution/volume trade-off, which limits their ability to extend to the smallest scales of the P1D and beat down the cosmic variance.

Our work builds on the PRIYA simulation suite [63], a state-of-the-art suite of large-volume, high resolution simulations of the Lyman-α forest which uniquely include patchy hydrogen and helium reionization. The PRIYA suite [63] addresses the resolution/volume challenge by employing a multi-fidelity emulator framework (see, e.g., Ref. [64–69]), integrating low- (1536^3) and high-resolution (3072^3) hydrodynamical simulations within $(120\,\text{Mpc}/h)^3$ volumes, which is so far the largest volume for a Lyman-α emulator. This strategy allows PRIYA to span a wide range of cosmological and reionization parameters without the need for ad hoc splicing corrections. Through multi-fidelity emulation, low-resolution outputs are corrected using high-resolution samples, preserving physical accuracy while extending parameter coverage.

In PRIYA, we vary the AGN feedback strength as one of the astrophysical parameters. As noted in Ref. [63], AGN feedback has little impact on the Lyman-α forest flux power spectrum within current survey errors, especially at $z > 2$ (see also Ref. [70]). In this work, however, we take the view that He II reionization is quasar-driven, and therefore we can view the He II reionization effect in PRIYA as a form of AGN (quasar) feedback.

The structure of this paper is as follows: In § 2, we describe the PRIYA simulation suite and the construction of the multi-fidelity emulator framework. § 3 outlines our inference pipeline, including introductions to KODIAQ-SQUAD and XQ100 P1Ds, high-column density absorber (HCDs) templates, likelihood function and parameter priors. We present the main results, including posterior constraints from the XQ100 and KODIAQ-SQUAD datasets, in § 4. In § 5, we discuss the implications of these findings, highlight potential limitations of the current P1D analysis, and suggest directions for future work. Finally, § 6 summarizes our findings and discusses prospects for next-generation Lyman-α forest studies.

2   Simulation Suite and Emulator

We use the PRIYA simulation suite [63] to predict the 1D flux power spectrum (or P1D) of the Lyman-α forest as a function of cosmology. In this section, we briefly recap the simulation

Parameter	Minimum	Maximum	Description
n_P	0.8	1.05	Scalar spectral index
A_P	1.2×10^{-9}	2.6×10^{-9}	Power amplitude at $k = 0.78 \, \mathrm{Mpc}^{-1}$
$\Omega_M h^2$	0.14	0.146	Total matter density
z_{Hei}	3.5	4.1	Start redshift of He II reionization
z_{Hef}	2.6	3.2	End redshift of He II reionization
α_q	1.3	2.5	Quasar spectral index during He II reionization
z_{Hi}	6.5	8	Median redshift of H I reionization
Prior-dominated			
h	0.65	0.75	Hubble parameter
ϵ_{AGN}	0.03	0.07	Thermal efficiency of black hole feedback
Post-processing			
τ_0	0.75	1.25	Mean optical depth at $z = 3$ in Eq. 2.2.
$d\tau_0$	-0.4	0.25	τ_0 redshift evolution in Eq. 2.2.

Table 2. Summary of emulator and post-processing parameters. We vary a total of 11 parameters: 4 for cosmology, 3 for the helium reionization model, 1 for the hydrogen reionization model, 1 for the strength of AGN feedback and 2 for the mean optical depth.

suite and the multi-fidelity emulator used to predict the P1D. Detailed descriptions of the simulation suite and emulator can be found in Refs. [63, 67]. For reviews of multi-fidelity emulation, see Ref. [65] for a thorough methodology review and Ref. [66, 68, 71–73] for cosmology applications.

The PRIYA suite[2] is run with the MP-Gadget[3] N-body and smoothed particle hydrodynamics (SPH) code. PRIYA suite is designed around the multi-fidelity emulator strategy, and combines simulations with two different resolutions to predict the P1D at high-fidelity through a machine learning model (see Ref. [67]). We run two sets of simulations, 60 low-fidelity (LF) and 3 high-fidelity (HF) simulations, with different particle loads, 1536^3 and 3072^3, in a $120 \, \mathrm{Mpc}/h$ co-moving box, as shown in Table 1. A multi-fidelity emulator is a machine learning model that interpolates over simulations with multiple fidelity levels. It predicts the output of a simulation at the highest fidelity for arbitrary input cosmological parameters within the parameter limits, just as a traditional emulator with homogeneous simulation input, but with a reduced cost training set. The accuracy of the multi-fidelity emulator is validated against high-fidelity simulation outputs using leave-one-out cross-validation (LOO-CV), as described in Ref. [63].

§ 2.1 describes and validates our Lyman-α forest 1D flux power spectrum emulator for $k = 0.003 - 0.06 \, \mathrm{s/km}$, which extends the emulator from Ref. [74] to model smaller scales. § 2.2 describes the cosmological and astrophysical parameters varied in the emulator, and the parameter prior limits.

2.1 P1D Emulator for XQ100 and KODIAQ-SQUAD

In Ref. [74], we trained a multi-fidelity emulator model which predicts the P1D at $\sim 1\%$ accuracy for the scales measured by eBOSS, $k = 0.001 - 0.02 \, \mathrm{s/km}$. In this work, we extend the emulator to predict the P1D on the smaller scales measured by XQ100 and KODIAQ-

---

[2] http://priya-sims.readthedocs.io
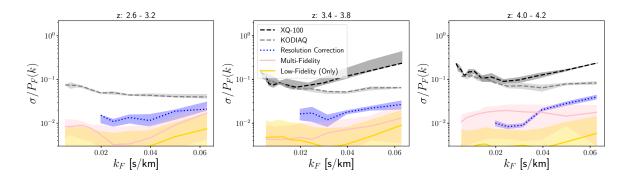[3] https://github.com/MP-Gadget/MP-Gadget

Figure 1. Leave-one-out cross-validation (LOO-CV) accuracy of the multi-fidelity emulator at $z = 2.6 - 3.2, 3.4 - 3.8$, and $4.0 - 4.2$. Pink lines (multi-fidelity) show the LOO-CV error against HF simulations (median and 68% percentiles), while yellow lines (LF only) show the LOO-CV of LF emulator against LF simulations. We overplot the square root of the diagonal elements of the covariance matrix of KODIAQ-SQUAD (gray dotted) and XQ100 (black dashed) for comparison. Blue lines show the convergence error estimated in Ref. [63] between HF and the extreme high-resolution simulation (2×768^3 particles in a $15\,\mathrm{Mpc}/h$ box).

SQUAD, $k = 0.003 - 0.06\,\mathrm{s/km}$. Our emulator construction, including multi-fidelity corrections and hyperparameters, is identical to Ref. [74].

Following Ref. [63], we validate our emulator using leave-one-out-cross-validation (LOO-CV) with each high fidelity simulation left out in turn. The LOO-CV results are shown in Figure 1 for three redshift ranges, $z = 2.6 - 3.2, 3.4 - 3.8$, and $4.0 - 4.2$. Note that we only have 3 HF simulations, so the LOO-CV would significantly over-estimate the HF interpolation error because we are discarding 1/3 of the HF training set. We find that the emulator interpolates the P1D at $k = 0.01 - 0.06\,\mathrm{s/km}$ at $\sim 1\%$ accuracy (median). At higher redshifts, $z = 4 - 4.2$, the interpolation error is slightly larger, $\sim 2\%$, as the Lyman-α forest probes smaller physical scales at higher redshifts [75].

The interpolation error is not the only source of uncertainty in the emulator, as the emulator is trained on simulations with a finite resolution. Ref. [63] checked the resolution convergence and found it to be $\sim 1\%$ for $k < 0.05\,\mathrm{s/km}$ and $\sim 3\%$ for $k = 0.05 - 0.06\,\mathrm{s/km}$, which is substantially smaller than the statistical error of 10% for KODIAQ-SQUAD and $\sim$ 20% for XQ100. Smaller scales are less well resolved, with on the order of $4 - 10\%$ convergence error, and for this reason we conservatively exclude these scales from our analysis. We attempt to improve resolution convergence by applying a cosmology-independent correction function in the emulator. Specifically, we multiply by P1D for $k = 0.02 - 0.06\,\mathrm{s/km}$ by the ratio between the two highest resolution simulations run by Ref. [63]. These are, respectively, 2×512^3 particles in a $15\,\mathrm{Mpc}/h$ box (thus a resolution equal to our high fidelity tier), and 2×768^3 particles in a $15\,\mathrm{Mpc}/h$ box.

2.2 Cosmological & Astrophysical Parameters

The cosmological and astrophysical parameters varied in the emulator are identical to Ref. [74], and are summarised in Table 2.[4] Our emulator contains two parameters related to the primordial (the pre-transfer function) power spectrum are varied: the scalar spectral index, n_P,

---

[4]We have constructed an online widget to visualize how the P1D depends on emulator parameters, available at https://jibancat.github.io/priya-widget/

and the power amplitude at $k = 0.78 \, \mathrm{Mpc}^{-1}$, A_P:

$$P(k) = A_P \left(\frac{k}{0.78 \, \mathrm{Mpc}^{-1}} \right)^{n_P - 1} . \tag{2.1}$$

The pivot scale $0.78 \, \mathrm{Mpc}^{-1}$ is chosen to match the scales probed by the Lyman-α forest. For eBOSS, this pivot scale minimises the degeneracy between n_P and A_P, although we shall see that some degeneracy exists for the high-resolution P1D we examine here, which probes smaller scales (see Figure 5). We also vary the Hubble parameter h and the total matter density $\Omega_M h^2$, although these have small $(1 - 4\%)$ effects on the Lyman-α forest and are not strongly constrained by the high resolution measurements. The dominant effect of h is to slightly change the mapping between simulated (Mpc/h) and observed (km/s) units.

There are three Helium reionization parameters, which control the thermal history of the IGM. The He II reionization model follows Ref. [41]: z_{Hei} and z_{Hef} are the redshifts for the start and end of He II reionization. α_q is the quasar spectral index, which scales the peak temperature during He II reionization, with a *lower* α_q value corresponding to a higher heating rate. z_{Hi} is the midpoint redshift of H I reionization, defined as the redshift at which 50% of the hydrogen is ionized. Note that our emulator includes scenarios where hydrogen reionization completes at $z < 6$, as supported by recent quasar dark gap measurements [76]; for $z_{Hi} = 6.5$, hydrogen is 95% ionized at $z = 5.5$.

Finally, we have two additional parameters for the Lyman-α effective optical depth, which measures the average amount of the Lyman-α forest absorption as a function of redshift and is set by the unknown UV photon background density. Higher redshifts have a larger effective optical depth, corresponding to a more neutral IGM. We parameterize the mean flux $\bar{\mathcal{F}} = \exp(-\tau_{\mathrm{H\,I}}^{\mathrm{eff}})$ as variations from the power law redshift evolution model of Ref. [77].

$$\begin{aligned} \tau_{\mathrm{H\,I}}^{\mathrm{Kim}}(z) &= 0.0023 \times (\tau_{z=3}) \times (1+z)^{3.65} , \\ \tau_{\mathrm{H\,I}}^{\mathrm{eff}}(z) &= \tau_0 \left(\frac{\tau_{\mathrm{H\,I}}^{\mathrm{Kim}}(z)}{\tau_{\mathrm{H\,I}}^{\mathrm{Kim}}(3)} \right)^{d\tau_0} \tau_{\mathrm{H\,I}}^{\mathrm{Kim}}(z) \\ &= \tau_0 \left(\frac{1+z}{4} \right)^{3.65 + d\tau_0} \tau_{\mathrm{H\,I}}^{\mathrm{Kim}}(z) . \end{aligned} \tag{2.2}$$

In practice, we rescale the optical depths in our simulated spectra by a constant factor in post-processing to match the desired $\bar{\mathcal{F}}$. We sample 10 linearly spaced values of the mean flux, independent of redshift, to cover the range of τ_0 and $d\tau_0$ at the extreme ($z = 2.2$ and $z = 4.6$) redshifts. The final set of P1D training set for the emulator is thus ten times the number of simulations, or 600 LF, and 30 HF simulated flux power spectra.

Our parameterization in Eq. 2.2 is based on Ref. [77], which is measured from 18 high resolution, high signal-to-noise VLT/UVES QSO spectra, and we allow the mean flux to vary (with a pivot point at $z = 3$) by τ_0 and the redshift evolution by $d\tau_0$ with a large enough prior. This allows us to capture a wide variety of the mean flux evolutions.[5]

3 Inference Scheme and Likelihood Function

We perform the likelihood analysis of the XQ100 and KODIAQ-SQUAD P1D using the emulator described in § 2. The inference scheme is as follows:

---

[5]We show the mean flux prior in Fig. 15 alongside the meanflux measurement from Ref. [77] and Ref. [78].

1. Predict the flux power spectrum (P1D) from the emulator for a set of input parameters (see Table 2).

2. Calculate a likelihood comparing these predictions to the XQ100 or KODIAQ-SQUAD P1D.

3. Run Markov Chain Monte Carlo (MCMC) chains using Cobaya [79, 80] to compute posterior parameter constraints.

Step 1 is straightforward, as the emulator trained on the PRIYA simulation suite can predict the P1D at the desired redshifts and scales. For Step 2, we directly take the covariance matrix from the XQ100 or KODIAQ-SQUAD data, and compute the multi-variate normal likelihood using Equation 3.2. With the likelihood in hand, we run MCMC chains using Cobaya.

§ 3.1 describes the covariance matrix used in the likelihood calculation. Likelihood function and prior are detailed in § 3.2 and § 3.3, respectively. A validation of our MCMC inference using simulated data is presented in § 3.4. § 3.5 discusses the flux power spectrum data from XQ100 and KODIAQ-SQUAD. Finally, § 3.6 describes the high-column density absorbers (HCDs) template used for marginalizing out the HCD contamination.

3.1 Covariance Matrix

Below we describe the covariance matrix used in the likelihood calculation.

$$\boldsymbol{K} = \boldsymbol{K}_{\mathrm{FPS}} + \boldsymbol{\sigma}_{CV} \cdot \boldsymbol{\sigma}_{CV}^T \,. \tag{3.1}$$

Here $\boldsymbol{K}_{\mathrm{FPS}}$ is the covariance matrix from either XQ100 data or KODIAQ-SQUAD data. $\boldsymbol{\sigma}_{CV}$ is residual sample variance from the finite box size. We include an estimate of sample variance using the leave-one-out errors discussed in Ref. [63], a technique made possible by the inclusion of h in our simulation suite. Gadget uses Mpc/h units, and the conversion of wavenumbers from h/Mpc to s/km changes the Fourier modes between bins depending on the value of h, which acting as a proxy for the sample variance from different initial phases. We approximate $\boldsymbol{\sigma}_{CV}$ with the averaged variance of the leave-one-out errors using the low fidelity simulations. Details of this approximation are discussed in Ref. [63].

3.2 Likelihood

The likelihood function is a multivariate normal likelihood summed over all redshifts and, for the flux power, all wavenumber bins:

$$\log \mathcal{L} = -\frac{1}{2} \sum_{z=z_{\mathrm{min}}}^{z=z_{\mathrm{max}}} \left(\left(\boldsymbol{P}_F^{\mathrm{diff}} \right)^\top \cdot \boldsymbol{K}^{-1} \cdot \boldsymbol{P}_F^{\mathrm{diff}} + \log \left(\det(\boldsymbol{K}) \right) \right) \,. \tag{3.2}$$

Here $\boldsymbol{P}_F^{\mathrm{diff}} = \boldsymbol{P}_F^{\mathrm{sim}} - \boldsymbol{P}_F^{\mathrm{obs}}$ is the vector difference between the simulation prediction and the observation. We chose to compute the $\boldsymbol{P}_F^{\mathrm{sim}}$ on the bins of the observed data. The covariance matrix, $\boldsymbol{K}$, is described in Equation 3.1. The minimum and maximum redshifts are set to $z = 3.4$ and $z = 4.2$ for the XQ100 data ($z = 2.6$ and $z = 4.2$ for the KODIAQ-SQUAD data). We use the same Cobaya MCMC sampler as in Ref. [74], which uses the Metropolis method discussed in Ref. [81], and a Gaussian + exponential proposal distribution that dynamically learns the proposal covariance.
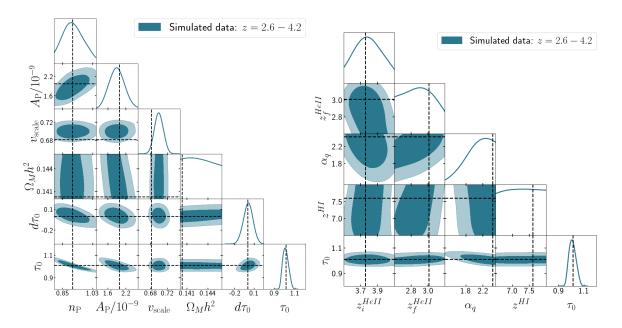
Figure 2. Posterior contours from our likelihood function and emulator, using simulated data and the KODIAQ-SQUAD covariance matrix for validation, restricted to $z = 2.6 - 4.2$. The left panel shows the cosmological parameters, and the right panel shows the reionization parameters. The dashed black lines indicate the correct input values of the parameters. Contours show the $1 - \sigma$ and $2 - \sigma$ confidence regions.
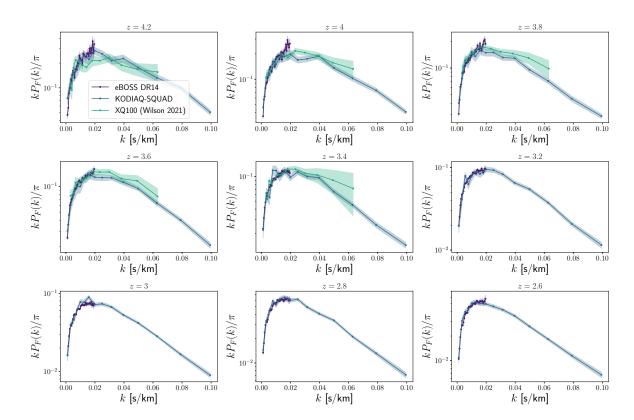
3.3 Priors

The parameter priors are shown in Table 2. We use an informative Gaussian prior on the AGN feedback parameter, ϵ_{AGN}, with $\mu = 0.05$ and $\sigma = 0.005$, as it is poorly constrained by the data and has minimal correlations with other parameters. We have shown in Ref. [63] that the AGN feedback parameter has minimal effect on the Lyman-α forest 1D flux power spectrum for the redshift range of interest. The ϵ_{AGN} changes the P1D at the level 2% at $k \sim 0.06\,\mathrm{s/km}$ rather than the sub-percent level for $k < 0.02\,\mathrm{s/km}$. However, this is still small compared to the statistical $\sim 10\%$ uncertainty of KODIAQ-SQUAD at $k \sim 0.06\,\mathrm{s/km}$. We experimented with using a uniform prior on ϵ_{AGN}. The mode shifts to $\epsilon_{AGN} \sim 0.03$, but there is negligible change to the other parameters.

We also place a weak Gaussian prior on the Hubble parameter, h, with $\mu = 0.70$ and $\sigma = 0.015$, as it is weakly constrained by the P1D. For all other parameters we use uniform priors within the parameter limits.

3.4 Inference Using Simulation Data from PRIYA

Figure 2 shows a validation of our inference scheme, using the simulation outputs in place of the observational data. Dashed black lines indicate the correct input parameters, and contours show the posterior outputs of the likelihood. We have a similar validation in Ref. [74], but here we use the covariance matrix and bins from KODIAQ-SQUAD instead of eBOSS.

With the simulated data as validation, we can test whether the likelihood function recovers the correct mode of the parameters, and we can also verify which parameters are not well constrained by this likelihood function. From Figure 2, we see the likelihood function

Figure 3. The flux power spectrum data from XQ100 (green) and KODIAQ-SQUAD (blue). We show eBOSS DR14 (purple) data for comparison. The XQ100 data is re-analyzed from Ref. [58], and the KODIAQ-SQUAD data is from the optimal estimator in Ref. [55]. The old XQ100 data from Ref. [53] is not used in the analysis. The XQ100 data is limited to $z = 3.4-4.2$ and $k = 0.003-0.065\,\mathrm{s/km}$, while the KODIAQ-SQUAD data is limited to $z = 2.6 - 4.2$ and $k = 0.0055 - 0.065\,\mathrm{s/km}$ in our analysis. The KODIAQ-SQUAD P1D span a wider range of k than XQ100 with a tighter 1σ errorbar.

recovers the correct mode of $n_P, A_P, z_i^{\mathrm{HeII}}, z_f^{\mathrm{HeII}}, \tau_0, d\tau_0$, and α_q. All parameters are falling within the $1-\sigma$ range of the correct input values.

There are some parameters that are not well constrained by the likelihood function, such as v_{scale}, which is the Hubble parameter h, and $\Omega_M h^2$. As we mentioned in Ref. [63], the MP-Gadget simulation is run with comoving units Mpc/h, so the h does not directly affect the gravitational evolution in our simulations. The primary effect of h in this case is the change of comoving unit to velocity units (km/s) in P1D, that is, shifting the Fourier modes between bins. This effect is dominated by the residuals of the sample variance due to our finite box size. Even with this residual, the P1D is only weakly sensitive to h at $< 2\%$ effect within the prior range.

z_{HI}, the mid-point of the redshift of hydrogen reionization, is also not well constrained by the likelihood function. This is because the P1D is not very sensitive to the redshift of hydrogen reionization, as it is higher redshift ($z \gtrsim 5$) than the redshift range of the P1D data we use ($z = 2.6 - 4.2$).

3.5 Flux Power Spectrum: XQ100 & KODIAQ-SQUAD

Figure 3 shows the P1D data used in this work. We use measurements from three high-resolution and high SNR quasar surveys: XQ100 [82], KODIAQ [83–85], and SQUAD [86].

The XQ100 survey comprises 100 quasars at $3.5 < z_{\rm QSO} < 4.5$ [82], observed with the X-Shooter spectrograph on the ESO Very Large Telescope (VLT) [87], with a resolving power of $R \sim 4\,000 - 7\,000$. We adopt the re-analysis by Ref. [58], which introduces a refined and improved covariance calculation using Lyβ regions. We restrict the usable redshift range to $z = 3.4 - 4.2$, following the re-analysis by Ref. [58].

The KODIAQ-SQUAD dataset combines archival spectra from Keck/HIRES [88] and VLT/UVES [89], both offering higher resolving power ($R \gtrsim 36\,000 - 40\,000$) over the redshift range $z = 2.0 - 5.0$ with 767 quasars in total. The joint measurement of P1D in Ref. [55] uses an optimal quadratic estimator, which is robust to masking and spectrograph inhomogeneities. For this work, we use only the KODIAQ and SQUAD subsamples from Ref. [55] (excluding XQ100 to avoid redundancy).

We have also made some decisions on the data cuts (redshift and k ranges). The k-range and redshift range of the XQ100 data is limited to $k = 0.003 - 0.064\,{\rm s/km}$ and $z = 3.4 - 4.2$, which are largely unchanged from the conservative data cuts already implemented by Ref. [58].

We limit the KODIAQ-SQUAD P1D to $k = 0.0055 - 0.065\,{\rm s/km}$ and $z = 2.6 - 4.2$. It was suggested in Ref. [55] to only use $0.004 < k < 0.1\,{\rm s/km}$ as a conservative cut. We remove additional bins, so that the lowest k bin is $k = 0.0055\,{\rm s/km}$. This is to avoid potential cosmic variance issues due to the small sample size, and because eBOSS already measures the low-k P1D well. Our focus here is on the small-scale P1D. The maximum k is set to $0.065\,{\rm s/km}$, due to the limited resolution of our simulation suite rather than the data itself. Scales above $0.065\,{\rm s/km}$ have $4 - 10\%$ resolution convergence error in our HF simulation suite at some redshift bins, which is comparable to the error of the KODIAQ-SQUAD P1D.

We also limit the analysed redshift range to $z = 2.6 - 4.2$, excluding data at high and low redshifts, as it is known that these ranges are difficult to handle data-wise. There have been discussions on the disagreement between eBOSS and other P1D measurements such as DESI DR1 and KODIAQ-SQUAD in bins with $z < 2.6$ (see, e.g., Ref. [18, 74]). These low-redshift bins are from quasar sightlines with shorter wavelength segments, which are more challenging for DLA finders. At high redshift $z > 4.2$, the sample size is small, which makes the covariance difficult to estimate, and implies limited impact on the overall cosmological inference in any case.

The KODIAQ-SQUAD P1D is a re-reduction of all available Keck and UVES spectra across a wide range of target selection strategies over 300+ PIs. Ref. [55] noted that KODIAQ targets O IV systems while UVES targets known DLAs, which introduce a selection bias toward higher HCD contamination. We will discuss the impact of this selection bias on our results in § 4 and § 5.

3.6 Template for DLAs and LLSs

Both the XQ100 and KODIAQ-SQUAD spectra have high resolution and SNR, enabling observers to visually identify and mask regions with prominent damped Lyman-α absorbers (DLAs), which are neutral hydrogen (HI) systems with column densities $N_{\rm HI} > 10^{20.3}\,{\rm cm}^{-2}$. However, sub-DLAs, which have column densities $10^{19.0} < N_{\rm HI} < 10^{20.3}\,{\rm cm}^{-2}$, are easier to miss. Lyman limit systems (LLSs), which have column densities $10^{17.2} < N_{\rm HI} < 10^{19.0}\,{\rm cm}^{-2}$ cannot be reliably identified on an absorber-by-absorber basis. To maintain consistency across datasets, we include the full four-parameter HCD template in the likelihood: $\alpha_{\rm LLS}$,

α_{subDLA}, $\alpha_{\mathrm{small-DLA}}$, and $\alpha_{\mathrm{large-DLA}}$.[6] We use the template from Ref. [90], which is based on simulated absorber populations in the Illustris simulation (which has a 106.5 Mpc box).

This differs from our modeling choice in Ref. [74], where we adopted a coarser two-parameter scheme separating systems above and below the DLA column density threshold (constant ratios for $\alpha_{\mathrm{LLS}}/\alpha_{\mathrm{subDLA}} = 1$ and $\alpha_{\mathrm{small-DLA}}/\alpha_{\mathrm{large-DLA}} = 1$). Since KODIAQ-SQUAD especially probes significantly smaller scales than eBOSS DR14, the contribution from LLSs and sub-DLAs becomes more relevant than that from DLAs ($N_{\mathrm{HI}} > 10^{20.3}\,\mathrm{cm}^{-2}$). It is therefore more appropriate to explicitly separate the LLS and sub-DLA terms, while retaining small- and large-DLA parameters to verify that masking was sufficiently effective.

The PRIYA simulations contain a realistic population of high-column-density systems. We follow the same masking procedure as in Ref. [14], masking all DLAs with $N_{\mathrm{HI}} \gtrsim 10^{20}\,\mathrm{cm}^{-2}$. Consequently, we do not expect DLAs to contribute significantly to the model P1D, and set the priors on $\alpha_{\mathrm{small-DLA}}$ and $\alpha_{\mathrm{large-DLA}}$ to be positive, accounting for potential residual contamination in the data. For sub-DLAs, which are only partially masked in PRIYA, we also expect α_{subDLA} to be positive. For LLSs, we only allow α_{LLS} to take positive values, which corresponds to adding extra LLSs from the Illustris-based template on top of the existing PRIYA prediction. We avoid negative values, since Ref. [90] only models the difference between the P1D with HCDs and the HCD-free P1D, and does not account for removing absorbers from an existing population.

Importantly, since KODIAQ-SQUAD reaches scales where LLS contamination becomes significant, the pixel-level suppression from unmasked systems, i.e., saturated absorption that washes out small-scale structure, is no longer negligible. While Ref. [90] suggests that the mean effect of HCD contamination can be absorbed into the mean flux calibration (effectively shifting the P1D vertically), we find the amplitude of this shift to be nontrivial. In KODIAQ-SQUAD, it can alter the inferred meanflux-redshift relation. We caution that this implies the inferred τ_0 from the survey may include a LLS component.

4 Results

This section presents the MCMC posterior computed from applying the PRIYA emulator to XQ100 P1D and KODIAQ-SQUAD P1D measurements.We limit the redshift range of XQ100 to $z = 3.4 - 4.2$, and KODIAQ-SQUAD to $z = 2.6 - 4.2$, as discussed in Section 3.5. For reference, we also include the PRIYA baseline MCMC chains from eBOSS DR14 [74], which use $z = 2.6 - 4.6$. All results incorporate data-related nuisance parameters, including contamination from HCDs (DLAs, sub-DLAs, LLSs) via the template model of Ref. [90].

The purpose of our analysis is to examine how the inferred posteriors respond across datasets with differing spectral resolution of Lyman-α forest sightlines. In particular, we are interested in whether high-resolution data, while less statistically powerful, reveals qualitatively different features in the posterior or better sensitivity to thermal parameters.

We begin by presenting the cross-parameter correlations in § 4.1, which provides some physical intuition before jumping into the parameter constraints. Then, in § 4.2, we compare the XQ100 P1D to the eBOSS P1D, focusing on cosmological parameters (A_P, n_P) in § 4.2.1 and astrophysical parameters, including the IGM thermal history, in § 4.2.2. Next, in § 4.3, we analyze the KODIAQ-SQUAD P1D, with the cosmological and astrophysical results detailed in § 4.3.1 and §4.3.2, respectively. We then examine the potential LLS contamination due

---

[6]We sometimes use the shorter names α_{lls}, α_{sub}, α_{sdla}, and α_{ldla}. These notations are interchangeable throughout this work.
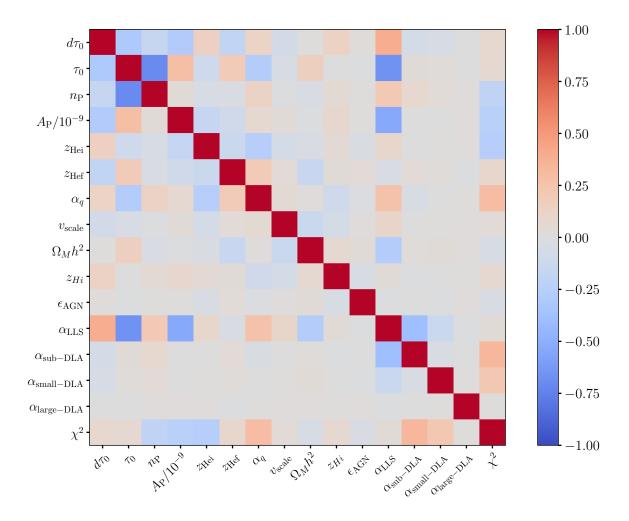
Figure 4. Correlation matrix computed from the MCMC chains for KODIAQ-SQUAD at $z = 2.6 - 4.2$. Positive correlations are shown in red, negative in blue, and zero correlation in white (or gray). We retain the upper triangle for visual clarity, even though the matrix is symmetric. χ^2 refers to the χ^2 values in the MCMC chains, i.e., the log-likelihood of the model prediction given the data.

to KODIAQ-SQUAD's selection bias in § 4.3.3. Finally, we discuss the mean flux model in § 4.4. We do not include maximum a posteriori (MAP) predictions in the main results, but they can be found in Appendix A.

4.1 Parameter Correlation Matrix

In this section, we examine how the cosmological and astrophysical parameters in PRIYA covary at the small scales probed by the P1D from high-resolution measurements, a regime that behaves quite differently from our previous eBOSS DR14 analysis (Ref. [74]). Figure 4 shows the correlation matrix for all parameters varied in the MCMC chains for the KODIAQ-SQUAD P1D ($z = 2.6-4.2$). We use the KODIAQ-SQUAD matrix here, as the redshift range of XQ100 is a subset of KODIAQ-SQUAD, and the correlations are qualitatively similar.

Starting with A_P: there is a strong negative correlation between A_P and a_{LLS}. As shown in Figure 5, the one-parameter variations of A_P (left panel) and a_{LLS} (right panel) produce similar responses in the P1D. The large-scale boost from LLSs comes from the damping
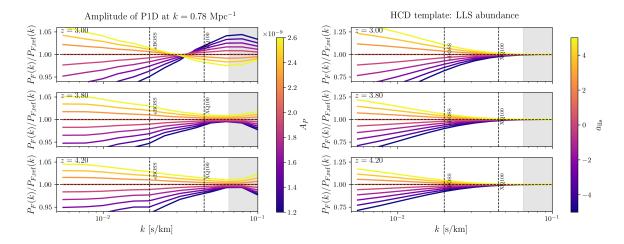
Figure 5. One-parameter variation of the PRIYA emulator, with all other parameters fixed to the eBOSS baseline chains from Ref. [74]. The left panel shows the variation of A_P (the amplitude of the P1D), and the right panel shows the variation of a_{LLS} (the LLS contribution). Note that the impact of LLS variation is significantly larger than that of A_P. We also label the scales relevant for eBOSS and XQ100: $k = 0.02\,\mathrm{s/km}$ and $k = 0.045\,\mathrm{s/km}$, respectively, which correspond to the maximum sensitive k-values for each dataset.

wings of their absorption profiles. By contrast, increasing A_P raises the P1D amplitude at $k < 0.02\,\mathrm{s/km}$, but the effect weakens on smaller scales because of peculiar velocity smoothing, causing a tilt rather than a uniform boost.
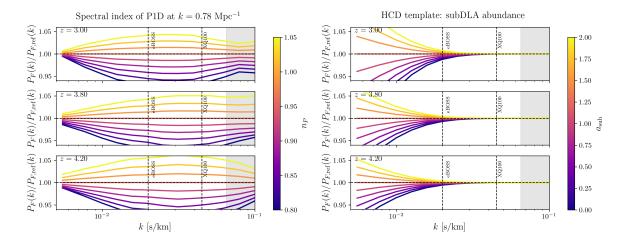
Because the HCD template does not include the saturated absorption suppression expected from LLSs, increasing a_{LLS} must be compensated by lowering τ_0 to keep the small-scale shape of the P1D consistent. This explains the negative correlation between τ_0 and a_{LLS}. Physically, an increase in A_P leads to a larger amount of overdense regions in the IGM, with strong correlated absorption. LLS arise from overdense, biased and partially shielded, regions of the IGM and have minimal damping wings. Their absorption signature thus strongly resembles the correlated overdensities that indicate increased A_P.

In terms of χ^2, interestingly, despite these strong parameter degeneracies, particularly for τ_0, and $d\tau_0$, the LLS amplitude a_{LLS} shows only a weak correlation with χ^2. This suggests that these degeneracies are largely unconstrained by the P1D alone.

Other than A_P, the parameters most decisive for the overall log-evidence are a_{sub}, α_q (along with HeII reionization parameters in general), and n_P. Figure 6 offers a clue as to why. The parameter n_P modifies the P1D slope at $k \sim 0.001\,\mathrm{s/km}$ and gradually bends it into an arc shape over the range $k \sim 0.005 - 0.06\,\mathrm{s/km}$. This is due to peculiar velocity effects smoothing out the slope change at $k > 0.02\,\mathrm{s/km}$. In the eBOSS range, n_P affects the slope; in the overlapping eBOSS-XQ100 range, it mostly changes the normalization; beyond XQ100 range ($k \sim 0.045\,\mathrm{s/km}$), it again alters the slope, but in the opposite direction. This is a fairly unique feature that other parameters cannot easily reproduce, which explains its correlation with χ^2.

The right panel of Figure 6 shows the variation of a_{sub}, the sub-DLA contribution. Unlike LLSs, sub-DLAs primarily affect the largest scales in the P1D relevant to XQ100 and KODIAQ-SQUAD, with a pivot scale around $k \sim 0.02\,\mathrm{s/km}$. This also provides a unique variation that other parameters cannot reproduce.

Figure 7 shows the one-parameter variations of the HeII reionization parameters: α_q,

Figure 6. One-parameter variation of the PRIYA emulator, with all other parameters fixed to the eBOSS baseline chains from Ref. [74]. The left panel shows the variation of n_P (slope of the P1D), and the right panel shows the variation of $a_{\rm sub}$ (sub-DLA contribution).
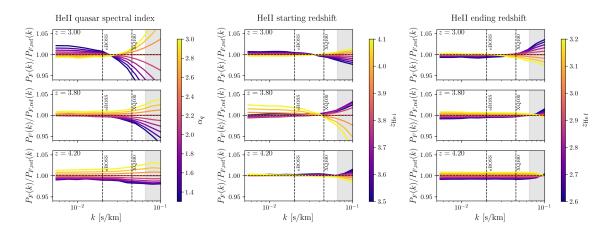

Figure 7. One-parameter variation of the PRIYA emulator, with all other parameters fixed to the eBOSS baseline chains in Ref. [74]. The left panel shows the variation of α_q (HeII reionization heating rate), the middle panel shows the variation of $z_i^{\rm HeII}$ (HeII reionization starting redshift), and the right panel shows the variation of $z_f^{\rm HeII}$ (HeII reionization ending redshift).

$z_i^{\rm HeII}$, and $z_f^{\rm HeII}$. These parameters exhibit very interesting and highly nonlinear behavior in terms of redshift evolution.

The fiducial values are approximately $\alpha_q \sim 1.3$, $z_i^{\rm HeII} \sim 4.12$, and $z_f^{\rm HeII} \sim 2.63$. A value of $\alpha_q \sim 1.3$ corresponds to a fairly large heating rate, which suppresses small-scale neutral hydrogen clustering beyond $k > 0.03\,{\rm s/km}$ after the onset of HeII reionization at $z_i^{\rm HeII}$. This redshift evolution is a unique feature of the HeII reionization parameters, so it is not surprising that they are strongly correlated with χ^2.

4.2   Consistency between XQ100 and eBOSS

We first compare the XQ100 results to the PRIYA baseline analysis of eBOSS DR14 from Ref. [74], which uses the P1D over $z = 2.6 - 4.6$ and incorporates external priors on the IGM thermal history, ultimately derived from KODIAQ DR2 Ref. [36]. Note that the IGM
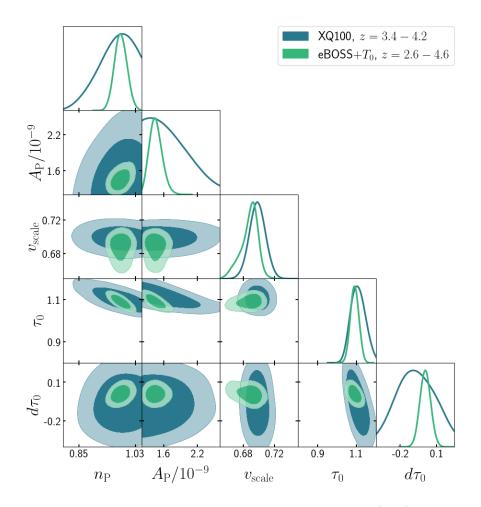
Figure 8. Posterior distributions for n_P, A_P, and v_{scale} from XQ100 (blue) and eBOSS (green) P1Ds. The XQ100 posteriors are broadly consistent with eBOSS, despite broader uncertainties due to smaller sample size and narrower redshift range. v_{scale} (the Hubble parameter) is already priored out to $v_{\mathrm{scale}} \sim \mathcal{N}(\mu = 0.7, \sigma = 0.15)$ in the MCMC chains, as it is not sensitive to the P1D as suggested in Ref. [74].

temperature measurements of Ref. [36] are derived assuming a fixed Planck cosmology. The purpose of this comparison is to determine the thermal history constraints available from the XQ100 P1D alone, without using external datasets or cosmological priors.

4.2.1 Cosmological Parameters: A_P, n_P

Figure 8 shows the marginalized posteriors for three cosmological parameters: the primordial amplitude A_P, the spectral index n_P (pivoted at $k = 0.78 \, \mathrm{Mpc}^{-1}$), and the velocity scale v_{scale}.

The XQ100 posteriors for A_P and n_P remain consistent with eBOSS baseline within 1σ. This is despite a narrower redshift range ($z = 3.4 - 4.2$), higher spectral resolution ($R \sim 5\,000$ versus eBOSS's $R \sim 2\,000$), and much larger statistical uncertainty. Although the constraints are broader due to increased cosmic variance, the n_P posterior mode aligns well with eBOSS, supporting recent results from PRIYA and Lyssa (see Refs. [59, 74]) that the tilt of the linear power spectrum inferred from the Lyman-α forest is consistent with
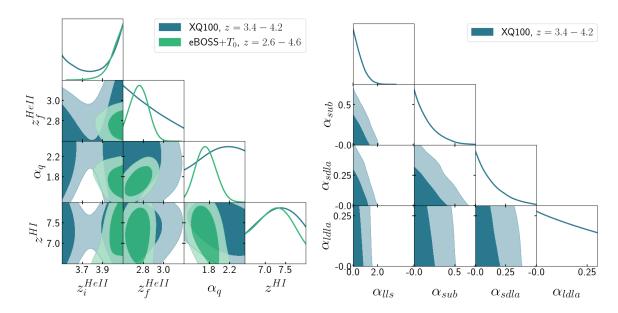
Figure 9. Posterior distributions for thermal history parameters (left) and DLA/LLS contamination parameters (right) for XQ100 (blue) and a combination of eBOSS and IGM mean temperature measurements (eBOSS+T_0, green). Note that eBOSS+T_0 constraints come predominantly from the IGM mean temperature prior (from Ref. [36]; for details see Ref. [74]), while XQ100 constraints come from the XQ100 P1D alone.
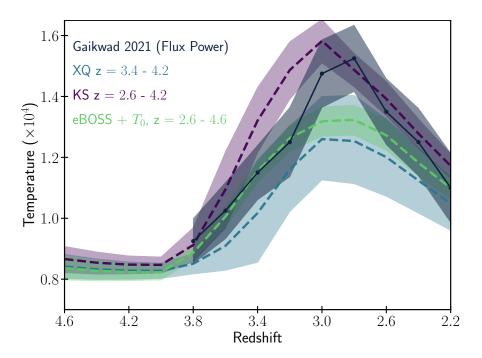
Planck. Interestingly, the A_P posterior from XQ100 also peaks near the PRIYA eBOSS value, albeit with a slight shift toward lower linear power amplitude ($A_P \sim 1.2 \times 10^{-9}$) compared to eBOSS ($A_P \sim 1.46 \times 10^{-9}$).

Note that XQ100 probes somewhat smaller physical scales ($k \sim 3\,h\,\mathrm{Mpc}^{-1}$) than eBOSS ($k \sim 1\,h\,\mathrm{Mpc}^{-1}$). The agreement in A_P across these datasets suggests that the small-scale cosmology from a high-resolution quasar survey like XQ100 and the large-volume eBOSS survey are at least consistent with each other, despite probing different scales. As we will see in § 5.3, the 1σ constraints from XQ100 on the amplitude and slope of the linear theory power spectrum parameters, Δ_L^2 and n_{eff}, are in fact consistent with both eBOSS and Planck, with the posterior mode slightly closer to eBOSS. The mean optical depth parameters τ_0 and $d\tau_0$ remain consistent with eBOSS, as we shall discuss further in §4.4.

4.2.2 Astrophysical Parameters: Thermal History and mean IGM temperature

Figure 9 presents the marginalized posteriors for the astrophysical parameters. The left panel focuses on thermal and ionization history parameters (z_i^{HeII}, z_f^{HeII}, α_q, z^{HI}), while the right panel shows the four nuisance parameters associated with residual DLA and LLS contamination, based on the HCD templates from Ref. [90].

Starting with the thermal parameters, both XQ100 and eBOSS+T_0 prefer an extended He II reionization history, with the starting redshift $z_i^{\mathrm{HeII}} \gtrsim 3.9$ and ending redshift $z_f^{\mathrm{HeII}} \lesssim 3.0$. The XQ100 posterior on z_i^{HeII} is bimodal, with a secondary peak around $z_i^{\mathrm{HeII}} \sim 3.6$, though we will see later in Figure 10 that this bimodality does not significantly affect the predicted IGM mean temperature. The spectral slope α_q, which controls the quasar heating rate, is consistent across datasets within 1σ, with XQ100 favoring slightly weaker heating
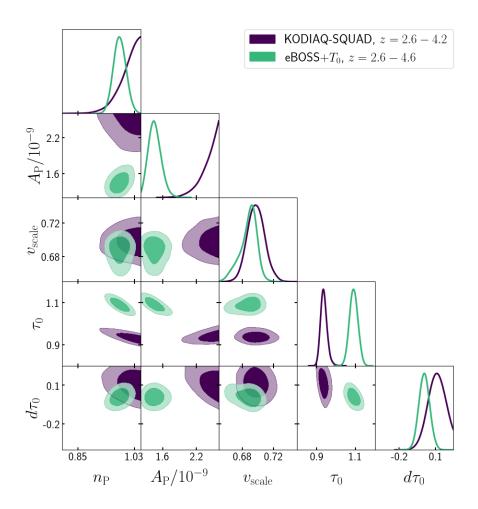
Figure 10. Mean IGM temperature posterior predictive distributions (shaded: 68% confidence interval) using the P1D from XQ100 (blue) and KODIAQ-SQUAD (purple), compared to the eBOSS+T_0 DR14 baseline (green). Note that eBOSS+T_0 includes data from an independent analysis of KODIAQ DR2.

(larger α_q implies less heating). As expected, z_f^{HeII} is only weakly constrained by XQ100 alone, since the ending redshift ($z \sim 2.8$) lies outside the redshift range probed.

Note that in Ref. [74], we showed that the eBOSS P1D alone *cannot* constrain thermal parameters without an external measurement of the IGM mean temperature. Figure 10 compares the predicted IGM mean temperature from the XQ100 MCMC chains to the eBOSS baseline. Despite *not* using any external temperature data, the XQ100 P1D prediction is consistent with the eBOSS baseline (which includes external data) within 1σ, suggesting that the small-scale P1D alone encodes substantial information about the thermal history.

The HI reionization redshift z^{HI} remains poorly constrained from P1D, as anticipated, since its imprint on the P1D is limited to residual memory effects in the IGM's temperature-density relation at high redshift. All chains remain within the prior range $z^{\mathrm{HI}} \sim 6.5 - 8$.

The right panel of Figure 9 shows the four HCD nuisance parameters: α_{LLS}, α_{subDLA}, $\alpha_{\mathrm{small-DLA}}$, and $\alpha_{\mathrm{large-DLA}}$. As described in § 3.6, we marginalize over these to account for residual contamination from unmasked HCDs. All HCD parameters are consistent with zero within their prior bounds, indicating no significant excess contamination. Notably, $\alpha_{\mathrm{large-DLA}}$ is unconstrained, implying the XQ100 P1D is insensitive to large DLAs. $\alpha_{\mathrm{small-DLA}}$ is also near zero, as small DLAs have already been masked in both data and simulation. Recall that $\alpha_{\mathrm{sub-DLA,LLS}} = 0$ implies exact agreement between the PRIYA model and the real Universe observed by XQ100. This suggests that the LLS and sub-DLA populations in PRIYA are broadly consistent with those inferred from XQ100.

Figure 11. Constraints for n_P, A_P, and v_{scale} from KODIAQ-SQUAD (purple) and eBOSS (green) P1Ds.

4.3 KODIAQ-SQUAD: Interpretation from PRIYA

We present the first cosmological interpretation of the KODIAQ-SQUAD P1D using the PRIYA suite. KODIAQ-SQUAD probes smaller scales even than XQ100. These scales are uniquely sensitive to the scale-dependent effects of Lyman Limit Systems, which have not been fully explored in earlier cosmological analyses.

As mentioned in § 3.5, the KODIAQ-SQUAD is biased due to the selection of quasar sightlines that preferentially target known DLAs and O IV absorbers, though Ref. [55] has masked all identified DLAs carefully, sub-DLAs and LLSs remain in the data. In this section, we will assess how this selection bias impacts the inferred cosmology and whether the HCD template model can adequately account for this bias.

As before, we start with the cosmological parameters in § 4.3.1, then move to the astrophysical parameters in § 4.3.2. We further examine the LLS bias and HCD template mismatch in § 4.3.3.

4.3.1 Cosmological parameters: A_P, n_P

Figure 11 shows the marginalized posteriors from KODIAQ-SQUAD for the primordial amplitude A_P, spectral index n_P, and velocity scale v_{scale}. We compare these results to the PRIYA

baseline analysis of eBOSS DR14 from Ref. [74]. Note that both eBOSS and KODIAQ-SQUAD span a similarly wide redshift range and include measurements down to $z = 2.6$.[7]

Somewhat unexpectedly, the KODIAQ-SQUAD posteriors for both A_P and n_P hit the upper boundary of the prior range, in clear tension with the posteriors from both eBOSS and XQ100 (as well as Planck [91]). One might initially expect that the elevated A_P inferred from KODIAQ-SQUAD simply boosts the overall P1D amplitude, as would be the case for the linear matter power spectrum. However, this intuition breaks down at small scales. Figure 5 shows the one-parameter variation of the P1D with respect to A_P and α_{LLS}, with all other parameters fixed to the eBOSS baseline chain values from Ref. [74]. In the left panel, we examine how varying A_P impacts the P1D. Rather than uniformly increasing the P1D, increasing A_P introduces a tilt: it boosts power at large scales ($k \lesssim 0.02\,\mathrm{s/km}$) but suppresses it at smaller scales. This counterintuitive behavior arises because the P1D becomes increasingly dominated by peculiar velocity effects at high-k, causing the response to A_P to change sign (see Appendix D).

Since the KODIAQ-SQUAD P1D has unusually small uncertainty at $k > 0.045\,\mathrm{s/km}$, it is more affected by this scale-dependent response. In other words, the high A_P inferred from KODIAQ-SQUAD is actually compensating for suppressed small-scale power and a steepened P1D slope, rather than signaling a higher amplitude in the flux power spectrum.

The leading suspicion for this behavior is selection effects. The KODIAQ-SQUAD spectra are selected from the literature, with a bias toward DLAs or metal absorbers. Since these are correlated with high density regions, they also select for an over-abundance of LLS, which we will show in § 4.3.3 produces a similar scale-dependent suppression to that seen from varying A_P (see again Figure 5). In addition, the correlation with higher density regions of the IGM may imply a biased region of the cosmological sky and lead to residual systematics that are difficult to fully marginalize out.

4.3.2 Astrophysical Parameters: Thermal History

Figure 12 shows the posteriors for the astrophysical parameters in our emulator, focusing on the thermal and ionization history (z_i^{HeII}, z_f^{HeII}, α_q, z^{HI}; left) and the four nuisance parameters associated with residual DLA and LLS contamination (right), based on the HCD template from Ref. [90], which we will discuss in § 4.3.3.

Starting with the thermal parameters, we find that KODIAQ-SQUAD prefers an extended HeII reionization history, with the start of HeII reionization $z_i^{\mathrm{HeII}} \gtrsim 3.9$. The end of HeII reionization prefers a slightly higher $z_f^{\mathrm{HeII}} \gtrsim 3.0$ value, shifting the best-fit peak temperature to a higher redshift than XQ100 and eBOSS. The posterior on α_q shows a preference for stronger heating rates, though the posterior hits the lower bound of the prior range.

4.3.3 LLS Abundance is not Consistent with HCD Template

The HCD contamination is shown in the right panel of Figure 12. KODIAQ-SQUAD favors a substantially higher abundance of LLSs than the PRIYA simulations, with $\alpha_{\mathrm{LLS}} \sim 2$, at high confidence. A higher α_{LLS} tilts the P1D downward at small scales, in a way that mimics the effect of increasing A_P (see Figure 5). $\alpha_{\mathrm{LLS}} \sim 2$ implies that KODIAQ-SQUAD is detecting triple the number of LLS present in PRIYA, i.e., that $2\times$ the LLSs from the Illustris-based HCD template [90] should be added to the LLS catalogue already in PRIYA.

---

[7] As discussed in § 3.5, we omit the $z = 2.2 - 2.4$ bins due to known inconsistencies across surveys in this range, based on the analyses in Refs. [59, 74]. We defer further investigation of $z = 2.2 - 2.4$ to future work following the release of DESI DR1.
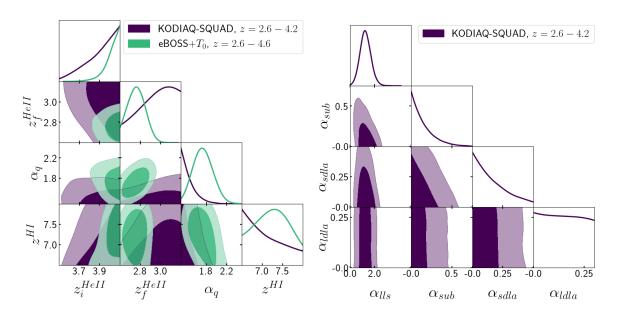
Figure 12. Posterior distributions for thermal history parameters (left) and DLA/LLS contamination parameters (right) for KODIAQ-SQUAD (purple), XQ100 (blue), and eBOSS with the IGM mean temperature prior (eBOSS+T_0, green). eBOSS constraints are tight for thermal history because it uses the IGM mean temperature prior (from Ref. [36]; details see Ref. [74]), while KODIAQ-SQUAD is P1D only.
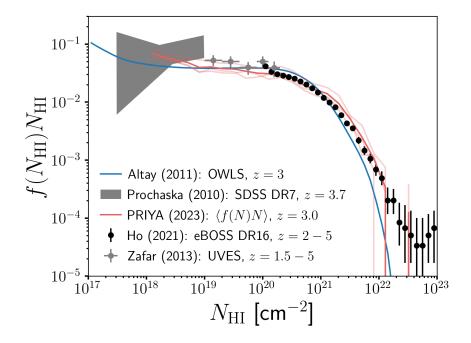
Figure 13. Column density distribution function (CDDF) moments (red solid line), $f(N_{\mathrm{HI}})N_{\mathrm{HI}}$, as a function of neutral hydrogen column density. We show the Ho 2021 measurement from eBOSS DR16 quasars (Ref. [92], $z = 2 - 5$), Altay (2011) simulations using OWLS at $z = 3$ (Ref. [93]), and observational constraints from Zafar 2013 (Ref. [94], UVES) and Prochaska 2010 (Ref. [95], SDSS DR7). The light red lines represent outputs from three independent simulations of the PRIYA suite at $z = 3.0$, with the thick line denoting the mean across realizations. All data are scaled by N_{HI} to highlight the first moment relevant for Ω_{DLA} calculations.

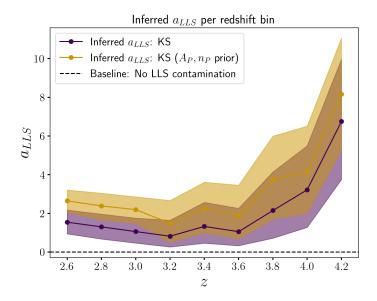There are two possible explanations. First, the PRIYA simulations may not accurately model the LLS population. Second, selection effects or systematics in the KODIAQ-SQUAD data may cause it to have a higher LLS abundance than a randomly selected part of the Universe, potentially with a redshift evolution not encoded in the LLS template we use.

We begin with the first possibility. Ref. [63] notes that the PRIYA includes galaxy physics models and a high enough resolution for a realistic population of DLAs with $N_{\rm HI} > 10^{20}\,{\rm cm}^{-2}$. As the simulation contains a realistic model of the IGM, of DLAs and of galaxies, it seems a priori likely that it would also contain a LLS population in reasonable agreement with observations. However, LLS present unique modeling challenges, as they trace lower-density circumgalactic gas, are physically small [96], and tend to lie further from halos than DLAs, so that accurate modeling requires capturing the CGM-IGM interface. Furthermore, they lie in a transition region between ionized gas in photoionization equilibrium with the UVB, and the self-shielded neutral gas which produces DLAs. Their neutral fraction thus depends sensitively on incident ionizing radiation and is sensitive to radiative transfer effects such as shadowing, as emphasized by previous simulation efforts (e.g., Refs. [97–101]). Although PRIYA implements inhomogeneous reionization and reproduces global quantities like the IGM temperature and DLA statistics well, it thus remains uncertain whether its LLS population is accurate. However, for P1D analyses, what matters most is that the total number density and the clustering of LLSs is statistically correct, as the P1D is averaged over sightlines.

We therefore examine the column density distribution function (CDDF) of DLAs/LLSs in PRIYA, shown in Figure 13. The CDDF is defined as the number of absorbers per unit redshift per unit column density and represents the abundance of DLAs and LLSs. The PRIYA CDDF is actually in good agreement with previous sub-DLA measurements (e.g., Ref. [94]). It also agrees with earlier OWLS simulation (Ref. [93]), and agrees with SDSS DR7 measurements (Ref. [95]) in the LLS regime ($N_{\rm HI} \sim 10^{17.2} - 10^{19}\,{\rm cm}^{-2}$). In addition, PRIYA matches the observed DLA abundance from eBOSS DR16 (CDDF in DLA GP, Ref. [92]). At the very least, it is unlikely that the LLS abundance in PRIYA is low by a factor of ~ 3, as would be required to explain the high confidence observations.

The second possibility is that the KODIAQ-SQUAD selection function is biased toward a higher LLS abundance. It is known [55, 102] that KODIAQ-SQUAD contains multiple spectra deliberately selected to contain large DLAs, and so a correlation between LLS and DLAs would imply that the selection function is also biased towards LLS. To test this, we impose eBOSS priors on A_P and n_P, which helps isolate the A_P constraint from small-scale astrophysical effects. The results, shown as yellow contours in Figure 14, indicate that with the eBOSS priors, the inferred $\alpha_{\rm LLS}$ shifts even higher, reaching ~ 3. This shift effectively compensates for the extra P1D slope tilt that would otherwise be absorbed by A_P. Interestingly, imposing the (A_P, n_P) priors has a negligible impact on the goodness of fit: the reduced χ^2 changes only slightly, from $\chi^2/{\rm dof} = 1.58$ to 1.57. This suggests that the apparent inconsistency of (A_P, n_P) with XQ100/eBOSS is just the MCMC sampler struggling to break the $A_P - \alpha_{\rm LLS}$ degeneracy, and that the KODIAQ-SQUAD P1D is still roughly consistent with the eBOSS A_P.

To further examine the complexity of the redshift-dependent LLS bias, we allow $\alpha_{\rm LLS}$ to vary independently in each redshift bin in the MCMC sampler. Figure 14 shows the inferred values of $\alpha_{\rm LLS}$ for each redshift bin, with and without the eBOSS (A_P, n_P) priors. The redshift evolution turns out to be notably non-trivial: $\alpha_{\rm LLS} \sim 3$ at $z = 2.6$, ~ 2 at $z = 3.2$, and rising to ~ 8 at $z = 4.2$. This redshift trend helps explain why A_P appears biasedly high

Figure 14. The inferred α_{LLS} for each redshift bin, with or without the eBOSS A_P, n_P priors.

in the MCMC chains: KODIAQ-SQUAD's selection function is skewed toward higher LLS abundance, and that bias is both non-trivial and redshift-dependent. When we marginalize over the HCD template (which assumes a different redshift evolution), the mismatch ends up being absorbed by a larger A_P.

This result suggests that in order to extract cosmological information from the KODIAQ-SQUAD P1D, one must either correct for the selection bias or mask LLSs in quasar spectra directly. However, both are challenging. Due to their higher abundance and relatively (compared to DLAs) subtle damping-wing features, there is currently no robust way to identify LLSs, apart from an early-stage random forest-based classifier (Ref. [103]). Even if such a tool is applied, the KODIAQ-SQUAD selection function is still biased toward regions with high neutral hydrogen clustering, since targeting quasars with known DLAs will tend to trace overdense environments near halos.

4.4 Mean Flux and Inferred Optical Depth

Figure 15 shows that the inferred mean flux models are broadly consistent with either Ref. [77] or Ref. [78] However, there is a normalization offset for KODIAQ-SQUAD.

This is not surprising. As discussed in § 3.6, pixel-level contamination from DLAs and LLSs is *not* explicitly modeled in the HCD template. As a result, the inferred τ_0 should be interpreted as an effective quantity that absorbs both the mean flux and residual contamination from unmasked HCDs. In other words, the HCD template accounts only for *damping wings*, not the *saturated absorption troughs* that suppress small-scale clustering in the Lyman-α forest and lower the overall P1D normalization. This suppression is indirectly captured by the τ_0 posterior.

To test the validity of the power-law slope assumption in PRIYA, we also ran separate MCMC chains for XQ100 and KODIAQ-SQUAD, allowing τ_0 to vary independently in each redshift bin. These results are shown in Figure 15 as dots and crosses. The independently inferred τ_0 values are broadly consistent with the global slope model (dashed lines; Equation 2.2), except for a mild discrepancy at $z = 4.2$. We do not assign strong physical meaning
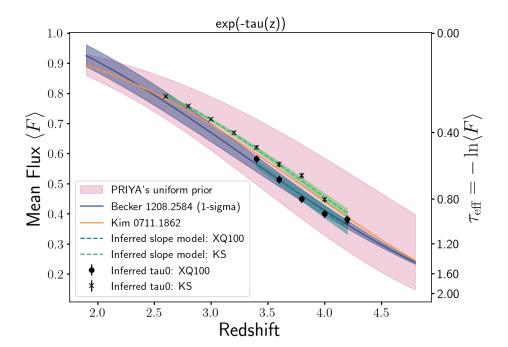
Figure 15. Inferred meanflux relation for XQ100 (blue dashed) and KODIAQ-SQUAD (green dashed). The blue-shaded region shows the measurement from Ref. [78] (Becker13), and the orange from Ref. [77] (Kim07). Dashed lines indicate the inferred power-law model parameters (τ_0, $d\tau_0$), while dots (XQ100) and crosses (KODIAQ-SQUAD) show the results from fits assuming τ_0 is independent for each redshift bin. Even though PRIYA does not impose an informed prior on the meanflux parameters (allowing them to be freely inferred from the data) we still find the inferred meanflux models to be broadly consistent with either Kim07 or Becker13. Any normalization shift in the τ_0 is actually correlated with HCD parameters and should not interpret as physically meaningful. We will discuss this in § 3.6.

to this deviation: τ_0 is a nuisance parameter entangled with both HCD contamination and n_P. At $z > 4$, quasar spectra become increasingly noisy and the sample size drops. As a result, τ_0 normalization shifts in the P1D are more easily absorbed by τ_0 than by parameters like n_P or α_{LLS}, which are more tightly constrained by lower-redshift data. Therefore, when jointly fitting multiple redshift bins via the per-bin τ_0 model, normalization shifts at high redshift are more likely to be absorbed into the per-z τ_0 rather than n_P or α_{LLS} (which are shared across redshift bins), explaining the offset at $z = 4.2$ seen in Figure 15.

5 Discussion

The results presented above illustrate several important points, regarding the small-scale Lyman-alpha forest P1D and its sensitivity to astrophysical/cosmological parameters, particularly the role of HCDs and the thermal history of the IGM through HeII reionization.

1. **XQ100 likelihood agrees with eBOSS DR14:** In Ref. [74], we derived cosmological parameters using an eBOSS DR14 likelihood, using external IGM temperature measurements, T_0, from Ref. [36] (based on KODIAQ), to calibrate the thermal history and constrain our inhomogeneous He II reionization model. However, Ref. [74] was not

able to assess the consistency between the eBOSS and external IGM temperature likelihoods. In this work, we find that the high-resolution P1D from XQ100 alone yields a thermal history consistent with the joint eBOSS DR14 and external IGM temperature ($FPS + T_0$) baseline. The posterior cosmological parameters are also in very good agreement, albeit with larger uncertainty in XQ100. This shows that combining the two datasets is a consistent choice and does not artificially bias the cosmological inference. Looking forward, high-resolution P1D datasets like XQ100 can be jointly analyzed with eBOSS or DESI to self-consistently constrain the thermal history, removing the need for external T_0 measurements altogether.

2. **Amplitude**: The power spectrum amplitude parameter, A_P, tilts the P1D slope rather than uniformly rescaling the amplitude (see Figure 5). It is strongly degenerate with the LLS contamination parameter, a_{LLS} on the scales probed by XQ100.

3. **Slope**: The small-scale P1D ($k > 0.02\,\mathrm{s/km}$) is more sensitive to the slope parameter, n_P, than to the amplitude parameter, A_P, due to n_P modifying the shape of the P1D in a distinct way.

4. **HCD Contamination**: All unmasked HCDs, such as LLSs and sub-DLAs, have a significant impact on the scales ($k = 0.005 - 0.10\,\mathrm{s/km}$) relevant to high-resolution P1D measurements. The LLS contamination in KODIAQ-SQUAD is substantially larger than expected in the HCD template from simulations (Ref. [90]), especially at high redshift ($z = 3.8 - 4.2$). This appears due to a redshift-dependent selection effect, which biases the KODIAQ-SQUAD selection function towards DLAs and thus LLS and overdense regions of gas.

5. **HeII Reionization**: The HeII reionization parameters, i.e., α_q (the heating rate), z_i^{HeII} (the start redshift), and z_f^{HeII} (the end redshift), strongly affect the total log-evidence of the fit. These parameters modify the small-scale ($k > 0.03\,\mathrm{s/km}$) P1D in a highly nonlinear and redshift-dependent way. Thus, any P1D inference at these small scales must account for HeII reionization.

6. **Subtlety of τ_0 with HCD Templates**: The effective mean flux parameter τ_0 is not a direct physical observable but instead absorbs residual contamination from HCD systematics, particularly at high k. Therefore, it should not be interpreted in isolation, nor should it be placed with a Gaussian prior unless co-varied with a_{HCD}.

5.1 Thermal history and cosmology are scale-separate

In § 4.3.3, we showed that the KODIAQ-SQUAD P1D is biased toward denser regions of the IGM, and that the degeneracy between HCD contamination and A_P is difficult to disentangle given KODIAQ-SQUAD's complex selection function. We also found that marginalizing over cosmology using eBOSS priors breaks the A_P-α_{LLS} degeneracy, while leaving the reduced χ^2 nearly unchanged. This suggests that the cosmological prior serves to disentangle that specific degeneracy, without significantly altering the thermal parameter constraints. If so, it implies that the cosmological and thermal information in the P1D are largely scale-separated: thermal history is encoded in the small-scale modes of P1D, while cosmological constraints are driven by large-scale modes.

In Figure 16 (left), we compare the inferred thermal histories with (yellow) and without (purple) the cosmological priors. The thermal posteriors remain nearly unchanged, with
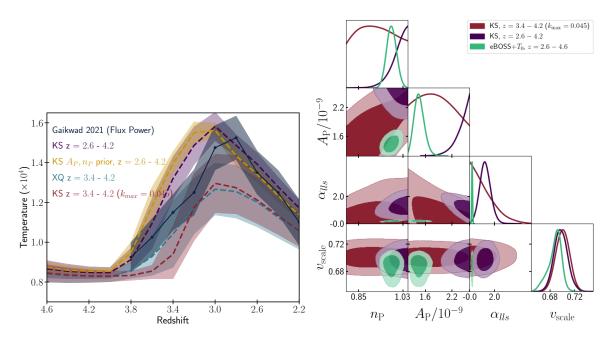
Figure 16. Left: Posterior predictions of the IGM mean temperature, T_0, from the MCMC chains shown on the left. Additional MCMC chains include the XQ100 (blue) and a reduced KODIAQ-SQUAD redshift range $z = 3.4 - 4.2$ (red). Right: Comparison of the KODIAQ-SQUAD posteriors at full redshift ($z = 2.6 - 4.2$, purple) and restricted redshift ($z = 3.4 - 4.2$, red).

both chains yielding consistent $T_0(z)$ evolution. The only noticeable shift occurs in α_{LLS}, which rises to ~ 3 when the cosmology is priored out. This supports the idea that thermal constraints are not significantly affected by cosmological information fixed at $k < 0.02\,\mathrm{s/km}$ (eBOSS).
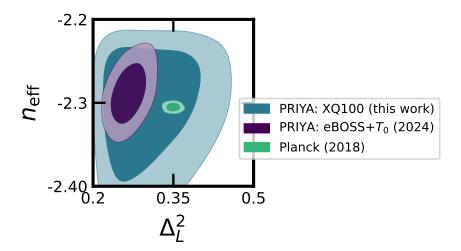
We further test this scale separation by varying the redshift and k range included in the fit. We find that thermal parameters are most sensitive to high-k modes ($k > 0.045\,\mathrm{s/km}$) and to low-redshift data ($z = 2.6 - 3.2$). As shown in Figure 16 (left; red line), using only $z = 3.4 - 4.2$ and $k < 0.045\,\mathrm{s/km}$ lowers the inferred T_0 by nearly 2σ, bringing it into agreement with the XQ100 thermal history. However, either including high-k data at high redshift, or including low-z bins with $k < 0.045\,\mathrm{s/km}$, is sufficient to recover the peak temperature of $T_0 \sim 1.5 \times 10^4$ K.

We conclude that thermal information in the P1D is largely separable in scale and redshift from cosmological information. In particular, most of the thermal sensitivity lies in $k \gtrsim 0.045\,\mathrm{s/km}$ and $z \lesssim 3.2$, while cosmological parameters are best constrained at lower k.

5.2 KS consistent with XQ100 when $z = 3.4 - 4.2$ and $k < 0.045\,\mathrm{s/km}$

We now examine the A_P degeneracy in this restricted chain, shown in the right panel of Figure 16. Interestingly, under this redshift and k cut, the KODIAQ-SQUAD posterior (yellow) becomes consistent with the eBOSS baseline cosmology (green), yielding $A_P \sim 1.85^{+0.24}_{-0.63} \times 10^{-9}$, though with large uncertainty. This agreement arises because the $k < 0.045\,\mathrm{s/km}$ cut removes part of the covariance between A_P and α_{LLS}.

As a result, the LLS contamination in these reduced chains is only weakly constrained, with an upper bound of $\alpha_{\mathrm{LLS}} < 1.49$, and no strong posterior mode near $\alpha_{\mathrm{LLS}} \sim 2$ as seen in the full redshift and scale range.

Figure 17. Reduced likelihood of the eBOSS P1D (purple) and XQ100 P1D (blue), compared to Planck (yellow).

Overall, these tests show that by restricting to high redshift and removing some small-scale modes, we can mitigate the some LLS contribution to the P1D and obtain cosmological constraints that are less sensitive to the biased selection function of KODIAQ-SQUAD. However, residual contamination likely persists, and it remains difficult to claim that the A_P measurement is fully unbiased, as the quasar sightlines still preferentially probe denser regions of the IGM. There are two connected results: first, because KODIAQ-SQUAD probes a biased sky, this specific measurement of the flux power spectrum on small scales is unreliable for estimating the LLS abundance and cosmological gas clustering that drives the A_P constraints. Second, any future analyses using quasars with an HI-blind selection function to obtain constraints on the small-scale power spectrum should include models for LLS contamination when probing scales $k > 0.045\,\mathrm{s/km}$.

5.3 Reduced likelihood: XQ100

As proposed in Ref. [104], we can assess the relative constraining power of different surveys by examining their reduced likelihood: the amplitude and slope of the linear theory power spectrum at $k = 1\,\mathrm{Mpc}^{-1}$ and $z = 3$. This choice of pivot scale and redshift is motivated by the fact that the Lyman-α forest P1D primarily probes Mpc-scale modes around this redshift. Note, however, that this $z = 3$ reference is slightly lower than the redshifts of XQ100, which has a midpoint near $z = 3.8$.

Figure 17 shows the reduced likelihood contour for XQ100 result of this work. As expected, the XQ100 posterior is broadly consistent with the eBOSS DR14 (FPS + T_0) baseline, and the uncertainties are wide enough to overlap with both the Planck baseline. This is consistent with the discussion before: while XQ100 probes smaller scales that are more sensitive to the IGM thermal history, its cosmological constraining power is limited by degeneracies with HCD contamination.

5.4 IGM temperature measurement from Gaikwad 2021

Figure 16 shows four measurements of the thermal history of the IGM from three distinct flux power spectra. We have argued that the KODIAQ-SQUAD dataset is biased. The high

peak temperature preferred by that measurement is consistent with a bias towards lower small-scale power, as high IGM temperatures erase small scale power (see Figure 7). Of our measurements, we thus prefer the inferred XQ100 thermal history. However, because the minimum redshift is $z = 3.4$, it cannot robustly infer the peak IGM temperature at $z \sim 3$. The thermal history measurements from Ref. [36] uses a similar KODIAQ dataset that we have argued is biased towards DLAs. However, they remove all spectra containing a DLA from the sample, rather than masking only the regions directly affected by damping wings as is done by Ref. [55]. This will substantially reduce the effect of the selection bias, although LLS are likely still over-represented as the sample will include spectra which attempted to select for DLAs but found only weaker absorbers. Figure 10 shows that in practice the thermal history measurement of Ref. [36] is consistent with the posterior from XQ100 at the 68% confidence level for $z > 3.0$. This suggests modest effects on the thermal history in the high redshift range most affected by over-sampling LLS.

6   Conclusions

In this work, we use the PRIYA simulation suite with a new emulator to analyze the small-scale Lyman-α forest flux power spectrum (P1D) from high-resolution quasar datasets, specifically XQ100 (Ref. [58]) and KODIAQ-SQUAD (Ref. [55]). The emulator enables simulation-based predictions across a broad parameter space, including cosmology and thermal history, while explicitly modeling the impact of HeII reionization on the P1D.

We find that the XQ100 P1D yields (A_P, n_P) constraints consistent with previous PRIYA results from eBOSS DR14 and Planck, albeit with broader uncertainties. Notably, this is achieved without using external IGM temperature data, demonstrating that the XQ100 P1D alone carries sufficient thermal information. In contrast, the KODIAQ-SQUAD P1D exhibits a biasedly high α_{LLS} that strongly degenerates with A_P. However, restricting the analysis to $k < 0.045\,\mathrm{s/km}$ and $z = 3.4 - 4.2$ significantly reduces sensitivity to LLSs, and A_P becomes consistent with XQ100 and eBOSS.

Two key features of the small-scale P1D are: first, it is highly sensitive to thermal history, particularly the He II heating rate α_q and its timing. Another is the contamination from LLSs and sub-DLAs, which alter the P1D slope at intermediate and small scales. We find strong degeneracies among LLSs, mean flux, and A_P. The LLS abundance inferred from KODIAQ-SQUAD appears biased: the LLS abundance increases with redshift and deviates from the HCD template. At $z = 3.8 - 4.2$, KODIAQ-SQUAD prefers $\sim 6\times$ more LLSs than PRIYA. Given that PRIYA's LLSs match empirical data at $z = 3$, this suggests the excess likely arises from KODIAQ-SQUAD's selection function, which uses quasars with known DLAs and metal absorbers, both strongly correlated with LLSs.

These findings show the importance of isolating LLS contamination when modeling the P1D at $k > 0.045\,\mathrm{s/km}$. However, masking LLSs on an absorber-by-absorber basis remains infeasible. Even for DLAs, robust masking is difficult at low S/N: the recent DESI DR2 DLA catalog [105] achieves 85% completeness and 80% purity, but only for $N_{\mathrm{HI}} > 10^{20.3}\,\mathrm{cm}^{-2}$, which are much easier to find than LLSs. Thus, the reliability of cosmological inference from high-resolution P1D fully depends on how accurately one can forward-model LLSs in simulations and build a robust HCD template.

Previous work on high-resolution P1D has largely focused on setting bounds on alternative dark matter models (see Refs. [43, 44, 48, 106]). However, these analyses are typically not discussed in the context of HCD contamination (except Ref. [56], in the context of comparing

with cluster cosmology). Ref. [104] also emphasized that the scale-dependent astrophysical uncertainties from HCDs might play a critical role in future Lyman-α DM analysis. In this work, we confirm that LLSs correlate strongly with the cosmological signal of the primordial amplitude in high-resolution P1D. This suggests that, since there is no practical way to mask LLSs, future Lyman-α DM analyses will need to be HI-blind and carefully marginalize over LLSs, especially when the DM signal is expected as a suppression of the linear power amplitude. Any residual LLS contamination in the data would show up as an extra suppression of high-k modes in the P1D.

An alternative path forward might be to measure the power spectrum of HCDs directly. For example, using Voigt profile decomposition rather than sinusoids in Fourier space, or employing wavelet scattering transforms (WST; see [107–110]) to isolate LLS abundance from the forest. Our work also motivates gathering samples of high resolution quasar spectra with an HI-blind selection function that extend to lower redshifts than XQ100.
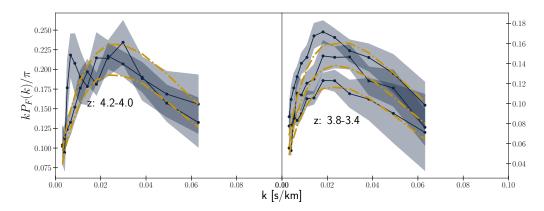
Taken together, this work highlights the value of small-scale Lyman-α forest P1D from high-resolution quasar spectra. XQ100 offers reliable thermal constraints while preserving cosmological consistency with eBOSS DR14. With upcoming DESI DR1 P1D measurements [18, 19] covering P1D until $k = 0.035\,\mathrm{s/km}$, the XQ100 P1D can serve as powerful constraints on thermal nuisance parameters in a joint cosmological fit.
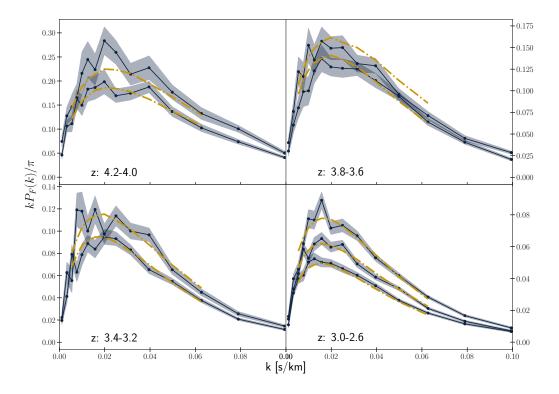
Acknowledgments

A   Maximum a Posteriori (MAP) Predictions

To visually assess how well the inferred posteriors explain the observed data, we compare the MAP predictions of the flux power spectrum with the actual measurements. Figures 18 and 19 show the MAP posterior predictions for XQ100 and KODIAQ-SQUAD, respectively. Notably, even though the raw KODIAQ-SQUAD P1D exhibits stronger small-scale suppression than XQ100, the MAP model reproduces this behavior via a combination of elevated A_P and a higher α_{LLS}. It remains a good fit up to the highest $k = 0.065\,\mathrm{s/km}$. As seen from the diagonal covariance, the KODIAQ-SQUAD posterior is primarily influenced by the lower

Figure 18. Posterior prediction of the flux power spectrum for XQ100 ($z = 3.4$–4.2). Shaded area represents the 1-σ confidence interval of the measurement.

Figure 19. Posterior prediction of the flux power spectrum for KODIAQ-SQUAD ($z = 2.6 - 4.2$). Dot-dashed lines denote MAP predictions for each redshift bin.

redshift bins ($z = 2.6 - 3.2$) rather than the higher ones ($z = 3.4 - 4.2$). We have seen that the separate redshift bins yield quite different posteriors in §5.2.

For XQ100, the diagonal covariance is more forgiving, so the MAP predictions comfortably sit within the 1σ confidence region of the observed P1D in all redshift bins.

KODIAQ-SQUAD, on the other hand, has much smaller covariance at high k, allowing the MAP prediction to be more sensitive to small-scale suppression. However, this is highly degenerate with various other parameters that alter the P1D slope, such as $A_P, \alpha_{\mathrm{LLS}}, \alpha_q$, which we have discussed in §5.1, §5.2 and §4.1.

Note that both Figures 18 and 19 only show the diagonal of the covariance matrix. Therefore, even if the MAP predictions do not lie within the 1σ confidence region, this does not imply inconsistency with the data because the off-diagonal terms of the covariance matrix within redshift bins are not being shown.

B   Table of Cosmological and Astronomical Parameter Constraints

Table 3 presents the posterior constraints on the model parameters from KODIAQ-SQUAD, XQ100, and eBOSS (FPS + T_0). Table 4 shows the posterior constraints on the LLS parameters from KODIAQ-SQUAD in different redshift bins, with (A_P, n_P) fixed to eBOSS prior.

Many parameters are constrained only with upper or lower limits, which is not surprising given the high-dimensional parameter space and the limited sensitivity of the P1D to some of them. As seen in Figure 4, χ^2 is not very sensitive to many parameters, and several are degenerate with each other. Nonetheless, this work represents an important step forward in the field of Lyman-α P1D inference: a self-consistent HeII reionization modeling across a wide range of cosmic evolution ($z = 2.6 - 4.2$), recognizing its importance for small-scale P1D modeling.

C   Comparison of MCMC Chains with Mean Flux Model in Independent Redshift Bins

To examine whether parameter degeneracies or constraints vary with redshift, we perform an alternative fit using a model in which the mean flux is allowed to vary independently in each redshift bin. Figures 20 and 21 show the resulting MCMC corner plots for the KODIAQ-SQUAD and XQ-100 datasets, respectively. This approach serves as a useful cross-check for our fiducial global mean flux model.

Allowing $\bar{F}(z)$ to vary independently can help isolate whether tension in certain parameter directions (e.g., between small-scale amplitude A_P and LLS contributions) is being absorbed by a global $\bar{F}$ assumption. In both datasets, the resulting posteriors are generally consistent with the global model.

D   Effects of Peculiar Velocities on the Flux Power Spectrum

Figure 22 shows the impact of turning off peculiar velocities in simulating Lyman-α sightlines. As expected, peculiar velocities suppress small-scale P1D by smoothing structure along the line of sight. While not directly varied in our inference, this figure illustrates a well-understood physical effect and helps explain why A_P changes the slope at $k > 0.02\,\mathrm{s/km}$ in the high-k P1D instead of simply boosting the overall amplitude.

Parameter	KS	KS (reduced)	XQ	eBOSS (FPS + T_0)
	$z = 2.6$–4.2	$z = 3.4$–4.2	$z = 3.4$–4.2	$z = 2.2$–4.6
	$k = 0.0055 - 0.065\,\mathrm{s/km}$	$k = 0.0055 - 0.045\,\mathrm{s/km}$	$k = 0.003 - 0.065\,\mathrm{s/km}$	$k = 0.001 - 0.02\,\mathrm{s/km}$
$d\tau_0$	$0.144^{+0.086}_{-0.044}$	> 0.0916	-0.07 ± 0.15	$0.009 \pm 0.045 \left(^{+0.089}_{-0.087}\right)$
τ_0	$0.944^{+0.014}_{-0.017}$	$1.032^{+0.034}_{-0.045}$	$1.108^{+0.035}_{-0.042}$	$1.090 \pm 0.022 \left(^{+0.043}_{-0.042}\right)$
n_P	> 1.01	> 0.906	$0.965^{+0.077}_{-0.030}$	$0.983 \pm 0.020 \left(^{+0.040}_{-0.039}\right)$
$A_\mathrm{P}/10^{-9}$	$> 2.26 \cdot 10^{-9}$	$\left(1.85^{+0.24}_{-0.63}\right) \cdot 10^{-9}$	$< 1.85 \cdot 10^{-9}$	$1.46^{+0.099}_{-0.13} \left(^{+0.22}_{-0.22}\right)$
z_i^{HeII}	> 3.92	< 3.85	—	$> 4.00 \, (> 3.87)$
z_f^{HeII}	> 3.00	—	< 2.91	$2.765^{+0.080}_{-0.093} \left(^{+0.14}_{-0.16}\right)$
α_q	< 1.48	> 1.99	2.16 ± 0.46	$1.74^{+0.18}_{-0.21} \left(^{+0.37}_{-0.38}\right)$
v_{scale}	0.695 ± 0.011	0.701 ± 0.011	0.698 ± 0.010	$0.688^{+0.013}_{-0.0074} \left(^{+0.018}_{-0.025}\right)$
$\Omega_M h^2$	> 0.143	—	< 0.143	$< 0.143 \, (--)$
z^{HI}	< 7.09	—	$7.28^{+0.55}_{-0.36}$	$7.24 \pm 0.38 \, (--)$
ϵ_{AGN}	0.0483 ± 0.0035	0.0501 ± 0.0035	0.0495 ± 0.0035	—
α_{lls}	$1.60^{+0.41}_{-0.49}$	< 1.49	< 0.659	—
α_{sub}	< 0.171	< 0.712	< 0.200	—
α_{sdla}	< 0.197	< 0.313	< 0.161	—
α_{ldla}	—	—	—	—

Table 3. Posterior constraints on model parameters from KODIAQ-SQUAD (full and restricted redshift/k ranges), XQ100, and eBOSS (FPS + T_0 from Ref. [74]). We report 68% confidence intervals or 95% limits where applicable.
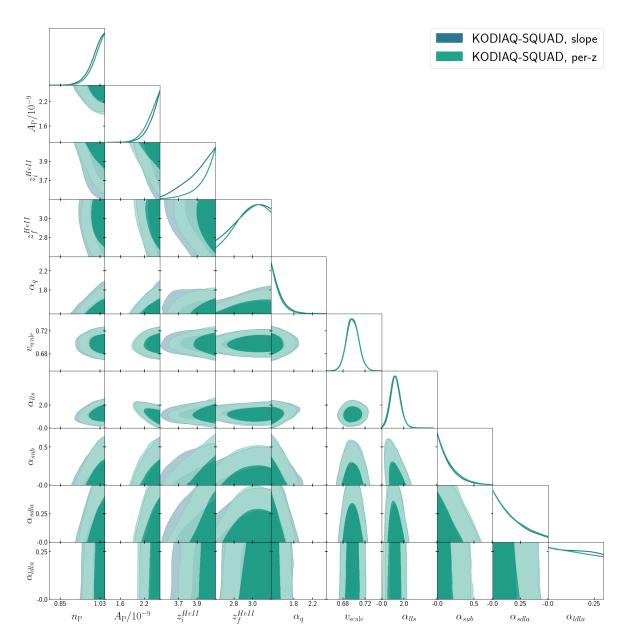
References

[1] J.E. Gunn and B.A. Peterson, *On the Density of Neutral Hydrogen in Intergalactic Space.*, *ApJ* **142** (1965) 1633.

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

[3] T. Theuns, A. Leonard, G. Efstathiou, F.R. Pearce and P.A. Thomas, *P$\hat{3}$M-SPH simulations of the Lyα forest*, *MNRAS* **301** (1998) 478 [astro-ph/9805119].

[4] P. McDonald, J. Miralda-Escudé, M. Rauch, W.L.W. Sargent, T.A. Barlow, R. Cen et al., *The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Lyα Forest*, *ApJ* **543** (2000) 1 [astro-ph/9911196].

[5] L. Hui, S. Burles, U. Seljak, R.E. Rutledge, E. Magnier and D. Tytler, *On Estimating the QSO Transmission Power Spectrum*, *ApJ* **552** (2001) 15 [astro-ph/0005049].

[6] M. Viel, S. Matarrese, H.J. Mo, M.G. Haehnelt and T. Theuns, *Probing the intergalactic medium with the Lyα forest along multiple lines of sight to distant QSOs*, *MNRAS* **329** (2002) 848 [astro-ph/0105233].

[7] X. Fan, M.A. Strauss, R.H. Becker, R.L. White, J.E. Gunn, G.R. Knapp et al., *Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z~6 Quasars. II. A Sample of 19 Quasars*, *AJ* **132** (2006) 117 [astro-ph/0512082].

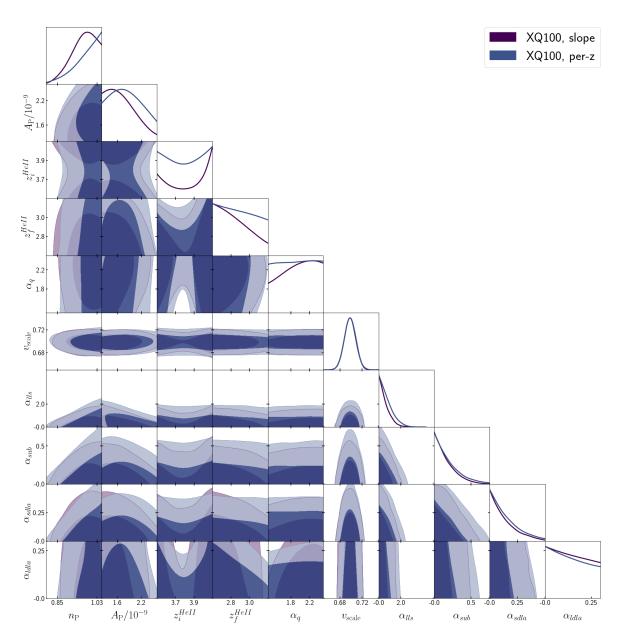| Parameter	KODIAQ-SQUAD $z = 2.6$–4.2 (eBOSS (A_p, n_P) prior)	KODIAQ-SQUAD $z = 3.8$–4.2 (eBOSS (A_p, n_P) prior)
$d\tau_0$	> 0.197	> 0.0840
τ_0	0.945 ± 0.013	0.971 ± 0.022
n_P	0.992 ± 0.014	0.985 ± 0.014
$A_P/10^{-9}$	$(1.532 \pm 0.078) \cdot 10^{-9}$	$(1.476 \pm 0.079) \cdot 10^{-9}$
z_i^{HeII}	> 4.24	> 4.14
z_f^{HeII}	> 3.10	> 2.62
α_q	< 1.37	< 2.00
v_{scale}	$0.691^{+0.011}_{-0.010}$	$0.697^{+0.011}_{-0.0098}$
$\Omega_M h^2$	> 0.145	> 0.143
z^{HI}	< 7.26	—
ϵ_{AGN}	0.0477 ± 0.0035	0.0489 ± 0.0035
α_{lls}	2.79 ± 0.41	$5.0^{+1.9}_{-1.4}$
α_{sub}	< 0.149	< 1.07
α_{sdla}	< 0.187	—
α_{ldla}	—	—

Table 4. Posterior constraints for KODIAQ-SQUAD in different redshift bins, with A_P and n_P priors applied.

[8] M. Viel and M.G. Haehnelt, *Cosmological and astrophysical parameters from the Sloan Digital Sky Survey flux power spectrum and hydrodynamical simulations of the Lyman α forest*, *MNRAS* **365** (2006) 231 [astro-ph/0508177].

[9] 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** (2006) 80 [astro-ph/0405013].

[10] 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*, *ApJ Supplement* **163** (2006) 80 [astro-ph/0405013].

[11] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. Le Goff, G. Rossi, M. Viel et al., *The one-dimensional Lyα forest power spectrum from BOSS*, *AAP* **559** (2013) A85 [1306.5896].

[12] A. Slosar, V. Iršič, D. Kirkby, S. Bailey, N.G. Busca, T. Delubac et al., *Measurement of baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9*, *JCAP* **2013** (2013) 026 [1301.3459].

[13] T. Delubac, J.E. Bautista, N.G. Busca, J. Rich, D. Kirkby, S. Bailey et al., *Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars*, *AAP* **574** (2015) A59 [1404.1801].

Figure 20. MCMC corner plot for the KODIAQ-SQUAD dataset using a mean flux model with independent redshift bins. This allows for redshift-dependent shifts in the amplitude and shape of the P1D, revealing any hidden degeneracies.
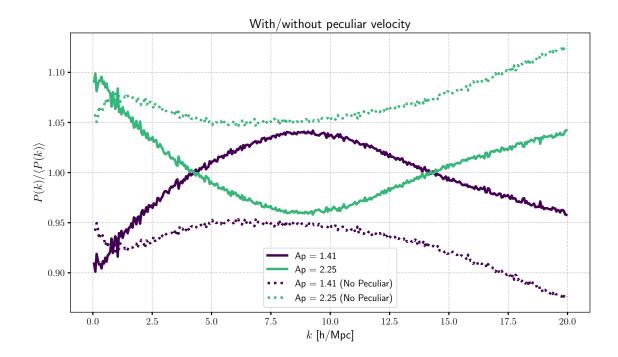
[14] S. Chabanier, N. Palanque-Delabrouille, C. Yèche, J.-M. Le Goff, E. Armengaud, J. Bautista et al., *The one-dimensional power spectrum from the SDSS DR14 Lyα forests*, *JCAP* **2019** (2019) 017 [1812.03554].

[15] B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander, R. Alfarsy et al., *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, *AJ* **164** (2022) 207 [2205.10939].

[16] C. Ravoux, M.L.A. Karim, E. Armengaud, M. Walther, N. Göksel Karaçaylı, P. Martini et al., *The Dark Energy Spectroscopic Instrument: One-dimensional power spectrum from first Lyman-α forest samples with Fast Fourier Transform*, *arXiv e-prints* (2023) arXiv:2306.06311

Figure 21. MCMC corner plot for the XQ100 dataset with redshift-independent $\bar{F}(z)$ assumptions relaxed. The results are broadly consistent with the fiducial global model.

[2306.06311].

[17] N. Göksel Karaçaylı, P. Martini, J. Guy, C. Ravoux, M.L.A. Karim, E. Armengaud et al., *Optimal 1D Lyα Forest Power Spectrum Estimation – III. DESI early data*, arXiv e-prints (2023) arXiv:2306.06316 [2306.06316].

[18] N.G. Karaçaylı, P. Martini, J. Aguilar, S. Ahlen, E. Armengaud, S. Bailey et al., *DESI DR1 Lyα 1D power spectrum: The optimal estimator measurement*, *arXiv e-prints* (2025) arXiv:2505.07974 [2505.07974].

[19] C. Ravoux, M.-L. Abdul-Karim, J.-M. Le Goff, E. Armengaud, J.N. Aguilar, S. Ahlen et al., *DESI DR1 Lyα 1D power spectrum: The Fast Fourier Transform estimator measurement*, *arXiv e-prints* (2025) arXiv:2505.09493 [2505.09493].

Figure 22. Effect of peculiar velocities on the flux power spectrum. Removing peculiar velocities leads to enhanced small-scale power by eliminating redshift-space broadening.

[20] A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander, M. Alvarez et al., *DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman alpha forest*, *JCAP* **2025** (2025) 124 [2404.03001].

[21] DESI Collaboration, M. Abdul-Karim, J. Aguilar, S. Ahlen, C. Allende Prieto, O. Alves et al., *DESI DR2 Results I: Baryon Acoustic Oscillations from the Lyman Alpha Forest*, *arXiv e-prints* (2025) arXiv:2503.14739 [2503.14739].

[22] M. Viel, M.G. Haehnelt and V. Springel, *Inferring the dark matter power spectrum from the Lyman α forest in high-resolution QSO absorption spectra*, *MNRAS* **354** (2004) 684 [astro-ph/0404600].

[23] P. McDonald, U. Seljak, R. Cen, D. Shih, D.H. Weinberg, S. Burles et al., *The Linear Theory Power Spectrum from the Lyα Forest in the Sloan Digital Sky Survey*, *ApJ* **635** (2005) 761 [astro-ph/0407377].

[24] M. Viel, M.G. Haehnelt and A. Lewis, *The Lyman α forest and WMAP year three*, *MNRAS* **370** (2006) L51 [astro-ph/0604310].

[25] U. Seljak, A. Makarov, P. McDonald, S.F. Anderson, N.A. Bahcall, J. Brinkmann et al., *Cosmological parameter analysis including SDSS Lyα forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy*, *Phys. Rev. D* **71** (2005) 103515 [astro-ph/0407372].

[26] U. Seljak, A. Slosar and P. McDonald, *Cosmological parameters from combining the Lyman-α forest with CMB, galaxy clustering and SN constraints*, *JCAP* **2006** (2006) 014 [astro-ph/0604335].

[27] C. Yèche, N. Palanque-Delabrouille, J. Baur and H. du Mas des Bourboux, *Constraints on neutrino masses from Lyman-alpha forest power spectrum with BOSS and XQ-100*, *JCAP* **2017** (2017) 047 [1702.03314].

[28] N. Palanque-Delabrouille, C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, S. Chabanier et al., *Hints, neutrino bounds, and WDM constraints from SDSS DR14 Lyman-α and Planck full-survey data*, *JCAP* **2020** (2020) 038 [`1911.09073`].

[29] M. Garny, T. Konstandin, L. Sagunski and M. Viel, *Neutrino mass bounds from confronting an effective model with BOSS Lyman-α data*, *JCAP* **2021** (2021) 049 [`2011.03050`].

[30] J.S. Bolton, M. Viel, T.S. Kim, M.G. Haehnelt and R.F. Carswell, *Possible evidence for an inverted temperature-density relation in the intergalactic medium from the flux distribution of the Lyα forest*, *MNRAS* **386** (2008) 1131 [`0711.2064`].

[31] J.S. Bolton, G.D. Becker, M.G. Haehnelt and M. Viel, *A consistent determination of the temperature of the intergalactic medium at redshift z = 2.4*, *MNRAS* **438** (2014) 2499 [`1308.4411`].

[32] F. Nasir, J.S. Bolton and G.D. Becker, *Inferring the IGM thermal history during reionization with the Lyman α forest power spectrum at redshift z ≃ 5*, *MNRAS* **463** (2016) 2335 [`1605.04155`].

[33] M. Walther, J. Oñorbe, J.F. Hennawi and Z. Lukić, *New Constraints on IGM Thermal Evolution from the Lyα Forest Power Spectrum*, *ApJ* **872** (2019) 13 [`1808.04367`].

[34] E. Boera, G.D. Becker, J.S. Bolton and F. Nasir, *Revealing Reionization with the Thermal History of the Intergalactic Medium: New Constraints from the Lyα Flux Power Spectrum*, *ApJ* **872** (2019) 101 [`1809.06980`].

[35] X. Wu, M. McQuinn, R. Kannan, A. D'Aloisio, S. Bird, F. Marinacci et al., *Imprints of temperature fluctuations on the z ∼ 5 Lyman-α forest: a view from radiation-hydrodynamic simulations of reionization*, *MNRAS* **490** (2019) 3177 [`1907.04860`].

[36] P. Gaikwad, R. Srianand, M.G. Haehnelt and T.R. Choudhury, *A consistent and robust measurement of the thermal state of the IGM at 2 ≤ z ≤ 4 from a large sample of Ly α forest spectra: evidence for late and rapid He II reionization*, *MNRAS* **506** (2021) 4389 [`2009.00016`].

[37] B. Villasenor, B. Robertson, P. Madau and E. Schneider, *Inferring the Thermal History of the Intergalactic Medium from the Properties of the Hydrogen and Helium Lyα Forest*, *ApJ* **933** (2022) 59 [`2111.00019`].

[38] L. Hui and N.Y. Gnedin, *Equation of state of the photoionized intergalactic medium*, *MNRAS* **292** (1997) 27 [`astro-ph/9612232`].

[39] M. McQuinn, A. Lidz, M. Zaldarriaga, L. Hernquist, P.F. Hopkins, S. Dutta et al., *He II Reionization and its Effect on the Intergalactic Medium*, *ApJ* **694** (2009) 842 [`0807.2799`].

[40] P.R. Upton Sanderbeck, A. D'Aloisio and M.J. McQuinn, *Models of the thermal evolution of the intergalactic medium after reionization*, *MNRAS* **460** (2016) 1885 [`1511.05992`].

[41] P. Upton Sanderbeck and S. Bird, *Inhomogeneous He II reionization in hydrodynamic simulations*, *MNRAS* **496** (2020) 4372 [`2002.05733`].

[42] M. Viel, J. Lesgourgues, M.G. Haehnelt, S. Matarrese and A. Riotto, *Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-α forest*, *Phys. Rev. D* **71** (2005) 063534 [`astro-ph/0501562`].

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

[44] V. Iršič, M. Viel, M.G. Haehnelt, J.S. Bolton, S. Cristiani, G.D. Becker et al., *New constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman-α forest data*, *Phys. Rev. D* **96** (2017) 023522 [`1702.01764`].

[45] A. Garzilli, A. Magalich, O. Ruchayskiy and A. Boyarsky, *How to constrain warm dark matter with the Lyman-α forest*, *MNRAS* **502** (2021) 2356 [1912.09397].

[46] B. Villasenor, B. Robertson, P. Madau and E. Schneider, *New constraints on warm dark matter from the Lyman-α forest power spectrum*, *PRD* **108** (2023) 023502 [2209.14220].

[47] K.K. Rogers and H.V. Peiris, *General framework for cosmological dark matter bounds using N-body simulations*, *Phys. Rev. D* **103** (2021) 043526 [2007.13751].

[48] K.K. Rogers and H.V. Peiris, *Strong Bound on Canonical Ultralight Axion Dark Matter from the Lyman-Alpha Forest*, *Phys. Rev. L* **126** (2021) 071302 [2007.12705].

[49] A. Font-Ribera, P. McDonald and A. Slosar, *How to estimate the 3D power spectrum of the Lyman-α forest*, *JCAP* **2018** (2018) 003 [1710.11036].

[50] R. de Belsunce, O.H.E. Philcox, V. Iršič, P. McDonald, J. Guy and N. Palanque-Delabrouille, *The 3D Lyman-α forest power spectrum from eBOSS DR16*, *MNRAS* **533** (2024) 3756 [2403.08241].

[51] N.G. Karaçaylı and C.M. Hirata, *Light in the dark forest – I. An efficient optimal estimator for 3D Lyman-alpha forest power spectrum*, *arXiv e-prints* (2025) arXiv:2503.15619 [2503.15619].

[52] N.G. Karaçaylı, P. Martini, J. Guy, C. Ravoux, M.L. Abdul Karim, E. Armengaud et al., *Optimal 1D Ly α forest power spectrum estimation - III. DESI early data*, *MNRAS* **528** (2024) 3941 [2306.06316].

[53] V. Iršič, M. Viel, T.A.M. Berg, V. D'Odorico, M.G. Haehnelt, S. Cristiani et al., *The Lyman α forest power spectrum from the XQ-100 Legacy Survey*, *MNRAS* **466** (2017) 4332 [1702.01761].

[54] A. Day, D. Tytler and B. Kambalur, *Power spectrum of the flux in the Lyman-alpha forest from high-resolution spectra of 87 QSOs*, *MNRAS* **489** (2019) 2536.

[55] N.G. Karaçaylı, N. Padmanabhan, A. Font-Ribera, V. Iršič, M. Walther, D. Brooks et al., *Optimal 1D Ly α forest power spectrum estimation - II. KODIAQ, SQUAD, and XQ-100*, *MNRAS* **509** (2022) 2842 [2108.10870].

[56] M. Esposito, V. Iršič, M. Costanzi, S. Borgani, A. Saro and M. Viel, *Weighing cosmic structures with clusters of galaxies and the intergalactic medium*, *MNRAS* **515** (2022) 857 [2202.00974].

[57] E. Boera, G.D. Becker, J.S. Bolton and F. Nasir, *Revealing Reionization with the Thermal History of the Intergalactic Medium: New Constraints from the Lyα Flux Power Spectrum*, *ApJ* **872** (2019) 101 [1809.06980].

[58] B. Wilson, V. Iršič and M. McQuinn, *A measurement of the Ly β forest power spectrum and its cross with the Ly α forest in X-Shooter XQ-100*, *MNRAS* **509** (2022) 2423 [2106.04837].

[59] M. Walther, N. Schöneberg, S. Chabanier, E. Armengaud, J. Sexton, C. Yèche et al., *Emulating the Lyman-Alpha forest 1D power spectrum from cosmological simulations: new models and constraints from the eBOSS measurement*, *JCAP* **2025** (2025) 099 [2412.05372].

[60] L. Cabayol-Garcia, J. Chaves-Montero, A. Font-Ribera and C. Pedersen, *A neural network emulator for the Lyman-α forest 1D flux power spectrum*, *MNRAS* **525** (2023) 3499 [2305.19064].

[61] S. Bird, K.K. Rogers, H.V. Peiris, L. Verde, A. Font-Ribera and A. Pontzen, *An emulator for the Lyman-α forest*, *JCAP* **2019** (2019) 050 [1812.04654].

[62] Z. Jin, M. Wolfson, J.F. Hennawi and D. González-Hernández, *Neural network emulator to constrain the high-z IGM thermal state from Lyman-α forest flux autocorrelation function*, *MNRAS* **536** (2025) 2277 [2410.06505].

[63] S. Bird, M. Fernandez, M.-F. Ho, M. Qezlou, R. Monadi, Y. Ni et al., *PRIYA: a new suite of Lyman-α forest simulations for cosmology*, *JCAP* **2023** (2023) 037 [2306.05471].

[64] P. Perdikaris, M. Raissi, A. Damianou, N.D. Lawrence and G.E. Karniadakis, *Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling*, *Proceedings of the Royal Society of London Series A* **473** (2017) 20160751.

[65] B. Peherstorfer, K. Willcox and M. Gunzburger, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, *SIAM Review* **60** (2018) 550 [https://doi.org/10.1137/16M1082469].

[66] M.-F. Ho, S. Bird and C.R. Shelton, *Multifidelity emulation for the matter power spectrum using Gaussian processes*, *MNRAS* **509** (2022) 2551 [2105.01081].

[67] M.A. Fernandez, M.-F. Ho and S. Bird, *A multifidelity emulator for the Lyman-α forest flux power spectrum*, *MNRAS* **517** (2022) 3200 [2207.06445].

[68] Y. Yang, S. Bird and M.-F. Ho, *Ten-parameter simulation suite for cosmological emulation beyond ΛCDM*, *PRD* **111** (2025) 083529 [2501.06296].

[69] K. Diao and Y. Mao, *Multi-fidelity emulator for large-scale 21 cm lightcone images: a few-shot transfer learning approach with generative adversarial network*, *arXiv e-prints* (2025) arXiv:2502.04246 [2502.04246].

[70] M.T. Tillman, B. Burkhart, S. Tonnesen, S. Bird and G.L. Bryan, *The Effects of Active Galactic Nuclei Feedback on the Lyα Forest Flux Power Spectrum*, *ApJ* **980** (2025) 72 [2410.05383].

[71] M.-F. Ho, S. Bird, M.A. Fernandez and C.R. Shelton, *MF-Box: multifidelity and multiscale emulation for the matter power spectrum*, *MNRAS* **526** (2023) 2903 [2306.03144].

[72] Y. Yang, S. Bird, M.-F. Ho and M. Qezlou, *Ten-dimensional neural network emulator for the nonlinear matter power spectrum*, *arXiv e-prints* (2025) arXiv:2507.07177 [2507.07177].

[73] Y. Yang, S. Bird, M.-F. Ho and M. Qezlou, *Design and optimization of neural networks for multifidelity cosmological emulation*, *arXiv e-prints* (2025) arXiv:2507.07184 [2507.07184].

[74] M.A. Fernandez, S. Bird and M.-F. Ho, *Cosmological constraints from the eBOSS Lyman-α forest using the PRIYA simulations*, *JCAP* **2024** (2024) 029 [2309.03943].

[75] J.S. Bolton and G.D. Becker, *Resolving the high redshift Lyα forest in smoothed particle hydrodynamics simulations*, *MNRAS* **398** (2009) L26 [0906.2861].

[76] Y. Zhu, G.D. Becker, S.E.I. Bosman, L.C. Keating, V. D'Odorico, R.L. Davies et al., *Long Dark Gaps in the Lyβ Forest at z ¡ 6: Evidence of Ultra-late Reionization from XQR-30 Spectra*, *ApJ* **932** (2022) 76 [2205.04569].

[77] T.S. Kim, J.S. Bolton, M. Viel, M.G. Haehnelt and R.F. Carswell, *An improved measurement of the flux distribution of the Lyα forest in QSO absorption spectra: the effect of continuum fitting, metal contamination and noise properties*, *MNRAS* **382** (2007) 1657 [0711.1862].

[78] G.D. Becker, P.C. Hewett, G. Worseck and J.X. Prochaska, *A refined measurement of the mean transmitted flux in the Lyα forest over $2 < z < 5$ using composite quasar spectra*, *MNRAS* **430** (2013) 2067 [1208.2584].

[79] J. Torrado and A. Lewis, *Cobaya: code for Bayesian analysis of hierarchical physical models*, *JCAP* **2021** (2021) 057 [2005.05290].

[80] J. Torrado and A. Lewis, "Cobaya: Bayesian analysis in cosmology." Astrophysics Source Code Library, record ascl:1910.019, Oct., 2019.

[81] A. Lewis, *Efficient sampling of fast and slow cosmological parameters*, *Phys. Rev. D* **87** (2013) 103529 [1304.4473].

[82] S. López, V. D'Odorico, S.L. Ellison, G.D. Becker, L. Christensen, G. Cupani et al., *XQ-100: A legacy survey of one hundred 3.5 ≲ z ≲ 4.5 quasars observed with VLT/X-shooter*, *AAP* **594** (2016) A91 [1607.08776].

[83] N. Lehner, J.M. O'Meara, A.J. Fox, J.C. Howk, J.X. Prochaska, V. Burns et al., *Galactic and Circumgalactic O VI and its Impact on the Cosmological Metal and Baryon Budgets at 2 ¡ z ≲ 3.5*, *ApJ* **788** (2014) 119 [1401.1811].

[84] J.M. O'Meara, N. Lehner, J.C. Howk, J.X. Prochaska, A.J. Fox, M.A. Swain et al., *The First Data Release of the KODIAQ Survey*, *AJ* **150** (2015) 111 [1505.03529].

[85] J.M. O'Meara, N. Lehner, J.C. Howk, J.X. Prochaska, A.J. Fox, M.S. Peeples et al., *The Second Data Release of the KODIAQ Survey*, *AJ* **154** (2017) 114 [1707.07905].

[86] M.T. Murphy, G.G. Kacprzak, G.A.D. Savorgnan and R.F. Carswell, *The UVES Spectral Quasar Absorption Database (SQUAD) data release 1: the first 10 million seconds*, *MNRAS* **482** (2019) 3458 [1810.06136].

[87] J. Vernet, H. Dekker, S. D'Odorico, L. Kaper, P. Kjaergaard, F. Hammer et al., *X-shooter, the new wide band intermediate resolution spectrograph at the ESO Very Large Telescope*, *AAP* **536** (2011) A105 [1110.1944].

[88] S.S. Vogt, S.L. Allen, B.C. Bigelow, L. Bresee, B. Brown, T. Cantrall et al., *HIRES: the high-resolution echelle spectrometer on the Keck 10-m Telescope*, in *Instrumentation in Astronomy VIII*, D.L. Crawford and E.R. Craine, eds., vol. 2198 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 362, June, 1994, DOI.

[89] H. Dekker, S. D'Odorico, A. Kaufer, B. Delabre and H. Kotzlowski, *Design, construction, and performance of UVES, the echelle spectrograph for the UT2 Kueyen Telescope at the ESO Paranal Observatory*, in *Optical and IR Telescope Instrumentation and Detectors*, M. Iye and A.F. Moorwood, eds., vol. 4008 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 534–545, Aug., 2000, DOI.

[90] 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 α forest flux power spectrum*, *MNRAS* **474** (2018) 3032 [1706.08532].

[91] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

[92] M.-F. Ho, S. Bird and R. Garnett, *Damped Lyman-α absorbers from Sloan digital sky survey DR16Q with Gaussian processes*, *MNRAS* **507** (2021) 704 [2103.10964].

[93] G. Altay, T. Theuns, J. Schaye, N.H.M. Crighton and C. Dalla Vecchia, *Through Thick and Thin—H I Absorption in Cosmological Simulations*, *ApJL* **737** (2011) L37 [1012.4014].

[94] T. Zafar, C. Péroux, A. Popping, B. Milliard, J.M. Deharveng and S. Frank, *The ESO UVES advanced data products quasar sample. II. Cosmological evolution of the neutral gas mass density*, *AAP* **556** (2013) A141 [1307.0602].

[95] J.X. Prochaska, J.M. O'Meara and G. Worseck, *A Definitive Survey for Lyman Limit Systems at z ~3.5 with the Sloan Digital Sky Survey*, *ApJ* **718** (2010) 392 [0912.0292].

[96] G.C. Rudie, C.C. Steidel, M. Pettini, R.F. Trainor, A.L. Strom, C.B. Hummels et al., *Column Density, Kinematics, and Thermal State of Metal-bearing Gas within the Virial Radius of z ∼ 2 Star-forming Galaxies in the Keck Baryonic Structure Survey*, *ApJ* **885** (2019) 61 [1903.00004].

[97] K. Kohler and N.Y. Gnedin, *Lyman Limit Systems in Cosmological Simulations*, *ApJ* **655** (2007) 685 [astro-ph/0605032].

[98] J. Fan, H. Zhu, C. Avestruz and N.Y. Gnedin, *Cosmic Reionization on Computers: Statistics, Physical Properties, and Environments of Lyman Limit Systems at z ∼ 6*, *ApJ* **963** (2024) 45 [2212.07033].

[99] I. Georgiev, G. Mellema and S.K. Giri, *The forest at EndEoR: the effect of Lyman limit systems on the end of reionization*, *MNRAS* **536** (2025) 3689 [2405.04273].

[100] M. McQuinn, S.P. Oh and C.-A. Faucher-Giguère, *On Lyman-limit Systems and the Evolution of the Intergalactic Ionizing Background*, *ApJ* **743** (2011) 82 [1101.1964].

[101] A. D'Aloisio, M. McQuinn, H. Trac, C. Cain and A. Mesinger, *Hydrodynamic Response of the Intergalactic Medium to Reionization*, *ApJ* **898** (2020) 149 [2002.02467].

[102] J.M. O'Meara, N. Lehner, J.C. Howk, J.X. Prochaska, A.J. Fox, M.S. Peeples et al., *The Second Data Release of the KODIAQ Survey*, *AJ* **154** (2017) 114 [1707.07905].

[103] M. Fumagalli, S. Fotopoulou and L. Thomson, *Detecting neutral hydrogen at z ≳ 3 in large spectroscopic surveys of quasars*, *MNRAS* **498** (2020) 1951 [2009.03322].

[104] K.K. Rogers and V. Poulin, *5σ tension between Planck cosmic microwave background and eBOSS Lyman-alpha forest and constraints on physics beyond ΛCDM*, *Physical Review Research* **7** (2025) L012018 [2311.16377].

[105] A. Brodzeller, M. Wolfson, D.M. Santos, M. Ho, T. Tan, M.M. Pieri et al., *Construction of the Damped Lyα Absorber Catalog for DESI DR2 Lyα BAO*, *arXiv e-prints* (2025) arXiv:2503.14740 [2503.14740].

[106] V. Iršič, M. Viel, M.G. Haehnelt, J.S. Bolton, M. Molaro, E. Puchwein et al., *Unveiling dark matter free streaming at the smallest scales with the high redshift Lyman-alpha forest*, *PRD* **109** (2024) 043511 [2309.04533].

[107] J. Bruna and S. Mallat, *Invariant Scattering Convolution Networks*, *arXiv e-prints* (2012) arXiv:1203.1513 [1203.1513].

[108] J. Anden and S. Mallat, *Deep Scattering Spectrum*, *IEEE Transactions on Signal Processing* **62** (2014) 4114 [1304.6763].

[109] M. Andreux, T. Angles, G. Exarchakis, R. Leonarduzzi, G. Rochette, L. Thiry et al., *Kymatio: Scattering Transforms in Python*, *arXiv e-prints* (2018) arXiv:1812.11214 [1812.11214].

[110] H.M. Tohfa, S. Bird, M.-F. Ho, M. Qezlou and M. Fernandez, *Forecast Cosmological Constraints with the 1D Wavelet Scattering Transform and the Lyman-α Forest*, *PRL* **132** (2024) 231002 [2310.06010].