

Prospects of a statistical detection of the 21-cm forest and its potential to constrain the thermal state of the neutral IGM during reionization

Tomáš Šoltínský^{1,2} ^{*}, Girish Kulkarni¹ , Shriharsh P. Tendulkar¹ , James S. Bolton³ ,

¹Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India

²INAF–Osservatorio Astronomico di Trieste, Via G.B. Tiepolo, 11, I-34143 Trieste, Italy

³School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

Accepted —. Received —; in original form —

ABSTRACT

The 21-cm forest signal is a promising probe of the Epoch of Reionization complementary to other 21-cm line observables and Ly α forest signal. Prospects of detecting it have significantly improved in the last decade thanks to the discovery of more than 30 radio-loud quasars at these redshifts, upgrades to telescope facilities, and the notion that neutral hydrogen islands persist down to $z \lesssim 5.5$. We forward-model the 21-cm forest signal using semi-numerical simulations and incorporate various instrumental features to explore the potential of detecting the 21-cm forest at $z = 6$, both directly and statistically, with the currently available (uGMRT) and forthcoming (SKA1-low) observatories. We show that it is possible to detect the 1D power spectrum of the 21-cm forest spectrum, especially at large scales of $k \lesssim 8.5 \text{ MHz}^{-1}$ with the 500 hr of the uGMRT time and $k \lesssim 32.4 \text{ MHz}^{-1}$ with the SKA1-low over 50 hr if the intergalactic medium (IGM) is 25% neutral and these neutral hydrogen regions have a spin temperature of $\lesssim 30 \text{ K}$. On the other hand, we infer that a null-detection of the signal with such observations of 10 radio-loud sources at $z \approx 6$ can be translated into constraints on the thermal and ionization state of the IGM which are tighter than the currently available measurements. Moreover, a null-detection of the 1D 21-cm forest power spectrum with only 50 hr of the uGMRT observations of 10 radio-loud sources can already be competitive with the Ly α forest and 21-cm tomographic observations in disfavoured models of significantly neutral and cold IGM at $z = 6$.

Key words: methods: numerical – intergalactic medium – quasars: absorption lines – dark ages, reionization, first stars

1 INTRODUCTION

The hyperfine 21-cm line arising from the spin-flip transition in neutral hydrogen (HI) residing in the intergalactic medium (IGM) is potentially a powerful probe of the Epoch of Reionization. This transition has a factor of $\sim 10^7$ smaller cross-section than the Ly α transition, and hence does not saturate at $z \gtrsim 6$ when the IGM is expected to be at least partially neutral. This is the main advantage of the 21-cm line over the Lyman-series absorption in the spectra of luminous quasars, which is currently the premier probe of the reionization era. The 21-cm line can be observed as a series of absorption lines in spectra of distant radio-loud sources, which is known as the 21-cm forest signal. Such sources include radio-loud quasars (Ciardi et al. 2013) or gamma-ray bursts (Ioka & Mészáros 2005; Ciardi et al. 2015). Given that other 21-cm line observables, such as 21-cm power spectrum from tomographic observations (e.g. Mertens et al. 2020; Trott et al. 2020; HERA Collaboration 2022) and sky-averaged 21-cm spectrum (e.g. Bowman et al. 2018; Singh et al. 2022), utilize Cosmic Microwave Background as the radio background source, the 21-cm forest signal is subject to different systematic uncertainties (Carilli et al. 2002; Furlanetto & Loeb 2002; Furlanetto et al. 2006;

Pritchard & Loeb 2012). Therefore, the 21-cm forest is complementary to these observables.

Not only is the 21-cm forest complementary to other probes of the Epoch of Reionization, it is also a unique probe as it is strongest in cold, neutral and slightly overdense gas (Šoltínský et al. 2021) as opposed to the Ly α forest. This makes it one of the few probes of cold and neutral regions in the IGM besides the Mg II line for example (Hennawi et al. 2021; Tie et al. 2024). Besides the diffuse IGM, minihaloes may contribute to the 21-cm forest absorption too (Meiksin 2011; Kadota et al. 2023; Naruse et al. 2024), and hence in principle one can probe such small-scale structure with this signal. Furthermore, it has been suggested that the 21-cm forest signal can aid in constraining the nature of dark matter (e.g. Shimabukuro et al. 2014, 2020, 2023; Kawasaki et al. 2021; Shao et al. 2023), primordial black holes (Villanueva-Domingo & Ichiki 2023), thermal state of the IGM (e.g. Xu et al. 2009, 2011; Mack & Wyithe 2012; Šoltínský et al. 2021; Shao et al. 2023), metal content in the IGM (Bhagwat et al. 2022) and supermassive black hole growth models (Šoltínský et al. 2023).

While there is no measurement of the cosmological 21-cm forest signal up to date, despite various attempts including the one by Carilli et al. (2007), the prospects of detecting it have improved in the last decade. Firstly, prospects of detecting the 21-cm forest have improved from the observational perspective in the form of tens of

* E-mail: tomas.soltinsky@inaf.it

recently confirmed radio-loud quasars (RLQSO) at high redshift. Since 2020, the number of $z \geq 6$ RLQSO has increased by a factor of 5 to ~ 15 while the number of $z \geq 5.5$ RLQSO has more than tripled to ~ 34 to date (Belladitta et al. 2020; Liu et al. 2021; Bañados et al. 2021, 2023, 2024; Ighina et al. 2021, 2023, 2024; Endsley et al. 2023; Gloude-mans et al. 2022, 2023; Wolf et al. 2024). This number is expected to increase by orders of magnitude. For example, the physics-driven models calibrated with recent ultra-violet (UV) observations developed by Niu et al. (2024) suggest that there are ≥ 50 RLQSO with an intrinsic flux density at 150 MHz, S_{150} , larger than 100 mJy and ≥ 1000 of $S_{150} \geq 10$ mJy at $z > 5.5$ over the whole sky. If large radio surveys such as The Low Frequency Array (LOFAR) Two-metre Sky Survey (LoTSS Shimwell et al. 2017; Kondapally et al. 2021), the Tata Institute of Fundamental Research (TIFR) Giant Metrewave Radio Telescope (GMRT) Sky Survey (TGSS; Intema et al. 2017), and the Galactic and Extragalactic All-sky Murchison Widefield Array (MWA) survey (GLEAM; Wayth et al. 2015) are combined with follow-up programmes such as the William Herschel Telescope Enhanced Area Velocity Explorer (WEAVE)-LOFAR survey (Smith et al. 2016), these high- z RLQSO should be detectable. In addition, the advancements on the instrumental front, including the recent completion of upgraded GMRT (uGMRT; Gupta et al. 2017) and plans for Square Kilometre Array (SKA), make the detection of the cosmological 21-cm forest even more encouraging because of their improved sensitivity.

Secondly, recent Ly α observations can be interpreted as reionization being completed by $z \leq 5.6$ (e.g. Kulkarni et al. 2019; Nasir & D’Aloisio 2020; Choudhury et al. 2021; Qin et al. 2021). Late-end reionization picture is consistent with the large spatial fluctuations in the Ly α forest opacity (Becker et al. 2015; Bosman et al. 2022), deficit of Ly α emitting galaxies around extended Ly α absorption troughs (Kashino et al. 2020; Keating et al. 2020; Christenson et al. 2021), clustering of Ly α emitters (Weinberger et al. 2019), thermal widths of Ly α forest transmission spikes at $z > 5$ (Gaikwad et al. 2020), Ly α equivalent widths (Nakane et al. 2024) and the detections of damping wings from neutral islands at $z < 6$ in the Ly α forest spectra (Spina et al. 2024; Becker et al. 2024; Zhu et al. 2024). Further support of the late-end reionization models comes from the measurements of long dark gaps in the Ly α forest (Zhu et al. 2021) and Ly β forest (Zhu et al. 2022), and mean free path of ionizing photons at $z = 6$ (Becker et al. 2021; Cain et al. 2021; Zhu et al. 2023; Gaikwad et al. 2023). In such late-end reionization scenario large H I islands are expected to persist until $z \sim 5.3$. Šoltinský et al. (2021) have shown that it might be possible to detect strong 21-cm forest absorbers at $z = 6$ with SKA1-low (or even LOFAR) in the case of reionization ending at $z \approx 5.3$ and the IGM being pre-heated by X-rays such that the spin temperature $T_S \lesssim 10^2$ K. On the other hand, they suggest that a null-detection of the 21-cm forest absorption can be utilized to put a model-dependent informative lower limit on the soft X-ray background radiation.

The goal of this study is to explore the potential of a direct and statistical detection of the 21-cm forest signal at $z \approx 6$ in the light of the above-mentioned advancements in the field. While different statistics of the 21-cm forest signal have been studied (e.g. Mack & Wyithe 2012; Ewall-Wice et al. 2014), here we focus on the 1D power spectrum calculated from the 21-cm forest spectrum, P_{21} (which has been investigated by Thyagarajan (2020), Shao et al. (2023, 2024) and Sun et al. (2024)). As opposed to the previous 21-cm forest 1D power spectrum studies listed above, we focus on mock observations of (a) rather small number of background radio sources; (b) at significantly lower redshift, $z = 6$ (end stages of reionization); (c) longer observational time. Furthermore, we aim to explore the

constraining power of the P_{21} , particularly if it can be used to put joint constraints on the thermal and ionization state of the IGM. To achieve this, we take advantage of semi-numerical simulations of the reionization era IGM, 21CMFAST (Mesinger et al. 2011), to explore a wide parameter space and Bayesian statistical methods to infer parameter estimates.

This paper is structured as follows. We begin by describing our numerical model of the IGM during the Epoch of Reionization and how we forward-model the synthetic 21-cm forest spectra in Sec. 2. We then present the statistical observables of the 21-cm forest signal, particularly the differential number density of the 21-cm forest flux and 1D power spectrum in Sec. 3.1 and 3.2, respectively. Sec. 4 focuses on the inference of parameters describing the thermal and ionization state of the IGM from the P_{21} . This study concludes with a summary in Sec. 5.

2 FORWARD-MODELLING THE 21-CM FOREST SIGNAL

In this section we present steps how the synthetic 21-cm forest spectra are constructed in this study. In particular, we describe how the semi-numerical simulations are used to generate line-of-sight skewers of various physical quantities considering a grid of parameter combinations in Sec. 2.1, how we compute the optical depths of the 21-cm forest signal in Sec. 2.2, and how we forward-model realistic 21-cm forest spectra which include instrumental features of different radio telescopes in Sec. 2.3.

2.1 Modelling the intergalactic medium at $z = 6$

To construct the synthetic 21-cm forest spectra at $z = 6$ we start by running 21CMFAST semi-numerical cosmological simulations (Mesinger et al. 2011; Murray et al. 2020). These simulations are based on the excursion set formalism and a perturbation theory approach to simulate the Epoch of Reionization universe, and hence are not computationally expensive. Therefore, this is a suitable tool for exploring a wide parameter space, which is one of the main goals of this study. The parameters of interest are the mean neutral fraction $\langle x_{\text{HI}} \rangle$ and the X-ray background radiation efficiency f_X . To vary the former one we fix the mean free path of ionizing photons (i.e. fix the R_BUBBLE_MAX simulation parameter) to 0.75 pMpc following the measurement at $z = 6$ by Becker et al. (2021) and calibrate the ionizing efficiency of high- z galaxies (i.e. varying HII_EFF_FACTOR simulation parameter) such that the $\langle x_{\text{HI}} \rangle$ of interest is achieved. For the ease of comparison with literature, we implement the parametrization of the X-ray background radiation luminosity introduced in Furlanetto (2006)

$$L_X = 3.4 \times 10^{40} \text{ erg s}^{-1} f_X \left(\frac{\text{SFR}}{1 \text{ M}_\odot \text{ yr}^{-1}} \right), \quad (1)$$

where the SFR is the star formation rate. In practice, we set L_X simulation parameter to $\log_{10} \left(3.4 \times 10^{40} \text{ erg s}^{-1} f_X \right)$.

We generate models for combinations of 21 values of $\langle x_{\text{HI}} \rangle = [0, 1]$ and 26 values of $\log_{10} f_X = [-4, 1]$. While the $\langle x_{\text{HI}} \rangle$ measurements from the Ly α forest at $z \approx 6$ including McGreer et al. (2015), Zhu et al. (2022), Gaikwad et al. (2023), Āurovčíková et al. (2024) and Greig et al. (2024) are consistent with $\langle x_{\text{HI}} \rangle \lesssim 0.38$, our simulations cover the whole physical range of $\langle x_{\text{HI}} \rangle$. The upper limits on the 21-cm power spectrum acquired by MWA, LOFAR and the Hydrogen Epoch of Reionization Array (HERA) at $z > 6.5$ (Greig et al. 2021a,b; HERA Collaboration 2023, respectively) disfavour scenarios of no IGM preheating. However, the X-ray background radiation

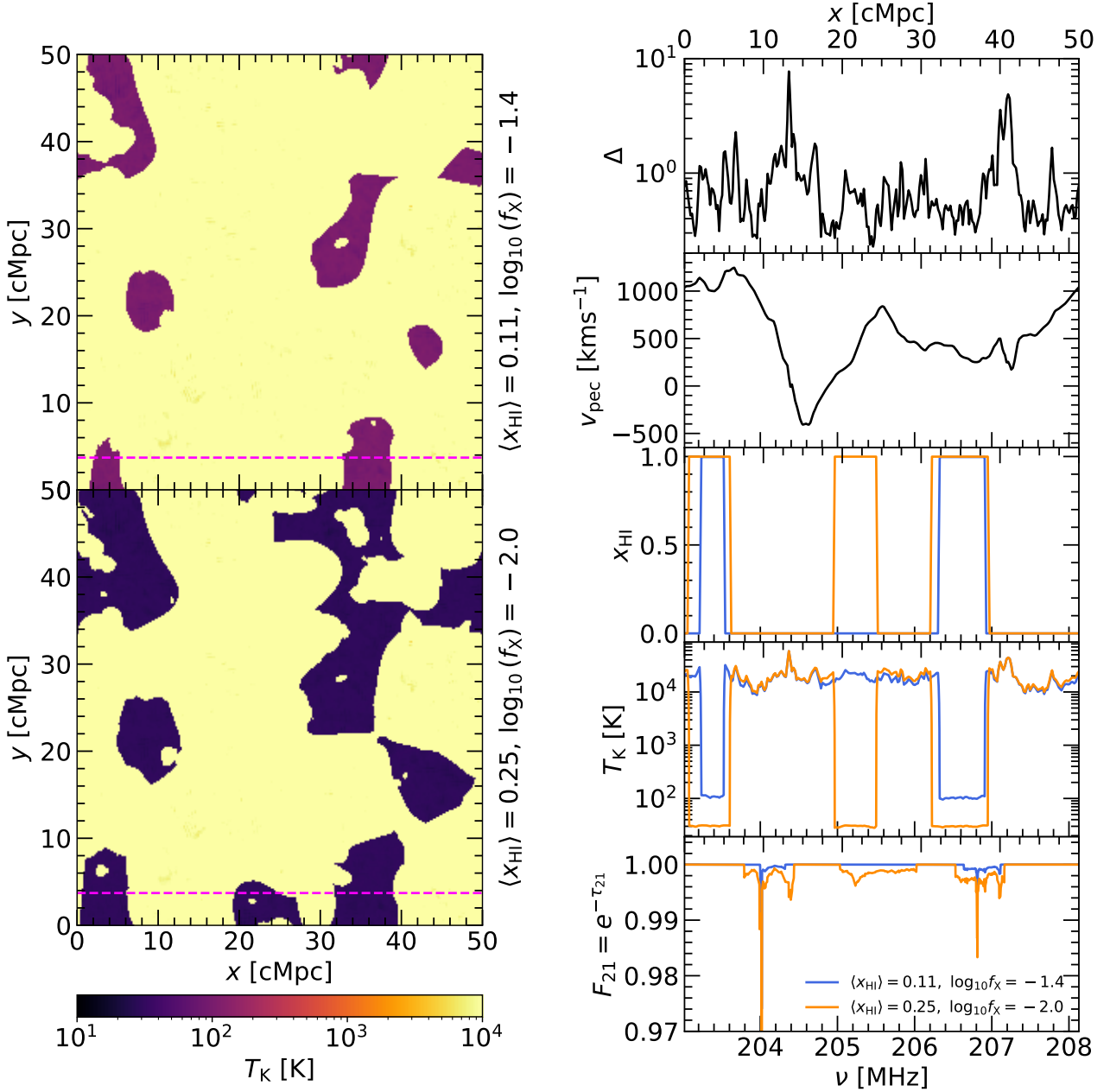


Figure 1. *Left:* 2D slices of the gas temperature at $z = 6$ for the combination of $\langle x_{\text{HI}} \rangle = 0.11$ and $\log_{10} f_X = -1.4$ (top panel) and $\langle x_{\text{HI}} \rangle = 0.25$ and $\log_{10} f_X = -2.0$ (bottom panel). *Right:* From top to bottom, gas overdensity, peculiar velocity, neutral hydrogen fraction, gas temperature and normalized 21-cm forest flux along the line-of-sight indicated by the dashed fuchsia lines in the left panels. The bottom three panels are shown for the cases of $\langle x_{\text{HI}} \rangle = 0.11$ and $\log_{10} f_X = -1.4$ (blue curves, corresponding to the top left panel) and $\langle x_{\text{HI}} \rangle = 0.25$ and $\log_{10} f_X = -2.0$ (orange curves, corresponding to the bottom left panel).

is not constrained well yet, and hence we consider a large range of $\log_{10} f_X$ values in this study. Also, note that X-ray photons can ionize HI too (Furlanetto & Stoever 2010) which results in $\langle x_{\text{HI}} \rangle$ significantly lower than 1 for high $\log_{10} f_X$ even if the `HII_EFF_FACTOR` of 0 is implemented. For example, if $\log_{10} f_X = 1$ at $z = 6$ in our models, we obtain $\langle x_{\text{HI}} \rangle \lesssim 0.82$, and to achieve $\langle x_{\text{HI}} \rangle > 0.99$ in our models $\log_{10} f_X \lesssim -0.4$ is required. This results in the fact that some of the high $\langle x_{\text{HI}} \rangle$ and high $\log_{10} f_X$ models are not acquired. In summary, we generate 534 models in total.

These simulations are run on a grid of 256^3 pixels with a volume of $(50 \text{ cMpc})^3$ at $z = 6$. Note that these simulations do not capture

coherent regions of HI islands on scales larger than the simulation box. The simulations are based on a flat Λ CDM cosmology consistent with the measurements of Planck Collaboration et al. (2014), namely $\Omega_\Lambda = 0.692$, $\Omega_m = 0.308$, $\Omega_b = 0.0482$, $\sigma_8 = 0.829$, $n_s = 0.961$ and $h = 0.678$. Furthermore, a primordial helium fraction by mass of $Y_p = 0.24$ (Hsyu et al. 2020) is adopted. Two examples of 2D slices of the gas kinetic temperature, are presented in the left panels of Fig. 1. The top panel corresponds to a warmer and more ionized IGM ($\log_{10} f_X = -1.4$, $\langle x_{\text{HI}} \rangle = 0.11$) than the bottom one ($\log_{10} f_X = -2$, $\langle x_{\text{HI}} \rangle = 0.25$). In both cases the ionized regions (yellow areas) have a temperature higher than 2×10^4 K, and hence the 21-cm forest is

expected to be suppressed completely. More importantly, in the top panel the mean temperature of neutral regions $T_{\text{HI}} = 102 \text{ K}^1$ and in the bottom panel $T_{\text{HI}} = 30 \text{ K}$. We extract 1000 periodic lines-of-sight parallel to the simulation box which contain overdensity $\Delta = 1 + \delta$, peculiar velocity v_{pec} , neutral hydrogen fraction x_{HI} and T_{K} fields. The right panels of Fig. 1 show these quantities from top to bottom, respectively, for the LOS piercing the IGM as indicated by the dashed fuchsia line in the left panels.

2.2 The 21-cm forest absorption spectrum

The ground state of neutral hydrogen atoms is split into two energy levels given by the relative orientation of the proton and electron spin. The energy difference corresponds to the energy of a photon with a rest-frame wavelength of $\lambda_{21} = 21.11 \text{ cm}$ or equivalently a frequency of $\nu_{21} = 1420.41 \text{ MHz}$. This hyperfine structure of HI atoms gives rise to the 21-cm line. Here we focus on the 21-cm forest signal. Following Šoltinský et al. (2021), we compute the normalized 21-cm forest flux, $F_{21} = e^{-\tau_{21}}$, where τ_{21} is the optical depth to the 21-cm photons. In the discrete form, the τ_{21} for pixel i is given by (e.g. Furlanetto & Loeb 2002)

$$\tau_{21,i} = \frac{3h_{\text{p}}c^3A_{10}}{32\pi^{3/2}\nu_{21}^2k_{\text{B}}H(z)} \frac{\delta\nu}{H(z)} \times \sum_{j=1}^N \frac{n_{\text{HI},j}}{b_jT_{\text{S},j}} \exp\left(-\frac{(v_{\text{H},i} - u_j)^2}{b_j^2}\right), \quad (2)$$

where h_{p} and k_{B} are the Planck and Boltzmann constants, respectively, c is speed of light, $A_{10} = 2.85 \times 10^{-15} \text{ s}^{-1}$ is the Einstein spontaneous emission coefficient for the hyperfine transition, $\delta\nu$ is the velocity width of the pixels², $H(z)$ is the Hubble parameter, n_{HI} is the HI number density and $b = (2k_{\text{B}}T_{\text{K}}/m_{\text{H}})^{1/2}$ is the Doppler parameter. We assume fully coupled spin temperature, T_{S} , to T_{K} (i.e. $T_{\text{S}} = T_{\text{K}}$). This is a valid assumption at $z = 6$ because Ly α background is expected to be strong at such low redshift. Note that Šoltinský et al. (2021) found differences in the 21-cm forest signal between the case of $T_{\text{S}} = T_{\text{K}}$ and $T_{\text{S}} \neq T_{\text{K}}$, however at $z = 6$ they were rather small. In Eq. 2 we convolve over the Gaussian line profile to include not only the effect of Hubble velocity, v_{H} , on the 21-cm forest, but also the effect of peculiar velocity of the gas along the LOS, in $u_j = v_{\text{H},j} + v_{\text{pec},j}$. In the bottom right panel of Fig. 1 we show an example synthetic 21-cm forest spectrum at $z = 6$ (without any instrumental features) for the case of $\log_{10}f_{\text{X}} = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$ (orange curve) and $\log_{10}f_{\text{X}} = -1$ and $\langle x_{\text{HI}} \rangle = 0.11$ (blue curve). It is clearly seen how the ionization bubbles completely eliminate the 21-cm forest absorbers while the higher spin temperature suppresses them.

2.3 Instrumental features of the uGMRT and SKA1-low

Before we describe how we incorporate instrumental features in our forward-modelled 21-cm forest spectra, we note that there is a substantial number of radio-loud sources at high z , particularly 34 RLQ-

¹ The 21cmFAST simulates reionization bubbles using excursion set principle, and hence the IGM can have neutral hydrogen fraction of $x_{\text{HI}} = 0$ (completely ionized bubbles) and $x_{\text{HI}} \approx 1$ (fully neutral islands) and no values between. Therefore, there is no arbitrary criterion for what is defined as a neutral region.

² We resample the LOS of IGM fields using linear interpolation to ensure that the 21-cm absorption line profiles are converged correctly. For more details we refer the reader to Šoltinský et al. (2021).

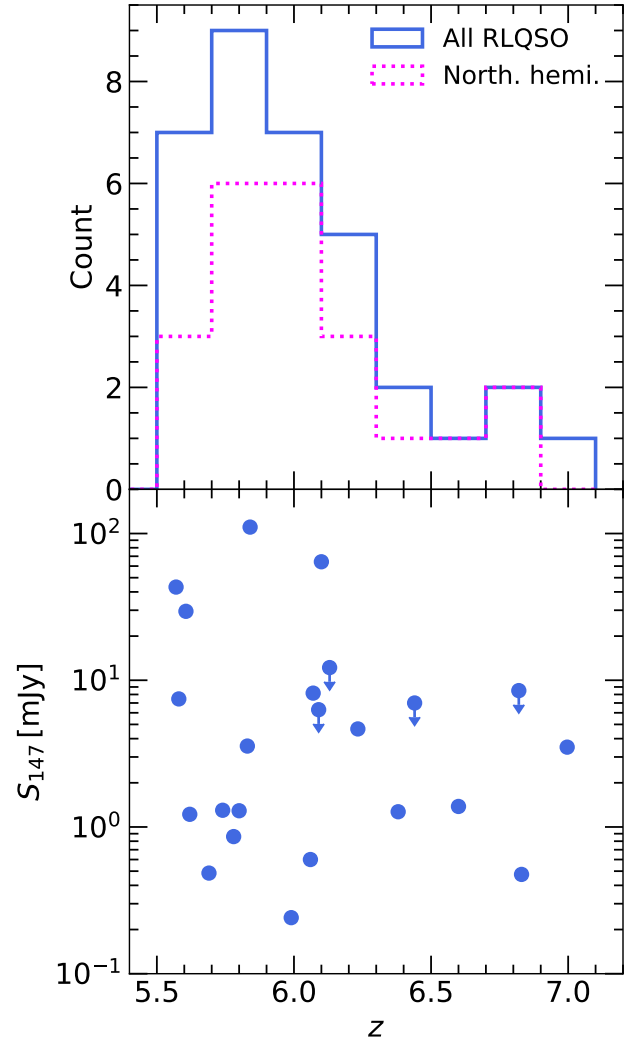


Figure 2. *Top panel:* Redshift distribution of radio-loud quasars identified at $z \geq 5.5$ across the whole sky (solid blue curve) and in the northern hemisphere (dashed fuchsia curve). The sample is compiled from Fan et al. (2001), McGreer et al. (2006), Willott et al. (2010), Zeimann et al. (2011), Bañados et al. (2015, 2018, 2021, 2023, 2024), Belladitta et al. (2020), Liu et al. (2021), Ighina et al. (2021, 2023, 2024), Endsley et al. (2023), Gloude-mans et al. (2022, 2023) and Wolf et al. (2024). *Bottom panel:* Intrinsic flux density at 147 MHz of these radio-loud quasars as a function of their redshift. The arrows indicate quasars for which only upper limits on S_{147} are available. Note that not all quasars have a S_{147} measurement.

SOs at $z \geq 5.5$ to date³. The redshift distribution of these RLQSOs is shown in Fig. 2 with the highest redshift reaching $z = 6.82$ (Bañados et al. 2021). Most of these RLQSOs are located in the northern hemisphere as indicated by the dashed fuchsia curve. Given that this sample lies within the redshift range of $5.5 \leq z \leq 7$, we will consider the observed frequencies of $177.5 \text{ MHz} \lesssim \nu_{\text{obs}} \lesssim 220 \text{ MHz}$.

In Fig. 3 we show the sensitivity of various telescopes assuming the whole array, $A_{\text{eff}}/T_{\text{sys}}$, as a function of observed frequency with the above frequency range indicated with the red shaded region. These values are taken from Braun et al. (2019) (see their fig. 8) and are shown for LOFAR (blue curve), band-2 of uGMRT (fuchsia curve)

³ The whole sample can be found at <https://tomassoltinsky.github.io>.

and SKA1-low (orange curve). For comparison, the green and grey shaded regions mark some of the past GMRT surveys, particularly the TGSS by [Intema et al. \(2017\)](#) and the one by [Gloude-mans et al. \(2023\)](#) in the uGMRT band-3, respectively. It is clear that at these frequencies the uGMRT is currently the most sensitive operational telescope. Therefore, we focus our analyses on mock observations by uGMRT in band-2 to explore the potential of the currently achievable observations of the 21-cm forest signal. However, we extend our analyses to the SKA1-low to probe future outlooks too.

Firstly, we model the effect of spectral resolution of the telescope on the observed 21-cm forest spectrum by convolving the spectrum with a boxcar function of $\Delta\nu$ width. The band-2 of uGMRT covers the frequency range of 125 – 250 MHz and contains 16384 spectral channels ([Gupta et al. 2017](#)). Splitting the whole bandwidth by the number of spectral channels, one can achieve the spectral resolution of $\Delta\nu = 8$ kHz, which we will use in our analyses. A smoothed spectrum for $\log_{10}f_X = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$ is shown as the dashed orange curve in Fig. 4. Comparing this with unsmoothed spectrum in Fig. 1, the main effect of smoothing is the suppression of the 21-cm forest absorption features depth. Note that we constructed longer spectra than in Fig. 1 by splicing 4 randomly chosen 50 cMpc long spectra. This way we forward model spectra of 200 cMpc and 22.1 MHz bandwidth at $z = 6$ which will be important in Section 3.2 and 4.

Secondly, we incorporate the telescope noise. In particular, we add a Gaussian white noise with the rms given by (cf. [Datta et al. 2007](#); [Ciardi et al. 2013](#))

$$\sigma_N = \left(\frac{A_{\text{eff}}}{T_{\text{sys}}} \right)_{N_d=1}^{-1} \frac{\sqrt{2}k_B}{\sqrt{N_d(N_d - 1)\Delta\nu t_{\text{int}}}}, \quad (3)$$

where $(A_{\text{eff}}/T_{\text{sys}})_{N_d=1}$ is the frequency-dependent sensitivity for a single dish taken from [Braun et al. \(2019\)](#)⁴, N_d is the number of dishes and t_{int} is the integration time⁵.

The 21-cm forest spectrum is then normalized based on the intrinsic spectrum of the background radio source. We model the radio spectrum of the RLQSO as a single power-law with index α_R and normalization given by the intrinsic flux density at 147 MHz, S_{147} , i.e. $S = S_{147} (\nu/147 \text{ MHz})^{\alpha_R}$. Note that this will make the normalized rms of the telescope noise frequency-dependent. Both S_{147} and α_R have been measured for various high- z RLQSO by [Gloude-mans et al. \(2022, 2023\)](#) and [Bañados et al. \(2023\)](#) to name few examples. The measurements of the S_{147} are provided in the bottom panel of Fig. 2. Note that most of the $z > 5.5$ RLQSOs are rather faint ($S_{147} \sim 1$ mJy), however there are 4 with $S_{147} > 29$ mJy. In what follows we will consider the PSO J0309+27 blazar, one of the brightest confirmed radio-loud sources at $z \approx 6$, which was measured to have $S_{147} = 64.2 \pm 6.2$ mJy and $\alpha_R = -0.44 \pm 0.11$ ([Belladitta et al. 2020](#)).

⁴ While we include the dependence of $A_{\text{eff}}/T_{\text{sys}}$ on the ν_{obs} , the $A_{\text{eff}}/T_{\text{sys}}$ does not vary significantly over our synthetic spectra given that their frequency range is narrow. Note also that [Braun et al. \(2019\)](#) gives the $A_{\text{eff}}/T_{\text{sys}}$ for the whole array of dishes instead of a single dish. Therefore, we define $(A_{\text{eff}}/T_{\text{sys}})_{N_d=1} = A_{\text{eff}}/T_{\text{sys}} \sqrt{N_{d,\text{tot}}(N_{d,\text{tot}} - 1)}$, where $N_{d,\text{tot}}$ is the number of dishes of the whole array of the observatory, i.e. 30 and 512 for the uGMRT and SKA1-low, respectively.

⁵ Instead, one can use GMRT Exposure Time Calculator (ETC, <http://www.ncra.tifr.res.in:8081/secr-ops/etc/rms/rms.html>) to compute the σ_N . We have compared our calculation with the GMRT ETC and found that we overestimate the σ_N by 22%. There is a sensitivity calculator for the SKA1-low too (<https://sensitivity-calculator.skao.int/low>) which results in $\sim 30\%$ larger σ_N than our calculation.

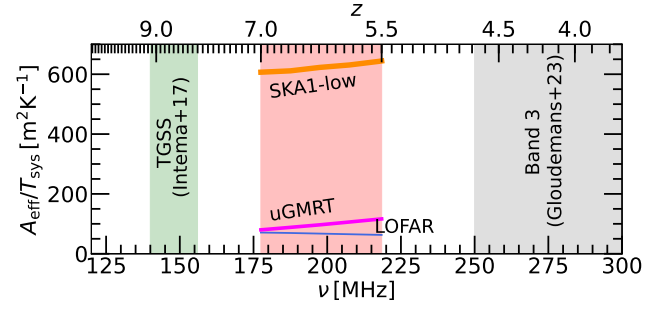


Figure 3. The sensitivity in terms of $A_{\text{eff}}/T_{\text{sys}}$ for the whole array at different observed frequencies for the LOFAR (thin blue curve), uGMRT (semi-thick fuchsia curve) and SKA1-low (thick orange curve) taken from [Braun et al. \(2019\)](#). The sensitivities of these three telescopes are shown only in the frequency range in which the 21-cm forest is expected to be currently observable, i.e. at $177.5 \text{ MHz} \lesssim \nu_{\text{obs}} \lesssim 218.5 \text{ MHz}$ (red shaded region). These frequencies correspond to the redshifts of the quasar sample shown in Fig. 2. The green shaded region indicates the frequencies at which the TGSS ([Intema et al. 2017](#)) was performed and the grey shaded region partially marks the band-3 of the uGMRT which was used for the survey in [Gloude-mans et al. \(2023\)](#).

The left panel of Fig. 4 shows a forward-modelled 21-cm forest spectrum implemented with the above described instrumental features assuming an $t_{\text{int}} = 500$ hr observation by the uGMRT (solid black curve) and the same LOS as for the case of no telescope noise (dashed orange curve, IGM model of $\log_{10}f_X = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$). We assumed the whole array of $N_d = 30$. The signal-to-noise ratio (SNR) of the deviation from the continuum (i.e. absorption signal) reaches 11.9 at $\nu \approx 224.4$ MHz while the highest SNR for all 1000 synthetic LOS is 36.5. However, frequency channels with such high SNR are > 20 times less abundant than the ones at more representative values of $\text{SNR} \approx 0.4$. More details about the SNR calculation can be found in Appendix C. Given this, it is obvious that the direct detection of individual absorption features is challenging. In addition, we do not consider the effect of the Radio Frequency Interference (RFI). RFI can introduce both narrow-band and broad-band features in the radio spectra which would complicate observations of the 21-cm forest signal. However, according to *The Exposure Time Calculator for the upgraded Giant Metrewave Radio Telescope* handbook⁶ the frequencies unusable due to the RFI span over 165 MHz – 190 MHz which are lower than the frequencies of interest as seen in Fig. 3.

We now assume the same LOS and instrumental behaviour for the SKA1-low observatory (i.e. using Eq. 3) and the whole interferometer with $N_d = 512$ but shorter observational time of $t_{\text{int}} = 50$ hr. The $A_{\text{eff}}/T_{\text{sys}}$ for the SKA1-low is taken from [Braun et al. \(2019\)](#) too. Given that the SKA1-low is expected to observe sky south from Dec. $+30^\circ$ ([Zheng et al. 2020](#)), the PSO J0309+27 should be observable by this telescope too. The observations at 72 MHz – 231 MHz by the MWA, which is located at the same site as the SKA1-low is planned to be built, show that this area is significantly less impacted by the RFI than LOFAR and uGMRT sites ([Offringa et al. 2015](#)). Besides this, we see in the right panel of Fig. 4 that the noise is not as severe even if the observational time is 10 times shorter than assumed for the uGMRT observation. However, while the highest SNR reaches 19.0 in this particular LOS and 63.3 in the whole sample of simulated

⁶ http://www.ncra.tifr.res.in:8081/~secr-ops/etc/etc_help.pdf

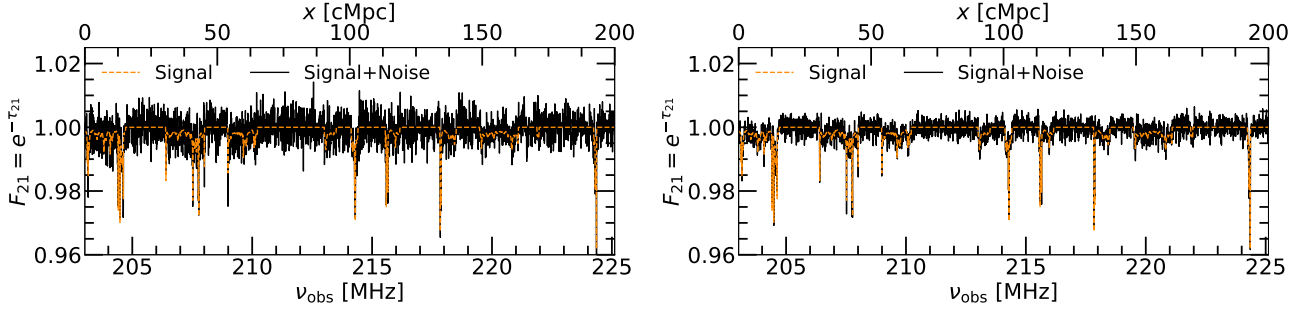


Figure 4. The forward-modelled 21-cm forest signal (solid black curves) in the spectrum of quasars with $S_{147} = 64.2$ mJy and $\alpha_R = -0.44$ at $z = 6$ as observed by the uGMRT over $t_{\text{int}} = 500$ hr (left panel) and SKA1-low over $t_{\text{int}} = 50$ hr (right panel). We also show the simulated physical signal without added noise in dashed orange curves. In both the cases, noisy spectrum and spectrum without the telescope noise, the spectrum has been smoothed by a boxcar function of width $\Delta\nu = 8$ kHz to model the spectral resolution of the telescopes. In all cases a model of $\log_{10} f_X = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$ has been used.

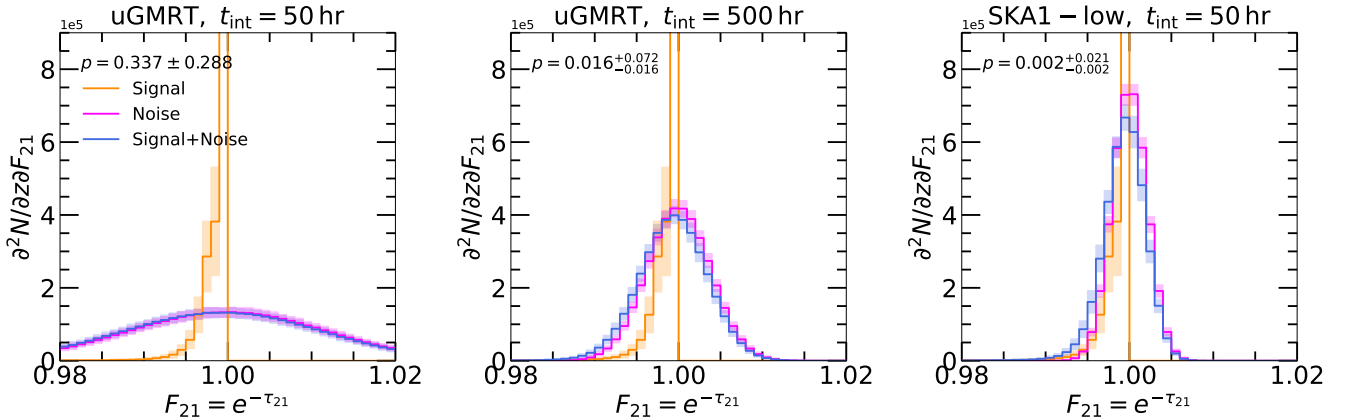


Figure 5. Differential number density distribution of 21-cm forest transmission by frequency channel for a spectrum containing purely the signal (orange curves), purely the noise (fuchsia curves) and an observed spectrum (signal and noise, blue curves). The solid lines correspond to the median values while the shaded regions indicate the 68 per cent range computed from 1000 mock spectra. Spectra were modelled for an observation of a background source with $S_{147} = 64.2$ mJy and $\alpha_R = -0.44$ by the uGMRT over $t_{\text{int}} = 50$ hr (left panel) and 500 hr (middle panel) and by the SKA1-low over 50 hr (right panel). A two-sided Kolmogorov-Smirnov test p -value for the observed and noise distribution is shown too.

LOS, similarly to the previous case, such frequency channels are > 20 times rarer than the bulk of the channels which has $\text{SNR} \approx 0.9$.

3 THE 21-CM FOREST STATISTICS

Motivated by the apparent difficulty of the direct detection of individual 21-cm forest absorption features in the spectra of even one of the brightest high- z radio-loud source we currently know we aim to investigate the prospects of a statistical detection of the 21-cm forest signal. In particular, we consider the differential number density of F_{21} in Sec. 3.1 and 1D power spectrum from the 21-cm forest spectrum in Sec. 3.2.

3.1 Differential number density distribution of the 21-cm forest transmission

Given that the background radio-source has a high brightness temperature, we expect the 21-cm forest signal always in absorption. Therefore, the frequency channel (or pixel) differential number density of the F_{21} of the spectrum containing purely just the signal will be at $F_{21} \leq 1$. This is shown in Fig. 5 where we show differential distributions of F_{21} by pixel of the 21-cm forest signal spectrum without (orange curves) and with the telescope noise included (blue curves)

and telescope noise only (without the signal, fuchsia curves). The solid curves are mean values over 1000 LOS and the shaded regions correspond to the 68 percent scatter around the mean. Note that this is different from the distributions shown in studies like Shimabukuro et al. (2014) and Šoltinský et al. (2021) as they considered distribution of individual absorption features and we are interested in the distribution of the F_{21} at each frequency channel. One can see that the signal only distribution (orange curves) extends only to $F_{21} \leq 1$ as expected.

On the other hand, the telescope noise is modelled as a Gaussian white noise as described in Section 2.3. Hence, it is distributed as a normal distribution function with a mean at $F_{21} = 1$ in both absorption and emission (higher and lower than $F_{21} = 1$). If a measurement of a radio spectrum at relevant frequencies is obtained, one would expect the distribution of the real F_{21} (signal and noise, blue curves) to deviate from the noise only distribution at $F_{21} < 1$ if such spectrum contained the 21-cm forest signal. Therefore, a deviation from the normal distribution can be considered to be a statistical detection of the 21-cm forest signal.

To test if the observed spectrum (signal and noise) distribution is consistent with the noise distribution, we use a two-sided Kolmogorov-Smirnov test. If the Kolmogorov-Smirnov test p -value is more than 0.05, it is likely that the observed spectrum distribution comes from the same distribution as the noise F_{21} , and therefore this

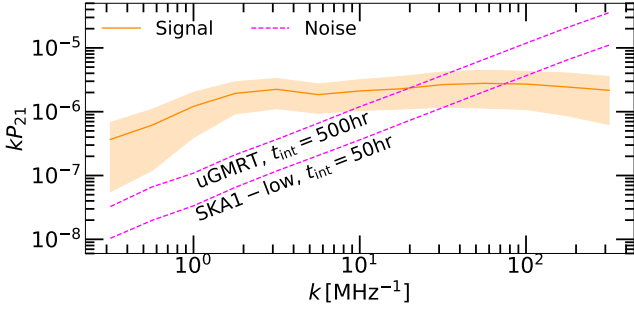


Figure 6. 1D power spectrum from a single 50 cMpc long 21-cm forest spectrum in the fiducial model of the IGM with $\log_{10} f_X = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$. The solid orange curve shows the median power spectrum of the signal and the orange shaded region marks 68 per cent range from 1000 mock spectra. The dashed fuchsia curve shows the power spectrum of the noise assuming an observation targeting $S_{147} = 64.2$ mJy and $\alpha_R = -0.44$ quasar with uGMRT over $t_{\text{int}} = 500$ hr (top curve) and SKA1-low over $t_{\text{int}} = 50$ hr.

would be considered a non-detection. An example of this is shown in the left panel of Fig. 5, where we considered an observation by the uGMRT over 50 hr. The mean $p = 0.337$ and $\approx 85\%$ of the 1000 synthetic spectra have $p > 0.05$. We infer that with such observational setup it is unlikely to achieve a detection this way. However, if the uGMRT observations are extended 10-fold in time (i.e. $t_{\text{int}} = 500$ hr), less than 10% of the 21-cm forest spectra results in a non-detection (the middle panel). This is further decreased to only few per cents of the spectra in the case of $t_{\text{int}} = 50$ hr with the SKA1-low. Not only that, but in the right panel of Fig. 5 it is clearly visible how the signal+noise curve deviates from the noise only distribution.

3.2 The 21-cm forest power spectrum

In this section we turn to the computation of the 1D power spectrum of the transmitted flux using the estimator $\delta_F = F_{21} - 1 = e^{-\tau_{21}} - 1$. This 1D 21-cm forest power spectrum, P_{21} , is defined as

$$P_{21}(k) \delta_D(k - k') = \langle \tilde{\delta}_F(k) \tilde{\delta}_F^*(k') \rangle, \quad (4)$$

where k is the wavenumber and δ_D is the Dirac delta function. We use discrete Fourier transforms of our estimator, $\tilde{\delta}_F$, by following Khan et al. (2024) (see their sec. 2 for more details). In particular, we estimate the P_{21} in Eq. 4 with

$$P_{21}(k_q) = \left(\frac{2\pi}{n\Delta\nu} \right) \langle |\tilde{\delta}_F(k_q)|^2 \rangle, \quad (5)$$

where $k_q = 2\pi q/n\Delta\nu$ for $q = 0, 1, \dots, n-1$ and the total number of pixels in the 21-cm forest spectrum is n .

We show an example of kP_{21} in Fig. 6 for our fiducial model (i.e. $\log_{10} f_X = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$). The solid orange curve and shaded region represent the median and the 68 per cent range of the kP_{21} from 1000 LOS. Throughout the manuscript we have binned the kP_{21} over the range of $10^{-0.5} \text{ MHz}^{-1} \leq k \leq 10^{2.5} \text{ MHz}^{-1}$ in the logarithmic k -bins of size 0.25 of dex, unless stated otherwise. The shape of the P_{21} broadly agrees with the literature (Thyagarajan 2020; Shao et al. 2023; Sun et al. 2024), particularly kP_{21} increasing with k at large scales and flattening at $k \gtrsim 2 \text{ MHz}^{-1}$. When comparing with previous works note that we present our results in terms of kP_{21} instead of P_{21} . The dashed fuchsia curves mark the noise limit for the same observational setups as in Fig. 4. The noise limits increase almost linearly with k as expected from a Gaussian white noise. The slight deviation from the linear behaviour reflects the frequency dependent sensitivity of the telescope and flux normalization due to

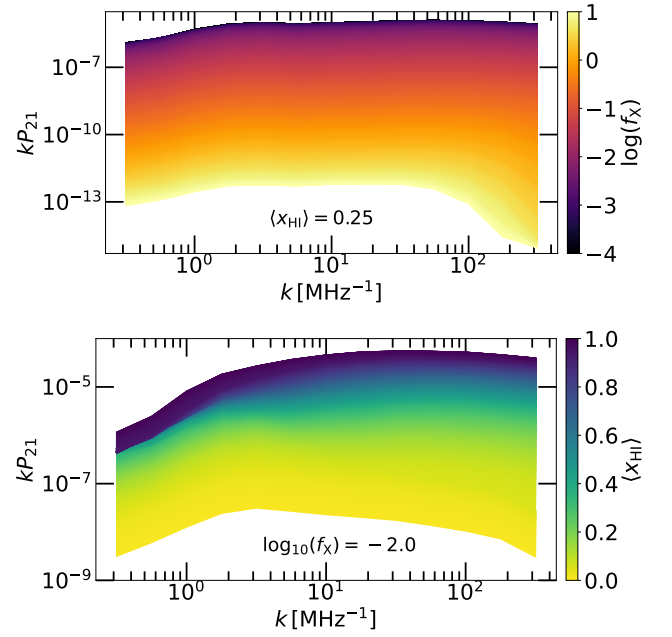


Figure 7. Top panel: Median 21-cm forest power spectrum, P_{21} from 1000 realizations of LOS for fixed $\langle x_{\text{HI}} \rangle = 0.25$ and varying $\log_{10} f_X$ (values shown in colorbar). Bottom panel: Same as the top panel but with fixed $\log_{10} f_X = -2$ and varying $\langle x_{\text{HI}} \rangle$.

the intrinsic spectrum of the background source described in Sec. 2.3. One can clearly see that the signal rises above these noise limits at large scales. Note that the noise limit for the SKA1-low is lower even if we assume 10 times shorter observational time.

Furthermore, we have tested the effect of the simulation resolution on our results. We find that the kP_{21} is boosted by up to 9% at $k \approx 6 \text{ MHz}^{-1}$ when the simulation resolution is increased to 512^3 pixels. However, this difference is significantly less than the uncertainty arising from the sample variance. More details can be found in Appendix A.

Besides the effect of instrumental features and simulation resolution on the detectability of the 21-cm forest power spectrum, we also explore the effect of various physical mechanisms. The position of individual absorption features in the 21-cm forest in the redshift space can be shifted and their depth can be significantly boosted by the peculiar velocity of the gas (Semelin 2016; Šoltínský et al. 2021). Both of these can potentially affect the P_{21} . We test this in Appendix B and find that while the P_{21} is not affected at $k \lesssim 20 \text{ MHz}^{-1}$ (for our fiducial model of $\log_{10} f_X = -2$, $\langle x_{\text{HI}} \rangle = 0.25$), at smaller scales the signal is boosted when we include v_{pec} in our calculation. However, the power spectrum that is affected by redshift space distortions significantly is either below the noise limit or even not accessible due to the limited spectral resolution.

To acquire the intuition of how the thermal and ionization state of the IGM affect the P_{21} , in Fig. 7 we plot the median values of the P_{21} from 1000 LOS with fixed value of one of the parameters governing the properties of the IGM while varying the other. In the top panel we fix $\langle x_{\text{HI}} \rangle = 0.25$ and the colormap shows the P_{21} for different $\log_{10} f_X$. As expected, the higher the f_X , and hence the more preheating of the IGM by X-rays, the more suppressed the signal becomes. Similarly, the signal is suppressed if the IGM is more ionized as can be seen in the bottom panel of Fig. 7, in which we fix the $\log_{10} f_X = -2$ and vary the $\langle x_{\text{HI}} \rangle$ (colormap). The shape

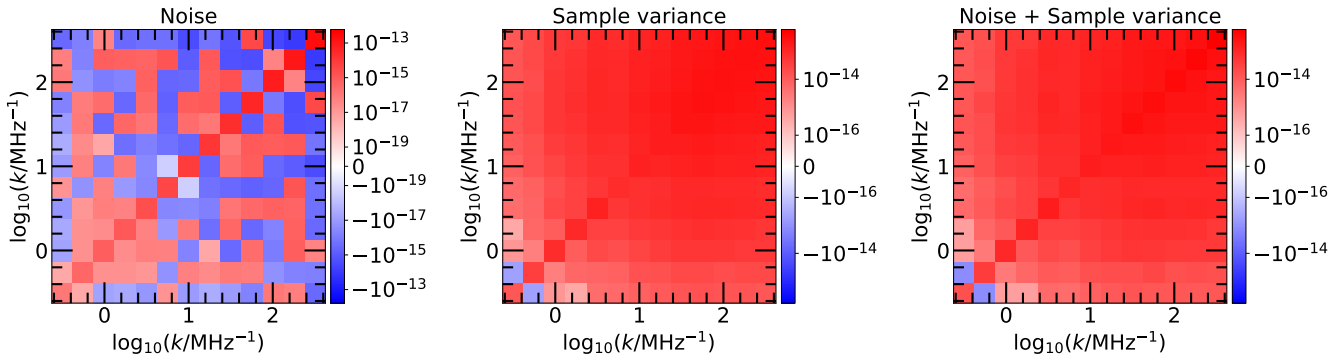


Figure 8. Covariance matrix for the noise only (left panel), sample variance only (middle panel) and their combined effect on the mock observation (right panel) assuming an IGM with $\langle x_{\text{HI}} \rangle = 0.25$ and $\log_{10} f_X = -2$ and $N_{\text{obs}} = 10$ observations of 200 cMpc spectra of a PSO J0309+27 like object (i.e. $S_{147} = 64.2$ mJy and $\alpha_{\text{R}} = -0.44$) at $z = 6$ with the uGMRT over $t_{\text{int}} = 500$ hr.

of the P_{21} as a function of k does not change significantly, however the slope changes at $k \gtrsim 50 \text{ MHz}^{-1}$ bin.

4 CONSTRAINING THE THERMAL AND IONIZATION STATE OF THE INTERGALACTIC MEDIUM

Given this strong sensitivity of the P_{21} on both the $\log_{10} f_X$ and the $\langle x_{\text{HI}} \rangle$, here we explore the potential of the P_{21} measurements to constrain both the thermal and ionization state of the IGM at $z = 6$. It is important to note that the 21-cm line is one of the few probes of the temperature of cold and neutral IGM regions. Similarly to Shao et al. (2023), who took the advantage of how differently the amplitude and shape of the 21-cm forest power spectrum⁷ vary with the efficiency of X-ray background radiation and the mass of warm dark matter particles to constrain these two quantities, we will explore the potential of the P_{21} to constrain the $\log_{10} f_X$ and the $\langle x_{\text{HI}} \rangle$. For this we will use Bayesian statistical methods described in Sec. 4.1. In Sec. 4.2 we discuss implications of our results.

4.1 Statistical inference procedure

In what follows we assume a measurement of 10 radio spectra of 22.1 MHz bandwidth each which corresponds to 200 cMpc long LOS at $z = 6$. As mentioned before, the uGMRT band-2 has a total bandwidth of 150 MHz, and therefore observation of longer spectra than considered here are possible. However, we are limited by the type of the simulation and its box size. We assume identical intrinsic spectra of the background sources for all 10 LOS, particularly $S_{147} = 64.2$ mJy and $\alpha_{\text{R}} = -0.44$. This corresponds to the PSO J0309+27 which is one of the brightest RLQSOs at $z > 5.5$ currently known as one can see in the bottom panel of Fig. 2. While this might be an optimistic observational requirement, we remind the reader that Niu et al. (2024) predicts more than 50 RLQSOs of $S_{147} = 100$ mJy at $z > 5.5$ in the whole sky. For this kind of observation we will define $\langle P_{21} \rangle_{10}$ which we construct by randomly selecting 10 synthetic 21-cm forest spectra, calculating their P_{21} and taking the mean at each k bin.

As explained above, our model is described by the X-ray background radiation efficiency and mean neutral hydrogen fraction.

⁷ Note that their power spectrum was computed from the 21-cm forest spectrum which was given in terms of differential brightness temperature as opposed to the transmitted flux as was done in this study.

Hence, let $\theta = \{\log_{10} f_X, \langle x_{\text{HI}} \rangle\}$. We assume uniform prior distributions, $\mathcal{P}(\theta)$, of the parameters. For the X-ray background radiation efficiency we have chosen a range of $\log_{10} f_X = [-4, 1]$. For the neutral hydrogen fraction we consider the whole physical range of $\langle x_{\text{HI}} \rangle = [0, 1]$ at $\log_{10} f_X < -0.6$ and $\langle x_{\text{HI}} \rangle = [0, 1 - 0.1125(\log_{10} f_X + 0.6)]$ at $\log_{10} f_X \geq -0.6$. The latter range is selected this way because in our simulations the IGM is at least partially ionized by the X-rays even if we set the background ionization efficiency to 0 as explained in Sec. 2.1. The highest $\langle x_{\text{HI}} \rangle$ in these models has slightly higher value than the one given by the relation for the upper bound on $\langle x_{\text{HI}} \rangle$.

To constrain these parameters, we follow Bayesian procedure with Gaussian likelihood defined as

$$\mathcal{L}(P_{21} | \theta) = \frac{1}{\sqrt{\det \mathbf{C}}} \exp\left(-\frac{1}{2} \mathbf{d}^T \mathbf{C}^{-1} \mathbf{d}\right), \quad (6)$$

where \mathbf{C} is the covariance matrix which is constructed from an ensemble of $10^4 \langle P_{21} \rangle_{10}$. Firstly, we consider the effect of the telescope noise by computing \mathbf{C} from the noisy radio spectra without the 21-cm forest signal. Such covariance matrix, for an observation by the uGMRT over $t_{\text{int}} = 500$ hr per source pointing at quasars with $S_{147} = 64.2$ mJy and $\alpha_{\text{R}} = -0.44$ (this corresponds to the upper dashed fuchsia line in Fig. 6), is shown in the left panel of Fig. 8. This shows a relatively weak (anti-)correlation between various k bins. In the middle panel of the same figure we show \mathbf{C} calculated from an ensemble of $\langle P_{21} \rangle_{10}$ incorporating only the 21-cm forest signal (not incorporating the telescope noise), which we define as $\langle P_{21}^{\text{sim}} \rangle_{10}(\theta)$, i.e. simulated/model $\langle P_{21} \rangle_{10}$. This is based on our fiducial IGM model with $\theta = [-2, 0.25]$. This panel shows the effect of the sample variance which is orders of magnitude stronger than the effect of the telescope noise.

To construct a mock observation, $\langle P_{21}^{\text{mock}} \rangle_{10}$, we need to incorporate both the effect of the telescope noise and sample variance. We do so by using 10 randomly selected synthetic 21-cm forest spectra including instrumental features (e.g., black curves in Fig. 4). An example of a single $\langle P_{21}^{\text{mock}} \rangle_{10}$ is presented as orange points in Fig. 9 considering the same observational setup and the same IGM model as above. The errorbars indicate 1σ uncertainty computed from an ensemble of $10^4 \langle P_{21}^{\text{mock}} \rangle_{10}$ (hence, incorporating uncertainty from both the telescope noise and sample variance). We compute the \mathbf{C} shown in the right panel of Fig. 8 from this ensemble of $10^4 \langle P_{21}^{\text{mock}} \rangle_{10}$. This is very similar to the middle panel, and hence we infer that the IGM parameter inference from the P_{21} at $z \approx 6$ will be dominated by the sample variance rather than the telescope noise.

The matrix $\mathbf{d} = \langle P_{21}^{\text{mock}} \rangle_{10} - \langle P_{21}^{\text{sim}} \rangle_{10}(\theta)$. With all of the above in

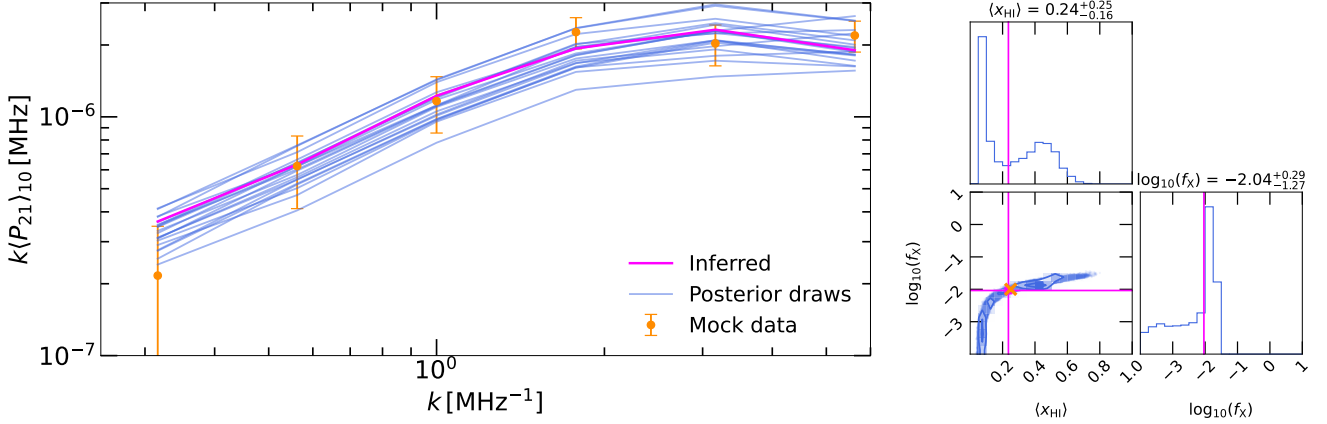


Figure 9. Bayesian inference of $\log_{10}f_X$ and $\langle x_{\text{HI}} \rangle$ from a mock observation of $S_{147} = 64.2$ mJy and $\alpha_R = -0.44$ quasar by uGMRT for 500 hr. The true values of the IGM parameters are $\log_{10}f_X = -2$, $\langle x_{\text{HI}} \rangle = 0.25$. *Left panel:* The $\langle P_{21} \rangle_{10}$ calculated from a spectrum including signal and noise. The orange data points indicate a mock observation of the 1D 21-cm forest power spectrum, $k\langle P_{21} \rangle_{10}$. The blue curves correspond to the $k\langle P_{21} \rangle_{10}$ randomly drawn from posterior acquired from the MCMC sampling. The fuchsia curve shows the $k\langle P_{21} \rangle_{10}$ for the inferred values of $\log_{10}f_X = -2.04$, $\langle x_{\text{HI}} \rangle = 0.23$. *Right panels:* Corner plot of the posterior distributions computed from the MCMC sampling (blue shaded region and blue curves). The true model parameters are indicated by the orange cross. Our analyses jointly recovers both parameters values of $\log_{10}f_X = -2.03^{+0.25}_{-1.19}$ and $\langle x_{\text{HI}} \rangle = 0.24^{+0.20}_{-0.17}$ as indicated by the fuchsia square and lines.

place, we can utilize the Bayes theorem

$$P(\theta|P_{21}^{\text{mock}}) \propto \mathcal{L}(P_{21}^{\text{mock}}|\theta)\mathcal{P}(\theta), \quad (7)$$

where $P(\theta|P_{21}^{\text{mock}})$ is the posterior probability distribution of the parameter values θ given the mock observation $\langle P_{21}^{\text{mock}} \rangle_{10}$. We implement the MCMC (Goodman & Weare 2010) using the EMCEE software (Foreman-Mackey et al. 2013) to sample the $\langle P_{21}^{\text{sim}} \rangle_{10}(\theta)$ ⁸. Therefore, we can neglect the normalization. Note that the $\langle P_{21}^{\text{sim}} \rangle_{10}(\theta)$ is calculated at any θ using linear interpolation depending on the MCMC sampling as long as the θ is within the ranges defining the $\mathcal{P}(\theta)$.

4.2 Results

Here we implement the procedure described in Sec. 4.1 to explore the potential of the P_{21} to constrain both the thermal and ionization state of the IGM. However, before we discuss the results of this procedure, we note that while the sample variance is the dominating effect in the parameter inference, the telescope noise also contributes to increasing the uncertainty on the inferred values of the IGM properties. Therefore, in our analyses we neglect the $k > 8.5$ MHz⁻¹ bins at which, as one can see in Fig. 6, the fiducial IGM model P_{21} is below or close to the noise limit. This significantly improves the parameter inference. After inspecting the P_{21} behaviour at low k , we suspect that adding bins with $k < 0.32$ MHz⁻¹ might improve our analyses. However, longer 21-cm forest spectra are required to do so.

The results of our Bayesian inference procedure are presented in Fig. 9 where we used $\langle P_{21}^{\text{mock}} \rangle_{10}$ assuming $\theta = [-2, 0.25]$ and an $t_{\text{int}} = 500$ hr per source observation by uGMRT. In the left panel, the blue curves represent 20 randomly drawn $\langle P_{21}^{\text{sim}} \rangle_{10}$ from the posterior distribution from the MCMC. They fit the errorbars of the mock observation (orange points) reasonably well. Our analyses infers the $\langle P_{21}^{\text{sim}} \rangle_{10}$ shown as the fuchsia curve. This corresponds to

⁸ We run 64 Markov chains over 100000 steps for each parameter. First 500 steps are disregarded given that the autocorrelation time is < 350 . The runs are initiated at the parameters initial guesses given by maximising the $\mathcal{L}(P_{21}^{\text{mock}}|\theta)$.

the inferred $\log_{10}f_X = -2.03^{+0.25}_{-1.19}$ and $\langle x_{\text{HI}} \rangle = 0.24^{+0.20}_{-0.17}$ indicated with the fuchsia square and horizontal and vertical lines in the corner plot on the right in Fig. 9. The true values of the IGM parameters (orange cross in the corner plot) lie very close to the mean inferred values. However, only upper limit on $\log_{10}f_X$ and lower limit on $\langle x_{\text{HI}} \rangle$ is acquired, as shown in the marginalized posterior distributions. Note that the P_{21} does not change significantly when one decreases $\log_{10}f_X$ below ≈ -3 which is shown in the top panel of Fig. 7. This range is smaller than the errorbars on the mock $\langle P_{21}^{\text{mock}} \rangle_{10}$ and as a consequence it is difficult for our analyses to distinguish between different $\log_{10}f_X$ in this parameter space. This results in a vertical tail of the posterior distribution in Fig. 9.

The constraining power of the P_{21} mock measurement with the same observational setup as shown in Fig. 9 over the whole considered parameter space is summarized in Fig. 10. We show the 1 and 2 σ credible intervals as the dark and light shaded regions with the crosses of the same colour indicating the true values used to generate mock observations. The more the IGM is neutral the tighter constraints can be acquired, as expected. Moreover, while colder IGM (i.e. lower $\log_{10}f_X$) would lead to tighter constraints, if the $\log_{10}f_X$ becomes too low, the credible intervals become elongated towards even lower values due to the diminutive changes in the P_{21} described above. However, if the $\log_{10}f_X$ is low, one can infer tighter constraints on the ionization state of the IGM.

On the other hand, if the IGM has experienced a significant heating by the X-ray background radiation and/or is significantly ionized such as in the case of the blue and teal crosses ($\log_{10}f_X = -1$ and -0.6 , and $\langle x_{\text{HI}} \rangle = 0.25$ and 0.79 , respectively), the telescope noise dominates over the signal and the $\langle P_{21}^{\text{mock}} \rangle_{10}$ is boosted significantly. Given that the $\langle P_{21}^{\text{sim}} \rangle_{10}(\theta)$ does not contain this boost from the telescope noise, the MCMC procedure is forced to find the best fitting $\langle P_{21}^{\text{sim}} \rangle_{10}(\theta)$ from stronger signal, and hence from lower $\log_{10}f_X$ and/or higher $\langle x_{\text{HI}} \rangle$. This is reflected in some of the posterior distributions deviating from the true parameter values of θ . Prime examples of this are the mock observations from the model of $\log_{10}f_X = -1$, $\langle x_{\text{HI}} \rangle = 0.25$ (blue cross and shaded region) and $\log_{10}f_X = -0.6$, $\langle x_{\text{HI}} \rangle = 0.79$ (teal cross and shaded region).

Naturally, in the case of telescope noise dominating over the signal one would consider this to be a null-detection of the signal. We define

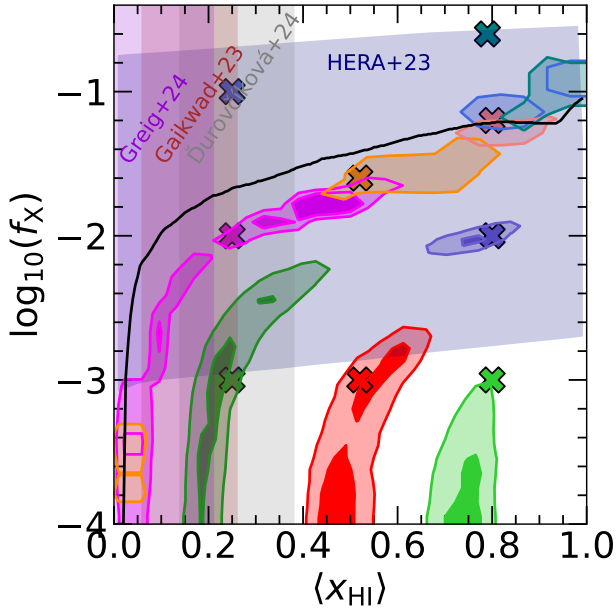


Figure 10. Posterior distributions corresponding to 1 and 2σ confidence regions indicated by contours for different true values (different colours) of the $\log_{10}f_X$ and $\langle x_{\text{HI}} \rangle$ indicated by the crosses. This shows the constraining power of the 21-cm forest power spectrum for an observation of $S_{147} = 64.2$ mJy and $\alpha_{\text{R}} = -0.44$ quasar by uGMRT for 500 hr. For comparison, Ly α forest measurements of the $\langle x_{\text{HI}} \rangle$ at $z \approx 6$, particularly $\langle x_{\text{HI}} \rangle = 0.17^{+0.09}_{-0.11}$ (brown shaded region, Gaikwad et al. 2023), $\langle x_{\text{HI}} \rangle = 0.21^{+0.17}_{-0.07}$ (grey shaded region, Đurovičková et al. 2024) and $\langle x_{\text{HI}} \rangle < 0.21$ (purple shaded region, Greig et al. 2024) is shown. The dark blue shaded region encompasses the $\log_{10}f_X$ which in our modelling corresponds to the spin temperature range inferred from the 21-cm power spectrum limits from HERA Collaboration (2023). The solid black curve indicates the disfavoured parameter space (lower $\log_{10}f_X$, higher $\langle x_{\text{HI}} \rangle$) in the case of a null-detection. Note that in the cases outside of this region the telescope noise dominates over the signal and as a consequence the posterior distribution deviates away from the true value.

a null-detection of the $\langle P_{21} \rangle_{10}$ as the case in which the $\langle P_{21}^{\text{sim}} \rangle_{10}$ in at least one of the k -bins at $< 8.5 \text{ MHz}^{-1}$ is below the 1D power spectrum of the noise only. However, similarly to Šoltinský et al. (2021) we suggest that a null-detection could be translated into joint lower limit on the $\log_{10}f_X$ and upper limit on the $\langle x_{\text{HI}} \rangle$. We show the 1σ limit on these parameters as the solid black curve in Fig. 10. To acquire this curve we generate an ensemble of $10^4 \langle P_{21}^{\text{sim}} \rangle_{10}(\theta)$ and find θ at which 32% of them result in a null-detection according to the definition described above. In other words, if the observation of 10 RLQSO by the uGMRT over 500 hr each results in a null-detection, one can disfavour the parameter space below this solid black curve. When comparing with the measurements of $\langle x_{\text{HI}} \rangle$ from the Ly α forest observations, particularly Gaikwad et al. (2023) (brown shaded vertical region), Greig et al. (2024) (purple shaded vertical region) and Đurovičková et al. (2024) (grey shaded vertical region), one can see that the $\langle P_{21} \rangle_{10}$ has a potential to constrain the $\langle x_{\text{HI}} \rangle$ even more than currently available measurements. This is even more pronounced in the case of f_X . The $\log_{10}f_X$ corresponding to $15.6 \text{ K} \leq T_{\text{K}} \leq 656.7 \text{ K}$, which is the spin temperature range inferred from the 21-cm power spectrum measurements by HERA Collaboration (2023), in our modelling is shown as a dark blue shaded region. It is obvious that according to the current observations the f_X is allowed

to range over few orders of magnitude which could be significantly constrained with the $\langle P_{21} \rangle_{10}$ observations.

We repeat the same analyses for a mock observation by the SKA1-low with $t_{\text{int}} = 50$ hr per source and observing quasars with the same properties at $z = 6$ ($S_{\text{min}} = 64.2$ mJy, $\alpha_{\text{R}} = -0.44$). We show the results in Fig. 11 where we again assumed a mock observation of the IGM with $\log_{10}f_X = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$ (orange points in the left panel). One can see in the corner plot on the right that the posterior distribution does not differ much from the case of the observation by the uGMRT and $t_{\text{int}} = 500$ hr per source. This is most likely caused by the sample variance effect dominating over the telescope noise as explained in Section 4.1. Moreover, the inferred values, $\log_{10}f_X = -2.22^{+0.36}_{-1.03}$ and $\langle x_{\text{HI}} \rangle = 0.15^{+0.18}_{-0.08}$, indicated by fuchsia deviate significantly more from the true values (orange cross) than in the previous case. This is caused by the $k = 0.562 \text{ MHz}^{-1}$ bin which is difficult to reproduce with the $\langle P_{21}^{\text{sim}} \rangle_{10}$ while fitting the other k bins. However, excluding the lowest k bin does not improve the constraining power of the $\langle P_{21} \rangle_{10}$ on θ . On the other hand, the constraints on the IGM parameters are slightly tighter. This is also shown in Fig. 12 in which the posterior distributions can encompass the true values at higher $\log_{10}f_X$ relatively to a case of the uGMRT observational setup considered previously. In addition, if the SKA1-low measurements of the $\langle P_{21} \rangle_{10}$ will result in a null-detection, it will lead to a tighter lower limit on the thermal state of the IGM at $z = 6$ even with 10 times shorter observational time. For instance, at $\langle x_{\text{HI}} \rangle = 0.25$ and 0.50 the lower limit on the X-ray background radiation efficiency would increase to $\log_{10}f_X \gtrsim -1.42$ and $\gtrsim -1.20$ for the SKA1-low, 50 hr per source observation, respectively, as compared to the $\log_{10}f_X \gtrsim -1.67$ and $\gtrsim -1.43$ in the case of the observation by the uGMRT, 500 hr per source.

Finally, we consider an observation by the uGMRT over $t_{\text{int}} = 50$ hr per source with the aim to test what is possible with currently operational instruments and is relatively not expensive. In this case, the telescope noise is 10 times larger (in terms of kP_{21}) than in the case of 500 hr per source with the uGMRT. Consequently, the MCMC results in posterior distributions that include the true values (crosses) within 2σ only in the cases of very cold and significantly neutral IGM, as can be seen in Fig. 13. Furthermore, the disfavoured parameter space with a null-detection utilizing such observations is much smaller. Despite this, the limits on the ionization state of the gas for a very cold IGM can reach $\langle x_{\text{HI}} \rangle \lesssim 0.18$ which is tighter than the currently available constraints on the $\langle x_{\text{HI}} \rangle$ from the Ly α forest, particularly from Gaikwad et al. (2023), Greig et al. (2024) and Đurovičková et al. (2024). The limits of $\log_{10}f_X \gtrsim -2.68$ and $\gtrsim -2.07$ if $\langle x_{\text{HI}} \rangle = 0.25$ and 0.50 , respectively. Also, note that we show only 1σ and 2σ credible intervals in Fig. 10, 12 and 13. We have checked that the full posterior distributions encapsulate their corresponding true values as long as they are within the solid black curve (even for the case of $\log_{10}f_X = -3.2$, $\langle x_{\text{HI}} \rangle = 0.25$ (dark green cross) and $\log_{10}f_X = -2$, $\langle x_{\text{HI}} \rangle = 0.8$ (purple cross)).

5 CONCLUSIONS

In this study we have explored the detectability of the 21-cm forest signal with focus on currently available options, namely targeting already confirmed radio-loud quasars at $z \sim 6$ with the uGMRT, but we also made forecasts for the SKA1-low telescope. We forward-modelled the 21-cm forest spectra using semi-numerical simulations of the IGM and incorporated several instrumental features. By considering a direct and statistical detection of the 21-cm forest signal

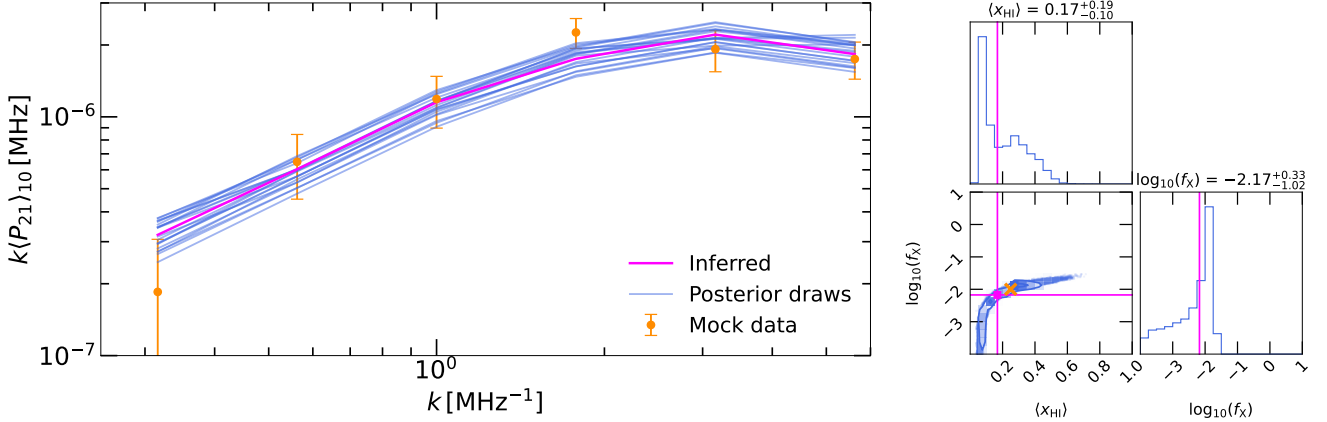


Figure 11. Same as Fig. 9 but for a mock observation by SKA1-low ant $t_{\text{int}} = 50$ hr. The jointly recovered parameters values are $\log_{10} f_X = -2.22^{+0.36}_{-1.03}$ and $\langle x_{\text{HI}} \rangle = 0.15^{+0.18}_{-0.08}$.

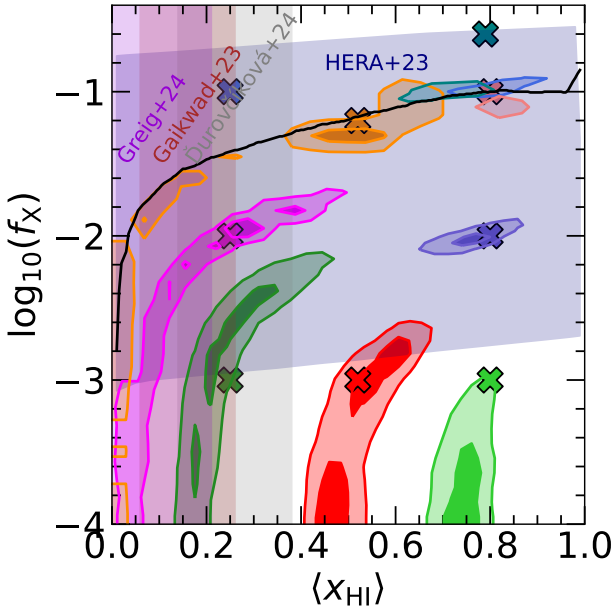


Figure 12. Same as Fig. 10 but for a mock observation by SKA1-low over 50 hr.

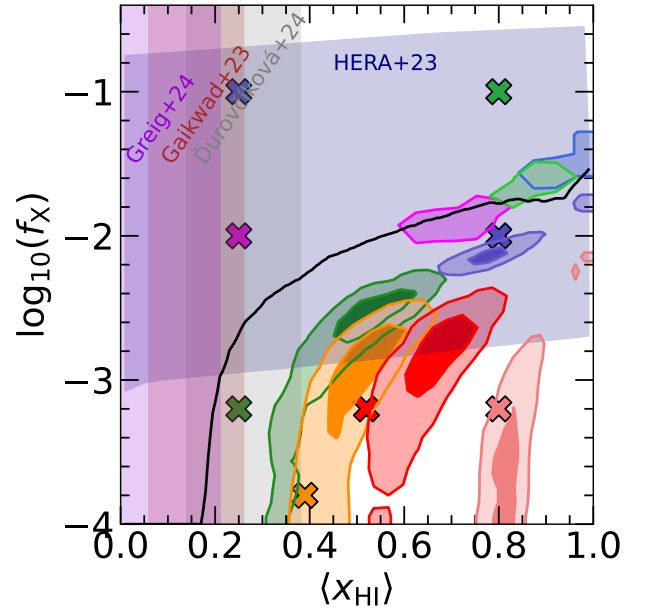


Figure 13. Same as Fig. 10 and 12 but for a mock observation by uGMRT over 50 hr.

and utilizing Bayesian techniques to constrain the thermal and ionization state of the IGM we found the following:

- In the case of late-end reionization and only modest IGM pre-heating by the X-ray background radiation, particularly $\langle x_{\text{HI}} \rangle = 0.25$ and $\log_{10} f_X = -2$, the SNR achieved by the uGMRT (SKA1-low) observing for 500 hr (50 hr) is 0.37 (1.2). Therefore, rather than aiming for a direct detection of individual absorption features we suggest that a statistical detection of the 21-cm forest is more plausible.

- One of the statistics we considered in this study is the differential number density of the 21-cm forest flux at each frequency channel (or pixel). We propose that a discrepancy between the distribution of the F_{21} of observed spectra and the noise distribution (normal distribution in our case), particularly at $F_{21} < 1$, can be considered as a statistical detection of the signal. Considering the same IGM model as in the previous point and using two-sided Kolmogorov-Smirnov test, we have found that while such detection was achieved only in $\approx 15\%$ of our synthetic spectra when observed by the uGMRT

over 50 hr, the fraction of detections increases greatly when either increasing the observational time by a factor of 10 or by using SKA1-low instead.

- Another promising strategy to acquire a detection of the 21-cm forest is the 1D power spectrum calculated from the normalized 21-cm forest flux, P_{21} . As can be seen in Fig. 6, the signal is above the noise limit at scales of $k \lesssim 8.5 \text{ MHz}^{-1}$ ($k \lesssim 32.4 \text{ MHz}^{-1}$) if measured by the uGMRT and $t_{\text{int}} = 500$ hr (SKA1-low and $t_{\text{int}} = 50$ hr) if $\langle x_{\text{HI}} \rangle = 0.25$ and $\log_{10} f_X = -2$, and hence should be detectable. In our modelling this corresponds to $T_{\text{HI}} \approx 30$ K.

- Making use of MCMC methods and exploring a wide range of parameter space, we have shown that a measurement of 10 P_{21} of 22.1 MHz bandwidth can be used to constrain both the thermal (i.e. $\log_{10} f_X$) and ionization (i.e. $\langle x_{\text{HI}} \rangle$) state of the IGM at the same time. While the uncertainty arising from the telescope noise plays an important role in our parameter inference, the sample variance is the dominating effect. Therefore, observations by the SKA1-low

over 50 hr per source do not significantly improve the constraining power of the P_{21} relatively to the uGMRT 500 hr per source observations. Despite this, the SKA1-low should still have a better chance of detecting the signal.

- On the other hand, we suggest that a null-detection of the P_{21} may provide interesting constraints on the properties of the high- z IGM by disfavouring parts of the parameter space. Fig. 10 and 12, particularly the solid black curve, indicate that a null-detection of the P_{21} may be utilized to: a) put a tighter upper limits on the $\langle x_{\text{HI}} \rangle$ than are currently available from the Ly α forest observations and b) put a lower limit on $\log_{10} f_X$ which is not yet constrained well. For example, a null-detection from the observation of 10 RLQSO 21-cm forest power spectra over 500 hr each by the uGMRT can disfavour $T_{\text{HI}} \lesssim 54$ K at $\langle x_{\text{HI}} \rangle \approx 0.25$, $T_{\text{HI}} \lesssim 80$ K at $\langle x_{\text{HI}} \rangle \approx 0.50$ and $T_{\text{HI}} \lesssim 104$ K at $\langle x_{\text{HI}} \rangle \approx 0.90$. With 50 hr of observations by the SKA1-low, one can disfavour somewhat larger parameter space than in the previous case, particularly $T_{\text{HI}} \lesssim 94$ K at $\langle x_{\text{HI}} \rangle \approx 0.25$, $T_{\text{HI}} \lesssim 141$ K at $\langle x_{\text{HI}} \rangle \approx 0.50$ and $T_{\text{HI}} \lesssim 181$ K at $\langle x_{\text{HI}} \rangle \approx 0.90$.

- Furthermore, even with $t_{\text{int}} = 50$ hr of uGMRT time on 10 RLQSO, one can already start disfavouring models with a relatively high neutral hydrogen fraction, e.g. $\langle x_{\text{HI}} \rangle \gtrsim 0.18$ for a very cold IGM, and a very low X-ray preheating of the IGM. Specifically, if such observation would result in a null-detection, a parameter space of $T_{\text{HI}} \lesssim 17$ K at $\langle x_{\text{HI}} \rangle \approx 0.25$, $T_{\text{HI}} \lesssim 26$ K at $\langle x_{\text{HI}} \rangle \approx 0.50$ and $T_{\text{HI}} \lesssim 34$ K at $\langle x_{\text{HI}} \rangle \approx 0.90$ can be disfavoured. This would be an important improvement in our understanding of the high- z IGM for a modest observational time requirement.

Note that we do not consider the contribution of minihaloes to the 21-cm forest absorption, however their significance in this matter is yet to be determined. Solving this is hindered by challenges of resolving such small scale structures in cosmological simulations and the fact that their contribution may be greatly reduced by feedback processes (Meiksin 2011; Park et al. 2016; Nakatani et al. 2020; Chan et al. 2024). In addition, even structures within the minihaloes, subhaloes, may enhance the 21-cm forest absorption (Kadota et al. 2023), but these are subject to tidal and/or ram pressure stripping (Naruse et al. 2024).

Given that the 21-cm forest is expected to be non-Gaussian, using a multivariate Gaussian likelihood as in Eq. 6 may be not an optimal method to extract information from the P_{21} (Wolfson et al. 2023). As alternatives one can utilize a different type of the likelihood or a likelihood-free techniques such as in Sun et al. (2024).

Furthermore, while we implement various effects that can affect the observability of the 21-cm forest signal, such as the intrinsic radio spectrum of the background quasar and instrumental features, there are other effects that need to be considered. For example, the radio frequency interference (RFI), can introduce both the narrow-band and broad-band features in the radio spectrum of the background source. These can be partially mitigated (e.g. parts of the 21-cm spectrum affected by the narrow-band RFI features can be filtered out), however a detailed study of these effects is beyond the scope of this manuscript. Another instrumental effect which is omitted in this study is the contamination caused by foreground objects in the observed field via mode-mixing between parallel and transverse k -modes (Thyagarajan 2020).

We note that there were attempts of detecting the 21-cm forest absorption in the past. For example, Carilli et al. (2007) measured a radio spectrum at frequencies where one would expect the 21-cm forest features of two radio-loud AGNs at $z = 5.11$ and $z = 5.2$ with the GMRT but they report no detection. This is not surprising because at these redshifts the IGM is already reionized and the signal is

completely suppressed. Furthermore, their spectra have bandwidths of $\Delta\nu < 5$ MHz. This means that they probe the IGM close enough to the host source that its radiation may heat up the gas even further (Šoltinský et al. 2023). In this manuscript we have modelled significantly longer spectra, particularly with $\Delta\nu = 22.1$ MHz. However, as one can see in Fig. 6, the kP_{21} from the signal does not fall faster than the noise limit at $k \leq 3.2 \text{ MHz}^{-1}$. From this we infer that longer spectra (lower k) may improve our Bayesian parameter inference. Note that for longer spectra the effect of redshift evolution along the LOS would be non-negligible. To capture this effect lightcone simulations would be more suitable than the simulations utilized here. The uGMRT band-2 spans to 250 MHz (Gupta et al. 2017) which could provide at least twice as long spectra for sources at $z \approx 6$ than assumed in our work. Unfortunately, at high enough frequencies we would start probing redshifts at which the reionization is already completed. This issue can be alleviated by observing higher redshift RLQSO. Note that there are now 3 RLQSO discovered at $z > 6.8$ (Bañados et al. 2021, 2024; Endsley et al. 2023). We leave this for the future work.

We believe that the forecasts in this study will aid in similar observational efforts to finally acquire the first detection of the 21-cm forest signal at high redshift. On the other hand, we suggest that a null-detection of the signal can be translated into lower limit on the f_X and the upper limit on the $\langle x_{\text{HI}} \rangle$. The former one can be instead changed to the lower limit on temperature of the neutral IGM at $z \gtrsim 6$, which so far cannot be traced by other observables. Therefore, both outcomes of such observations would be very useful for the high- z IGM studies.

Despite the caveats and challenges listed above, the prospects of achieving a statistical detection of the 21-cm forest are improving on theoretical (late-end reionization), observational (rising number of identified radio-loud quasars at high redshift) and instrumental (improved sensitivity of the uGMRT, forthcoming construction of the SKA) fronts. These improvements in the field in combination with the potential of constraining the IGM properties at $z \geq 6$ increase the importance of the 21-cm forest signal.

ACKNOWLEDGEMENTS

We thank Arvind Balasubramanian, Silvia Belladitta, Abhirup Datta, Frederick Davies, Dominika Ďurovčiková, Keith Grainge, Nissim Kanekar, Nabendu Kumar Khan, Shikhar Mittal and Jose Oñorbe for helpful discussions. Furthermore, we thank the anonymous referee for constructive comments which improved this manuscript. We appreciate the feedback during the Astronomical Society of India meeting 2024 (<https://astron-soc.in/asi2024/>), Cosmology in the Alps 2024 (<https://indico.skatelescope.org/event/1098/>), Cosmic Dawn at High Latitudes (<https://indico.fysik.su.se/event/8499/>), The First Gigayear(s) (<https://noirlab.edu/science/events/websites/first-gigayears-2024>) and SKA Cosmology SWG Meeting 2024 (<https://indico.ict.inaf.it/event/3054/>) conferences which also shaped this manuscript. TŠ and GK gratefully acknowledge support by the Max Planck Society via a partner group grant. TŠ and GK are also partly supported by the Department of Atomic Energy (Government of India) research project with Project Identification Number RTI 4002. TŠ also expresses his gratitude for being supported by the Istituto Nazionale di Astrofisica - Osservatorio Astronomico di Trieste (INAF-OATs) under the Theory grant - 'Cosmological Investigation of the Cosmic Web' (C93C23006820005). We note that this work is heavily based on the 21_{CMFAST} simulations (Mesinger et al. 2011; Murray et al. 2020). The authors also thank the developers

of *ASTROPY* (Astropy Collaboration et al. 2013), *CORNER* (Foreman-Mackey 2016), *EMCEE* (Foreman-Mackey et al. 2013), *MATPLOTLIB* (Hunter 2007), *NUMPY* (Harris et al. 2020) and *SCIPY* (Virtanen et al. 2020) packages which were implemented in the analyses in this work. This work has been performed as part of the DAE-STFC Technology and Skills Programme ‘Building Indo-UK collaborations towards the Square Kilometre Array’.

DATA AVAILABILITY

The codes used for our analyses and for generating data and plots are publicly available at https://github.com/tomassoltinsky/21cm-forest_IDPS. All data used in this study can be requested from the first author.

REFERENCES

- Astropy Collaboration et al., 2013, *A&A*, **558**, A33
- Bañados E., et al., 2015, *ApJ*, **804**, 118
- Bañados E., Carilli C., Walter F., Momjian E., Decarli R., Farina E. P., Mazzucchelli C., Venemans B. P., 2018, *ApJ*, **861**, L14
- Bañados E., et al., 2021, *ApJ*, **909**, 80
- Bañados E., et al., 2023, *ApJS*, **265**, 29
- Bañados E., et al., 2024, *Nature Astronomy*,
- Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, *MNRAS*, **447**, 3402
- Becker G. D., D’Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, *MNRAS*, **508**, 1853
- Becker G. D., Bolton J. S., Zhu Y., Hashemi S., 2024, *MNRAS*, **533**, 1525
- Belladitta S., et al., 2020, *A&A*, **635**, L7
- Bhagwat A., Ciardi B., Zackrisson E., Schaye J., 2022, *MNRAS*, **517**, 2331
- Bosman S. E. I., et al., 2022, *MNRAS*, **514**, 55
- Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, *Nature*, **555**, 67
- Braun R., Bonaldi A., Bourke T., Keane E., Wagg J., 2019, *arXiv e-prints*, p. [arXiv:1912.12699](https://arxiv.org/abs/1912.12699)
- Cain C., D’Aloisio A., Gangolli N., Becker G. D., 2021, *ApJ*, **917**, L37
- Carilli C. L., Gnedin N. Y., Owen F., 2002, *ApJ*, **577**, 22
- Carilli C. L., Wang R., van Hoven M. B., Dwarkanath K., Chengalur J. N., Wyithe S., 2007, *AJ*, **133**, 2841
- Chan T. K., Benítez-Llambay A., Theuns T., Frenk C., Bower R., 2024, *MNRAS*, **528**, 1296
- Choudhury T. R., Paranjape A., Bosman S. E. I., 2021, *MNRAS*, **501**, 5782
- Christenson H. M., Becker G. D., Furlanetto S. R., Davies F. B., Malkan M. A., Zhu Y., Boera E., Trapp A., 2021, *ApJ*, **923**, 87
- Ciardi B., et al., 2013, *MNRAS*, **428**, 1755
- Ciardi B., et al., 2015, *MNRAS*, **453**, 101
- Datta K. K., Bharadwaj S., Choudhury T. R., 2007, *MNRAS*, **382**, 809
- Đurovičková D., et al., 2024, *ApJ*, **969**, 162
- Endsley R., et al., 2023, *MNRAS*, **520**, 4609
- Ewall-Wice A., Dillon J. S., Mesinger A., Hewitt J., 2014, *MNRAS*, **441**, 2476
- Fan X., et al., 2001, *AJ*, **122**, 2833
- Foreman-Mackey D., 2016, *The Journal of Open Source Software*, **1**, 24
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, **125**, 306
- Furlanetto S. R., 2006, *MNRAS*, **371**, 867
- Furlanetto S. R., Loeb A., 2002, *ApJ*, **579**, 1
- Furlanetto S. R., Stoever S. J., 2010, *MNRAS*, **404**, 1869
- Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, **433**, 181
- Gaikwad P., et al., 2020, *MNRAS*, **494**, 5091
- Gaikwad P., et al., 2023, *MNRAS*, **525**, 4093
- Garaldi E., et al., 2024, *MNRAS*, **530**, 3765
- Gloude-mans A. J., et al., 2022, *A&A*, **668**, A27
- Gloude-mans A. J., et al., 2023, *A&A*, **678**, A128
- Goodman J., Weare J., 2010, *Communications in Applied Mathematics and Computational Science*, **5**, 65
- Greig B., Trott C. M., Barry N., Mutch S. J., Pindor B., Webster R. L., Wyithe J. S. B., 2021a, *MNRAS*, **500**, 5322
- Greig B., et al., 2021b, *MNRAS*, **501**, 1
- Greig B., et al., 2024, *MNRAS*, **530**, 3208
- Gupta Y., et al., 2017, *Current Science*, **113**, 707
- HERA Collaboration 2022, *ApJ*, **925**, 221
- HERA Collaboration 2023, *ApJ*, **945**, 124
- Harris C. R., et al., 2020, *Nature*, **585**, 357
- Hennawi J. F., Davies F. B., Wang F., Oñorbe J., 2021, *MNRAS*, **506**, 2963
- Hsyu T., Cooke R. J., Prochaska J. X., Bolte M., 2020, *ApJ*, **896**, 77
- Hunter J. D., 2007, *Computing in Science and Engineering*, **9**, 90
- Ighina L., Belladitta S., Caccianiga A., Broderick J. W., Drouart G., Moretti A., Seymour N., 2021, *A&A*, **647**, L11
- Ighina L., Caccianiga A., Moretti A., Belladitta S., Broderick J. W., Drouart G., Leung J. K., Seymour N., 2023, *MNRAS*, **519**, 2060
- Ighina L., et al., 2024, *A&A*, **692**, A241
- Intema H. T., Jagannathan P., Mooley K. P., Frail D. A., 2017, *A&A*, **598**, A78
- Ioka K., Mészáros P., 2005, *ApJ*, **619**, 684
- Kadota K., Villanueva-Domingo P., Ichiki K., Hasegawa K., Naruse G., 2023, *J. Cosmology Astropart. Phys.*, **2023**, 017
- Kashino D., Lilly S. J., Shibuya T., Ouchi M., Kashikawa N., 2020, *ApJ*, **888**, 6
- Kawasaki M., Nakano W., Nakatsuka H., Sonomoto E., 2021, *J. Cosmology Astropart. Phys.*, **2021**, 019
- Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, *MNRAS*, **491**, 1736
- Khan N. K., Kulkarni G., Bolton J. S., Haehnelt M. G., Iršič V., Puchwein E., Asthana S., 2024, *MNRAS*, **530**, 4920
- Kondapally R., et al., 2021, *A&A*, **648**, A3
- Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, *MNRAS*, **485**, L24
- Lewis J. S. W., et al., 2022, *MNRAS*, **516**, 3389
- Liu Y., et al., 2021, *ApJ*, **908**, 124
- Mack K. J., Wyithe J. S. B., 2012, *MNRAS*, **425**, 2988
- McGreer I. D., Becker R. H., Helfand D. J., White R. L., 2006, *ApJ*, **652**, 157
- McGreer I. D., Mesinger A., D’Odorico V., 2015, *MNRAS*, **447**, 499
- Meiksin A., 2011, *MNRAS*, **417**, 1480
- Mertens F. G., et al., 2020, *MNRAS*, **493**, 1662
- Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, **411**, 955
- Murray S., Greig B., Mesinger A., Muñoz J., Qin Y., Park J., Watkinson C., 2020, *The Journal of Open Source Software*, **5**, 2582
- Nakane M., et al., 2024, *ApJ*, **967**, 28
- Nakatani R., Fialkov A., Yoshida N., 2020, *ApJ*, **905**, 151
- Naruse G., Hasegawa K., Kadota K., Tashiro H., Ichiki K., 2024, *J. Cosmology Astropart. Phys.*, **2024**, 091
- Nasir F., D’Aloisio A., 2020, *MNRAS*, **494**, 3080
- Niu Q., Li Y., Xu Y., Guo H., Zhang X., 2024, *arXiv e-prints*, p. [arXiv:2407.18136](https://arxiv.org/abs/2407.18136)
- Oñorbe J., Davies F. B., Lukić Z., Hennawi J. F., Sorini D., 2019, *MNRAS*, **486**, 4075
- Offringa A. R., et al., 2015, *Publ. Astron. Soc. Australia*, **32**, e008
- Park H., Shapiro P. R., Choi J.-h., Yoshida N., Hirano S., Ahn K., 2016, *ApJ*, **831**, 86
- Planck Collaboration et al., 2014, *A&A*, **571**, A16
- Pritchard J. R., Loeb A., 2012, *Reports on Progress in Physics*, **75**, 086901
- Puchwein E., et al., 2023, *MNRAS*, **519**, 6162
- Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021, *MNRAS*, **506**, 2390
- Semelin B., 2016, *MNRAS*, **455**, 962
- Shao Y., Xu Y., Wang Y., Yang W., Li R., Zhang X., Chen X., 2023, *Nature Astronomy*, **7**, 1116
- Shao Y., Sun T.-Y., Zhao M.-L., Zhang X., 2024, *arXiv e-prints*, p. [arXiv:2411.17094](https://arxiv.org/abs/2411.17094)
- Shimabukuro H., Ichiki K., Inoue S., Yokoyama S., 2014, *Phys. Rev. D*, **90**, 083003
- Shimabukuro H., Ichiki K., Kadota K., 2020, *Phys. Rev. D*, **102**, 023522

- Shimabukuro H., Ichiki K., Kadota K., 2023, *Phys. Rev. D*, **107**, 123520
- Shimwell T. W., et al., 2017, *A&A*, **598**, A104
- Singh S., et al., 2022, *Nature Astronomy*, **6**, 607
- Smith D. J. B., et al., 2016, in Reylé C., Richard J., Cambrésy L., Deleuil M., Pécontal E., Tresse L., Vauglin I., eds, *SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*. pp 271–280 (arXiv:1611.02706)
- Spina B., Bosman S. E. I., Davies F. B., Gaikwad P., Zhu Y., 2024, *A&A*, **688**, L26
- Sun T.-Y., Shao Y., Li Y., Xu Y., Zhang X., 2024, *arXiv e-prints*, p. arXiv:2407.14298
- Šoltinský T., et al., 2021, *MNRAS*, **506**, 5818
- Šoltinský T., Bolton J. S., Molaro M., Hatch N., Haehnelt M. G., Keating L. C., Kulkarni G., Puchwein E., 2023, *MNRAS*, **519**, 3027
- Thyagarajan N., 2020, *ApJ*, **899**, 16
- Tie S. S., Hennawi J. F., Wang F., Onorato S., Yang J., Bañados E., Davies F. B., Oñorbe J., 2024, *MNRAS*, **535**, 223
- Trott C. M., et al., 2020, *MNRAS*, **493**, 4711
- Villanueva-Domingo P., Ichiki K., 2023, *PASJ*, **75**, S33
- Virtanen P., et al., 2020, *Nature Methods*, **17**, 261
- Wayth R. B., et al., 2015, *Publ. Astron. Soc. Australia*, **32**, e025
- Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, *MNRAS*, **485**, 1350
- Willott C. J., et al., 2010, *AJ*, **140**, 546
- Wolf J., et al., 2024, *A&A*, **691**, A30
- Wolfson M., Hennawi J. F., Davies F. B., Lukić Z., Oñorbe J., 2023, *arXiv e-prints*, p. arXiv:2309.05647
- Xu Y., Chen X., Fan Z., Trac H., Cen R., 2009, *ApJ*, **704**, 1396
- Xu Y., Ferrara A., Chen X., 2011, *MNRAS*, **410**, 2025
- Zeimann G. R., White R. L., Becker R. H., Hodge J. A., Stanford S. A., Richards G. T., 2011, *ApJ*, **736**, 57
- Zheng Q., Wu X.-P., Guo Q., Johnston-Hollitt M., Shan H., Duchesne S. W., Li W., 2020, *MNRAS*, **499**, 3434
- Zhu Y., et al., 2021, *ApJ*, **923**, 223
- Zhu Y., et al., 2022, *ApJ*, **932**, 76
- Zhu Y., et al., 2023, *ApJ*, **955**, 115
- Zhu Y., et al., 2024, *MNRAS*, **533**, L49

APPENDIX A: CONVERGENCE TEST OF THE IGM SIMULATION FOR THE 21-CM FOREST POWER SPECTRUM

We have performed a convergence test to investigate the effect of the simulation resolution on the 1D 21-cm forest power spectrum. This is presented in Fig. A1 where we change the resolution of the simulation by fixing the simulated volume at 50 cMpc and varying the number of pixels to 64^3 (green), 128^3 (blue), 256^3 (orange, used in this study) and 512^3 (fuchsia). The top panel shows the kP_{21} with the same k -binning as used in the paper while the bottom panel shows the residuals relative to the simulation used in the paper defined as $(P_{21,i} - P_{21,256^3})/P_{21,256^3}$, where $P_{21,i}$ is the 1D 21-cm forest power spectrum for a simulation with the number of pixels equal to i . The residuals indicate that our fiducial simulation underestimates the signal by up to 4% at $k \approx 1 \text{ MHz}^{-1}$ and up to 9% at $k \approx 6 \text{ MHz}^{-1}$ compared to the simulation with the highest resolution considered (i.e. 512^3 pixels, fuchsia). However, note that the discrepancy between the 512^3 and 256^3 simulations is significantly smaller than the uncertainty due to the sample variance (indicated by the shaded regions).

We also emphasize that this study focuses on taking the advantage of the advancements in the field which improve the prospects of detecting the 21-cm forest signal and exploration of what can be implied from the (null)-detection of this signal rather than the detailed modelling of the physics in the simulation. A more accurate modelling, such as the use of hydrodynamics/radiative

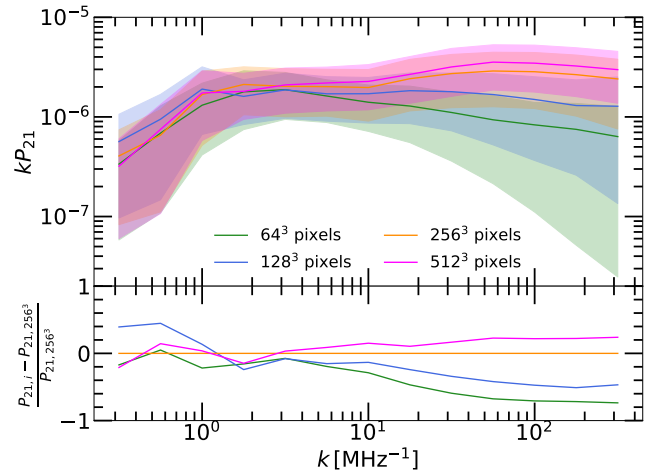


Figure A1. *Top:* The convergence test of the 21cmFAST simulations. Particularly, we show the mean 21-cm forest 1D power spectrum, kP_{21} (solid curves) and 68% scatter (shaded regions with corresponding colours) calculated from 1000 mock spectra for the fiducial model of the IGM (i.e. $\log_{10} f_X = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$). We compare simulations of 50 cMpc volume and vary the number of pixels to 64^3 (green), 128^3 (blue), 256^3 (orange, used in the paper) and 512^3 (fuchsia). *Bottom:* The fractional residuals computed relatively to the simulation used in the paper (i.e. 256^3 pixels, orange curve).

transfer simulations including NYX (Oñorbe et al. 2019), CoDA III (Lewis et al. 2022), SHERWOOD-RELICS (Puchwein et al. 2023) and THESAN (Garaldí et al. 2024) to name a few. Unfortunately, even high resolution simulations with similar volumes as considered in our study do not resolve minihaloes due to large dynamic range they are required to capture. This is one of the big challenges currently in the numerical cosmology community. Furthermore, such simulations would be impossible to run for as large parameter space as desired for our study due to computational costs. Hence, we decided to utilize the 21cmFAST instead. Running the 21cmFAST simulation with finer resolution would be computationally too expensive too and would not improve the analysis very much. In spite of this, we plan to use better simulations (more detailed physics, wider dynamic range) for such analysis in the future.

APPENDIX B: THE EFFECT OF REDSHIFT SPACE DISTORTIONS ON THE 21-CM FOREST POWER SPECTRUM

The redshift space distortions can affect the 21-cm forest signal by both shifting the absorption lines relative to the location of the absorbers from which they arise in redshift space and significantly boosting the optical depth of the 21-cm line absorption features (Semelin 2016; Šoltinský et al. 2021). In this section we test if the redshift space distortions have a substantial effect on the 21-cm forest power spectrum or if the effect of peculiar velocities of the gas can be neglected. In Fig. B1 we show the P_{21} for the case in which we neglect the peculiar velocity of the gas (i.e. $v_{\text{pec}} = 0$, blue) and if we include it in our computation (orange). The effect of redshift space distortions, if any, is expected to be more pronounced at small scales due to shifts of absorption features in the redshift space. Therefore, we deliberately do not smooth our synthetic 21-cm forest spectra such that we access the largest k values. One can clearly see in Fig. B1 that the redshift space distortions increase the P_{21} at $k \gtrsim 20 \text{ MHz}^{-1}$

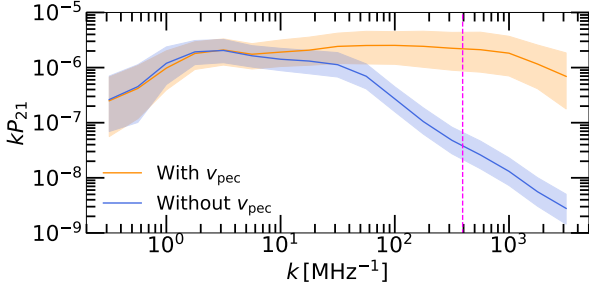


Figure B1. 21-cm forest 1D power spectrum including (orange curve) and neglecting (blue curve) peculiar velocities. The solid lines represent median values and the shaded regions mark 68 per cent range from 2000 mock spectra. The vertical dashed fuchsia line marks the highest k that is accessible by a telescope given a spectral resolution $\Delta\nu = 8$ kHz.

as they force the absorption features to cluster more. This might lead to misinterpretation of the observations, especially the dark matter constraints as described in [Shao et al. \(2023\)](#), who neglect the redshift space distortions, which come from the large k part of the P_{21} . In addition, the discrepancy between the $v_{\text{pec}} = 0$ and $v_{\text{pec}} \neq 0$ case extends to $k = 393 \text{ MHz}^{-1}$, corresponding to the spectral resolution of $\Delta\nu = 8 \text{ kHz}^{-1}$ assumed in this study (vertical dashed fuchsia line). Hence, neglecting v_{pec} would affect the analyses of observations like we describe in this work too.

APPENDIX C: SIGNAL-TO-NOISE RATIO OF THE 21-CM FOREST

To quantify the difficulty of a direct detection of individual absorption features of the 21-cm forest we compute the signal-to-noise ratio (SNR) for each frequency channel in the synthetic spectra. We define $\text{SNR}(\nu) = (1 - F_{21}(\nu))/\sigma_{\text{N, norm}}(\nu) = (1 - e^{-\tau_{21}(\nu)})/\sigma_{\text{N, norm}}(\nu)$, where $\sigma_{\text{N, norm}}$ is the noise rms computed from Eq. 3 normalized with the intrinsic spectrum of the background radio source. Note that the $\sigma_{\text{N, norm}}(\nu)$ depends on the frequency ν given the dependency of the telescope sensitivity on the ν and the normalization by the background source continuum.

In Fig. C1 we show the differential number density of the SNR per frequency channel for the IGM model of $\log_{10} f_X = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$. Firstly, we consider a single LOS, specifically the one shown in Fig. 4 (dashed curves). We see that the highest SNR reached using the observational setup of the uGMRT, $t_{\text{int}} = 500$ hr is 11.9 and for the SKA1-low, $t_{\text{int}} = 50$ hr it is 19.0. While the absorption features with $\text{SNR} \approx 10$ could be detectable, frequency channels with such SNR are ~ 20 times less abundant than the ones at the peak of the distribution which are at $\text{SNR} \approx 0.4$ and 0.9 for the uGMRT, $t_{\text{int}} = 500$ hr and the SKA1-low, $t_{\text{int}} = 50$ hr, respectively. There is a similar trend if we consider all 1000 simulated LOS in the same IGM model (solid curves). There are some LOS that reach SNR of 36.5 if observed by the uGMRT over 500 hr and 63.3 if observed by the SKA1-low over 50 hr. Unfortunately, these are still very rare. Hence, the direct detection of the individual absorption features of the 21-cm forest seems very difficult.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.

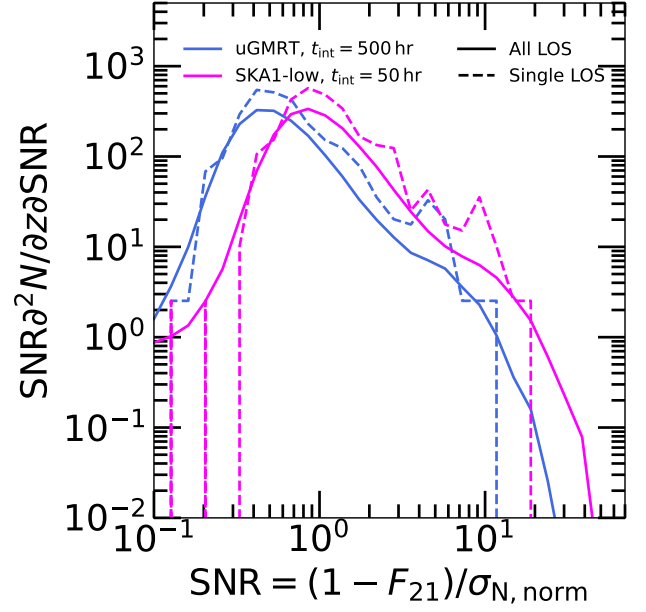


Figure C1. The differential number density of the SNR of the simulated 21-cm forest spectrum per frequency channel. The blue curves correspond to the noise assuming an observation by the uGMRT over 500 hr and the fuchsia curves correspond to the observation by the SKA1-low over 50 hr. In both cases we consider that the radio telescopes target a $z = 6$ quasar with the spectrum defined by $S_{147} = 64.2$ mJy and $\alpha_{\text{R}} = -0.44$ in the IGM with $\log_{10} f_X = -2$ and $\langle x_{\text{HI}} \rangle = 0.25$. The solid curves show distribution for 1000 synthetic LOS while the dashed curves are for a single LOS shown in Fig. 4.