Measuring the photo-ionization rate, neutral fraction and mean free path of HI ionizing photons at $4.9 \leq z \leq 6.0$ from a large sample of XShooter and ESI spectra

Prakash Gaikwad[1] *, Martin G. Haehnelt[2], Fredrick B. Davies[1], Sarah E. I. Bosman[1], Margherita Molaro[3], Girish Kulkarni[4], Valentina D'Odorico[5,6,7], George D. Becker[8], Rebecca L. Davies[9,10], Fahad Nasir[1], James S. Bolton[3], Laura C. Keating[11], Vid Iršič[2], Ewald Puchwein[12], Yongda Zhu[8], Shikhar Asthana[2], Jinyi Yang[13], Samuel Lai[14] and Anna-Christina Eilers[15]

[1] *Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[2] *Kavli Institute for Cosmology and Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA, UK*
[3] *School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, NG7 2RD, UK*
[4] *Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*
[5] *INAF-Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34143 Trieste, Italy,*
[6] *Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy,*
[7] *IFPU-Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy*
[8] *Department of Physics and Astronomy, University of California, Riverside, CA, 92521, USA*
[9] *Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia,*
[10] *ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*
[11] *Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK*
[12] *Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany*
[13] *Steward Observatory, University of Arizona, 933 N Cherry Avenue, Tucson, AZ 85721, USA*
[14] *Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia*
[15] *MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Ave., Cambridge, MA 02139, USA*

ABSTRACT

We measure the mean free path ($\lambda_{\mathrm{mfp,HI}}$), photo-ionization rate ($\langle\Gamma_{\mathrm{HI}}\rangle$) and neutral fraction ($\langle f_{\mathrm{HI}}\rangle$) of hydrogen in 12 redshift bins at $4.85 < z < 6.05$ from a large sample of moderate resolution XShooter and ESI QSO absorption spectra. The fluctuations in ionizing radiation field are modeled by post-processing simulations from the Sherwood suite using our new code "EXtended reionization based on the Code for Ionization and Temperature Evolution" (EX-CITE). EX-CITE uses efficient Octree summation for computing intergalactic medium attenuation and can generate large number of high resolution Γ_{HI} fluctuation models. Our simulation with EX-CITE shows remarkable agreement with simulations performed with the radiative transfer code *Aton* and can recover the simulated parameters within 1σ uncertainty. We measure the three parameters by forward-modeling the Lyα forest and comparing the effective optical depth ($\tau_{\mathrm{eff,HI}}$) distribution in simulations and observations. The final uncertainties in our measured parameters account for the uncertainties due to thermal parameters, modeling parameters, observational systematics and cosmic variance. Our best fit parameters show significant evolution with redshift such that $\lambda_{\mathrm{mfp,HI}}$ and $\langle f_{\mathrm{HI}}\rangle$ decreases and increases by a factor ~ 6 and $\sim 10^4$, respectively from $z \sim 5$ to $z \sim 6$. By comparing our $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\langle f_{\mathrm{HI}}\rangle$ evolution with that in state-of-the-art *Aton* radiative transfer simulations and the *Thesan* and *CoDa-III* simulations, we find that our best fit parameter evolution is consistent with a model in which reionization completes by $z \sim 5.2$. Our best fit model that matches the $\tau_{\mathrm{eff,HI}}$ distribution also reproduces the dark gap length distribution and transmission spike height distribution suggesting robustness and accuracy of our measured parameters.

Key words: cosmology: large-scale structure of Universe - methods: numerical - galaxies: intergalactic medium - quasars: absorption lines

* E-mail: gaikwad@mpia.de

1 INTRODUCTION

The ionization of neutral hydrogen (H I) at $z > 5$ by ultraviolet photons from astrophysical sources is one of the important phase transitions in the Universe (see review by Rauch 1998; Meiksin 2009; McQuinn 2016; Gnedin & Madau 2022). The transmission spikes and dark gaps in the spectra of background Quasi Stellar Objects (QSO) at $z > 5$ are useful probes to study the end stages of H I reionization (Chardin et al. 2018a; Garaldi et al. 2019; Gaikwad et al. 2020). The observed long dark troughs (Fan et al. 2001; Gallerani et al. 2006; Becker et al. 2015), the scatter in the observed effective optical depth ($\tau_{\rm eff,HI}$) (Becker et al. 2018; Eilers et al. 2018; Bosman et al. 2018; Yang et al. 2020; Bosman et al. 2022) and the measurement of a rather short mean free path suggests that H I reionization is late and inhomogeneous (Worseck et al. 2014; Becker et al. 2021). State-of-the-art cosmological radiative transfer simulations confirm this late and patchy H I reionization scenario by demonstrating that spatial fluctuations in the amplitude of ionizing radiation field are needed to reproduce the observed properties of the Lyα forest at $z > 5$ (Gnedin 2014; Chardin et al. 2017; Gnedin et al. 2017; Rosdahl et al. 2018; Nasir & D'Aloisio 2020; Kulkarni et al. 2019; Ocvirk et al. 2020; Qin et al. 2021; Kannan et al. 2022). There are two main consequences of late and patchy reionization. First, the late end of reionization results in a higher (lower) neutral fraction $f_{\rm HI}$ (photo-ionization rate, $\Gamma_{\rm HI}$) at $z < 6$ than previously thought. Second, the patchiness of reionization affects the mean free path of ionizing photons, $\lambda_{\rm mfp,HI}$. Thus, for a late and patchy reionization scenario, one expects to see rapid evolution of $\lambda_{\rm mfp,HI}$, $\Gamma_{\rm HI}$ and $f_{\rm HI}$ at $5 \leq z \leq 6$.

Recently Becker et al. (2021) measured $\lambda_{\rm mfp,HI}$ at $z = 5.1, 6.0$ using composite spectra at the Lyman limit edge (912 Å). These measurements suggests that $\lambda_{\rm mfp,HI}$ is significantly decreasing from $z = 5.1$ to $z = 6.0$ (see also Bosman 2021). The $\lambda_{\rm mfp,HI}$ evolution in these observations is found to be steeper than the extrapolated power-law from lower redshift, $\lambda_{\rm mfp,HI} \propto (1 + z)^{-5.4}$, suggesting a late and rapid end of H I reionization at $5 \leq z \leq 6$. This strong evolution of $\lambda_{\rm mfp,HI}$ with redshift is, however, difficult to reproduce in cosmological radiative transfer simulations (Keating et al. 2020a; Cain et al. 2021, 2023) but also see Lewis et al. (2022). Direct measurements of $\lambda_{\rm mfp,HI}$ at $z > 5$ are challenging and are slightly uncertain due to uncertainty in QSO environment and proximity zone sizes (see Becker et al. 2021, for details). It is thus important to measure the mean free path using alternative methods that complement the existing measurements.

The photo-ionization rate ($\Gamma_{\rm HI}$) set by the ionizing sources in the Universe controls the amount of neutral hydrogen present at a given epoch since $f_{\rm HI} \propto \Gamma_{\rm HI}^{-1}$ once reionization is completed (Weinberg et al. 1998; Gaikwad et al. 2019). The photo-ionization rate is expected to be strongly correlated with the mean free path of ionizing photons for a given emissivity ($\lambda_{\rm mfp,HI} \propto \Gamma_{\rm HI}$, Haardt & Madau 2012). In order to get realistic constraints on $\lambda_{\rm mfp,HI}$, it is essential to measure $\Gamma_{\rm HI}$ and $\lambda_{\rm mfp,HI}$ simultaneously. The photo-ionization rate, $\Gamma_{\rm HI}$ at $5 \leq z \leq 6$ has been measured in

the literature by matching observed and simulated mean opacities (Bolton & Haehnelt 2007; Becker & Bolton 2013; D'Aloisio et al. 2018; Choudhury et al. 2021) or using high redshift QSO near zones (Wyithe & Bolton 2011; Calverley et al. 2011). The $\Gamma_{\rm HI}$ measurements in QSO near zones are performed by modeling the ionizing radiation field from the QSO and the background photo-ionization rate in the IGM. The number of spectra in previous analyses at $z \sim 6$ were limited to < 10. Furthermore, the $\Gamma_{\rm HI}$ measurements are complicated by uncertainty in the QSO environment, the uncertainty in the thermal state of the gas and the QSO spectral energy index (Bolton et al. 2012). On the other hand, the $\Gamma_{\rm HI}$ measurements based on the mean opacity need to model the evolution of the ionizing radiation field as well as $\lambda_{\rm mfp,HI}$ (D'Aloisio et al. 2018; Choudhury et al. 2021). In D'Aloisio et al. (2018) $\lambda_{\rm mfp,HI}$ is assumed to evolve as a power-law with redshift, $\lambda_{\rm mfp,HI} \propto (1 + z)^{-5.4}$. However, marginalization over $\lambda_{\rm mfp,HI}$ has not been accounted for in the $\Gamma_{\rm HI}$ measurements. It is important to account for the uncertainties of both $\lambda_{\rm mfp,HI}$ and $\Gamma_{\rm HI}$.

By varying the mean free path and photo-ionization rate in simulations, one also naturally constrains the evolution of the neutral fraction $f_{\rm HI}$. It is thus possible to infer the neutral fraction for a given $\lambda_{\rm mfp,HI}$-$\Gamma_{\rm HI}$ measurements in simulations. Most of the $f_{\rm HI}$ measurements at $z > 5.5$ are usually in the form of lower or upper limits. The neutral fraction $f_{\rm HI}$ at $z < 6$ has been estimated by matching the mean opacity (Fan et al. 2006) or measuring the dark pixel fractions (Mesinger 2010; McGreer et al. 2011). McGreer et al. (2015) constrained $f_{\rm HI}$ by performing a model independent analysis of the dark pixel fraction in the QSO absorption spectra (see also Jin et al. 2023). The $f_{\rm HI}$ measurements from dark pixels are upper limits because even a small residual (e.g., 10^{-4}) neutral fraction is sufficient to saturate the absorption. The lower limits for $f_{\rm HI}$ are generally obtained by comparing the Lyα forest opacity modelled by simulations with a spatially homogeneous UV background with the observed mean opacity (Becker et al. 2015; Yang et al. 2020; Bosman et al. 2022). At high redshift the photo-ionization rate can be chosen such that the opacity in these simulations matches the observed mean opacity but not its scatter because spatial fluctuations in the ionizing radiation field are not modeled. As a result, the true $f_{\rm HI}$ is generally larger than the measured $f_{\rm HI}$. The $f_{\rm HI}$ measurements then only place lower limits (but see Choudhury et al. 2021). Note that recently Zhu et al. (2022) have placed upper limits on $f_{\rm HI}$ at $5.5 < z < 6.0$ using the dark gap length distribution in the Lyβ forest using cosmological simulations that model spatial fluctuations of the ionizing radiation. These constraints were found to be in agreement with previous constraints from the dark pixel fraction (McGreer et al. 2015).

Our main aim in this work is to accurately measure the evolution of the spatially averaged photo-ionization rate $\langle \Gamma_{\rm HI} \rangle$, mean free path of H I ionizing photons $\lambda_{\rm mfp,HI}$ and spatially averaged neutral fraction $\langle f_{\rm HI} \rangle$ in the redshift range $5 \leq z \leq 6$ by modeling the spatial fluctuations in the ionizing radiation field and comparing the predictions for the Lyα forest with observations. Varying these parameters in radiative transfer simulations is challenging as one would need to perform a large suite of simulations with different ionization histories. Performing radiative hydrodynamic simulations is computationally very expensive. The

other challenge is to produce large spatial variation in mean free path and photo-ionization rate. Current cosmological Radiative Transfer (hereafter RT) simulations have difficulty in producing a large variation in mean free path (Kulkarni et al. 2019; Keating et al. 2020a). To obtain a short mean free path in large cosmological simulations, addition of neutral sinks has been proposed. These sinks mimic the effect of photo-evaporation of minihalos during reionization (Shapiro et al. 2004; Iliev et al. 2005a; Cain et al. 2021). These additional sinks are attributed to unresolved density structure on scales of $1h^{-1}$ ckpc (Shapiro et al. 2004; Furlanetto & Oh 2005; McQuinn et al. 2007; Alvarez & Abel 2012; Mesinger et al. 2014; D'Aloisio et al. 2020; Nasir et al. 2021). Such sinks slow down the ionization fronts and consume ionizing photons until the minihalos are photo-evaporated (Iliev et al. 2005b; Shapiro et al. 2006; Park et al. 2016, 2021). However, the approach of adding IGM sinks must assume the sub-grid density distribution which is not known a priori. We present here an efficient alternative method that varies both mean free path and photo-ionization rate in cosmological radiative transfer simulations.

Davies & Furlanetto (2016) have proposed a theoretical framework for fluctuations of the mean free path for modelling fluctuations in the ionizing radiation field (see also Mesinger & Furlanetto 2009; Davies & Furlanetto 2014; Davies et al. 2017). In this approach the local mean free path of each cell is determined by the amount of neutral hydrogen that depends on the local photo-ionization rate in the cell. The photo-ionization rate in a given cell is calculated by adding the contribution of ionizing radiation from all sources and by taking into account the attenuation from the IGM. The IGM attenuation depends on the local mean free path of all cells along the sightline towards the source. Thus photo-ionization rate and mean free path are coupled together by two non-linear equations that need to be solved iteratively for all the cells in the simulation box simultaneously until convergence is achieved. The main advantage of this framework is that one can vary mean free path and photo-ionization rate as free parameters. Due to the numerical complexity of the iterative method, previous applications of this framework have been limited to few 10s of models with somewhat limited resolution of $> 3\ h^{-1}$ cMpc ($\sim 64^3$ to 128^3 grids, Davies et al. 2017; D'Aloisio et al. 2018). As a result, previously it was challenging to constrain $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ simultaneously from such simulations.

In this work, we present an efficient numerical method to model EXtended reionization, based on our Code for Ionization and Temperature Evolution (CITE), which we will call EX-CITE. Our new EX-CITE code captures the fluctuations in Γ_{HI} using the Octree method (Barnes & Hut 1986). With the EX-CITE code, we generate $\sim 650\ \Gamma_{\mathrm{HI}}$ fluctuation models at a much higher resolution of $0.31\ h^{-1}$ cMpc (512^3 grids). Compared to previous studies, EX-CITE allows us to improve the number of models and the resolution by a factor ~ 35 and 10 respectively. We then compare our models with unprecedented quality QSO absorption spectra taken with the XShooter and ESI instruments and present measurements of $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\langle f_{\mathrm{HI}}\rangle$. The method of measuring $\lambda_{\mathrm{mfp,HI}}$ presented in this work, is complementary to that employed by Becker et al. (2021).

The paper is organized as follows. In §2 we describe our observational sample. In §3 and §4 we lay out the theoret-

ical framework and describe the simulations, respectively. We describe our method of recovering and measuring the parameters using EX-CITE simulations from our fiducial RT simulation and observations in §5. We describe the main result of our analysis in §6. Finally we summarize our main findings in §7. We suggests that readers who are less interested in the numerical methods to go through the sections §2, 4 and 6 for the main analysis of this work. Throughout this work we use a flat ΛCDM cosmological parameters with value $\Omega_\lambda = 0.692$, $\Omega_{\mathrm{m}} = 0.308$, $\Omega_{\mathrm{b}} = 0.0482$, $h = 0.678$, $Y = 0.24$, $n_{\mathrm{s}} = 0.961$, $\sigma_8 = 0.829$ (Planck Collaboration et al. 2014). The photo-ionization rate expressed in units of 10^{-12} s^{-1} is denoted by Γ_{12}. For comoving and physical distances we use prefix symbols 'c' and 'p', respectively.

2 OBSERVATIONS

We primarily use a sample of 67 high redshift ($z > 5.5$) QSO absorption spectra from Bosman et al. (2022). Here we briefly summarize our observational sample that consists of 25 spectra from the XQR-30 program (D'Odorico et al. 2023), 26 archival XShooter spectra and 16 archival spectra taken with the ESI instrument (Sheinis et al. 2002; Vernet et al. 2011). All spectra were reduced in an identical manner with the same custom pipeline to ensure that the systematics arising from the different instruments and data reduction pipelines are minimal (Becker et al. 2019; Zhu et al. 2021; Bosman et al. 2022; Chen et al. 2022; Lai et al. 2022; Bischetti et al. 2022; D'Odorico et al. 2023). The spectral resolution of XShooter depends on the seeing conditions and is different in the two arms (VIS and NIR) of the instrument. Typically, the XShooter spectral resolution in our observed sample varies from 22 to 30 km s^{-1} (27 to 39 km s^{-1}) for the visible (near infrared) arms (Davies et al. 2022).

The spectral resolution of the ESI spectra is ~ 60 km s^{-1}. We use appropriate spectral resolutions when forward-modeling the spectra from the simulations and fitting the transmission spikes in observations/simulations. The signal-to-noise ratio (hereafter S/N) per pixel in our sample is typically larger than 10 and varies from 10.3 to 560.5. Note that the XShooter and ESI spectra reductions used different pixel size. The S/N per pixel values described in this work are same as that in Bosman et al. (2022).

Note that the intrinsic QSO continuum is uncertain, especially at $z > 5$, due to the lack of a significant number of pixels where the flux recovers to the continuum. The QSO continuum has been estimated using the Principle Component Analysis (PCA) method (see Suzuki et al. 2005; Davies et al. 2018; Bosman et al. 2022, for details). The continuum fitting uncertainty has been accounted for when estimating the uncertainty in the observed $\tau_{\mathrm{eff,HI}}$. Fig. 1 shows the redshift coverage of the Lyα forest in our observed sample of QSOs. We exclude a proximity zone with the size of 10 pMpc blueward of the QSO Lyα emission. We also exclude the region of the spectra that correspond to rest wavelengths $\lambda < 1080$Å to avoid contamination by Lyβ or lower redshift Lyα forest.

For measuring the mean free path and photo-ionization rate, we use $\tau_{\mathrm{eff,HI}}$ measurements from all the spectra in the observed sample. The $\tau_{\mathrm{eff,HI}}$ measurements are calculated using a redshift interval of $dz = 0.1$. The consistency of mea-
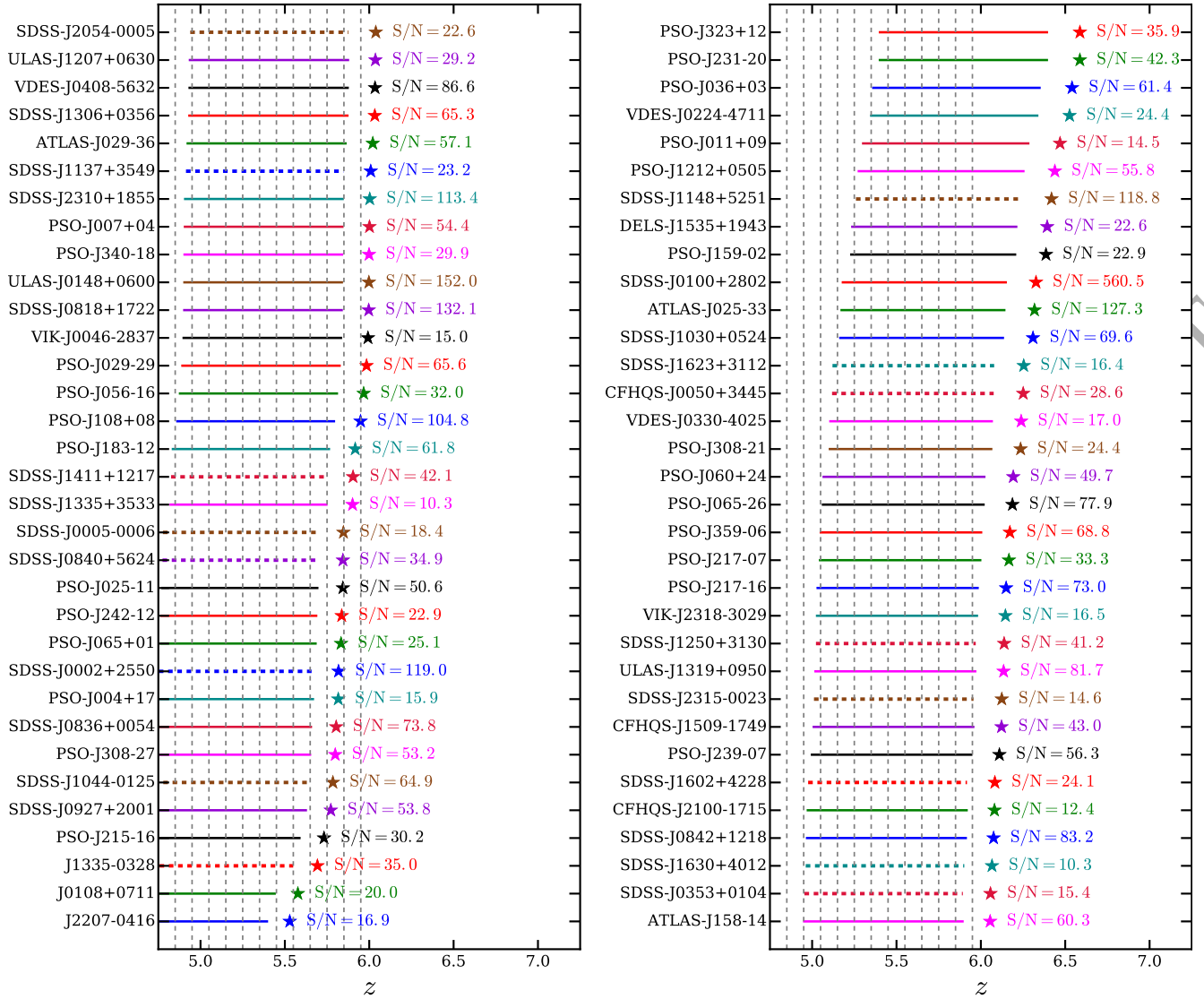
Figure 1. The panels show the redshift coverage of QSO sightlines in our sample. Our sample consists of absorption spectra taken with the XShooter and ESI spectrographs. The redshift coverage of spectra obtained using XShooter are shown by solid lines while the corresponding ESI spectral coverage is shown by dashed lines. We divide the observational sample in 12 redshift bins as shown by vertical dashed lines. We use both XShooter and ESI spectra to calculate the $\tau_{\mathrm{eff,HI}}$ CDF and the dark gap length CDF. We use only XShooter sightlines to derive transmission spike statistics because of the higher resolution of XShooter. The emission redshift of each QSO is indicated by stars. A proximity zone of size 10 pMpc is excluded blueward of each QSO redshift. Similarly, the lower limit of redshift coverage corresponds to 1080 Å to exclude the Lyβ emission line. The S/N per pixel is given for each sightline and varies from 10.3 to 560.5. Our observed sample is very similar to that presented in Bosman et al. (2022).

sured best fit parameters is then tested using the dark gap length statistics and the pseudo-Column Density Distribution Function (pCDDF, Gaikwad et al. 2020). Similarly as for the $\tau_{\mathrm{eff,HI}}$ CDF statistics, we use all spectra in our sample to calculate the dark gap length statistics. The observed pCDDF statistics are computed only from high resolution XShooter spectra (see §6.6 for details). This is because the pCDDF statistics are obtained by decomposing the spikes into multi-component inverted Voigt profiles. The internal structure of the transmission spikes is washed out if the resolution of observed spectra is too low (see Gaikwad et al. 2020, for details). To reduce the systematics in the number

of Voigt components for a given $\log \widetilde{N}_{\mathrm{HI}}$ bin, we opt to use XShooter spectra only. It is noteworthy here that the total number of XShooter spectra is a factor ~ 3 times larger than that of the ESI spectra. Hence the results would be similar if we include the ESI spectra. Finally, we refer the reader to Zhu et al. (2021); Bosman et al. (2022) and references therein for detailed information on the observational sample and the data reduction (D'Odorico et al. 2023).

3 THEORETICAL FRAMEWORK

The spatial fluctuations in H I photo-ionization rates are substantial during the end stages of H I reionization ($z > 5$). These fluctuations are due to the discrete nature and the clustering of the ionizing sources during reionization and persist for some time after the percolation of the ionized regions (Chardin et al. 2018a). In this section, we first lay out the theoretical framework with which we capture these fluctuations. We then discuss the numerical implementation of the theoretical framework in our code EX-CITE. We verify the consistency of our approach by recovering the mean free path and H I photo-ionization rate of a full radiative transfer simulation run with the code *Aton* at $5 < z < 6$ in §5.

We use smooth particle hydrodynamic simulations and generate halo catalogs at the redshifts of interest. Emissivity weights are assigned to each halo based on halo mass. These emissivity weights correspond to the contribution of each halo to the total ionizing emissivity in a given simulation volume. For H I reionization, we assume that the H I ionizing photons are contributed by star-forming galaxies. Consensus is emerging that the contribution of QSOs to the H I ionizing background is moderate due to the rapid decline of the space density of luminous QSOs at high redshift (Chardin et al. 2015; Jiang et al. 2022). Our simulation does not account for the complex processes of galaxy and supermassive black hole formation. We instead assign the H I ionizing emissivity to dark matter haloes as follows,

$$
\begin{aligned}
\epsilon_{\mathrm{halo,i}} &= \epsilon_0 \times M_{\mathrm{halo,i}}^{\beta} && \text{for} \quad M_{\mathrm{halo}} \geq M_{\mathrm{cutoff}} \\
&= 0 && \text{for} \quad M_{\mathrm{halo}} < M_{\mathrm{cutoff}},
\end{aligned} \tag{1}
$$

where ϵ_0 is a normalization factor that is independent of the halo mass, β is the emissivity power-law index and M_{cutoff} is the minimum halo mass above which halos can contribute to the H I ionizing emissivity. We use values of $M_{\mathrm{cutoff}} = 10^9 \, \mathrm{M_\odot}$ and $\beta = 1.0$ for our fiducial models (Kulkarni et al. 2019). We find that the effect of varying $M_{\mathrm{cutoff}}(10^8 - 10^{10} \, \mathrm{M_\odot})$ and β (0.5 - 1.5, see Table 1) on our measured parameters is small. For the simulations used in this work, the halo catalogs are 99 percent complete above $M_{\mathrm{cutoff}} > 10^9 \mathrm{M_\odot}$. It is important to note that the observational constraints on the nature of the ionizing sources are still limited. In particular, it is not clear that the ionizing emissivity should monotonically increase with halo mass. In reionization simulations that model the complex processes governing galaxy, star formation and the resulting emission of ionizing photons are highly intermittent (Rosdahl et al. 2018).

However, our simplistic approach of assigning ionizing emissivity to halos can nevertheless capture the large scale fluctuations in ionizing background required to reproduce the rather large Lyα opacity fluctuations (Kulkarni et al. 2019). We find that the UVB fluctuations depend on the modelling of both the sources and the sinks, but the dependence is rather weak. As we show later the main uncertainty in constraining $\lambda_{\mathrm{mfp,HI}}$ and Γ_{HI} is contributed by the uncertainty in thermal parameters of the IGM.

Note that the value of ϵ_0 characterising the ionizing emissivity affects the absolute value of the photo-ionization rate $\langle \Gamma_{\mathrm{HI}} \rangle$ (which is a free parameter in our formalism), but does not affect the fluctuations $\Gamma_{\mathrm{HI}} / \langle \Gamma_{\mathrm{HI}} \rangle$. This is because

fluctuations in $\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle$ are mainly sensitive to the relative location of sources with respect to sinks and the distribution of densities along the sightlines in our model. Since we are interested in producing $\Gamma_{\mathrm{HI}} /\langle \Gamma_{\mathrm{HI}} \rangle$ maps, the value of ϵ_0 is not important. In the rest of our formalism, we therefore use halo emissivity weights $w_{\mathrm{halo,i}}$, that describe the dependence of H I ionizing emissivity on halo mass,

$$
\begin{aligned}
w_{\mathrm{halo,i}} &= M_{\mathrm{halo,i}}^{\beta} \; / \; \sum_{i=1}^{N_{\mathrm{halo}}} M_{\mathrm{halo,i}}^{\beta} && \text{for } M_{\mathrm{halo}} \geq M_{\mathrm{cutoff}} \\
&= 0 && \text{for } M_{\mathrm{halo}} < M_{\mathrm{cutoff}}.
\end{aligned} \tag{2}
$$

We note here that we have varied the normalization factor, ϵ_0, with redshift in the radiative transfer simulations with *Aton* to achieve a given reionization history. However in this work, we do not specify a redshift evolution of ϵ_0, but rather vary $\langle \Gamma_{\mathrm{HI}} \rangle$ in our method. This allows us to explore a large $\lambda_{\mathrm{mfp,HI}} - \langle \Gamma_{\mathrm{HI}} \rangle$ parameter space that is independent of redshift and facilitates an efficient comparison with observations.

Given a set of emissivity weights $w_{\mathrm{halo,i}}$ one can calculate the fluctuations in photo-ionization rate ($\delta \Gamma_{\mathrm{HI,j}}$) at a cell j in the simulation box as,

$$
\frac{\Gamma_{\mathrm{HI,j}}}{\langle \Gamma_{\mathrm{HI}} \rangle} = f_{\mathrm{norm}} \sum_{\mathrm{i}=1, \mathrm{i} \neq j}^{N_{\mathrm{source}}} \frac{w_{\mathrm{halo,i}}}{(4\pi r_{\mathrm{ij}}^2)} \exp \left[- \int_{r_i}^{r_j} \frac{dx}{\lambda(x)} \right], \tag{3}
$$

where the summation index i is over all the sources in the simulation box (N_{source}), r_{ij} is the distance between source at cell i and cell j at which photo-ionization rate fluctuations need to be calculated (Davies & Furlanetto 2016). The factor $4\pi r_{\mathrm{ij}}^2$ is the flux dilution factor at a distance r_{ij} away from the source. The exponential term in the above equation accounts for the IGM attenuation that corresponds to absorption of ionizing photons along the sightline from source to sink. The dimensional factor f_{norm} is a normalization that ensures that the average of $\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle$ over all the simulation volume is 1. It is noteworthy that the value of photo-ionization rates (Γ_{HI}) explicitly depends on the source spectral energy distribution (usually assumed to be a blackbody with $T \sim 50000$ K for O-type stars), scale factor (a) and H I photo-ionization cross-section (σ_{HI}, see Choudhury et al. 2021, for details). However, when we take the ratio of the photo-ionization rate in a cell and the spatially averaged photo-ionization rate (i.e., the relative fluctuation, $\delta \Gamma_{\mathrm{HI}}$), these dependencies cancel out leaving the simple expression shown in Eq. 3.

The optical depth $\tau_{\mathrm{ij}} = \int dx/\lambda(x)$ encountered by H I ionizing photons depends on the mean free path $\lambda(x)$ in each cell along sightlines. To compute the mean free path in each cell we follow an approach similar to that of Davies & Furlanetto (2016); D'Aloisio et al. (2018). The mean free path in a given cell is assumed to depend on the local overdensity Δ and photo-ionization rate fluctuations as,

$$
\lambda(x) = \lambda_0 \, \Delta^{-1} \left[\frac{\Gamma_{\mathrm{HI}}(x)}{\langle \Gamma_{\mathrm{HI}} \rangle} \right]^{\zeta} \left[\frac{E_{\mathrm{bin}}}{E_{\mathrm{ion,HI}}} \right]^{0.9}, \tag{4}
$$

where λ_0 is a 'spatially averaged estimate of the mean free path' (a free parameter in our formalism) and ζ is the power-law index that describes the relation between fluctuations in the H I photo-ionization rate and the mean free path.

$E_{\rm bin}, E_{\rm ion,HI}$ are the average energy of the photons in frequency bin and the ionization potential of H I, respectively (Haardt & Madau 1996; Miralda-Escudé et al. 2000; Muñoz et al. 2016). We use a fiducial value of $\zeta = 2/3$ consistent with Davies & Furlanetto (2016). We have explicitly checked whether the choice of ζ affects our measurement or the correlations between parameters. We assume a mono-frequency scenario and use an average energy $E_{\rm bin} = 20.62$ eV to represent a blackbody spectrum with $T \sim 50000$ K (Kulkarni et al. 2019).

The spectral energy distribution (SED) of galaxies could of course be significantly different from a simple blackbody spectrum. However, the differences in galaxy SED will nevertheless not significantly affect the parameter estimation in this work. This is because the differences in SED mainly changes the number of ionizing photons and the average energy of photons in frequency bins. The change in number of ionizing photons mainly drive the variation in amplitude of the spatially averaged photo-ionization rate $\langle\Gamma_{\rm HI}\rangle$. Changes in the energy of ionizing photons changes the thermal state of the IGM (i.e., the temperature rise due to reionization). In our formalism, $\langle\Gamma_{\rm HI}\rangle$ and the thermal parameters (T_0, γ) are free parameters. Thus, instead of varying the source SED, we vary $\langle\Gamma_{\rm HI}\rangle$ and the thermal parameters in our model. We emphasize that our aim is to simultaneously measure $\lambda_{\rm mfp,HI}$ and $\langle\Gamma_{\rm HI}\rangle$. Hence we use a range of values in the thermal parameters measured in our earlier work (Gaikwad et al. 2020). The $\langle\Gamma_{\rm HI}\rangle$ and $\lambda_{\rm mfp,HI}$ measurements presented in this paper are marginalized over the thermal parameter uncertainty.

It is important to note that the spatially averaged mean free path parameter, λ_0, defined above is just a proxy for the mean free path $\lambda_{\rm mfp,HI}$ as defined in the literature or inferred from observations. λ_0 is a convenient parameter used to generate fluctuations in $\Gamma_{\rm HI}$. The actual mean free path in observations and simulations is calculated by stacking/averaging the Ly-continuum flux at the H I ionizing wavelength, 912 Å. The flux is then fitted with an exponentially decreasing profile with $\lambda_{\rm mfp,HI}$ as a free parameter. For each λ_0 and $\langle\Gamma_{\rm HI}\rangle$ combination, we calculate the true mean path $\lambda_{\rm mfp,HI}$ as described in §5.2. We then use the $\lambda_{\rm mfp,HI}$-$\langle\Gamma_{\rm HI}\rangle$ parameter space to measure the photo-ionization rate and mean free path from observations. In §5.7, we show that we can recover the mean free path in radiative transfer simulation from the $\lambda_{\rm mfp,HI}$-$\langle\Gamma_{\rm HI}\rangle$ grids generated using EX-CITE.

Eq. 3 together with Eq. 4 forms the basis of our formalism. Eq. 3, describes how the mean free path globally affects fluctuations of the photo-ionization rate while Eq. 4 describes how the mean free path depends on local photo-ionization rate fluctuations. Because of the interdependence of λ and $\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle$ in Eq. 3 and Eq. 4, one needs to solve these equations iteratively.

We have developed a code EXtended reionization based on the Code for Ionization and Temperature Evolution (EX-CITE) to solve these equations in a post-processing step of state-of-the art cosmological hydrodynamic simulation. In EX-CITE, the contribution of ionizing source to a given location is calculated using efficient Octree methods similar to that used in the gravity solver of codes like P-GADGET-3 (Barnes & Hut 1986; Springel 2005). EX-CITE allows us to efficiently explore the large parameter space. In EX-CITE, we start with the cosmological density field, halo catalog. We

assume a mean free path parameter λ_0 to generate the fluctuations in photo-ionization rate. We use following steps to compute the fluctuations in the photo-ionization rate.

(i) We assign emissivity weights to the halos (Eq. 2). and construct an octree from the emissivity field by dividing the computational domain in to sub-cubes (also known as tree nodes).

(ii) For each grid cell in our simulation, we traverse the octree from top to bottom. We check if the tree node satisfies the criterion $s/r < \theta$, where s is the size of the sub-cube, r is the distance between grid cell and center of the sub-cube. We take $\theta = 0.7$ as a trade-off between accuracy and speed.

(iii) For tree-nodes satisfying the above criterion, we cast a ray from the tree-nodes to grid cells and compute the IGM attenuation (Eq. 4) and photo-ionization rate fluctuations (Eq. 3). If a tree-node does not satisfy the above criterion, we traverse down the tree and continue until the criterion is satisfied.

(iv) We add the contribution of all such tree-nodes to a given grid cell. Steps (ii) to (iii) are repeated for all the grid cells in the simulation box.

(v) As Eq. 3 Eq. 4 are non-linearly coupled together, we iterate steps (ii) to (iv) until the maximum absolute difference between current and previous $\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle < 10^{-6}$.

We refer the reader to the online appendix A for the details of the numerical implementation of EX-CITE.

4   SIMULATIONS

In this work, our modelling is based on the Sherwood[1] simulation suite performed with the P-GADGET-3[2] code (Springel 2005; Bolton et al. 2017). The details of all the simulations used in this work are summarized in Table 1. Our primary simulation contains 2×2048^3 particles in a volume of 160 h^{-1} cMpc (denoted by L160N2048). The simulation has been performed with a spatially uniform but time varying Haardt & Madau (2012) ultra-violet background (UVB) model. The ionization and thermal evolution equations are solved for primordial abundances using equilibrium equations (see Puchwein et al. 2015, 2019; Gaikwad et al. 2019, for non-equilibrium effects). The simulation outputs were saved at 40 Myr intervals starting from redshift $z = 40$. In this work, we use simulation outputs stored at 6 redshifts $z = 5.11, 5.26, 5.41, 5.58, 5.76$ and 5.95. Kulkarni et al. (2019) used the same simulation for post-processing with the radiative transfer code *Aton*. The choice of number of particles and box size is to account for the mean free path of H I ionizing photons, which is typically large at lower redshift ($z \sim 5$), while at the same time ensuring the H I Lyα forest spectra have resolution similar to that in observations (~ 34 or ~ 60 km s^{-1}). In order to check the effect of box size, mass resolution and initial conditions on the Lyα forest statistics, we also use simulations based on L40N512, L40N2048, L80N1024, L80N2048, L160N512 and L160N1024. Unlike our primary simulation box, the outputs for these models are stored at $z = 4.8, 5.4$ and 6.0. The

---

[1] https://www.nottingham.ac.uk/astronomy/sherwood/
[2] https://wwwmpa.mpa-garching.mpg.de/gadget/

Table 1. Summary of EX-CITE models performed in this work

Simulation	$N_{\mathrm{Grid},\Gamma_{\mathrm{HI}}}$	N_{model}	M_{cutoff}	β	ζ	Purpose
L160N2048	512	648	10^9	1.0	2/3	Default model for parameter measurements (Fig. 7, 8)
L160N2048 (*Aton*)	2048	1	10^9	1.0	-	Accuracy test: Parameter recovery (Fig. 6 and Fig. G1)
L40N512	512	6	10^9	1.0	2/3	Convergence test: Box size (left panel of Fig. C1, C3)
L80N1024	512	6	10^9	1.0	2/3	Convergence test: Box size (left panel of Fig. C1, C3)
L160N512	512	6	10^9	1.0	2/3	Convergence test: Mass resolution (middle panel of Fig. C1, C3)
L160N1024	512	6	10^9	1.0	2/3	Convergence test: Mass resolution (middle panel of Fig. C1, C3)
L40N2048	512	6	10^9	1.0	2/3	Convergence test: Initial conditions (right panel of Fig. C1, C3)
L80N2048	512	6	10^9	1.0	2/3	Convergence test: Initial conditions (right panel of Fig. C1, C3)
L160N2048	64	6	10^9	1.0	2/3	Convergence of $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ maps (Fig. C4 and Fig. C5)
L160N2048	128	6	10^9	1.0	2/3	Convergence of $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ maps (Fig. C4 and Fig. C5)
L160N2048	256	6	10^9	1.0	2/3	Convergence of $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ maps (Fig. C4 and Fig. C5)
L160N2048	1024	6	10^9	1.0	2/3	Convergence of $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ maps (Fig. C4 and Fig. C5)
L160N2048	512	6	10^8	1.0	2/3	Modeling uncertainty: Effect of M_{cutoff} (Fig. 9 and Fig. F1)
L160N2048	512	6	10^{10}	1.0	2/3	Modeling uncertainty: Effect of M_{cutoff} (Fig. 9 and Fig. F1)
L160N2048	512	6	10^9	0.5	2/3	Modeling uncertainty: Effect of β (Fig. 9 and Fig. F1)
L160N2048	512	6	10^9	1.5	2/3	Modeling uncertainty: Effect of β (Fig. 9 and Fig. F1)
L160N2048	512	6	10^9	1.0	1/3	Modeling uncertainty: Effect of ζ (Fig. 9 and Fig. F1)
L160N2048	512	6	10^9	1.0	3/4	Modeling uncertainty: Effect of ζ (Fig. 9 and Fig. F1)
L160N2048 (*uniform*)	-	6	-	-	-	Effect of Γ_{HI} fluctuations (Fig. 8 and Fig. 14)

Table 2. EX-CITE time consumption (in cpu hours) per model. Cumulative time refers to the total time required to generate Γ_{HI} maps at refinement level $(< R)$

$N_{\mathrm{Grid},\Gamma_{\mathrm{HI}}}$	Time in cpu hours	Cumulative time
64	2	2
128	42	44
256	640	684
512	1152	1836
1024	3800	5636

convergence tests are thus performed at $z \sim 5.0$, 5.4 and 6.0. We find that the Sherwood simulation L160N2048 post-processed with EX-CITE is well converged with respect to box size, mass resolution, halo mass function and the number of grids used to generate the $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ field fluctuations. We refer the reader to the online appendix C for details.

All the simulations mentioned above employ a simplified star formation prescription in which particles with $\Delta > 1000$ and $T < 10^5$ K are converted to star particles (Viel et al. 2004). Our simulations do not model astrophysical processes of galaxy formation such as stellar or AGN feedback or metal enrichment. In order to capture fluctuations in Γ_{HI}, we need the location of ionizing sources (galaxies) in our simulations. We use the halo catalogs that have been generated on-the-fly while performing the simulations. The halo sample in all of our models is complete above 10^9 $\mathrm{M_\odot}$. We use the halo catalogs to assign emissivity weights (see §3 for details). For all the models, we grid the density, velocity (3 components) and temperature fields on Cartesian grids of 64^3, 128^3, 256^3, 512^3, 1024^3 and 2048^3 at all the redshifts of

interest. As we show in §3, this is necessary for our iterative method to achieve fast convergence. Even though temperature fields are available for the Sherwood simulation suite, we impose temperature-density relations to account for the uncertainties in observed thermal parameters (see §5.4)

5 METHOD

In this section we describe our method of generating models, forward-modeling of Lyα forest spectra, the description of Lyα forest statistics, method of parameter estimation and the consistency checks of our approach with an *Aton* radiative transfer simulation by recovering $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$.

5.1 Model generation

Our main aim in this work is to simultaneously measure the mean free path of H I ionizing photons, $\lambda_{\mathrm{mfp,HI}}$ and the spatially averaged H I photo-ionization rate $\langle\Gamma_{\mathrm{HI}}\rangle$. We generate $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ fields at 6 redshifts where Sherwood snapshots are available (see §4). The formalism of EX-CITE (see §3) accounts for spatial fluctuations in the photo-ionization rate i.e., $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$. We first vary the mean free path parameter (in h^{-1} cMpc) in equally spaced logarithmic bins of $\log\langle\lambda_0\rangle = -1.50, -1.46, \cdots, 2.78, 2.82$. We use then EX-CITE to generate these 108 $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ models at each redshift. For brevity we denote these fields by the symbol $[\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle]_{\mathrm{EX-CITE}}$. In total we generate $108 \times 6 = 648$ $[\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle]_{\mathrm{EX-CITE}}$ models for our default simulation L160N2048. Generating the $[\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle]_{\mathrm{EX-CITE}}$ models is the most computationally expensive part of our analysis.

The variation of other parameters (e.g., $\langle\Gamma_{\rm HI}\rangle$ and thermal parameters) is performed in a post-processing step and hence is relatively less expensive.

For each $[\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle]_{\rm EX-CITE}$ model (i.e., for given λ_0, z), we then vary $\langle\Gamma_{\rm HI}\rangle$ in 81 equi-spaced logarithmic bins of $\log\langle\Gamma_{\rm HI}\rangle = -15, -14.95, \cdots, -11.05, -11$. In total there are $108 \times 81 = 8748$ $\lambda_0 - \langle\Gamma_{\rm HI}\rangle$ models of the H I Lyα forest at a single redshift bin. We effectively treat λ_0 and $\langle\Gamma_{\rm HI}\rangle$ as independent parameters. As a consequence, the actual mean free path $\lambda_{\rm mfp,HI}$ in our model can be different from the mean free path parameter λ_0. Hence we recalculate $\lambda_{\rm mfp,HI}$ for each $\lambda_0 - \langle\Gamma_{\rm HI}\rangle$ model and use the $\lambda_{\rm mfp,HI} - \langle\Gamma_{\rm HI}\rangle$ parameter space to measure mean free path and photo-ionization rate in the observations. We refer the reader to §5.2 for details. This approach of generating $[\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle]_{\rm EX-CITE}$ models independent of $\langle\Gamma_{\rm HI}\rangle$ allows us to efficiently explore a large parameter space.

All the $[\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle]_{\rm EX-CITE}$ models are generated for 64^3, 128^3, 256^3 and 512^3 grids. The size of $[\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle]_{\rm EX-CITE}$ grid cells ($\sim 300h^{-1}$ckpc) is larger than the smallest mean free path ($\sim 30h^{-1}$ ckpc) assumed in this work. We restrict our models to a maximum grid size of 512^3 in order to probe the large parameter space with our available computational resources. Table 2 shows the CPU time consumption per model for various refinement levels. In total we use $744 \times 1836 \sim 1.36$ million CPU hours to run all the models presented in this work The resolution of our $\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle$ fields obtained is a factor 4 and 2 times higher than that in Davies et al. (2018); Nasir & D'Aloisio (2020), respectively. In EX-CITE, we feed back the output of lower grids to higher grids as initial guess. Hence the CPU time required in subsequent step scale non-linearly with grid size as the maps converges faster requiring less number of iterations. The number of models simulated here is larger by a factor of 200 than in previous works. While the $\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle$ fields are generated on 512^3 grids, we linearly interpolate them on 2048^3 grids when extracting skewers from the simulation box. The other fields from the simulation box such as density and velocity are gridded on a 2048^3 grid. The interpolation of the $\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle$ fields is necessary to match the observed resolution of the Lyα forest. We have checked the effect of using such an interpolation on the Lyα forest spectra in simulations. We find that as long as the resolution is as high or higher than that of the 512^3 simulation, the statistics of the Lyα forest are converged within 2 percent. We refer the reader to online appendix C for a detailed comparison of $\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle$ and $f_{\rm HI}$ fields with $N_{\rm Grid,\Gamma_{\rm HI}} = 64, 128, 256, 512$ and 1024. Fig. C4 and C5 shows that our default model that uses $N_{\rm Grid,\Gamma_{\rm HI}} = 512$ shows good convergence of the $\Gamma_{\rm HI}/\langle\Gamma_{\rm HI}\rangle$ and $f_{\rm HI}$ fields.

5.2 Calculating the true mean free path $\lambda_{\rm mfp,HI}$ as for observations

In the previous section, the fluctuations in the ionizing radiation field are parameterized by the free parameter λ_0 while the strength of the ionizing radiation field is set by our second free parameter, $\langle\Gamma_{\rm HI}\rangle$. In our approach λ_0 is a convenient parameter to generate the fluctuations in the $\Gamma_{\rm HI}$ field that accounts for the source location, source properties and density distribution in our simulation box. As discussed in §3), the mean free path parameter λ_0 will be different from the

true mean free path $\lambda_{\rm mfp,HI}$. Hence we need to calculate the true mean free path $\lambda_{\rm mfp,HI}$ for a given λ_0-$\langle\Gamma_{\rm HI}\rangle$ parameter combination. In other words, we transform from λ_0-$\langle\Gamma_{\rm HI}\rangle$ parameter space to $\lambda_{\rm mfp,HI}$-$\langle\Gamma_{\rm HI}\rangle$ parameter space.

First, we calculate the Lyman continuum optical depth along large number of skewers (2048^2) in our simulation box as

$$\tau_{\rm Lyc} = \int n_{\rm HI} \ \sigma_{\rm HI,ion} \ dx, \tag{5}$$

where $n_{\rm HI}, \sigma_{\rm HI,ion} = 6.34 \times 10^{-18}$ cm^2 are H I number density and H I photo-ionization cross-section, respectively (Verner et al. 1994).

We calculate $\tau_{\rm Lyc}$ along all skewers using cumulative summation. We then convert $\tau_{\rm Lyc}$ to Ly-continuum flux as $F_{\rm Lyc} = e^{-\tau_{\rm Lyc}}$. The mean Ly-continuum transmission is then calculated by averaging 2048^2 $F_{\rm Lyc}$ profiles. This averaging operation is equivalent to the method of stacking QSO spectra at 912 Å in observations. The average Ly-continuum transmission profile is fitted with two free parameters ($F_0, \lambda_{\rm mfp,HI}$),

$$\langle F_{\rm Lyc}\rangle(x) = F_0 \exp\left[-\frac{x}{\lambda_{\rm mfp,HI}}\right], \tag{6}$$

where x is the distance in h^{-1}cMpc, F_0 is the normalization of the profile and $\lambda_{\rm mfp,HI}$ is the actual mean free path of H I ionizing photons in the simulation box. We use this $\lambda_{\rm mfp,HI}$-$\langle\Gamma_{\rm HI}\rangle$ parameter space to measure mean free path and photo-ionization rate in observations. The mapping from parameters $\lambda_0 - \Gamma_{\rm HI}$ to $\lambda_{\rm mfp,HI}$-$\langle\Gamma_{\rm HI}\rangle$ is non-linear. This introduces a physical but presumably somewhat model dependent correlation between $\lambda_{\rm mfp,HI}$ and $\langle\Gamma_{\rm HI}\rangle$. Note that we calculate the mean free path $\lambda_{\rm mfp,HI}$ in a similar way as is usually done in observations. We refer the reader to online appendix B for a detailed discussion on the difference between λ_0 and $\lambda_{\rm mfp,HI}$.

Finally we would like to emphasize that our approach of varying the mean free path independently of $\langle\Gamma_{\rm HI}\rangle$ is based on the widely used heuristic scaling of the mean free path with density and photo-ionization rate given in Eq. 4. This or similar approximations are necessary as the dynamical range of even the largest simulations is not sufficient to capture at the same time the large scales at which ionized bubbles overlap and fully resolve the sinks of ionizing radiation. In reality, mean free path and photo-ionization rate evolve simultaneously as reionization progresses (Haardt & Madau 2012). However, our approach is efficient in exploring the unknown large parameter space. Furthermore, we also ignore the wavelength dependence of Lyman continuum opacity in Eq. 5 that could be important at large mean free paths (see Worseck et al. 2014, for details). In §5.7, we demonstrate that our method can recover the $\lambda_{\rm mfp,HI}$ and $\langle\Gamma_{\rm HI}\rangle$ parameters from our fiducial *Aton* radiative transfer simulation.

5.3 Calculating the neutral fraction $f_{\rm HI}$

Spatial fluctuations in $\Gamma_{\rm HI}$ amplify the fluctuations in the neutral fraction $f_{\rm HI}$ that is crucial for the Lyα optical depth

calculation. We calculate the neutral hydrogen fraction f_{HI} for a given $[\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle]_{\mathrm{EX-CITE}}$ model and $\langle \Gamma_{\mathrm{HI}} \rangle$ value as,

$$f_{\mathrm{HI}} = \frac{\mu_e \, n_H \, \alpha_{\mathrm{HI}}(T)}{\langle \Gamma_{\mathrm{HI}} \rangle \times [\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle]_{\mathrm{EX-CITE}}}, \qquad (7)$$

where $\mu_e = \left[(1 - Y) \, f_{\mathrm{HII}} + Y/4 \, (f_{\mathrm{HeII}} + 2 \, f_{\mathrm{HeIII}})\right]/(1 - Y)$ is the mean molecular weight of electrons ($Y = 0.24$) for singly ionized helium ($f_{\mathrm{HeIII}} \sim 0$, $f_{\mathrm{HeII}} \sim 1$) and hydrogen ($f_{\mathrm{HII}} \sim 1$), n_{H} is the hydrogen number density, $\alpha_{\mathrm{HI}}(T)$ is the recombination rate coefficient of H I, $[\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle]_{\mathrm{EX-CITE}}$ is the photo-ionization rate fluctuation field generated using EX-CITE and $\langle \Gamma_{\mathrm{HI}} \rangle$ is the spatially averaged photo-ionization rate to be measured. It is important to note that above equation can give $f_{\mathrm{HI}} > 1$, especially in the regions where Γ_{HI} is very small. So while applying above equation we impose a time scale criterion that cells with $\Gamma_{\mathrm{HI}}^{-1} > t_{\mathrm{Hubble}}$ are neutral $f_{\mathrm{HI}} = 1$ where t_{Hubble} is the Hubble time. This ensures that the neutral fraction ≤ 1.

In Eq. 7, we assume photo-ionization equilibrium. The photo-ionization equilibrium approximation is valid when photo-ionization (t_{ion}) and recombination time scales ($f_{\mathrm{HI}} \, t_{\mathrm{rec}}$) are comparable to each other. The regions well inside ionized bubbles typically satisfy this condition. Regions within the ionization front have large gradients in the photo-ionization rates and photo-ionization equilibrium is then not a good approximation. However, the volume filling factor of ionization fronts is small compared to that of ionized bubbles or neutral regions. Hence for practical purposes, photo-ionization equilibrium is overall a good approximation when calculating volume-weighted neutral fractions. The assumption of photo-ionization equilibrium has been shown to be reasonably correct in self-consistent radiative transfer simulations at $z < 5$ (Molaro et al. 2022). From the f_{HI} field, it is straightforward to calculate the H I Lyα optical depth by integrating, $\tau_{\mathrm{HI}} = \int n_{\mathrm{H}} \, f_{\mathrm{HI}} \, \sigma_{\mathrm{Ly}\alpha}(\nu, b, v) \, dx$, along sightlines where n_{H} is the number density of total hydrogen and $\sigma_{\mathrm{Ly}\alpha}$ is the Lyα absorption cross-section. The observable field, the transmitted Lyα flux, is then obtained as $F = e^{-\tau_{\mathrm{HI}}}$.

5.4 Thermal parameter variation

The H I Lyα optical depth (τ_{HI}) and neutral fraction (f_{HI}) depend explicitly on the temperature field through the Doppler parameter (b) and the recombination rate coefficient. Uncertainty in the modelling of the temperature field lead to variations in τ_{HI} and translate into uncertainty in the measurement of $\lambda_{\mathrm{mfp,HI}} - \langle \Gamma_{\mathrm{HI}} \rangle$. In homogeneous reionization models, the temperature and density are well correlated as a power-law $T = T_0 \Delta^{\gamma - 1}$ (at $\Delta < 10$) where T_0 is the normalization and γ is the slope of the temperature-density relation (TDR, Hui & Gnedin 1997). Due to the fluctuations in Γ_{HI} in inhomogeneous reionization models, one expects to see significant scatter in temperature for a given density (Keating et al. 2018; Gaikwad et al. 2020; Nasir & D'Aloisio 2020). Regions that ionize early are expected to have a steeper TDR with lower T_0 as these regions experience significant cooling due to Hubble expansion. On the other hand recently ionized regions are expected to have a flatter TDR with higher T_0 as the photo-heating is independent of density and there is not enough time for Hubble, Compton and collisional excitation cooling to take effect

(Puchwein et al. 2019). To model these spatial fluctuations of the TDR, one needs to evolve the ionization and thermal state of the IGM using the emissivity evolution of sources. However, our approach is static in the sense that we only model the spatial fluctuations in Γ_{HI} at a given redshift. The approach of evolving the emissivity, even though more physical, has practical difficulties. It is computationally expensive, the spectral energy distribution of sources is uncertain and it is not straightforward to vary $\lambda_{\mathrm{mfp,HI}}$ in this approach.

In order to obtain a temperature field, we use a 'two zone model' (neutral and ionized zones). We define the neutral zone as a region with $\Gamma_{12,\mathrm{HI}} < 10^{-1.6}$, while ionized regions correspond to $\Gamma_{12,\mathrm{HI}} \geq 10^{-1.6}$ (see Cain et al. 2021, for a similar definition). This choice of $\Gamma_{12,\mathrm{HI}} = 10^{-1.6}$ corresponds to $\tau_{\mathrm{Ly}\alpha} > 15$ i.e., the pixels in these regions are in the saturated parts of the spectrum. Our definition of neutral and ionized zones thus corresponds to whether the pixels are in the saturated or in the transmission spike regions of the spectra. The photo-ionization rate in neutral regions is small and they have not experienced any photo-heating due to reionization. We assume the temperature of the gas to be similar to the CMB temperature (~ 20 K) for neutral regions. We have checked that even if the temperature in neutral regions is ~ 1000 K, the properties of the Lyα forest at $z < 6$ are not significantly different. This is because the neutral region produce dark troughs in the spectra and do not affect the overall flux level of the Lyα transmission.

For ionized regions that have experienced photo-heating due to reionization, we assume that the gas follows a TDR. We thereby assume 10 percent scatter in the TDR similar to the scatter seen in homogeneous UVB simulations. Effectively, we assume that the gas in ionized regions is ionized at the same time. The TDR is described by the normalization (T_0) and slope (γ), that are free parameters in our approach. We have chosen a wide range of thermal parameters consistent with recent measurements of Gaikwad et al. (2020) as discussed in online appendix D (see Fig. D1). In particular, we chose three combination of thermal parameters: (i) default T_0, γ evolution, (ii) $T_0 - \delta T_0$, $\gamma + \delta\gamma$ and (iii) $T_0 + \delta T_0$, $\gamma - \delta\gamma$. The parameter combination $T_0 + \delta T_0$, $\gamma - \delta\gamma$ corresponds to a model where all the gas in ionized regions are recently ionized. The parameter combination $T_0 - \delta T_0$, $\gamma + \delta\gamma$ corresponds to earlier ionized regions that are able to cool down because of Hubble expansion. The two zone model does not account for the shock heating of the gas due to structure formation (Puchwein et al. 2023). The effect of shock heating mostly affects the high density IGM with $\Delta > 100$. The Lyα forest at $z > 5$ on the other hand is mostly sensitive to $\Delta < 10$ (Gaikwad et al. 2020). The effect of shock heating on our measurements is thus small. Our two-zone model of temperature is a somewhat extreme case where we assume that all the cells in ionized regions are ionized either very recently (hot case, $T_0 + \delta T_0, \gamma - \delta\gamma$) or ionized very early (cold case, $T_0 - \delta T_0, \gamma + \delta\gamma$). We assume a fairly large uncertainty on the thermal parameters T_0 and γ. Scatter in the temperature density relation, will lead to some cells being ionized early and some ionized late depending on the timing of ionization. The parameter constraints from such temperature fluctuation model would lie somewhere between the two extremes where all cells are assumed to be ionized either early or late. We show in §5.7 that our

two zone model is reasonable and can recover $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$ even when there is scatter in the temperature density relation.

We have chosen thermal parameter combinations that produce the largest possible uncertainty in the $\langle \Gamma_{\mathrm{HI}} \rangle$ $-\lambda_{\mathrm{mfp,HI}}$ parameters. As discussed in §5.1, at each redshift we generate $108 \times 81 = 8748$ models. Since we are using three combinations of thermal parameters in our models, we effectively generate $108 \times 81 \times 3 = 26244$ models at any given redshift. All the 1σ constraints shown in this work account for the uncertainty in thermal parameters. The final measurements presented in this work are marginalized over the uncertainty in observed T_0 and γ values (see §6 for details).

5.5 Lyα forest statistics

We forward model the simulated Lyα forest spectra such that they match the properties of the observational sample as discussed in §2. Our approach of forward modeling the Lyα forest and generating mocks is similar to that described in Gaikwad et al. (2017a, 2018). For a given redshift range, we use the same redshift path length as the observed sample. We do not need to concatenate lines of sight as our box size ($160\,h^{-1}$ cMpc, $\delta z > 0.2$ at $5 < z < 6$) is larger than the redshift range over which parameters are measured. For a given observed redshift bin, we extract the skewers from the nearest available Sherwood simulation snapshot. We convolve the simulated flux with a Gaussian line spread function of given full width at half maximum. For the observed XShooter and ESI spectra the FWHM of the line spread function (LSF) depends on observing conditions. The LSF is determined for each spectrum individually (see §2 for details). The FWHM typically varies from 22 to 39 km s^{-1} (Davies et.al .in prep) for XShooter and 40 to 60 km s^{-1} for ESI. We use the respective Gaussian FWHM for each of the spectra in our modeling. The wavelengths in the simulated spectra are resampled with the same spacing as in the observations. Finally, uncorrelated Gaussian noise is added to the spectra using the S/N per pixel array from observations. We generate $n \times 1000$ simulated spectra in a given redshift range, where n is the number of observed spectra in the same redshift range. We generate 1000 mock samples each containing n spectra that mimic the observed sample. The Lyα forest statistics is then calculated for all the 1000 mock samples individually.

We derive and compare three statistics of the Lyα forest from simulations with observations, (i) the cumulative distribution function of effective optical depth (hereafter $\tau_{\mathrm{eff,HI}}$ CDF), (ii) the cumulative distribution function of dark gap lengths (hereafter dark gap statistics) and (iii) the pseudo-Column Density Distribution Function (hereafter pCDDF). We use the $\tau_{\mathrm{eff,HI}}$ CDF to measure $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$. The dark gap statistics and pCDDF are then used to further check the consistency of our best fit models with the observations. The accuracy of the $\tau_{\mathrm{eff,HI}}$ measurements are usually limited by the noise properties of the spectra. We calculate the effective optical depth for each sightline as $\tau_{\mathrm{eff,HI}} = -\ln\langle F \rangle$, where $\langle F \rangle = \langle F_{\mathrm{unnorm}}/F_{\mathrm{cont}} \rangle$ is the mean of the normalized flux. The $\tau_{\mathrm{eff,HI}}$ uncertainties are calculated by considering the uncertainties of the mean flux. The uncertainties of the mean flux also account for the uncertainty in continuum placement. The final uncertainties of

the measured $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$ accounts for the uncertainty in observed $\tau_{\mathrm{eff,HI}}$ (and hence continuum). Cases where we do not detect significant transmitted flux compared to the noise level, we treat as non-detections and calculate $\tau_{\mathrm{eff,HI}}$ using twice the mean flux uncertainty (Becker et al. 2015; Bosman et al. 2018, 2022).

The $\tau_{\mathrm{eff,HI}}$ CDF is one of the most robust statistics that can be derived from Lyα forest spectra. Since $\tau_{\mathrm{eff,HI}}$ is calculated by taking the mean of the transmitted flux along the sightline, the detailed information within the spectra such as number and height of transmission spikes and the occurrence of dark gaps are not captured explicitly. As a further consistency check, the best fit model that matches the $\tau_{\mathrm{eff,HI}}$ CDF should thus also match the statistics of transmission spikes and dark gaps. We use dark gap and pCDDF statistics to perform such consistency checks (see §6.6). We derive the pCDDF and dark gap statistics in a way similar to that described in Gaikwad et al. (2020) and Zhu et al. (2021) respectively (see online appendix E for details). We use all the observed spectra (XShooter + ESI) to calculate the $\tau_{\mathrm{eff,HI}}$ CDF and the dark gap statistics. However, for the pCDDF statistics we use only XShooter spectra. This is because Voigt profile decomposition (and hence the pCDDF) of transmission spikes is sensitive to the resolution of the spectra. In order to homogenize the data set, we prefer to use the better resolution and larger sample of the XShooter data for the pCDDF. We follow identical procedures to derive Lyα forest statistics in simulations and observations.

We now discuss how the statistical properties of the Lyα forest spectra are affected by the choice of $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$. Fig. 2 shows the effect of changing $\lambda_{\mathrm{mfp,HI}}$ on Γ_{HI} fluctuations (i.e, $\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle$, top panel) and H I neutral fraction (bottom panel) fields. If $\lambda_{\mathrm{mfp,HI}}$ is small, high Γ_{HI} values are confined to smaller regions around the ionizing sources. As the mean free path increases, regions with higher Γ_{HI} extend to larger distances away from the sources. For larger mean free path, the $\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle$ fields look more homogeneous compared to corresponding fields with shorter mean free path. This is expected because as the mean free path increases, the models approach the limit of a uniform UVB model, where the photo-ionization rate is spatially homogeneous. It is also evident from Fig. 2 that $\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle$ changes gradually away from the sources. The morphology of the f_{HI} fields are closely related to that of the $\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle$ fields. The regions with higher $\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle$ are at average more ionized hence have smaller f_{HI} and vice versa. The spatial fluctuations in f_{HI} as shown in the bottom row of Fig. 2 are mainly responsible for the large scatter in the observed properties of the H I Lyα forest.

In Fig. 3, we compare the line of sight flux from models with variation in $\langle \Gamma_{\mathrm{HI}} \rangle$ (top) and $\lambda_{\mathrm{mfp,HI}}$ (middle and bottom panels) parameters. The main effect of varying $\langle \Gamma_{\mathrm{HI}} \rangle$ (for fixed value of $\lambda_{\mathrm{mfp,HI}}$) is the change in transmitted flux along the sightlines. The location of transmission spikes remains relatively similar. On the other hand, with increasing $\lambda_{\mathrm{mfp,HI}}$, the probability of a sightline intersecting an ionized region increases. As a result, additional transmission spikes appear at several new locations (middle panel). This is also the case if we vary $\langle \Gamma_{\mathrm{HI}} \rangle$ such that the mean flux of the mock samples is matched (bottom panel in Fig. 3). A direct consequence of this variation of flux with $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$ is responsible for the scatter in $\tau_{\mathrm{eff,HI}}$ as shown in Fig. 4.
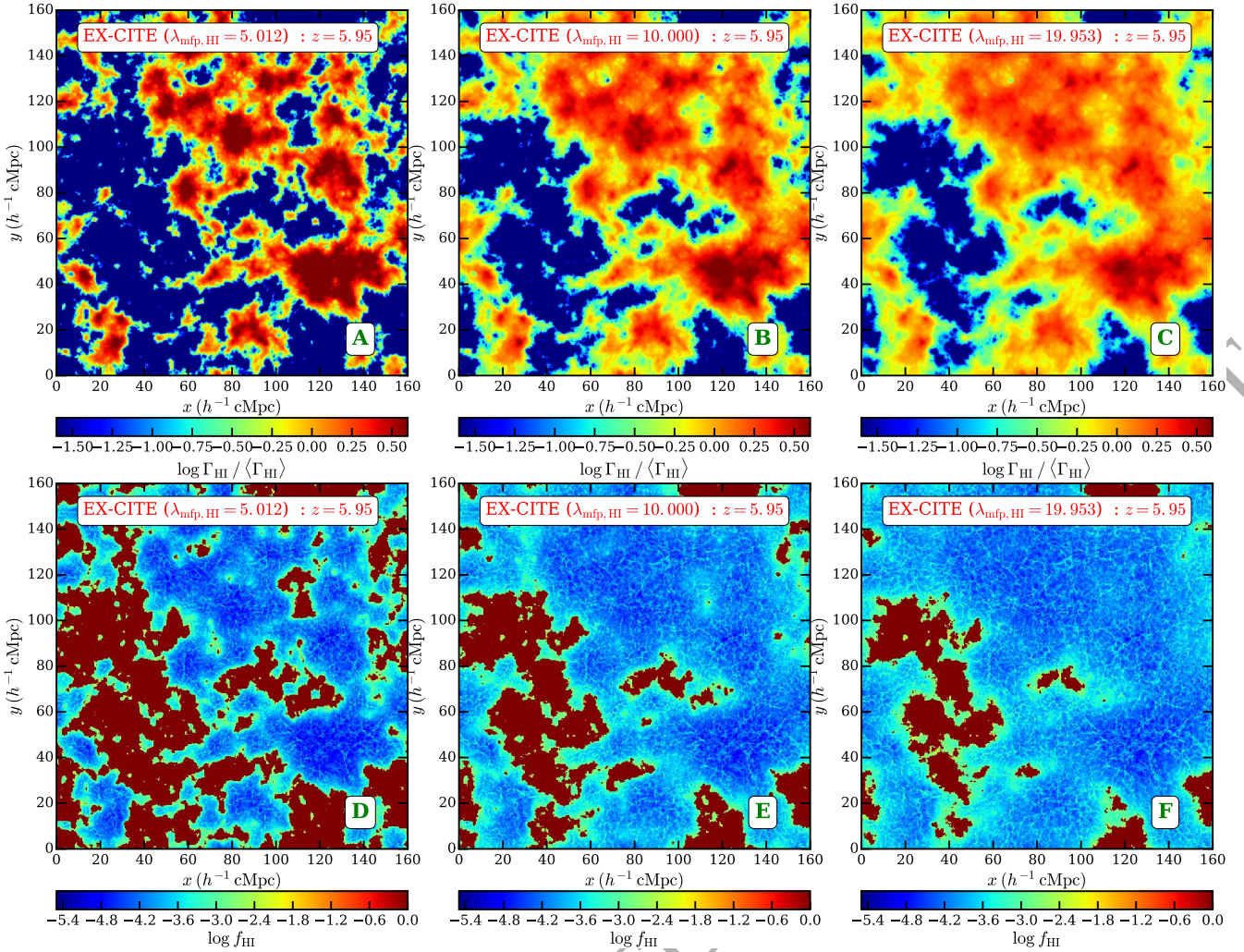
Figure 2. Panel A, B and C show comparisons of $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ slices for three different values of mean free path, $\log\lambda_{\mathrm{mfp,HI}} = 0.7, 1.0$ and 1.3. The mean free path has been varied using EX-CITE and the fields are calculated with $N_{\mathrm{Grid},\Gamma_{\mathrm{HI}}} = 512$. Panel D, E and F shows the corresponding comparison of the neutral hydrogen fractions f_{HI}. While calculating f_{HI}, we assume identical values for $\langle\Gamma_{\mathrm{HI}}\rangle$, T_0 and γ. All the slices are shown for the L160N2048 simulation at $z = 5.95$ for a slice with a thickness of $0.3125\ h^{-1}$ cMpc. With the increase in mean free path the ionizing radiation field percolates more in to the low-density IGM. As a result, the $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ field extends to larger distances. The ionizing radiation field gradually decreases from the center of the ionized regions outwards. The attenuation between sources and cell is calculated using octree summations. The directional dependence of the ionizing radiation field relative to the source locations is thus preserved in EX-CITE. In the regions with larger $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$, the IGM is highly ionized leaving small neutral fractions of $\sim 10^{-4}$. The regions that are yet to receive ionizing radiation are still neutral. The fluctuations in ionizing radiation field and neutral fraction lead to the large scatter in the $\tau_{\mathrm{eff,HI}}$ distribution that is observed and that optically thin (uniform UVB) models fail to reproduce. All the EX-CITE models shown in this figure assume $\langle\Gamma_{\mathrm{HI}}\rangle = 10^{-13}\ \mathrm{s}^{-1}$, the same value as in our fiducial *Aton* simulation.

Fig. 4 shows the sensitivity of the $\tau_{\mathrm{eff,HI}}$ CDF to $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$. The left panel illustrates that the $\tau_{\mathrm{eff,HI}}$ CDF is systematically shifted such that the median $\tau_{\mathrm{eff,HI}}$ is lower for a model with higher $\langle\Gamma_{\mathrm{HI}}\rangle$. However, the shape of the $\tau_{\mathrm{eff,HI}}$ CDF remains similar. This is because a change in $\langle\Gamma_{\mathrm{HI}}\rangle$ changes the flux in a systematic way, while the location at which transmission spikes occur remain the same. On the other hand the middle panel in Fig. 4 shows that the shape and median of $\tau_{\mathrm{eff,HI}}$ are both affected by changes in $\lambda_{\mathrm{mfp,HI}}$ for constant $\langle\Gamma_{\mathrm{HI}}\rangle$. To better understand this, we vary $\langle\Gamma_{\mathrm{HI}}\rangle$ in the right panel such that the mean flux for the three models is constant. We find that the model with small $\lambda_{\mathrm{mfp,HI}}$ shows more scatter in $\tau_{\mathrm{eff,HI}}$ and vice versa while the mean flux is the same. This is expected because

the probability of transmission spikes occurring along different sightlines varies significantly in the model with lower $\langle\lambda_{\mathrm{mfp,HI}}\rangle$. The transmission spikes occur more uniformly and are distributed evenly along different sightlines in the larger $\lambda_{\mathrm{mfp,HI}}$ model. Thus, the median of the $\tau_{\mathrm{eff,HI}}$ distribution is mainly sensitive to $\langle\Gamma_{\mathrm{HI}}\rangle$, while the shape of the $\tau_{\mathrm{eff,HI}}$ CDF is primarily sensitive to $\lambda_{\mathrm{mfp,HI}}$. Because of these two effects we can use the $\tau_{\mathrm{eff,HI}}$ CDF to measure $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ from the observations. Similar to $\tau_{\mathrm{eff,HI}}$ CDF, we show the sensitivity of pCDDF and dark gap statistics in online appendix E. In summary, we find that the normalization and shape of the pCDDF are sensitive to $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$, respectively. The median and shape of the dark gap length CDF are likewise sensitive to $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$.
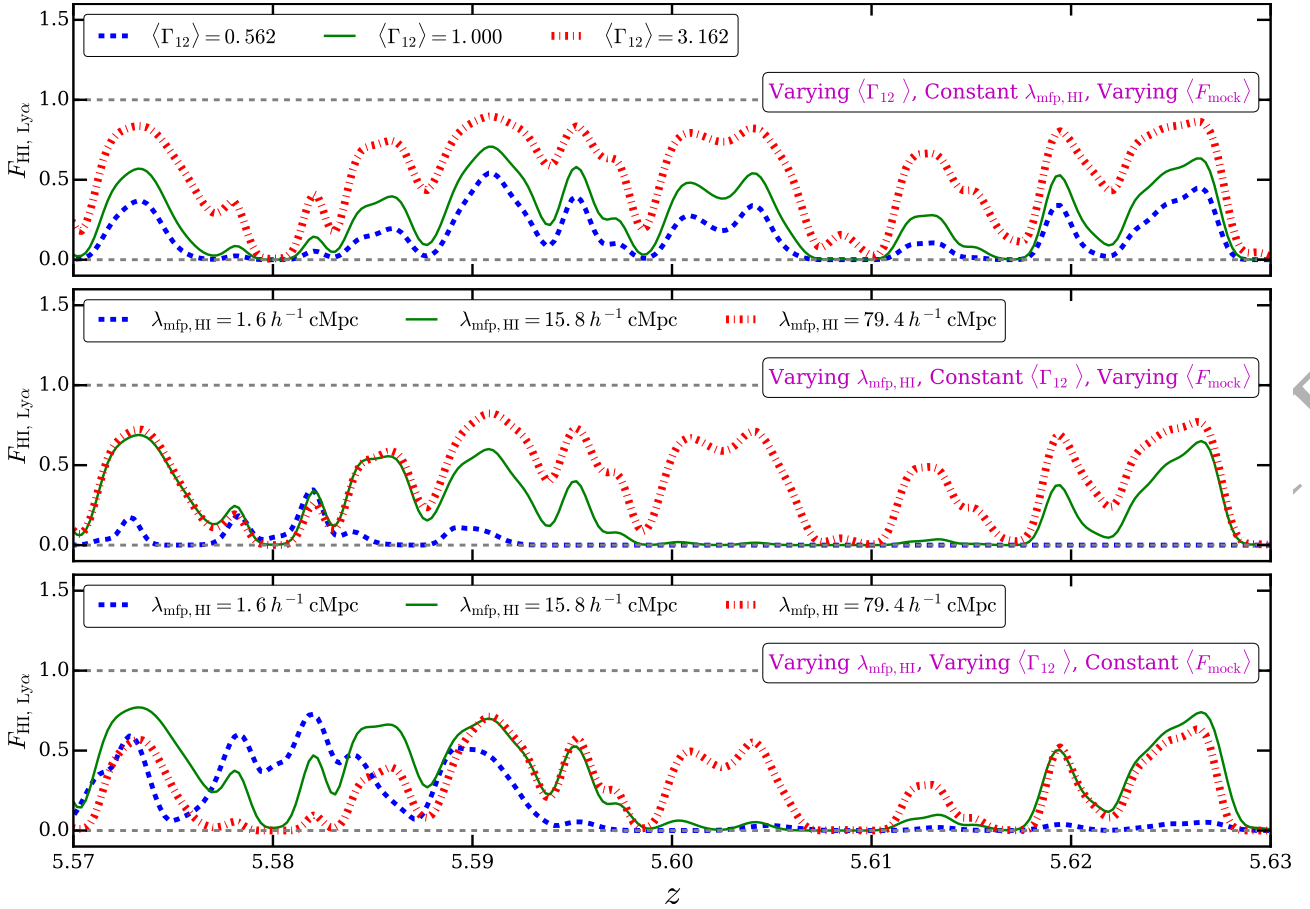
Figure 3. The top and middle panel show the sensitivity of H I Lyα forest spectra to H I photo-ionization rate and mean free path, respectively, keeping the other parameters fixed. The bottom panel shows the effect of varying the mean free path on a H I Lyα forest spectrum where $\langle \Gamma_{HI} \rangle$ is varied in such a way that the mean flux of the mock spectra is constant for the three models. The top panel shows how with increasing $\langle \Gamma_{HI} \rangle$, the flux level increases as the IGM is more ionized. Note that the location of the transmission spikes remains relatively unchanged. The middle panel illustrates that with increasing mean free path, the sightline intersects more frequently higher $\langle \Gamma_{HI} \rangle$ regions leading to a larger number of transmission spikes at several locations. The bottom panel qualitatively illustrates that even if the mean flux of the mock sample is the same, the number of transmission spikes and their clustering is sensitive to changes in the mean free path. For illustration purposes, the spectra in this figure are shown without adding any noise. Note further that the quantitative results presented in this paper are using simulations with noise properties similar to observations.

5.6 Method of parameter estimation

In this section, we describe our method of estimating / recovering $\langle \Gamma_{HI} \rangle$ and $\lambda_{mfp,HI}$ using the statistics of $\tau_{eff,HI}$. In the literature, parameters have been estimated from Lyα forest statistics by using the probability distribution function (Bolton et al. 2008; Rollinde et al. 2013; Gaikwad et al. 2021). One can calculate and use the PDF of $\tau_{eff,HI}$ to estimate $\langle \Gamma_{HI} \rangle$ and $\lambda_{mfp,HI}$. There are two main challenges; first the number of sightlines in a given redshift bin are limited (< 40) and second many of the $\tau_{eff,HI}$ measurements at $z > 5.6$ are lower limits due to non-detections. In principle, we can still use kernel density estimation (KDE) to determine the $\tau_{eff,HI}$ PDF. However, the shape of the PDF will be affected by the functional form of the KDE function and the choice of smoothing parameters for small samples. Estimating the error of the $\tau_{eff,HI}$ PDF is challenging because of the non-detections. We also find that using KDE can artificially reduce the scatter in the $\tau_{eff,HI}$ distribution which is undesirable as $\lambda_{mfp,HI}$ is sensitive to the $\tau_{eff,HI}$ scatter. To

circumvent these difficulties we chose to estimate $\langle \Gamma_{HI} \rangle$ and $\lambda_{mfp,HI}$ using non-parametric tests.

The comparison of the $\tau_{eff,HI}$ CDFs can be performed with non-parametric tests such as the Kolmogorov-Smirnov (hereafter KS) or Anderson-Darling test (hereafter AD). These tests do not make any assumption about the underlying distribution of $\tau_{eff,HI}$. The KS and AD statistics are more suitable for small samples. Furthermore, the tests are sensitive to both the median and the shape of the distribution. The KS test is mostly sensitive to the middle of the CDF and less sensitive to the tail of the distribution. The AD test is equally sensitive to middle and tail of the distribution (Press et al. 1992). As discussed in §5.5, it is important to use statistics that are sensitive to the median as well as the scatter of the $\tau_{eff,HI}$ distribution since both properties of the CDF are sensitive to $\langle \Gamma_{HI} \rangle$ and $\lambda_{mfp,HI}$.

We generate a total of 26244 $\lambda_{mfp,HI}$-$\langle \Gamma_{HI} \rangle$ models for the 108 Γ_{HI} fluctuation maps, 81 $\langle \Gamma_{HI} \rangle$ values and 3 ther-
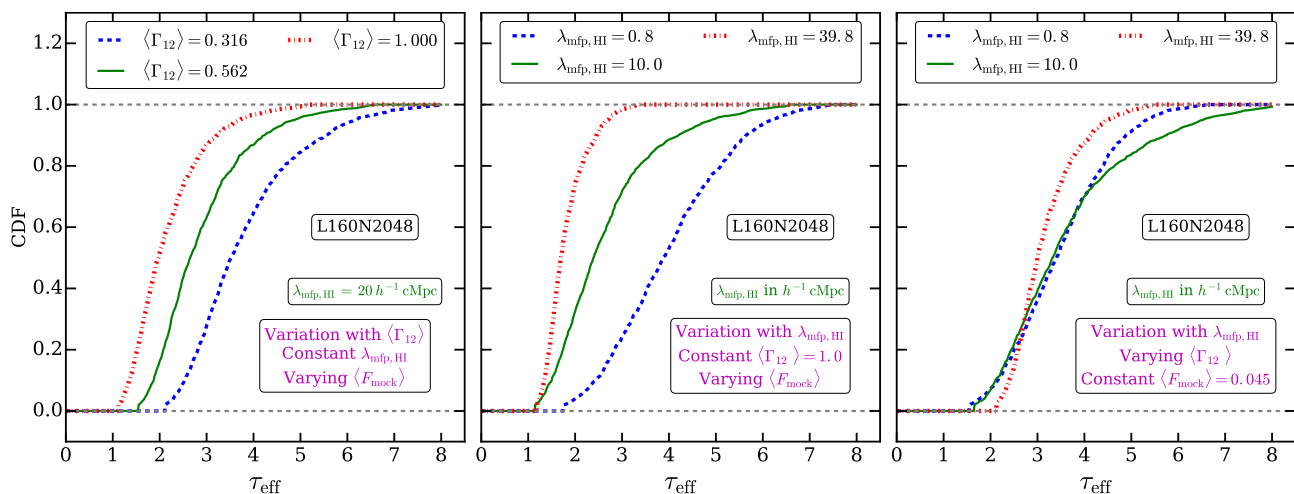
Figure 4. The left and middle panel show the sensitivity of the $\tau_{\mathrm{eff,HI}}$ CDF to $\langle \Gamma_{\mathrm{HI}} \rangle$ and $\lambda_{\mathrm{mfp,HI}}$ at $5.5 \leq z \leq 5.7$, respectively, while keeping the other parameters fixed. The right panel shows the variation of the $\tau_{\mathrm{eff,HI}}$ CDF with $\lambda_{\mathrm{mfp,HI}}$ when $\langle \Gamma_{\mathrm{HI}} \rangle$ is varied such that the mean flux of the mock absorption spectra is constant for the three models. With increasing $\langle \Gamma_{\mathrm{HI}} \rangle$, the $\tau_{\mathrm{eff,HI}}$ CDF systematically shifts to lower values as the neutral fraction decreases. The shape of the $\tau_{\mathrm{eff,HI}}$ CDF remains similar. The middle and right panel illustrate that the increase in $\lambda_{\mathrm{mfp,HI}}$ makes the $\tau_{\mathrm{eff,HI}}$ CDF narrower. The scatter in $\tau_{\mathrm{eff,HI}}$ CDF is due to the density and photoionization rate fluctuations. With increasing $\lambda_{\mathrm{mfp,HI}}$, the ionizing radiation field morphology becomes more homogeneous and uniform reducing the scatter in $\tau_{\mathrm{eff,HI}}$ (see Fig. 2). It is noteworthy that for small $\lambda_{\mathrm{mfp,HI}}$ values, the high $\tau_{\mathrm{eff,HI}}$ tail end of the CDF is significantly affected while the small $\tau_{\mathrm{eff,HI}}$ end are relatively similar. We assume noise properties similar to the observations.

mal parameter combinations as described in §5.1 and §5.4[3]. For each model (i.e., $\lambda_{\mathrm{mfp,HI}}$-$\langle \Gamma_{\mathrm{HI}} \rangle$ parameter combination), we generate 1000 mock samples by extracting Lyα forest spectra along random skewers. Each mock sample contains the same number of spectra as the observations for a given redshift bin ensuring the redshift path length is the same in the two cases. Each mock sample mimics the observations as discussed in §5.5.

In order to obtain the best fit parameters and their uncertainties, we use the null hypothesis that the $\tau_{\mathrm{eff,HI}}$ distribution in each mock sample and the observations have the same underlying distribution. The null hypothesis is rejected at 1σ significance if the AD (or KS) test statistics is greater than its critical value. We can then reformulate the above statement in terms of probability (p) values. The null hypothesis is rejected at 1σ significance if the p value between mock sample and observation is less than a threshold value p_{th}. We chose a value of $p_{\mathrm{th}} = 0.32$ and the null hypothesis is rejected if $p < p_{\mathrm{th}}$. We calculate the p-value between each mock sample and the observations for a given statistic. Since there are 1000 mocks samples for each $\lambda_{\mathrm{mfp,HI}}$-$\langle \Gamma_{\mathrm{HI}} \rangle$ parameter combination, we obtain 1000 p-values. The 50th percentile (median) of the p-distribution (p_{med}) is used to obtain best fit parameters. We assign the median p-value to each grid model (i.e., each $\lambda_{\mathrm{mfp,HI}}$-$\langle \Gamma_{\mathrm{HI}} \rangle$ parameter pair). The best fit value corresponds to the $\lambda_{\mathrm{mfp,HI}}$-$\langle \Gamma_{\mathrm{HI}} \rangle$ parameter that has the maximum p_{med}. The 1σ statistical uncertainty for $\lambda_{\mathrm{mfp,HI}}$-$\langle \Gamma_{\mathrm{HI}} \rangle$ are obtained by drawing contours at the $p_{\mathrm{med}} = 0.32$ level. This means all the models with

---

[3] The three thermal parameter combinations with the same λ_0 $-\langle \Gamma_{\mathrm{HI}} \rangle$ do not necessarily have the same $\lambda_{\mathrm{mfp,HI}}$ $-\langle \Gamma_{\mathrm{HI}} \rangle$. This is because $\lambda_{\mathrm{mfp,HI}}$ depends on the neutral fraction that in turn depends on thermal parameters (see Eq. 7). Thus, all the 26244 $\lambda_{\mathrm{mfp,HI}}$ $-\langle \Gamma_{\mathrm{HI}} \rangle$ parameter combinations are unique.

$p_{\mathrm{med}} > 0.32$ are consistent with the data within 1σ uncertainty. Our approach of computing the best fit parameters and their uncertainties from the $\tau_{\mathrm{eff,HI}}$ distribution is similar to that used by Worseck et al. (2019).
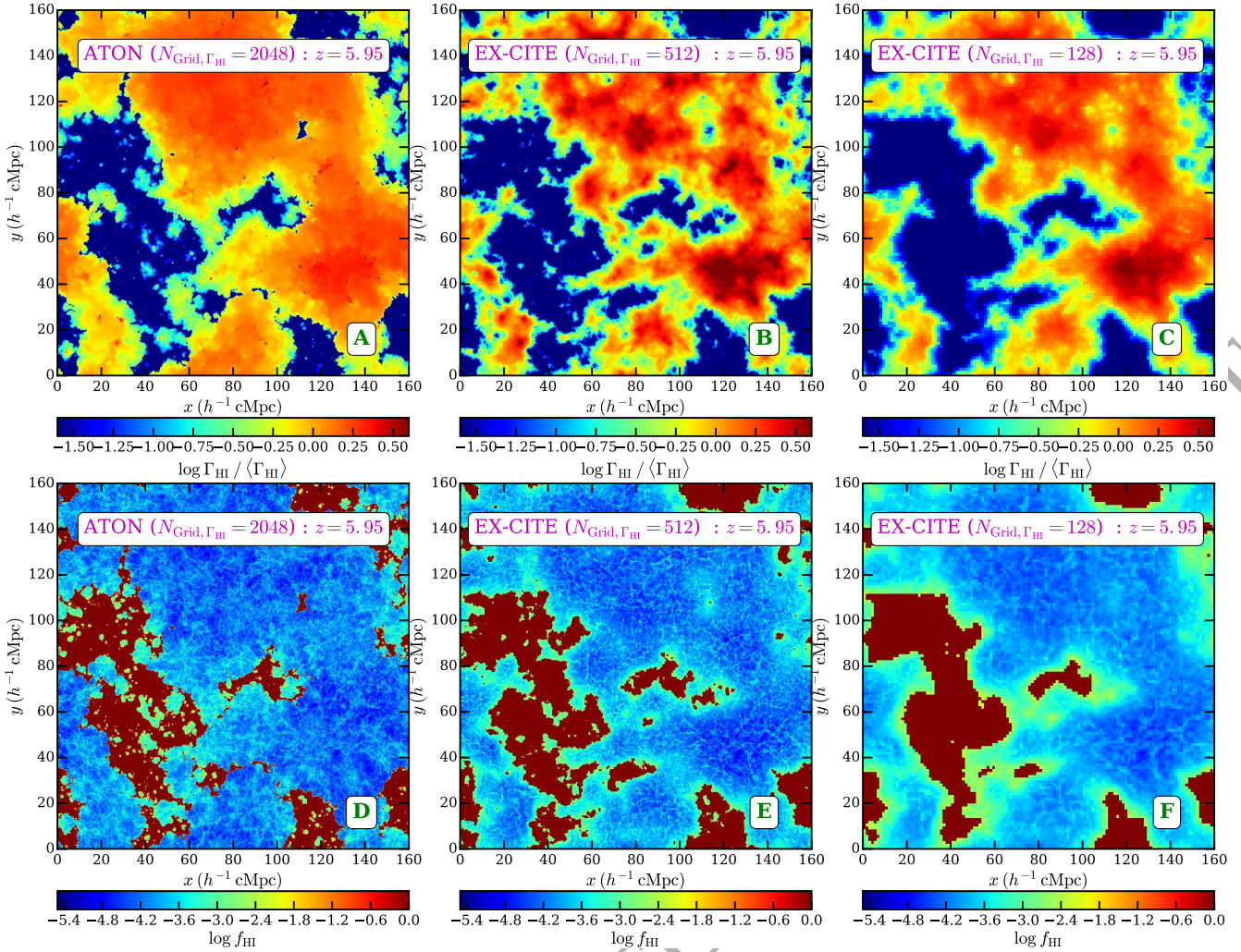
We have tested if our way of defining 1σ contours is statistically meaningful. For this we generate 10000 different fiducial mocks for a fiducial *Aton* simulation. We find that in ~ 68 percent of cases the true value is recovered if we use the criterion $p_{\mathrm{med}} > 0.32$ validating our definition. The 16th and 84th percentile of the p-values is a measure of statistical uncertainty due to cosmic variance (large scale density variation) and astrophysical variance (large scale ionization field fluctuations). Hence the scatter in 1000 p-values is used to estimate the uncertainty due to cosmic variance (see §6 for details). We emphasize that our way of defining the 1σ constraints accounts for the uncertainty in the thermal parameters as the $\lambda_{\mathrm{mfp,HI}}$ $-\langle \Gamma_{\mathrm{HI}} \rangle$ parameter space explicitly depends on T_0 and γ.

5.7 Comparison of EX-CITE modelling with a radiative transfer simulation using *Aton*

Before performing any measurements from observations, it is important to check how our method of capturing spatial fluctuations in photo-ionization rate and neutral fraction compares to radiative transfer (RT) simulations. This should allow us to get a idea of the accuracy of EX-CITE in modeling and recovering the parameters of interest. We compare photo-ionization rate and neutral fraction fields from EX-CITE with that from a radiative transfer simulation with *Aton* in Fig. 5.

In order to do a fair comparison, we consider an EX-CITE model that has similar $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$ parameters to those in our fiducial *Aton* simulation. The initial conditions are identical in the two models. We use $M_{\mathrm{cutoff}} =$

Figure 5. Panel A, B and C show fluctuations in the H I photo-ionization rate ($\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$) for *Aton* ($N_{\mathrm{grid}} = 2048$), EX-CITE ($N_{\mathrm{Grid},\Gamma} = 512$) and EX-CITE ($N_{\mathrm{Grid},\Gamma} = 128$) models respectively. Panel D, E and F are similar to panels A, B and C except that maps of H I fractions are shown. The thickness of the slice shown in panel B (C) is 4 (16) times that of slice A. The color scheme for panel A, B, C (and also D, E, F) are identical to each other for a fair comparison. On large scales, both EX-CITE models are qualitatively similar to the *Aton* models. On small scales, the EX-CITE models show more fluctuations compared to *Aton* models. The EX-CITE model shows a gradual decrease (spatial gradient) in Γ_{HI} away from sources while *Aton* shows a more uniform distribution of Γ_{HI} in ionized regions. This is partly because *Aton* is a moment based radiative transfer code while EX-CITE is similar to a ray-tracing code. Since the angular dependence of the ionizing radiation field is averaged when solving the RT equations in *Aton*, there will be artificial smoothing on small scales. There will also be additional smoothing on small scales due to the global Lax-Friedrich Riemann solver used in *Aton*. To generate f_{HI} fields in EX-CITE models, we assume $\langle\Gamma_{12}\rangle = 0.2$, $T_0 = 12000$ K and $\gamma = 1.1$ similar to that of the *Aton* model at $z = 5.95$. Previous work in the literature using semi-numerical methods generated Γ_{HI} and f_{HI} maps at lower resolution like in panel C ($N_{\mathrm{Grid},\Gamma} = 128$). Such low resolution maps generally over-estimate the mean free path as some neutral regions are absent in the low resolution maps. In this work, all the models are generated with $N_{\mathrm{Grid},\Gamma} = 512$. Our results are well converged for $N_{\mathrm{Grid},\Gamma} = 512$. We refer the reader to online appendix C (Fig. C4 and C5) where we show comparisons of $N_{\mathrm{Grid},\Gamma} = 64, 128, 256, 512$ and 1024 maps. It is important to note that $\langle\Gamma_{\mathrm{HI}}\rangle$ in *Aton* is predicted by using source emissivity models while in our method it is a free parameter. The main motivation of our code EX-CITE is to explore parameter space, not to perform a detailed comparison between moment-based and ray-tracing approaches.

10^9 M$_\odot$, $\beta = 1$ and assume a $T = 50000$ K black body spectrum with one mono-frequency bin (for the average energy calculation), the same as in the *Aton* simulation (Kulkarni et al. 2019). We calculate the thermal parameters T_0, γ from the median temperature-density relation in the *Aton* simulation and use the same in EX-CITE. There are two main differences between the *Aton* and EX-CITE models (i) the two fields in the *Aton* simulation were produced on 2048^3 grids while in EX-CITE these fields are obtained for 512^3 grids and (ii) *Aton* properly accounts for fluctuations in temperature

while EX-CITE assumes a 'two zone model' of temperature as described in §5.4.

Fig. 5 shows a comparison of the Γ_{HI} and f_{HI} fields from *Aton* (left panel) with that from EX-CITE. We show the Γ_{HI} and f_{HI} fields from EX-CITE for 512^3 (middle panel) and 128^3 (right panel) grids. On large scales the Γ_{HI} and f_{HI} from EX-CITE are reasonably similar to that obtained with *Aton*. The large scale size and morphology of ionizing bubbles in the two models look very similar. Such similarity is expected as the large scale cosmic density field, source
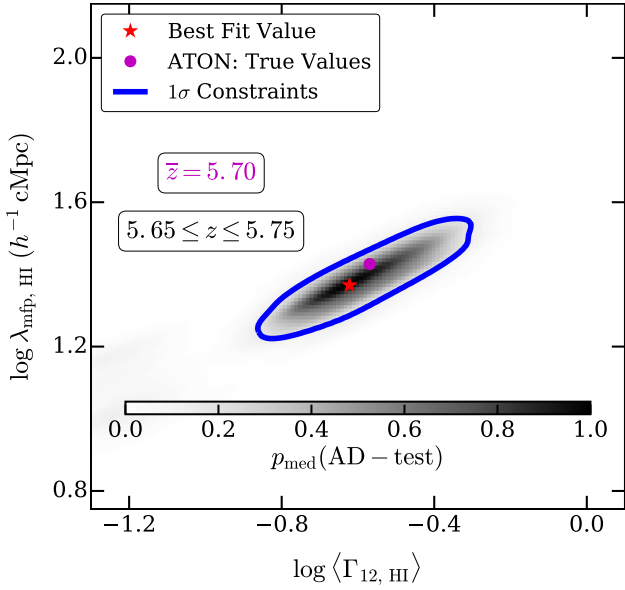
Figure 6. The figure shows the recovery of the H I photo-ionization rate and mean free path of H I ionizing photons using the $\tau_{\mathrm{eff,HI}}$ CDF statistics at $5.65 \leq z \leq 5.75$. We calculate the $\tau_{\mathrm{eff,HI}}$ CDF from the *Aton* model for random sightlines and treating them as our fiducial model. The true value of $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$ in this model are shown by the magenta circle. We use EX-CITE to generate models by varying $\langle \Gamma_{\mathrm{HI}} \rangle$ and $\lambda_{\mathrm{mfp,HI}}$. We compare the $\tau_{\mathrm{eff,HI}}$ CDF between each EX-CITE model and that of our fiducial *Aton* model using Anderson-Darlington (AD) statistics. The color scheme in each panel shows median p-values between EX-CITE model and the fiducial *Aton* model using AD statistics. At all redshifts, the true $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$ (red stars) are recovered within the 1σ contours shown by the blue curve. The total modeling uncertainty (mainly due to thermal parameters) has been accounted for when plotting the blue curve (see Fig. 9 and §6.2 for details). We also tried Kolmogorov-Smirnov statistics (KS) for comparing the $\tau_{\mathrm{eff,HI}}$ CDFs. We find that the AD statistics constrains parameters somewhat better than the KS statistics. This is expected as KS statistics are mostly sensitive to the middle region of the CDF while AD statistics are equally sensitive to the middle and the tail of the CDF. We follow a similar procedure when measuring the $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$ parameters from observations. The recovery of parameters in other redshift bins is shown in Fig. G1 of the online appendix.

distribution, source properties and mean free path of H I ionizing photons are similar in those two models.

On small scales EX-CITE shows more fluctuations in Γ_{HI} than *Aton*. *Aton* computes ionizing radiation fields by taking the angular moments of the radiative transfer equations. The set of moment equations are very similar to fluid equations such that photons are treated like a fluid without gravity. Due to the presence of the radiation pressure tensor in the moment form of the RT equation, the radiation field on small scale appears to be smoother. This is qualitatively very similar to the property of fluids to exert pressure on small scales causing smoothing (Aubert & Teyssier 2008; Rosdahl et al. 2013). In other words, the angular averaging of the radiative transfer equation introduces directional symmetry and smoothing of the ionizing radiation field[4].

Furthermore, the advection equation solved in *Aton* using a Riemann solver can introduce numerical diffusion which adds to the small scale smoothing of the ionizing field. EX-CITE, on the other hand, preserves the angular dependence of the ionizing radiation field. Our method of producing Γ_{HI} fields in EX-CITE is similar to the ray-tracing approach. In EX-CITE's octree implementation, the source contribution at a given location is explicitly calculated from all directions. Regions close to the ionizing sources show large amplitude of Γ_{HI} fluctuations. As one goes away from the sources, the Γ_{HI} fluctuations decrease gradually. Thus, the small scale fluctuations in our Γ_{HI} fields are mainly due to the angular distribution of sources around a given location. We also emphasize that the gradual decrease in Γ_{HI} fluctuations in our model could be partly due to the static rather than dynamical modelling of ionization fronts. It is noteworthy that other approaches using excursion set (Choudhury et al. 2021; Maity & Choudhury 2022a,b) and ray tracing methods (Wu et al. 2021) also find that there are more fluctuations in ionizing radiation fields on small scales. The small scale fluctuations in these models is attributed to source shot noise which is equivalent to the angular dependence of the ionizing radiation (see also Meiksin 2020).

We emphasize that our aim in this work is not to replace self-consistent radiative transfer codes, but rather to understand the difference between them and EX-CITE that can influence our parameter estimation. EX-CITE as developed in this work should be seen as a complementary tool to efficiently model and explore parameter space of inhomogeneous reionization which would be difficult with self-consistent codes like *Ramses-RT*, *Aton*, *Arepo-RT* and hybrid methods as described in Puchwein et al. (2023).

We perform a quantitative comparison between the two methods by recovering the $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$ parameters. The recovery of parameters from self-consistent RT models is important to assess the accuracy of EX-CITE and understanding pppoj the systematics (if any). We generate a fiducial mock sample from the *Aton* simulations at $4.9 < z < 6.1$ assuming properties of the H I Lyα forest similar to that of our observed sample[5]. We forward model the fiducial mock sample consisting of ~ 66 spectra with S/N, resolution and LSF properties similar to observations as discussed in §2 (see also Bosman et al. 2022). We use the same method as described in §5.6 to recover the parameters with the difference that the observed $\tau_{\mathrm{eff,HI}}$ CDF is replaced by the $\tau_{\mathrm{eff,HI}}$ CDF from our fiducial *Aton* simulation.

Fig. 6 and Fig. G1 shows the recovery of $\lambda_{\mathrm{mfp,HI}}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$. The best fit value corresponds to the $\lambda_{\mathrm{mfp,HI}}$-$\langle \Gamma_{\mathrm{HI}} \rangle$ parameter that has maximum p_{med}. The 1σ uncertainty for $\lambda_{\mathrm{mfp,HI}}$-$\langle \Gamma_{\mathrm{HI}} \rangle$ is obtained by drawing contours at the $p_{\mathrm{med}} = 0.32$ level. The largest systematic uncertainty in recovering the $\lambda_{\mathrm{mfp,HI}} - \langle \Gamma_{\mathrm{HI}} \rangle$ parameters is due to the thermal parameters (T_0, γ) and has been accounted for in the 1σ contours. This also demonstrates that the effect of spatial fluctuations in temperature does not significantly affect

---

[4] *Aton* takes two moments of the RT equation and uses the M1

condition to close the hierarchy of equations. Ideally one need to take large number of moments to capture the full angular dependence which is, however, not convenient for computation.

[5] We have also checked that the parameters can be recovered by using 100 different fiducial samples generated from different skewers

our recovered parameters. The temperature fluctuations are modeled self-consistently in the *Aton* simulation. With our two-zone temperature model we recover the mean free path and photo-ionization rate well. This illustrates that the two-zone temperature model with observed uncertainty in T_0, γ is a reasonable assumption. Similar results have been found by Nasir & D'Aloisio (2020), where authors find that the effect of temperature fluctuations on the scatter in effective optical depth is not significant. Fig. G1 shows that the true value is within the 1σ contour in all the redshift bins, perhaps suggesting that we somewhat overestimates the errors. The median $p-$value (p_{med}) systematically increases (black regions) near the true value of $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ (red stars) and p_{med} systematically decreases as one moves away from the true values. We find that the constraints on $\lambda_{\mathrm{mfp,HI}}$ are slightly less tight with KS test statistics compared to AD test statistics. This is because KS test statistics are sensitive to the middle part of the $\tau_{\mathrm{eff,HI}}$ CDF while AD statistics are sensitive to the middle and the tail of the $\tau_{\mathrm{eff,HI}}$ CDF. Therefore, we use AD test statistics for measuring $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ in this work. Fig. G1 also demonstrates that the highest redshift bin, $5.95 \leq z \leq 6.05$, is dominated by non-detections due to the finite S/N of the spectra. As a result constraints on $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ are less tight using current properties of the observed dataset at this redshift. The good recovery of $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ parameters from *Aton* using EX-CITE validates our approach of measuring these parameters from observations.

6 RESULTS

In this section, we present the main result of this work, measurements of $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ from the observed sample. We first discuss the best fit parameter estimation accounting for thermal parameter uncertainty. We then describe the source of observational and modeling uncertainties on our measurement and how we account for them in the final measurements. We then compare our measurements with previous work and discuss implications of these measurements for models of reionization. Finally, we show the consistency of our best fit models with observations using the dark gap length and pCDDF statistics.

6.1 Parameter estimation ($\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$)

Fig. 7 summarizes the main result of this work, the measurement of $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$ in 12 redshift bins at $4.9 \leq z \leq$ 6.1. The method of modeling Γ_{HI} fluctuations, generating Lyα forest spectra and measuring the parameters with appropriate statistical uncertainty has already been described in §3, §4 and §5. In short, we generate 1000 mock samples for each of 26244 $\lambda_{\mathrm{mfp,HI}}$-$\langle\Gamma_{\mathrm{HI}}\rangle$ parameter combinations. We calculate the AD statistics p value between data and each mock sample. The median p value (p_{med}) describes the level of agreement between data and each model. The best fit $\lambda_{\mathrm{mfp,HI}}$-$\langle\Gamma_{\mathrm{HI}}\rangle$ corresponds to the model that has maximum p_{med}. The 1σ uncertainty corresponds to the region with $p_{\mathrm{med}} > 0.32$ and accounts for uncertainty in thermal parameter estimation. (see §5.6 for details.). Fig. 7 shows that the best fit $\lambda_{\mathrm{mfp,HI}}$-$\langle\Gamma_{\mathrm{HI}}\rangle$ values are evolving with redshift. The best fit $\lambda_{\mathrm{mfp,HI}}$ significantly decreases with increasing

redshift from 57.4 h^{-1}cMpc at $z = 5.0$ to 8.4 h^{-1} cMpc at $z = 6$. The $\langle\Gamma_{\mathrm{HI}}\rangle$ also decreases from 0.56 at $z = 5.0$ to 0.145 at $z = 6.0$ with increasing redshift but not as drastically as $\lambda_{\mathrm{mfp,HI}}$. Fig. 7 also illustrates that $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ are strongly correlated with each other as expected from physical models of reionization. We emphasize that the correlation between $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ at $z > 5.2$ is driven by the data whereas at $z \leq 5.2$, it is more likely to be driven by the physical prior (see §5.2 and online appendix B). The size of the 1σ uncertainty (shown by the blue contours), is similar at $z < 5.7$ while it increases at $z > 5.7$. The less stringent constraints on $\lambda_{\mathrm{mfp,HI}}$-$\langle\Gamma_{\mathrm{HI}}\rangle$ at $z > 5.8$ are due to the combined effect of the limited number of spectra and the finite S/N of the spectra. Many of the $\tau_{\mathrm{eff,HI}}$ measurements at $z > 5.7$ are lower limits due to non-detections. Due to the presence of non-detections, the constraining power of the $\tau_{\mathrm{eff,HI}}$ CDF decreases. As a result, contour sizes increase at $z > 5.7$.

Fig. 8 shows the comparison of the $\tau_{\mathrm{eff,HI}}$ CDF from the best fit EX-CITE models with observations and a uniform UVB model in the 12 redshift bins. Each $\lambda_{\mathrm{mfp,HI}}$-$\langle\Gamma_{\mathrm{HI}}\rangle$ model consists of 1000 mocks. The scatter in the τ_{eff} CDF due to individual mocks (shown by gray curves) represents the cosmic variance. The combined $\tau_{\mathrm{eff,HI}}$ CDF of all 1000 mocks (red curves) is shown for illustration purposes. We use the p value calculated from individual $\tau_{\mathrm{eff,HI}}$ mocks for parameter estimation. The comparison of the best fit $\tau_{\mathrm{eff,HI}}$ CDF and observations shows that our best fit model is in good agreement with observations at all the redshifts. The scatter in the $\tau_{\mathrm{eff,HI}}$ measurements is one of the main characteristics of the observations. For comparison, we also show the $\tau_{\mathrm{eff,HI}}$ CDF for uniform UVB models in Fig. 8. The uniform UVB models fail to reproduce the scatter in the observed $\tau_{\mathrm{eff,HI}}$ at $z > 5.4$ (similar to the findings in Becker et al. 2015; Bosman et al. 2022). However, in our fluctuating UVB model with varying mean free path, the scatter as well as the median of the $\tau_{\mathrm{eff,HI}}$ measurements are well reproduced. At $z \leq 5.2$, the uniform UVB model can reproduce the scatter in the $\tau_{\mathrm{eff,HI}}$ measurements. This suggests that the sensitivity of the $\tau_{\mathrm{eff,HI}}$ distribution to the mean free path is reduced at $z \leq 5.2$. This is also evident in Fig. H1 where the mean free path estimated from uniform UVB model matches well with $\lambda_{\mathrm{mfp,HI}}$ measured from EX-CITE. Hence the mean free path measurements at $z < 5.2$ should be treated as lower limits. At $z > 5.2$ the $\tau_{\mathrm{eff,HI}}$ distribution is clearly sensitive to $\lambda_{\mathrm{mfp,HI}}$. The good match between uniform UVB model and EX-CITE models with the observed $\tau_{\mathrm{eff,HI}}$ distribution at $z \leq 5.2$ suggests that H I reionization is largely complete by $z \sim 5.2$, consistent with the results of Kulkarni et al. (2019); Bosman et al. (2022).

Even though the simulations with uniform UVB models reproduce the $\tau_{\mathrm{eff,HI}}$ distribution at $z < 5.3$, these models are still missing important aspects. Uniform UVB models assume the mean free path of H I ionizing photons to be larger than the horizon. In reality, the mean free path of H I ionizing photons in the post-reionization universe is mainly set by the average distance between self-shielded neutral regions that manifest themselves as (super-)Lyman Limit Systems and damped Lyman-alpha systems. Such neutral regions do not exist in simulations of the post-reionization Universe with uniform UVB models that do not account for self-shielding (Chardin et al. 2018b; Cain et al. 2022). We
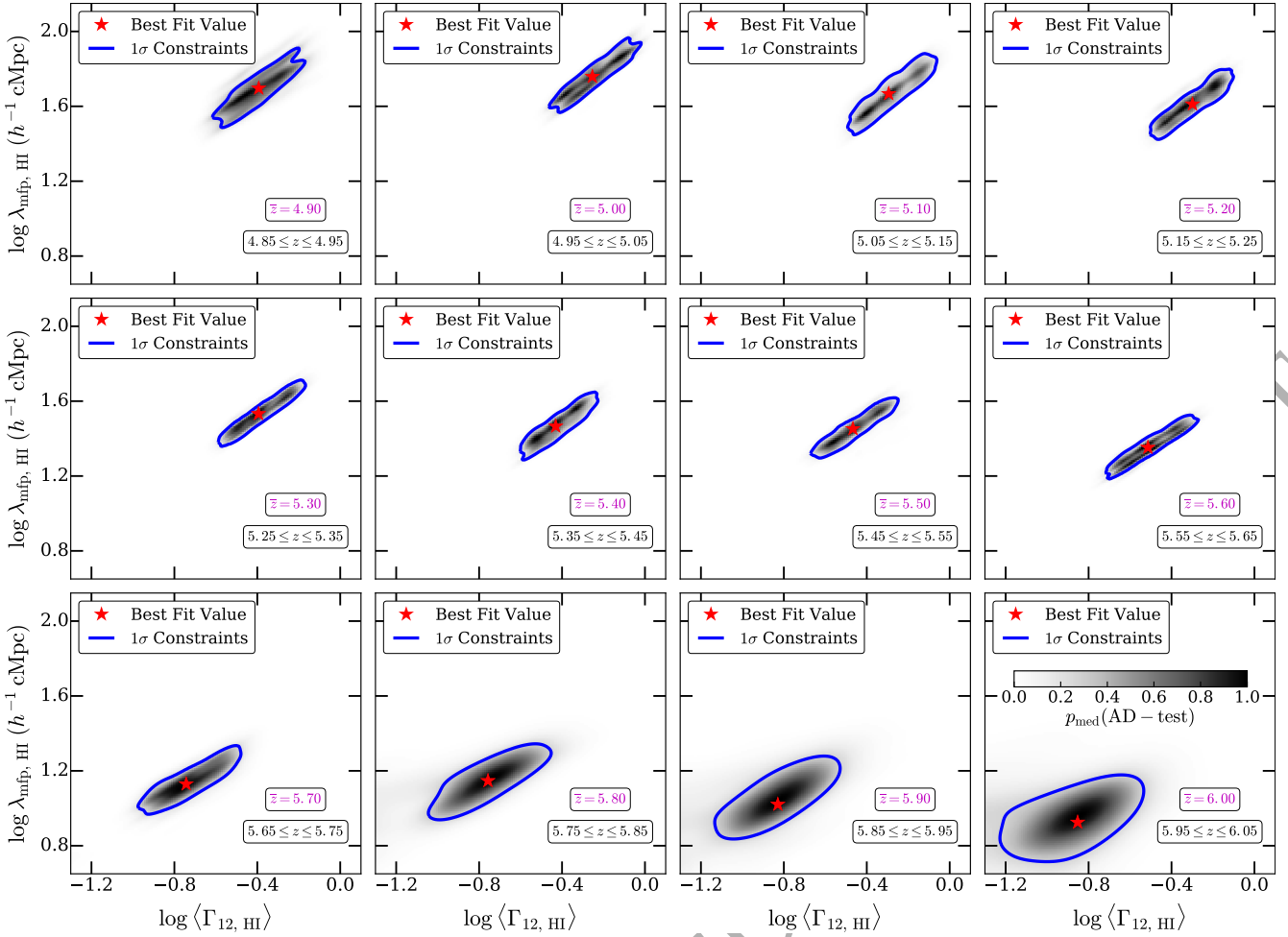
Figure 7. The panels show constraints on $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ obtained by comparing the observed $\tau_{\mathrm{eff,HI}}$ CDF with that from EX-CITE simulations. The $\tau_{\mathrm{eff,HI}}$ CDFs are compared using the non-parametric Anderson-Darlington (AD) test that is sensitive to the median as well as the tail of the distribution. The best fit model parameters are shown by red stars in each panel. The 1σ constraints on the model parameters are obtained by demanding $p_{\mathrm{med}} > 0.32$. The color scheme is the same for all the panels and represents the median p value obtained between model and observed $\tau_{\mathrm{eff,HI}}$ CDF. The constraints on $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ at $z > 5.7$ are less stringent because of the limited number of sightlines and the limited S/N. The 1σ contours shown account for the total modeling uncertainty (i.e., including thermal parameter uncertainty, see Fig. 9 and §6.2 for details). We have checked that the 1σ contours are similar at $z < 5.4$ when we use the simulation with $L_{\mathrm{box}} = 320\,h^{-1}\mathrm{cMpc}$.

further caution the reader that the unlike the neutral gas in the extended neutral islands, the neutral self-shielded gas in collapsed haloes/galaxies is not properly captured in our simulations because the cell size ($\sim 300\,h^{-1}\,ckpc$) is too coarse.

6.2 Parameter uncertainty

Any proper measurement of $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$ should account for the sources of uncertainties in the modeling and observations. There are three main sources of uncertainties in our measured parameters (i) modeling uncertainty, (ii) statistical uncertainty due to cosmic variance and (iii) observational uncertainties. While modeling the fluctuations in the ionizing radiation field, we have fixed various parameters to their fiducial values e.g., halo mass cutoff (M_{cutoff}), emissivity power law index (β), power law index between Γ_{HI} fluctuations and mean free path (ζ) etc (see §3). However the chosen values of these parameters could be different

than what we have assumed and this needs to be taken into account while estimating the uncertainty in $\langle\Gamma_{\mathrm{HI}}\rangle - \lambda_{\mathrm{mfp,HI}}$ parameters.

The left panel in Fig. 9 shows the effect of the modeling uncertainty on $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$ measurements. The main contribution to the modeling uncertainty is due to the thermal parameter uncertainties. To illustrate this, we show the 1σ contours of $\lambda_{\mathrm{mfp,HI}} - \langle\Gamma_{\mathrm{HI}}\rangle$ measurements for three thermal parameter combinations (i) default T_0, γ evolution (blue contours), (ii) $T_0 - \delta T_0, \gamma + \delta\gamma$ i.e. all cells in ionized regions are assumed to ionize early (green contours), and (iii) $T_0 + \delta T_0, \gamma - \delta\gamma$ i.e. all cells in ionized regions are assumed to ionize recently (red contours). We chose a wide range for the thermal parameters consistent with recent measurements of Gaikwad et al. (2020) as shown in Fig. D1.

Fig. 9 (left panel) and Fig. F1 show that the parameter combination $T_0 - \delta T_0, \gamma + \delta\gamma$ produces systematically larger $\langle\Gamma_{\mathrm{HI}}\rangle - \lambda_{\mathrm{mfp,HI}}$ while $T_0 + \delta T_0, \gamma - \delta\gamma$ produces systematically lower $\langle\Gamma_{\mathrm{HI}}\rangle - \lambda_{\mathrm{mfp,HI}}$. This is expected as the recombination
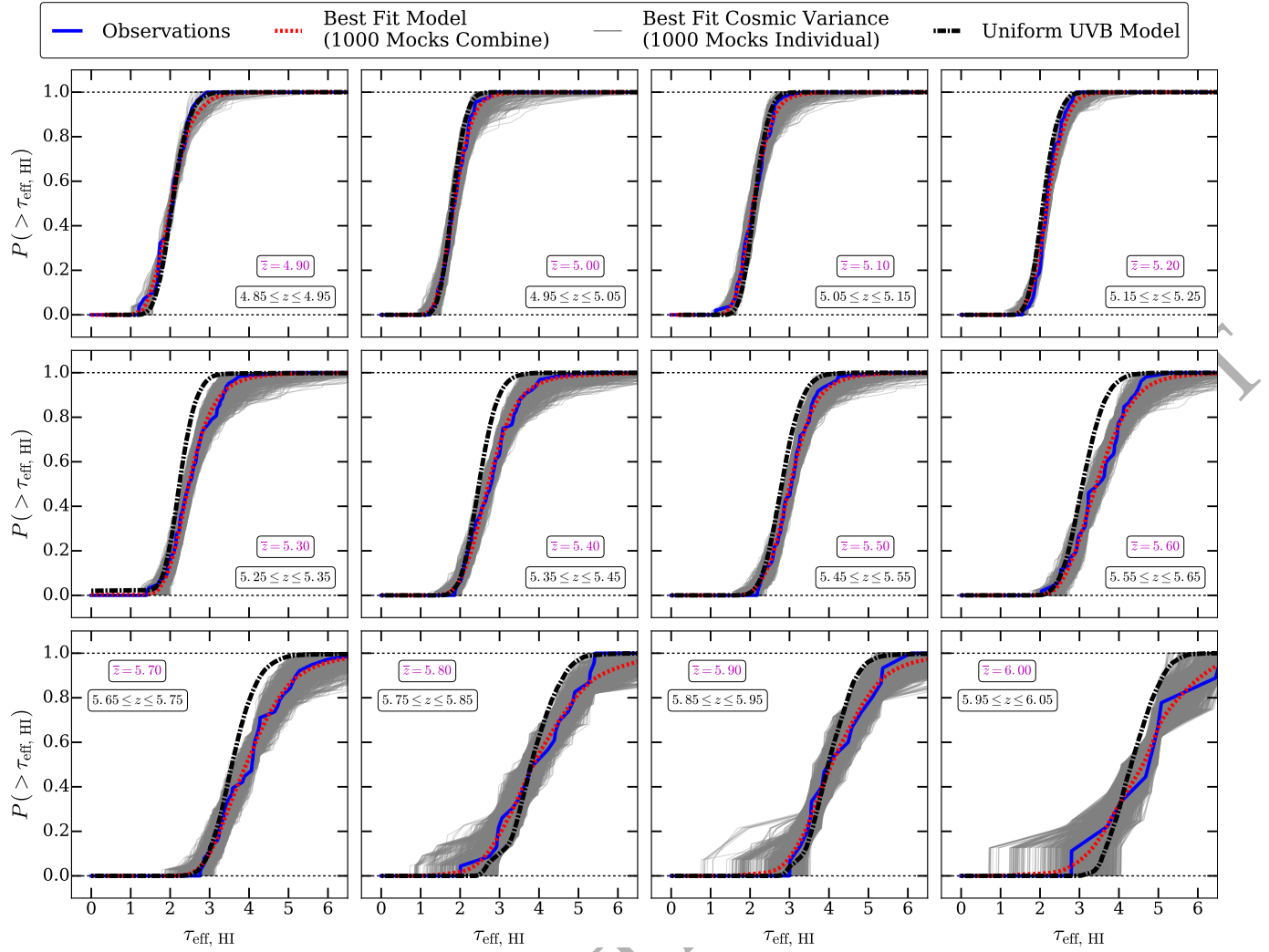
Figure 8. The figure compares the observed $\tau_{\rm eff,HI}$ CDF (blue curve) with the best fit $\tau_{\rm eff,HI}$ CDF (red and gray curves) in 12 redshift bins at $4.9 \leq z \leq 6.0$. The $\tau_{\rm eff,HI}$ CDF of 1000 individual mocks is shown by the gray curves. The scatter in $\tau_{\rm eff,HI}$ in these models represents the cosmic variance. Each mock sample has the same redshift path length as the observations. The combined $\tau_{\rm eff,HI}$ CDF from all 1000 mocks is shown by the red dashed curve. The $\tau_{\rm eff,HI}$ CDF from uniform UVB models (1000 mocks combined) is shown by the black dash-dot curve. The photo-ionization rate in the uniform UVB model is rescaled to match the mean flux of the observed sample. The combined $\tau_{\rm eff,HI}$ is shown for visual purposes only. We do not use the red curves in measuring the parameters. The uniform UVB model can not reproduce the scatter in $\tau_{\rm eff,HI}$ at $z > 5.2$. The best fit that includes fluctuations in the ionizing radiation field reproduces the observed $\tau_{\rm eff,HI}$ distribution remarkably well. The fact that the uniform UVB model reproduces the observed $\tau_{\rm eff,HI}$ distribution at $z \leq 5.2$ suggests that IGM is fully ionized by $z \sim 5.2$.

rate, α, scales with temperature as $T_0^{-0.7}$ and $\Gamma_{\rm HI} \propto \alpha(T)$. In order to estimate the total uncertainty due to the thermal parameters we use the maximum of the $p_{\rm med}$ values between observed and model $\tau_{\rm eff,HI}$ CDF for three combinations of thermal parameters (magenta dotted curve). Similar to the thermal parameters, we vary other parameters in EX-CITE, $M_{\rm cutoff} = [10^8\,{\rm M_\odot}, 10^{10}\,{\rm M_\odot}]$, $\beta = [0.5, 1.5]$ and $\zeta = [1/3, 3/4]$. These parameter combinations are physically motivated e.g., a small (large) $M_{\rm cutoff}$ corresponds to reionization by small (large) mass galaxies, the range in β is sensitive to galaxy-halo bias and the escape fraction of ionizing photons, ζ values are motivated from scaling relation between $\langle \Gamma_{\rm HI} \rangle$ and $\lambda_{\rm mfp,HI}$ seen in radiative transfer simulations (Muñoz et al. 2016). The maximum differences in $\langle \Gamma_{\rm HI} \rangle - \lambda_{\rm mfp,HI}$ produced by changing these parameter

combinations is less than 2.4 percent and is shown by the black solid curve. Throughout this work, we show the contours of $\langle \Gamma_{\rm HI} \rangle - \lambda_{\rm mfp,HI}$ measurements that account for this modeling uncertainty (except in the left panel of Fig. 9 and F1).

We would like to stress that our way of accounting for the uncertainty in thermal parameters may not be fully rigorous because there will be significant scatter in the temperature at a given density depending on the timing of reionization. However, our aim in this work is not to model the temperature fluctuations but rather to assess its effect on the uncertainty in our $\langle \Gamma_{\rm HI} \rangle - \lambda_{\rm mfp,HI}$ parameter measurements. Our estimated uncertainty in $\langle \Gamma_{\rm HI} \rangle - \lambda_{\rm mfp,HI}$ could be somewhat smaller than the actual uncertainty. Note that we assume a fairly large uncertainty in thermal parameters.
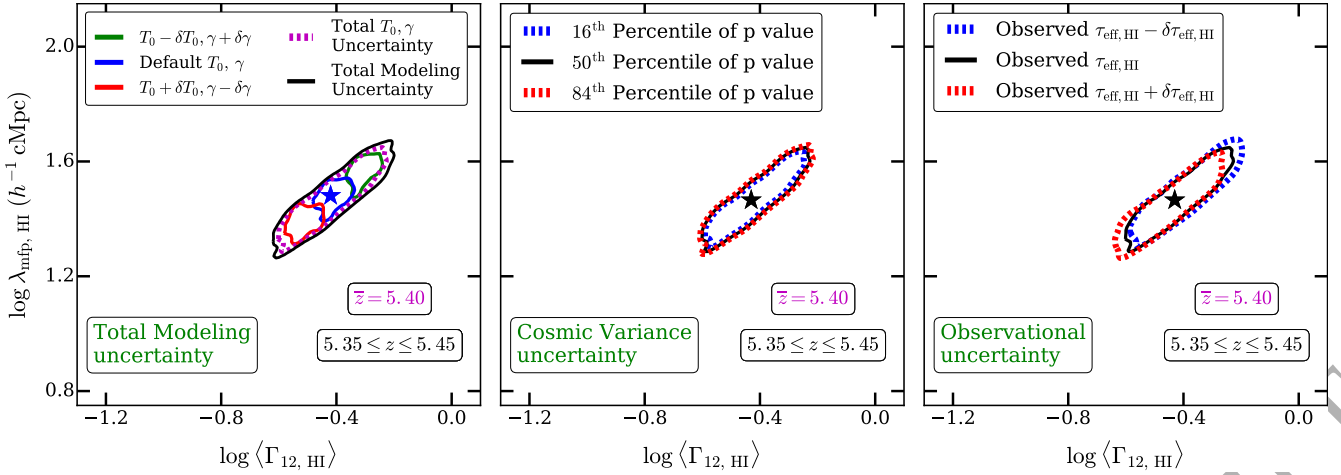
Figure 9. The left, middle and right panel show the effect of modeling, cosmic variance and observational uncertainty on the measurements of mean free path ($\lambda_{\mathrm{mfp,HI}}$) and spatially averaged photo-ionization rate ($\langle \Gamma_{\mathrm{HI}} \rangle$), respectively, at $5.65 \leq z \leq 5.75$. The left panel shows that a smaller value of T_0 ($T_0 - \delta T_0$) and larger value of γ ($\gamma + \delta\gamma$) systematically result in higher $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$ (magenta contours) and vice-versa (red contours). The magenta dashed contours show the total uncertainty in $\langle \Gamma_{\mathrm{HI}} \rangle$ and $\lambda_{\mathrm{mfp,HI}}$ due to accounting for uncertainty in thermal parameters. The total assumed uncertainty in thermal parameters ($T_0 \pm \delta T_0$, $\gamma \pm \delta\gamma$) is shown in Fig. D1. The black contours show the total modeling uncertainty i.e. including uncertainty in halo mass cutoff (M_{cutoff}), emissivity power-law index (β) and power-law index (ζ) between $\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle$ and λ_0. All the contours in this work (except the left panel and Fig. F1) are shown after accounting for the total modeling uncertainty. The middle panel shows that the effect of cosmic variance uncertainty on the measured parameters is moderate and is typically less than 1.7 percent. The blue dashed, black solid and red dashed curves in the middle panel correspond to constraints on $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$ when we use the 16th, 50th and 84th percentile of p values between observations and 1000 mock samples. The good convergence of all the contours shows that the PDF of p values from 1000 mocks is sharply peaked and does not contribute significantly to parameter uncertainty at this redshift. The right panel shows the effect of observational uncertainty in the $\tau_{\mathrm{eff,HI}}$ measurements on the measurements of $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$. A systematically higher value of $\tau_{\mathrm{eff,HI}}$ corresponds to a smaller value of $\langle \Gamma_{\mathrm{HI}} \rangle$ and $\lambda_{\mathrm{mfp,HI}}$ (red dashed contour) and vice-versa (blue dashed contour). The $\tau_{\mathrm{eff,HI}}$ uncertainty accounts for the uncertainty in the continuum fitting of the observed spectra. The total reported uncertainty of our measured parameters accounts for all the uncertainties mentioned above. The final uncertainty in each individual parameters is obtained by marginalizing over the other nuisance parameters. The effect of modeling uncertainty, cosmic variance and observational uncertainty on measurements of $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$ in other redshift bins is illustrated in the online appendix in Fig. F1, F2 and F3, respectively.

The middle panel of Fig. 9 shows the effect of cosmic variance on the $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$ parameters. In each redshift bin, we generate 1000 mocks such that each mock sample is similar to the observed sample. We calculate the p value between each mock and observed sample. Since the mocks are drawn randomly along different sightlines, the scatter in them corresponds to cosmic/astrophysical variance in our simulation box. For brevity, we refer to this as cosmic variance. We use the median p value (p_{med}) from 1000 mocks samples to determine best fit parameters. We estimate the effect of cosmic variance by using the 16th and 84th percentile of p values in Fig. 9 and Fig. F2. The uncertainty in $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$ due to cosmic variance increases with increasing redshift. This is expected as fluctuations in the ionizing radiation fields are larger at higher redshift, and that in turn produces more scatter along different sightlines. However, the uncertainty due to cosmic variance is typically smaller than 1.7 percent.

The right panel of Fig. 9 shows the effect of observational systematics on the measured $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$ parameters. The finite S/N of the observed spectra, uncertainty in the sky subtraction and uncertainty in the QSO continuum estimation lead to significant uncertainty in the observed $\tau_{\mathrm{eff,HI}}$. We estimate the effect of observational uncertainty by measuring $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$ with observed $\tau_{\mathrm{eff,HI}} - \delta\tau_{\mathrm{eff,HI}}$ and $\tau_{\mathrm{eff,HI}} + \delta\tau_{\mathrm{eff,HI}}$ as shown in Fig. 9 and Fig. F3. A larger (smaller) value of observed $\tau_{\mathrm{eff,HI}}$ corresponds to lower

(higher) measurements of $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$. We find that typically the uncertainty in measured $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$ is smaller than 4 percent at any given redshift.

The modeling uncertainties and observational uncertainties are systematic in nature i.e. $\langle \Gamma_{\mathrm{HI}} \rangle - \lambda_{\mathrm{mfp,HI}}$ constraints change systematically if these parameters are varied. The uncertainty due to cosmic variance on the other hand is random in nature. We add cosmic variance uncertainties in quadrature while modeling and observational uncertainties are added directly to the total uncertainty. Table 3 summarizes our best fit parameters with total uncertainty i.e., including cosmic variance, observational and modeling uncertainties. In the next section, we compare our measured parameter evolution with other measurements from the literature.

6.3 Comparison to previous work

In this section, we compare the evolution of $\langle \Gamma_{\mathrm{HI}} \rangle$, $\lambda_{\mathrm{mfp,HI}}$ and $\langle f_{\mathrm{HI}} \rangle$ measured in this work with measurements in the literature using similar or different techniques.

6.3.1 $\langle \Gamma_{\mathrm{HI}} \rangle$ measurement comparison

The left panel in Fig. 10 compares our measurements of $\langle \Gamma_{\mathrm{HI}} \rangle$ with those from Bolton & Haehnelt (2007); Wyithe

Table 3. The table shows our measurements of the photo-ionization rate ($\langle\Gamma_{\mathrm{HI}}\rangle$, in 10^{-12} s^{-1}), mean free path ($\lambda_{\mathrm{mfp,HI}}$, in h^{-1} cMpc), neutral fraction ($\langle f_{\mathrm{HI}}\rangle$), emissivity at 912 Å ($\epsilon_{912}$, in erg s^{-1} cMpc^{-3} Hz^{-1}) and ionizing photon number density per unit time ($\dot{n}$, in s^{-1} cMpc^{-3}) with total 1σ uncertainty (i.e., including modeling, cosmic variance and observational uncertainties). The uncertainty in ϵ_{912} and $\dot{n}$ also accounts for the uncertainty in the spectral energy distribution index $\alpha_s = 2.0 \pm 0.6$ and slope of H I column density distribution $\beta_{\mathrm{HI}} = 1.3 \pm 0.05$ at $16 < \log \mathrm{N_{HI}} < 18$. The uncertainty in each parameter is obtained by marginalizing over the uncertainty in the other parameters.

Redshift	$\langle\Gamma_{12,\mathrm{HI}}\rangle$	$\lambda_{\mathrm{mfp,HI}}$	$\langle f_{\mathrm{HI}}\rangle$	ϵ_{912}	$\dot{n}$
4.90 ± 0.05	$0.501^{+0.275}_{-0.232}$	$50.119^{+33.058}_{-19.919}$	$2.534^{+0.585}_{-0.444} \times 10^{-5}$	$0.795^{+0.043}_{-0.189} \times 10^{25}$	$0.600^{+0.304}_{-0.248} \times 10^{51}$
5.00 ± 0.05	$0.557^{+0.376}_{-0.218}$	$57.412^{+38.088}_{-21.104}$	$2.267^{+0.840}_{-0.449} \times 10^{-5}$	$0.746^{+0.103}_{-0.132} \times 10^{25}$	$0.563^{+0.352}_{-0.207} \times 10^{51}$
5.10 ± 0.05	$0.508^{+0.324}_{-0.192}$	$46.452^{+31.173}_{-18.909}$	$2.668^{+1.343}_{-0.553} \times 10^{-5}$	$0.813^{+0.196}_{-0.162} \times 10^{25}$	$0.614^{+0.475}_{-0.236} \times 10^{51}$
5.20 ± 0.05	$0.502^{+0.292}_{-0.193}$	$40.832^{+22.264}_{-15.713}$	$2.758^{+0.811}_{-0.555} \times 10^{-5}$	$0.886^{+0.162}_{-0.145} \times 10^{25}$	$0.668^{+0.461}_{-0.238} \times 10^{51}$
5.30 ± 0.05	$0.404^{+0.272}_{-0.147}$	$34.041^{+18.440}_{-12.163}$	$5.100^{+8.044}_{-4.036} \times 10^{-4}$	$0.827^{+0.142}_{-0.093} \times 10^{25}$	$0.624^{+0.421}_{-0.198} \times 10^{51}$
5.40 ± 0.05	$0.372^{+0.217}_{-0.126}$	$29.242^{+15.427}_{-10.187}$	$3.533^{+15.084}_{-2.457} \times 10^{-3}$	$0.858^{+0.171}_{-0.131} \times 10^{25}$	$0.648^{+0.462}_{-0.225} \times 10^{51}$
5.50 ± 0.05	$0.344^{+0.219}_{-0.130}$	$28.907^{+18.053}_{-11.124}$	$7.246^{+27.311}_{-3.506} \times 10^{-3}$	$0.778^{+0.153}_{-0.022} \times 10^{25}$	$0.587^{+0.417}_{-0.149} \times 10^{51}$
5.60 ± 0.05	$0.319^{+0.194}_{-0.120}$	$22.961^{+11.712}_{-7.826}$	$1.630^{+2.544}_{-0.834} \times 10^{-2}$	$0.883^{+0.108}_{-0.115} \times 10^{25}$	$0.666^{+0.402}_{-0.221} \times 10^{51}$
5.70 ± 0.05	$0.224^{+0.223}_{-0.112}$	$16.596^{+9.707}_{-7.476}$	$5.596^{+7.141}_{-3.362} \times 10^{-2}$	$0.831^{+0.066}_{-0.026} \times 10^{25}$	$0.627^{+0.340}_{-0.145} \times 10^{51}$
5.80 ± 0.05	$0.178^{+0.194}_{-0.078}$	$13.183^{+9.205}_{-5.769}$	$9.364^{+6.182}_{-6.391} \times 10^{-2}$	$0.807^{+0.148}_{-0.131} \times 10^{25}$	$0.609^{+0.420}_{-0.141} \times 10^{51}$
5.90 ± 0.05	$0.151^{+0.151}_{-0.079}$	$10.471^{+9.027}_{-4.976}$	$1.282^{+1.260}_{-0.736} \times 10^{-1}$	$0.840^{+0.066}_{-0.105} \times 10^{25}$	$0.634^{+0.343}_{-0.207} \times 10^{51}$
6.00 ± 0.05	$0.145^{+0.157}_{-0.087}$	$8.318^{+7.531}_{-4.052}$	$1.744^{+0.925}_{-1.089} \times 10^{-1}$	$0.929^{+0.052}_{-0.050} \times 10^{25}$	$0.701^{+0.357}_{-0.191} \times 10^{51}$

& Bolton (2011); Calverley et al. (2011); D'Aloisio et al. (2018). Overall our $\langle\Gamma_{\mathrm{HI}}\rangle$ measurements are in very good agreement (within 1σ) with those from all the measurements in the literature at $z < 5.5$. The best fit $\langle\Gamma_{\mathrm{HI}}\rangle$, is systematically somewhat lower at $z > 5.5$ than that from D'Aloisio et al. (2018), but is still consistent within 1σ. Note that D'Aloisio et al. (2018) obtain their $\langle\Gamma_{\mathrm{HI}}\rangle$ measurements using three different mean free path models: fiducial, intermediate and short. Their best fit $\langle\Gamma_{\mathrm{HI}}\rangle$ is based on their fiducial model in which the mean free path is consistent with the Worseck et al. (2014) power-law scaling. However, as we show in the next section the best fit $\lambda_{\mathrm{mfp,HI}}$ deviates from this power-law at $z > 5.5$. This explains the systematically lower values of $\langle\Gamma_{\mathrm{HI}}\rangle$ in this work compared to D'Aloisio et al. (2018). Given that the $\lambda_{\mathrm{mfp,HI}}$ evolution is not known apriori, it is important to vary both $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ when comparing the $\tau_{\mathrm{eff,HI}}$ distribution in simulations and observations. Note that our uncertainty on $\langle\Gamma_{\mathrm{HI}}\rangle$ is larger and more realistic as our $\langle\Gamma_{\mathrm{HI}}\rangle$ measurements are marginalized over $\lambda_{\mathrm{mfp,HI}}$. In addition, we also account for the observational and modeling uncertainties in our analysis. Finally, our $\langle\Gamma_{\mathrm{HI}}\rangle$ measurements at $z = 6.0$ are in very good agreement with those measured by Bolton & Haehnelt (2007); Wyithe & Bolton (2011); Calverley et al. (2011). Note further that due to the limited number of sightlines and significant number of non-detections, our uncertainty on $\langle\Gamma_{\mathrm{HI}}\rangle$ becomes significantly larger at $5.95 \leq z \leq 6.05$.

6.3.2 $\lambda_{\mathrm{mfp,HI}}$ *measurement comparison*

The middle panel of Fig. 10 compares our mean free path measurements with those from Bolton & Haehnelt (2007); Worseck et al. (2014); Becker et al. (2021); Bosman (2021).

Our $\lambda_{\mathrm{mfp,HI}}$ measurements at $z < 5.1$ are again in good agreement with those from Bolton & Haehnelt (2007); Worseck et al. (2014); Becker et al. (2021). The $\lambda_{\mathrm{mfp,HI}}$ in Worseck et al. (2014); Becker et al. (2021) is measured by stacking quasar spectra at 912 Å and then fitting the attenuation profile with an exponential function (see §5.2 for details) with $\lambda_{\mathrm{mfp,HI}}$ as a free parameter. Note that the size of the proximity zone of quasars (becoming increasingly prominent at $z \sim 6$) has been accounted for. Our method of using the $\tau_{\mathrm{eff,HI}}$ distribution of the Lyα forest to measure $\lambda_{\mathrm{mfp,HI}}$ is complementary to the stacking methods. The agreement of $\lambda_{\mathrm{mfp,HI}}$ at $z < 5.1$ from this work with those from Worseck et al. (2014); Becker et al. (2021) suggests that the two methods are consistent. Note that the size of proximity zones at $z < 5.1$ are typically smaller than the mean free path of ionizing photons (private communication with G. Worseck). At $z > 5.1$, our best fit $\lambda_{\mathrm{mfp,HI}}$ measurements show a gradual evolution with redshift, $\lambda_{\mathrm{mfp,HI}}$ decreases systematically with increasing redshift. Our best fit $\lambda_{\mathrm{mfp,HI}}$ is thereby consistent with that measured by Becker et al. (2021) within 1.2σ at $z = 6$. Our $\lambda_{\mathrm{mfp,HI}}$ measurements at $z = 6.0$ are also consistent with the upper and the lower limits obtained by Bolton & Haehnelt (2007) and Bosman (2021), respectively. The $\lambda_{\mathrm{mfp,HI}}$ measurements from Bolton & Haehnelt (2007) are systematically larger than our best fit values at $z = 5, 6$. Bolton & Haehnelt (2007) measure $\lambda_{\mathrm{mfp,HI}}$ using an analytic expression for the mean free path from Miralda-Escudé et al. (2000) that depends on the fraction of gas at a given redshift below the self-shielding overdensity. Note that Bolton & Haehnelt (2007) ignore the contribution of the intervening gas between self-shielded regions. If we correct for the contribution of such intervening systems, their best fit
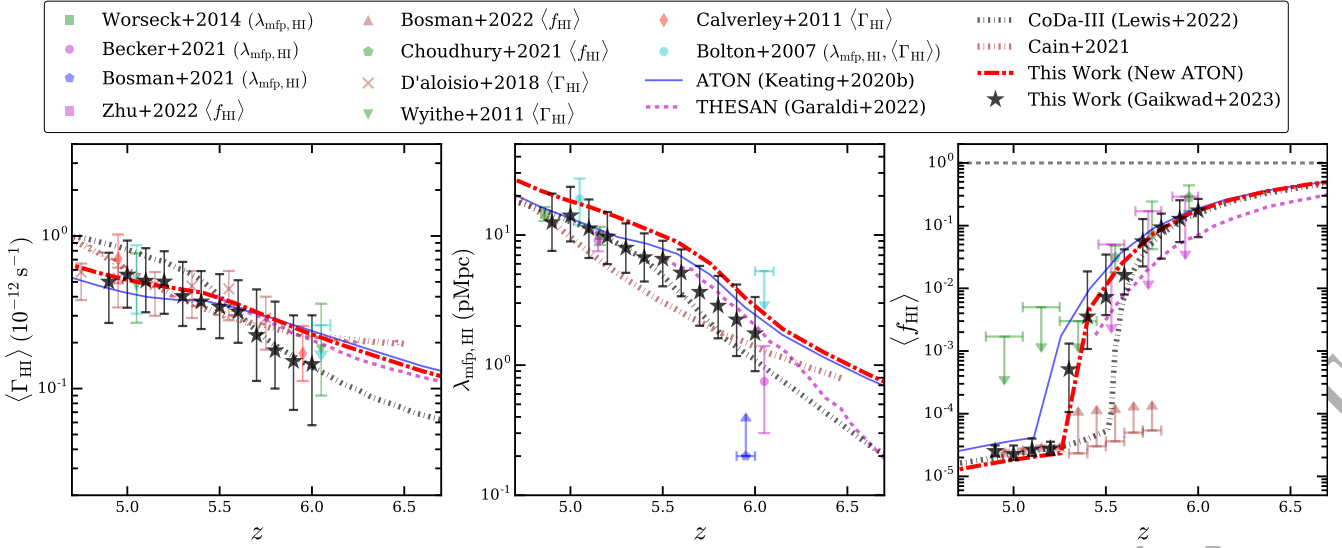
Figure 10. The left, middle and right panel show the evolution of $\langle\Gamma_{\rm HI}\rangle$, $\lambda_{\rm mfp,HI}$ and $\langle f_{\rm HI}\rangle$ obtained from this work with measurements from the literature. Our $\langle\Gamma_{\rm HI}\rangle$, $\lambda_{\rm mfp,HI}$ measurements are in good agreement with that from the literature at $4.9 \leq z \leq 6.0$. Our $\langle f_{\rm HI}\rangle$ measurements are in good agreement with the limits / measurements of Choudhury et al. (2021); Bosman et al. (2022); Zhu et al. (2022). The uncertainties in our measured $\langle\Gamma_{\rm HI}\rangle$, $\lambda_{\rm mfp,HI}$ and $\langle f_{\rm HI}\rangle$ are typically larger than those from the literature because we account for fluctuations in the ionizing background, observational, modeling uncertainties and we marginalize over the other nuisance parameters. Our best fit $\langle\Gamma_{\rm HI}\rangle$ are consistent with measurements/upper limits from Bolton & Haehnelt (2007); Wyithe & Bolton (2011); Calverley et al. (2011) at $5.9 \leq z \leq 6.1$. Similarly, the best fit $\lambda_{\rm mfp,HI}$ at $5.9 \leq z \leq 6.1$ is consistent with that from Bolton & Haehnelt (2007); Becker et al. (2021); Bosman et al. (2022). The blue solid, magenta dashed, black dotted and brown dotted curves show the evolution of $\langle\Gamma_{\rm HI}\rangle$, $\lambda_{\rm mfp,HI}$ and $\langle f_{\rm HI}\rangle$ parameters obtained from simulations with *Aton* (Keating et al. 2020b, their low $\tau_{\rm CMB}$ hot model) and the *Thesan* (Garaldi et al. 2022), *CoDa-III* (Lewis et al. 2022) and Cain et al. (2021) radiative transfer simulations respectively. The $\langle\Gamma_{\rm HI}\rangle$ evolution (left panel) in all the models is consistent within 1σ of our measurements. The *Aton* simulation shows a systematically higher $\langle\Gamma_{\rm HI}\rangle$ at $z > 5.6$ while $\langle\Gamma_{\rm HI}\rangle$ in *CoDa-III* is higher at $z < 5.2$ owing to their evolution in emissivity. The mean free path from all the models also show reasonable agreement with our measured $\lambda_{\rm mfp,HI}$. The $\lambda_{\rm mfp,HI}$ (middle panel) in *Thesan* in general passes through the best fit $\lambda_{\rm mfp,HI}$. The *Aton* and *CoDa-III* simulations predict slightly higher and lower $\lambda_{\rm mfp,HI}$ at $z > 5.5$ respectively. The $\lambda_{\rm mfp,HI}$ evolution in the Cain et al. (2021) model is systematically under-predicted at $5 \leq z \leq 6$ suggesting that reionization is rather late in their model. The $\langle f_{\rm HI}\rangle$ evolution in *Thesan* is within 1σ of our measurements . Reionization in the *Aton* simulation completes around $z \sim 5.1$ slightly later than suggested by the observations. The *CoDa-III* $\langle f_{\rm HI}\rangle$ evolution shows better agreement with our measurements at $z > 5.6$ but the reionization is slightly too rapid and completes at $z = 5.5$ instead. Also shown is a new *Aton* model that was run independently and prior to analysis of this work. The evolution of $\langle\Gamma_{\rm HI}\rangle$, $\lambda_{\rm mfp,HI}$ and $\langle f_{\rm HI}\rangle$ in this new *Aton* model matches very well with our best fit parameters. This new *Aton* model is consistent with a uniform UVB model at $z < 5.2$ corroborating that the observations favour a scenario where H I reionization is only completed by $z \sim 5.2$.

$\lambda_{\rm mfp,HI}$ reduces by a factor of $\sqrt{\pi}$ which brings their measurement into very good agreement with our measurements.

D'Aloisio et al. (2018) find that the $\tau_{\rm eff,HI}$ distribution can be well reproduced if the mean free path is $\sim 10h^{-1}$ cMpc (~ 2.16 pMpc) at $z = 5.8$. Our best fit value at the same redshift is larger by a factor of 1.5 but is consistent within 1.1σ. Note, however, that the resolution of the $\Gamma_{\rm HI}$ fluctuation maps is important here and may be the main reason for the somewhat lower values of $\lambda_{\rm mfp,HI}$ in D'Aloisio et al. (2018). The $\lambda_{\rm mfp,HI}$ evolution in Fig. H1 shows a deviation from a simple power-law ($\lambda_{\rm mfp,HI} \propto (1 + z)^{-5.4}$) at $z > 5.5$. Our best fit $\lambda_{\rm mfp,HI}$ decreases more rapidly with redshift. This suggests that the sizes of ionized regions are rapidly evolving between $5 \leq z \leq 6$, again corroborating previous suggestions that H I reionization is still in progress, i.e. late H I reionization ending at $z < 6$ is favoured by the observations.

6.3.3 Comparison of $\langle f_{\rm HI}\rangle$ measurements

In our models, the spatially averaged neutral fraction, $\langle f_{\rm HI}\rangle$, is uniquely determined by the combination of the two parameters $\lambda_{\rm mfp,HI}$-$\langle\Gamma_{\rm HI}\rangle$. Low values of $\lambda_{\rm mfp,HI}$ and $\langle\Gamma_{\rm HI}\rangle$ correspond to a low value of $\langle f_{\rm HI}\rangle$ and vice-versa. As a result once $\lambda_{\rm mfp,HI}$ and $\langle\Gamma_{\rm HI}\rangle$ are measured from the observations, we also have measurements on the corresponding $\langle f_{\rm HI}\rangle$ from our models. We calculate a best fit $\langle f_{\rm HI}\rangle$ value using the model with best fit $\lambda_{\rm mfp,HI}$ and $\langle\Gamma_{\rm HI}\rangle$. Similarly, the uncertainty in $\langle f_{\rm HI}\rangle$ is calculated using the combination of parameters $\lambda_{\rm mfp,HI}$ $\pm\delta\lambda_{\rm mfp,HI}$ and $\Gamma_{\rm HI}$ $\pm\delta\Gamma_{\rm HI}$. We use that combination of $\lambda_{\rm mfp,HI}$ $-\langle\Gamma_{\rm HI}\rangle$ which give the largest uncertainty for the $\langle f_{\rm HI}\rangle$ measurements. This means the lower (upper) limit of $\langle f_{\rm HI}\rangle$ correspond to $\lambda_{\rm mfp,HI}$ $-\delta\lambda_{\rm mfp,HI}$ ($\lambda_{\rm mfp,HI}$ $+\delta\lambda_{\rm mfp,HI}$) and $\langle\Gamma_{\rm HI}\rangle$ $-\delta\langle\Gamma_{\rm HI}\rangle$ ($\langle\Gamma_{\rm HI}\rangle$ $+\delta\langle\Gamma_{\rm HI}\rangle$), respectively.

The right panel in Fig. 10 compares the evolution of the spatially averaged neutral fraction of hydrogen $\langle f_{\rm HI}\rangle$ obtained in this way and the literature. At $z \leq 5.2$, our best fit $\langle f_{\rm HI}\rangle$ measurements are in good agreement with those from Bosman et al. (2022). However at $z > 5.2$, our best fit

$\langle f_{\mathrm{HI}} \rangle$ are systematically higher than Bosman et al. (2022). Note that most of the $\langle f_{\mathrm{HI}} \rangle$ measurements in the literature are lower limits at $z > 5.2$ (Fan et al. 2006; Yang et al. 2020; Bosman et al. 2022). This is because they generally estimate the $\langle f_{\mathrm{HI}} \rangle$ evolution from the data using simulations with a a uniform UVB. The simulations generally also do not account for self-shielding in dense regions. Note again that in our models, we capture the spatial fluctuations in the ionizing background. We calculate $\langle f_{\mathrm{HI}} \rangle$ from the entire simulation box for a given set of $\lambda_{\mathrm{mfp,HI}}$-$\langle \Gamma_{\mathrm{HI}} \rangle$ parameter combinations. So our $\langle f_{\mathrm{HI}} \rangle$ measurements contain the contribution of neutral regions as well as ionized regions. As a result, we can measure $\langle f_{\mathrm{HI}} \rangle$ and its uncertainty at $z > 5.2$. A $\langle f_{\mathrm{HI}} \rangle$ value beyond the quoted uncertainty corresponds to $\lambda_{\mathrm{mfp,HI}}$-$\langle \Gamma_{\mathrm{HI}} \rangle$ parameters that would produce $\tau_{\mathrm{eff,HI}}$ distributions inconsistent with observations. Note that the uncertainty of our $\langle f_{\mathrm{HI}} \rangle$ measurements at $z < 5.2$ is larger than that of previous work using uniform UVB models where the mean free path was not varied. This is because we vary mean free path but also because we assume a realistic uncertainty for the thermal parameters. Our $\langle f_{\mathrm{HI}} \rangle$ measurements account for the uncertainty in mean free path as well as thermal parameters. As a result, the uncertainty of our $\langle f_{\mathrm{HI}} \rangle$ measurements is also larger and more realistic. McGreer et al. (2015) have placed upper limits on $\langle f_{\mathrm{HI}} \rangle$ at $z > 5.5$ using the dark pixel statistics for Lyα forest observations. Our $\langle f_{\mathrm{HI}} \rangle$ evolution is consistent with the upper limits from McGreer et al. (2015); Zhu et al. (2022) and the lower limits from Fan et al. (2006); Yang et al. (2020); Bosman et al. (2022). Recently, Choudhury et al. (2021) measured the ionization fraction $\langle f_{\mathrm{HII}} \rangle$ by modeling the fluctuations in the ionizing radiation field with their code SCRIPT (see also Maity & Choudhury 2022b,a). They measured $\langle f_{\mathrm{HI}} \rangle = 1 - \langle f_{\mathrm{HII}} \rangle$ at $z > 5.2$. Our $\langle f_{\mathrm{HI}} \rangle$ constraints are in good agreement with theirs but we again report somewhat larger and we think more realistic errors. Note further that Choudhury et al. (2021) used the Bosman et al. (2018) $\tau_{\mathrm{eff,HI}}$ measurements whereas in this work the $\tau_{\mathrm{eff,HI}}$ we used is taken from the large number of high quality XQR-30 spectra from Bosman et al. (2022). We refer the reader to online appendix H and Fig. H3 for a more extensive comparison of our $\langle f_{\mathrm{HI}} \rangle$ measurements to measurements in the literature at $5 \leq z \leq 8$. Fig. 10 also illustrates that the best fit $\langle f_{\mathrm{HI}} \rangle$ shows a rapid evolution from z=5.4 to z=5.2 and $\langle f_{\mathrm{HI}} \rangle$ is consistent with uniform UVB models at $z < 5.3$. This is also what we found from Fig. 8 where the $\tau_{\mathrm{eff,HI}}$ scatter is equally well reproduced by inhomogeneous as well as uniform UVB models. Here, we define the end of reionization as the highest redshift at which the uniform UVB models can reproduce the observed $\tau_{\mathrm{eff,HI}}$ distribution (by rescaling Γ_{HI}). Thus based on our definition, we conclude that the observations suggest that H I reionization is only fully completed by $z \lesssim 5.2$, consistent with the interpretation of Kulkarni et al. (2019); Bosman et al. (2022).

6.4 Ionizing photon emission rate ($\dot{n}$)

With measurements of the evolution of $\lambda_{\mathrm{mfp,HI}}$, $\langle \Gamma_{\mathrm{HI}} \rangle$ and f_{HI} with realistic uncertainties, we can also estimate the evolution of the ionizing emissivity at the Lyman limit edge (ϵ_{912}) and the ionizing photon number density per unit time ($\dot{n}$). These quantities are important ingredients of cosmolog-

ical radiative transfer simulations of reionization. Furthermore, surveys of high-redshift galaxies are used to constrain these parameters from observations (Bouwens et al. 2015). Estimates of these quantities from QSO absorption spectra provide important complementary measurements.

During the late stages of reionization, the mean free path of ionizing photons is shorter than the horizon size. Meiksin & White (2003) show that with this 'absorption limited approximation', the mean free path and the angle averaged UVB intensity $J(\nu)$ are related by $4\pi J(\nu) = \epsilon(\nu)\,\lambda(\nu)$, where we explicitly write the dependence of mean free path on frequency as $\lambda(\nu)$. Integrating the above equation with respect to frequency gives the spatially averaged photoionization rate,

$$\langle \Gamma_{\mathrm{HI}} \rangle = \int_{\nu_L}^{\infty} \frac{\epsilon(\nu)\,\lambda(\nu)\,\sigma(\nu)}{h_p \nu}\,d\nu, \qquad (8)$$

where $\sigma(\nu) = \sigma_L\,(\nu/\nu_L)^{-3}$ is the photo-ionization cross-section, $\sigma_L = 6.34 \times 10^{-18}$ cm^2 for H I, h_p is Planck constant and ν_L is frequency corresponding to 912 Å (Verner et al. 1994; Ferland et al. 1998). We assume here that emissivity scales as $\epsilon(\nu) = \epsilon_{912}\,(\nu/\nu_L)^{-\alpha_s}$ and that the mean free path scales as $\lambda(\nu) = \lambda_{\mathrm{mfp,HI}}\,(\nu/\nu_L)^{3(\beta_{\mathrm{HI}}-1)}$. We vary α_s in the range 2.0 ± 0.6 and the slope of the column density distribution function in the range $\beta_{\mathrm{HI}} = 1.3 \pm 0.05$, consistent with Becker & Bolton (2013). Note that the range in α_s and β_{HI} used in this work is similar to that assumed in the literature (see Becker & Bolton 2013, for details). Hence our uncertainty on $\dot{n}$ should be considered conservative. The emissivity at 912 Å (in units of 10^{24} ergs s^{-1} cMpc^{-3} Hz^{-1}) is then given by,

$$\epsilon_{912} = \left(\frac{4.608}{\mathcal{I}_{\epsilon,\nu}} \right) \left(\frac{\langle \Gamma_{\mathrm{HI}} \rangle}{10^{-12}\,s^{-1}} \right) \left(\frac{10\,\mathrm{pMpc}}{\lambda_{\mathrm{mfp,HI}}} \right) \left(\frac{6}{1+z} \right)^3, \quad (9)$$

where $\mathcal{I}_{\epsilon,\nu} = (\alpha_s - 3\beta_{\mathrm{HI}} + 6)^{-1}$ and $\langle \Gamma_{\mathrm{HI}} \rangle$, $\lambda_{\mathrm{mfp,HI}}$ are the measured photo-ionization rate and the mean free path expressed in units of s^{-1} and pMpc, respectively. It is important to note that the emissivity measured above is the average intergalactic emissivity of all ionising sources and accounts for the escape fraction. The emissivity in UV background modeling is calculated by integrating the luminosity of sources and spectral energy distribution of the sources (Haardt & Madau 2012; Puchwein et al. 2019).

Given the emissivity, it is straightforward to calculate the ionizing photon emission rate at a given epoch,

$$\dot{n}(t) = \int_{\nu_L}^{\infty} \frac{\epsilon(\nu)}{h_p \nu}\,d\nu = \frac{\epsilon_{912}(t)}{h_{\mathrm{p}}\,\alpha_s}, \qquad (10)$$

where we assume that $\epsilon(\nu) \propto \nu^{-\alpha_s}$.

Fig. 11 compares the evolution of $\dot{n}$ obtained in this way with that from observations and simulations from the literature. Our measured $\dot{n}$ evolution is within 1σ of that from Bouwens et al. (2015); Becker & Bolton (2013); Mason et al. (2019); Becker et al. (2021). The $\dot{n}$ measured by Becker et al. (2021) at $z \sim 6$ is systematically higher than our measurements. This is possibly due to the mean free path being assumed to be smaller and the photo-ionization rate being larger in Becker et al. (2021) than our measurements (since $\dot{n} \propto \Gamma_{\mathrm{HI}} \lambda_{\mathrm{mfp,HI}}^{-1}$). The $\dot{n}$ estimated from galaxy
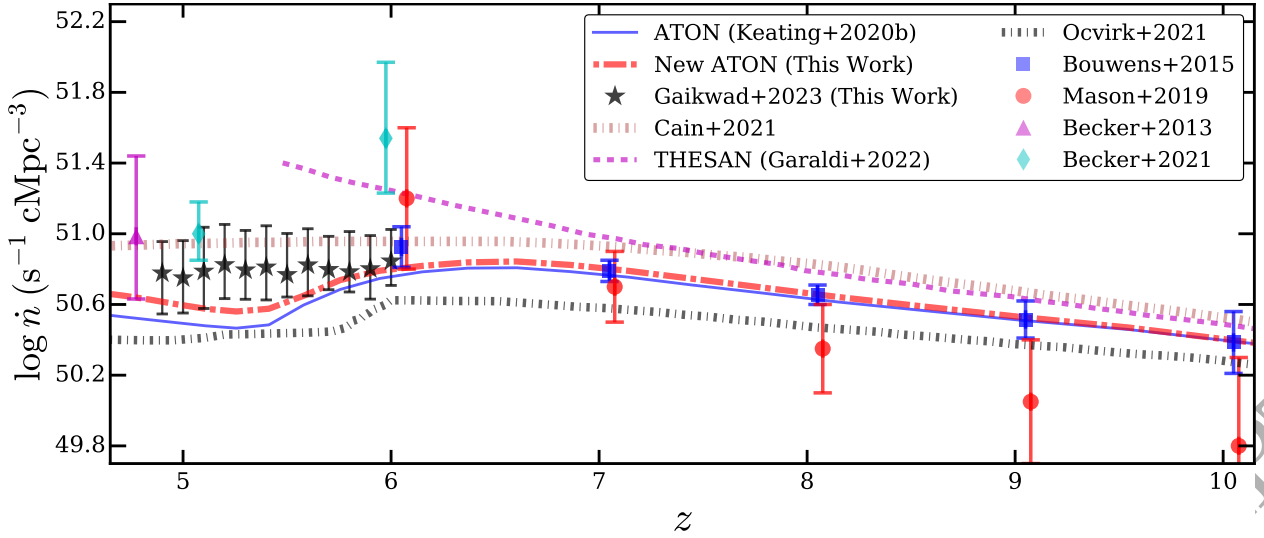
Figure 11. The figure compares the ionizing photon emission rate $\dot{n}$ calculated in this work with that from observations of Becker & Bolton (2013, magenta triangles), Bouwens et al. (2015, blue squares), Mason et al. (2019, red circles) and Becker et al. (2021, cyan diamonds). The uncertainty in our measured $\dot{n}$ accounts for the uncertainty in measured $\lambda_{\rm mfp,HI}$, $\langle\Gamma_{\rm HI}\rangle$, spectral energy distribution index of ionizing source ($\alpha_{\rm s} = 2.0 \pm 0.6$) and slope of column density distribution function. at $16 < \log N_{\rm HI} < 18$ ($\beta_{\rm HI} = 1.3 \pm 0.05$). Our $\dot{n}$ measurements do not show a significant evolution with redshift at $4.9 \leq z \leq 5.9$. Our $\dot{n}$ evolution at $z = 5.1$ is consistent with that from Becker et al. (2021). The $\dot{n}$ at $z \sim 6$ in Becker et al. (2021) is somewhat higher (but still within 1σ) than our best-fit $\dot{n}$ at the same redshift. The differences are mainly due to the differences in the best fit mean free path measurements in the two analysis. Similarly our $\dot{n}$ measurements at $z \sim 6$ are consistent with those from Bouwens et al. (2015); Mason et al. (2019). The evolution of $\dot{n}$ used in the *Aton* simulation (Keating et al. 2020b, solid blue curve, their low $\tau_{\rm CMB}$ hot model) simulation and the *CoDa-III* (Ocvirk et al. 2021, black dot curve), Cain et al. (2021, brown dot curve), *Thesan* Garaldi et al. (2022, magenta dashed curve) simulations are also shown. The $\dot{n}$ evolution from the *Aton* (at $< 1.4\sigma$) and Cain et al. (2021, at $< 1\sigma$) model are also in good agreement with our measurements. The *CoDa-III* simulation has systematically lower $\dot{n}$ while the *Thesan* simulation has systematically higher $\dot{n}$. The $\dot{n}$ evolution from our new *Aton* model is shown by the red dot dashed curve. Our new *Aton* model that matches $\langle\Gamma_{\rm HI}\rangle$ and $\lambda_{\rm mfp,HI}$ evolution also matches the measured $\dot{n}$ evolution at $< 1.2\sigma$.

surveys assume values of ionizing photon production efficiency $\log_{10} \xi_{\rm ion}$ in the range of 25.2 to 25.46 and escape fractions $f_{\rm esc}$ in the range of 10-30 percent (Bouwens et al. 2015; Mason et al. 2019).

6.5 Implications for HI reionization

In this section, we discuss implications of our measured $\langle\Gamma_{\rm HI}\rangle$, $\lambda_{\rm mfp,HI}$ and $\langle f_{\rm HI}\rangle$ evolution for reionization by comparing with the evolution in radiative transfer simulations. We consider four main radiative transfer simulations in the literature that attempt to model the fluctuations in the ionizing radiation field during HI reionization. These are the *Aton* simulations by Kulkarni et al. (2019); Keating et al. (2020a,b), *Thesan* (Kannan et al. 2022; Garaldi et al. 2022), *CoDa-III* (Ocvirk et al. 2020, 2021; Lewis et al. 2022) and a simulation by Cain et al. (2021). These simulations differ in box size, mass resolution, implementation of solving the radiative transfer equation, coupling to hydrodynamics and sub-grid galaxy formation physics etc. These simulations all self-consistently predict the evolution of $\langle\Gamma_{\rm HI}\rangle$, $\lambda_{\rm mfp,HI}$ and $\langle f_{\rm HI}\rangle$. In Fig. 10, we show these three parameters from the *Aton*, *Thesan* and *CoDa-III* and Cain et al. (2021) simulations. The *Thesan* simulation has been performed only down to $z \sim 5.4$. The residual neutral fraction evolution has not been tracked self-consistently in the Cain et al. (2021) simulation, hence we do not show $\langle f_{\rm HI}\rangle$ for this simulation. The $\langle\Gamma_{\rm HI}\rangle$ evolution from the four models are consistent with

our measurements usually within 1.2σ at all redshifts. The $\langle\Gamma_{\rm HI}\rangle$ evolution in *CoDa-III* is consistent with our measurements at $z > 5.5$, while $\langle\Gamma_{\rm HI}\rangle$ in *CoDa-III* is systematically higher than our best fit $\langle\Gamma_{\rm HI}\rangle$ at $z < 5.5$. Note that HI reionization in *CoDa-III* is more rapid and $\langle\Gamma_{\rm HI}\rangle$ is higher than that measured from observations at $z < 5.0$. The $\langle\Gamma_{\rm HI}\rangle$ evolution in Cain et al. (2021) and *Thesan* is in good agreement with our measurements at $5 \leq z \leq 6$ and $z > 5.4$, respectively.

The middle panel in Fig. 10 shows the comparison of $\lambda_{\rm mfp,HI}$ from the four RT models and from this work. Similar to $\langle\Gamma_{\rm HI}\rangle$, the $\lambda_{\rm mfp,HI}$ evolution in the four RT simulations is usually within 1.5σ of our best fit measurements with some small differences. The $\lambda_{\rm mfp,HI}$ from the *Aton* simulations is slightly higher than the best fit values at $z < 5.6$. This could be mainly due to the fact that reionization in the *Aton* simulations completes slightly late. *CoDa-III* predicts a systematically lower $\lambda_{\rm mfp,HI}$ at $z > 5.5$, but is in good agreement with Becker et al. (2021) at $z = 6.0$. Note that the mean free path in *CoDa-III* is measured in ionized regions only. This may lead to biased measurements compared to observations. The observed Lyα forest shows significant numbers of long dark gaps expected to signpost neutral regions (see Zhu et al. 2021) along with the occurrence of transmission spikes from ionized regions (see Gaikwad et al. 2020). It is difficult to compare a mean free path that is measured in ionized regions only in simulations to the observations. The $\lambda_{\rm mfp,HI}$ evolution in both *Aton* and *CoDa-III* simulations match

well our best fit measurements at $z < 5.5$. The Cain et al. (2021) simulation systematically underpredicts the $\lambda_{\mathrm{mfp,HI}}$ at $z < 5.9$ and is in $> 2\sigma$ tension with the Becker et al. (2021) measurements at $z = 5$. This suggests that reionization in the Cain et al. (2021) simulation is somewhat later than indicated by the observations. The *Thesan* simulation agrees well with the best fit $\lambda_{\mathrm{mfp,HI}}$ evolution at $z > 5.4$. One can also see correlations between $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$ in the three RT models in Fig. 10. The RT model that predicts larger (smaller) $\langle\Gamma_{\mathrm{HI}}\rangle$ compared to the best fit values, also predict larger (smaller) $\lambda_{\mathrm{mfp,HI}}$ at $z > 5.5$. The *Thesan* model that matches well the $\langle\Gamma_{\mathrm{HI}}\rangle$ evolution also matches well the $\lambda_{\mathrm{mfp,HI}}$ evolution.

The right panel in Fig. 10 compares the reionization history as probed by the $\langle f_{\mathrm{HI}}\rangle$ evolution in the three RT models and our best fit measurements. The $\langle f_{\mathrm{HI}}\rangle$ evolution in the *Aton* simulation is in good agreement (1.2σ) with our measurements at all redshifts except at $z = 5.2$. This is because the reionization in the *Aton* simulation appears to complete slightly late compared to what is suggested by the data as our best fit $\langle f_{\mathrm{HI}}\rangle$ at $z < 5.5$ is systematically lower than that from the *Aton* simulation. This is expected as the $\langle\Gamma_{\mathrm{HI}}\rangle$ evolution in the *Aton* is also systematically lower at $z < 5.5$, while the $\langle f_{\mathrm{HI}}\rangle$ values are systematically higher in the corresponding redshift range. The $\langle f_{\mathrm{HI}}\rangle$ evolution in *CoDa-III* is in very good agreement with our best fit $\langle f_{\mathrm{HI}}\rangle$ measurements at $z > 5.5$. However, $\langle f_{\mathrm{HI}}\rangle$ drops very quickly from $z = 5.6$ to $z = 5.5$ suggesting that the reionization is more rapid in *CoDa-III* compared to what is suggested by the observations. Reionization in *CoDa-III* is already largely completed by $z \leq 5.6$. This is inconsistent (at $> 2.5\sigma$) with our best fit $\langle f_{\mathrm{HI}}\rangle$ measurements at $z = 5.4, 5.5$. The *Thesan* simulation shows good agreement of the $\langle f_{\mathrm{HI}}\rangle$ evolution with our measurements at all redshifts. At $z \geq 5.7$, *Thesan* shows systematically lower $\langle f_{\mathrm{HI}}\rangle$ compared to our best fit values. This is consistent with their $\langle\Gamma_{\mathrm{HI}}\rangle$ evolution being higher at $z > 5.7$ compared to our best fit measurements in the left panel of Fig. 10.

In Fig. 10, we show results from a new *Aton* simulation that is performed with a slight change in the emissivity evolution and spectral energy distribution of sources. The new *Aton* simulation was performed independently and prior to the analysis presented in this work. The $\langle\Gamma_{\mathrm{HI}}\rangle$, $\lambda_{\mathrm{mfp,HI}}$ and $\langle f_{\mathrm{HI}}\rangle$ evolution in the new *Aton* model are in good agreement with the best fit measurements. This shows that the observed evolution of $\langle\Gamma_{\mathrm{HI}}\rangle$, $\lambda_{\mathrm{mfp,HI}}$ and $\langle f_{\mathrm{HI}}\rangle$ can be reproduced with the current set of radiative transfer simulations albeit one needs to tune the sub-grid physics to match the observed mean flux. We plan to present a detailed analysis of the new *Aton* simulations together with other models in future work (in prep.).

In Fig. 11, we compare the $\dot{n}$ evolution used in the *Aton*, *Thesan*, *CoDa-III* and Cain et al. (2021) simulations. Our $\dot{n}$ evolution seems to be consistent with that from Cain et al. (2021) within $< 1\sigma$. The $\dot{n}$ evolution in *Thesan* (*CoDa-III*) is systematically larger (smaller) than our $\dot{n}$ measurements at $4.9 \leq z \leq 6$. The $\dot{n}$ evolution in the Keating et al. (2020b, their low τ_{CMB} hot model) *Aton* model is calibrated to match the observed Lyα transmitted mean flux. Their $\dot{n}$ evolution is consistent with the $\dot{n}$ measurements at $< 1.4\sigma$. The $\dot{n}$ evolution in our new *Aton* model is in slightly better agreement $(< 1.2\sigma)$ with our $\dot{n}$ measurements than the

Keating et al. (2020b) model. It is also noteworthy that the $\dot{n}$ evolution in the *Aton* models is consistent with the measurements from Bouwens et al. (2015) at $z > 6$. Thus, the new *Aton* model that matches observed $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\langle f_{\mathrm{HI}}\rangle$ evolution also produces a $\dot{n}$ evolution consistent with our measurements.

In summary, the *Aton*, *Thesan*, *CoDa-III* and Cain et al. (2021) RT models show good agreement with our measured $\langle\Gamma_{\mathrm{HI}}\rangle$, $\lambda_{\mathrm{mfp,HI}}$ and $\langle f_{\mathrm{HI}}\rangle$ evolution with some distinct differences. (i) The end of reionization in the *Aton* model is slightly later than that suggested by the latest observations, (ii) The reionization in the *CoDa-III* simulation is rapid and completes at $z \leq 5.6$ whereas the observations suggest it to be completed by $z \leq 5.2$, (iii) *Thesan* shows good agreement for all three parameters, but it is unclear when reionization completes in *Thesan*, (iv) the $\lambda_{\mathrm{mfp,HI}}$ evolution in Cain et al. (2021) is systematically lower compared to our best fit evolution and (v) our latest new *Aton* simulation shows good agreement with all the best fit parameters measured in this work.

## 6.6 Consistency of our best fit model with transmission spike and dark gap statistics

In this work, we have primarily used the $\tau_{\mathrm{eff,HI}}$ CDF statistics to measure the $\langle\Gamma_{\mathrm{HI}}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$ parameters because $\tau_{\mathrm{eff,HI}}$ CDF is relatively straightforward to calculate and it is one of the most robust statistics that can be derived from absorption spectra. However, the the count and location of transmitted spikes provides important additional information (Zhu et al. 2021, 2022). Gaikwad et al. (2020) in particular showed that the number of transmission spikes per unit redshift interval (as characterized by the pseudo column density distribution function pCDDF) is sensitive to photo-ionization rates. Deriving these statistics from observations and simulations is usually challenging. The definition of dark gaps is usually based on the observed noise threshold (S/N per pixel) while fitting Voigt profiles to a large number of (inverted) transmitted flux is computationally expensive. In this section, we therefore just check whether our simulations with best fit parameters derived from the $\tau_{\mathrm{eff,HI}}$ CDF are consistent with observed dark gap statistics and pCDDF statistics.

Fig. 12 shows lightcones of overdensity Δ, mean free path $(\lambda_{\mathrm{mfp,HI}})$, ionizing radiation field fluctuations $(\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle)$ and neutral hydrogen fraction (f_{HI}) from our simulations. The IGM at high redshift $(z > 5)$ traces linear $(\Delta \sim 1)$ or mildly non-linear $(\Delta < 10)$ overdensities. As we have already discussed the fluctuations in the cosmic density field alone are not sufficient to reproduce the large scatter seen in the observed $\tau_{\mathrm{eff,HI}}$ distribution (see Fig. 8). We have used the best fit evolution of mean free path $\lambda_{\mathrm{mfp,HI}}$ and spatially average photo-ionization rate $\langle\Gamma_{\mathrm{HI}}\rangle$ parameters to generate the lightcones of $\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$ and f_{HI} shown in Fig. 12. To construct lightcones, (i) we extract multiple slices from a given simulation box at an angle of $20°$ exploiting the periodicity of the boundary conditions, (ii) we extract these multiple slices at 6 different redshifts for which the simulation outputs are stored, (iii) we calculate the redshift axis spanning $4.9 \leq z \leq 6.1$ and with a dz that corresponds to the spatial resolution of the simulation box and (iv) we calculate the co-evolution of the quantity of interest by map-
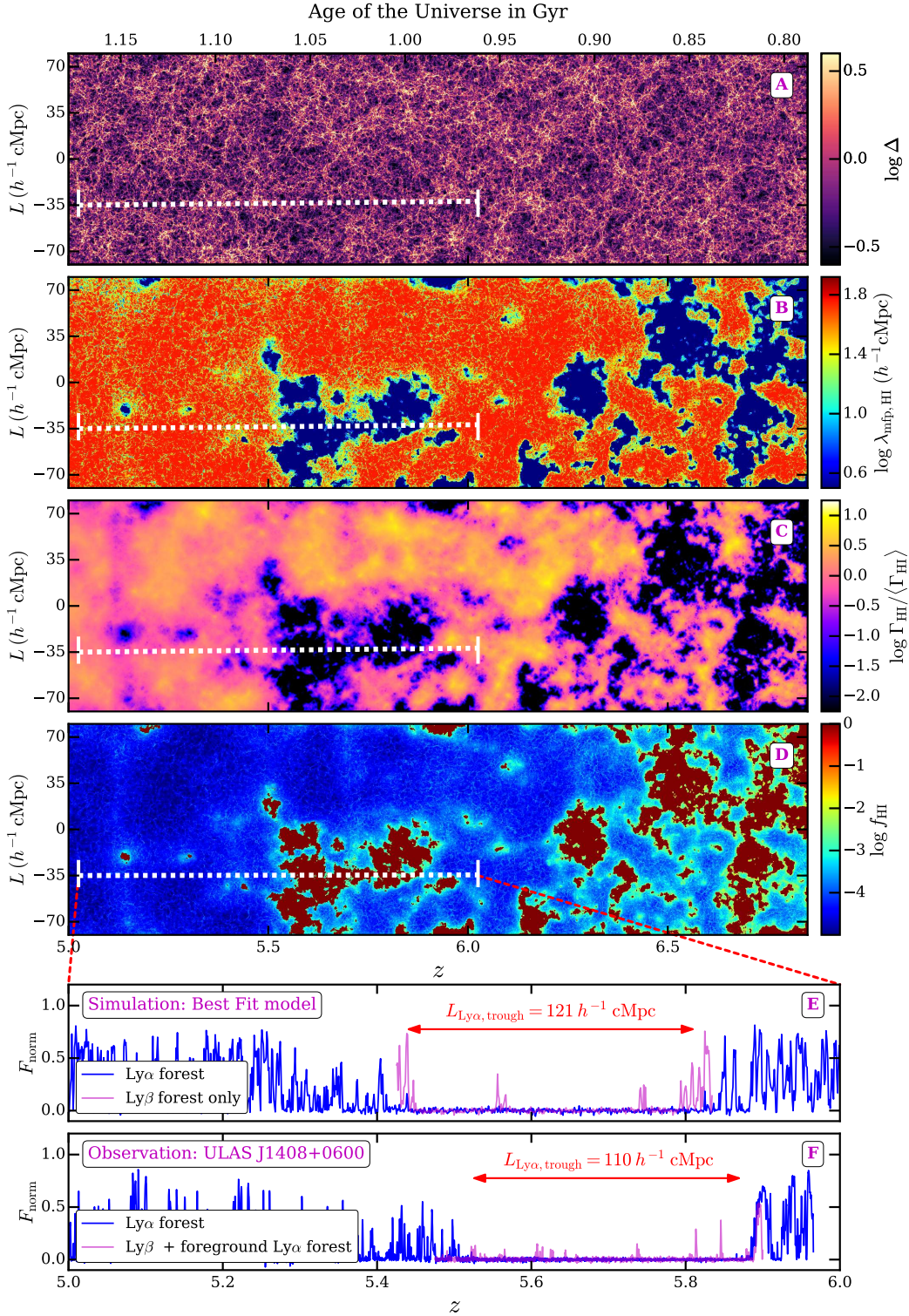
Figure 12. Panel A, B, C and D show lightcones of overdensity (Δ), spatial fluctuations in mean free path ($\lambda_{\mathrm{mfp,HI}}$), Γ_{HI} fluctuations ($\Gamma_{\mathrm{HI}}/\langle\Gamma_{\mathrm{HI}}\rangle$) and neutral hydrogen fraction (f_{HI}) respectively. The redshift evolution of these quantities is calculated by interpolating the respective quantities from simulation snapshots at 6 different redshifts. The evolution of $\lambda_{\mathrm{mfp,HI}}$ fluctuations, Γ_{HI} fluctuations and f_{HI} corresponds to a model with the best fit measured $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{\mathrm{HI}}\rangle$ parameters. Panel A shows that the typical intergalactic densities are linear with $\Delta \sim 1$ to mildly non-linear with $\Delta < 10$. The fluctuations in the cosmic density field, alone, can not reproduce the $\tau_{\mathrm{eff,HI}}$ distribution and long dark gaps in observed spectra. The evolution of mean free path (panel B) produces the fluctuations in ionizing radiation fields (panel C) and thereby fluctuations in neutral fraction (panel D). The regions that are yet to receive ionizing radiation are neutral and persists down to $z \sim 5.4$. The sightlines (white dash line) passing through a neutral region that produces the long trough of length $\sim 121\,h^{-1}$ cMpc (as shown in panel E). A similarly long trough ($\sim 110\,h^{-1}$ cMpc) has been observed in the absorption spectrum towards QSO ULAS J1408+0600 shown in panel F. The corresponding Lyβ forest in the trough region is shown by magenta line. Our best fit models, that is consistent with observations, shows a small number of neutral regions at $z < 5.4$. This figure qualitatively nicely demonstrates that reionization is only fully completed by $z \leq 5.2$ in models that fit the data.
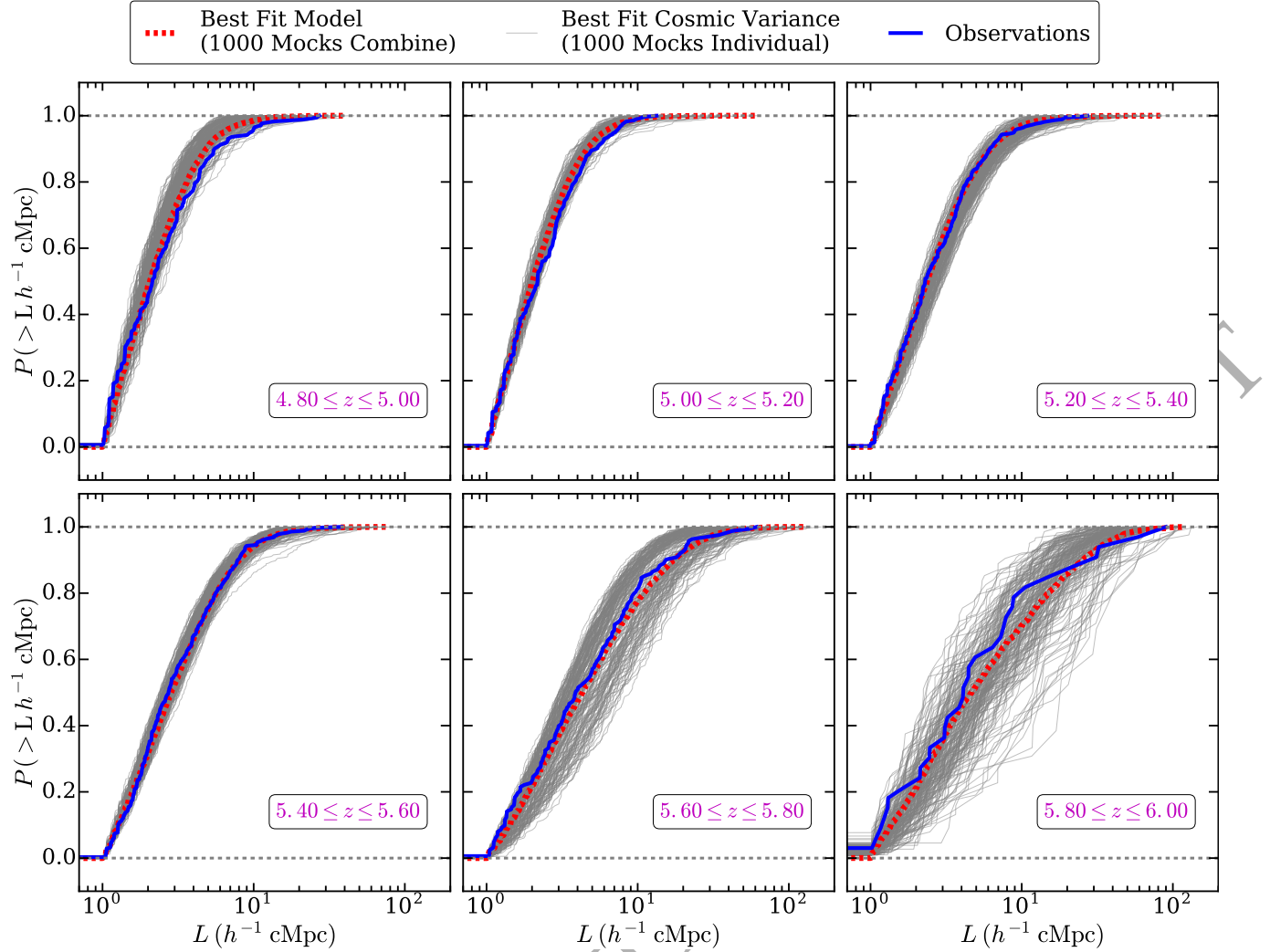
Figure 13. The panels show a comparison of the cumulative distribution function of dark gap lengths in observation with that from our best fit model at $4.8 \leq z \leq 6.0$. The dark gaps are chosen from the Lyα forest regions only, excluding proximity region and Lyβ emission (i.e. rest frame $\lambda < 1180$ Å). We have chosen a slightly larger redshift bin size of $\Delta z = 0.2$ here because the largest observed long dark gap has a length of 110 h^{-1} cMpc. The dark gaps are defined as the contiguous regions of the spectra with $F < F_{\text{threshold}}$ where $F_{\text{threshold}} = 0.05$ corresponding to the lowest S/N in our observed sample. The blue curve shows the observed dark gap length CDF, while the gray curve shows the dark gap length CDF from individual best fit mocks. Each mock has a redshift path length similar to the observed sample in the given redshift bin. The red dashed curve shows the dark gap length CDF calculated from all the mocks. The observed dark gap CDF is well reproduced by our best fit model.

ping on to the redshift axis using linear interpolation. For $z > 6$, we have not measured $\lambda_{\text{mfp,HI}}$, $\langle\Gamma_{\text{HI}}\rangle$. In this case we use $\lambda_{\text{mfp,HI}}$, $\langle\Gamma_{\text{HI}}\rangle$ values similar to that of the *Aton* late reionization model discussed in the previous section. However, the fluctuations in Γ_{HI} and f_{HI} are calculated using EX-CITE. We emphasize that the lightcones and spectra are shown for qualitative analysis and illustration purposes. The lightcones are not used to derive any statistics or perform any parameter estimation.

Fig. 12 shows that the fluctuations in Γ_{HI} correlate well with the fluctuations in f_{HI}. There is a significant number of regions at $z > 5.4$ that are yet to receive ionizing radiation and are still neutral. When a sightline passes through such neutral regions, a long dark gap or trough occurs in the absorption spectrum (Keating et al. 2020a). Panel D in Fig. 12 shows an example of a long trough of length $126 h^{-1}$cMpc

in the lightcones. A similarly long trough has been observed along the sightline towards QSO ULAS J1408+0600 (panel E). Fig. 12 also nicely illustrates the disappearance of the last neutral islands at $z < 5.2$ and that reionization is only fully completed by $z \leq 5.2$ in models that fit the data. This is consistent with the observations of the $\tau_{\text{eff,HI}}$ CDF as shown in Fig. 8 (see also Bosman et al. 2022).

Recently Zhu et al. (2021, 2022) quantified the statistics of such dark gaps in the observations and showed that they are a useful diagnostic statistics to constrain the properties of reionization. In this work, we define the dark gap in the same way as done by Zhu et al. (2021, see also§5.5 for details). We use the dark gap statistics only in the Lyα forest region of the spectra i.e. between the Lyβ and Lyα wavelength of the QSO emission. The proximity region and the Lyβ emission line profile are excluded from the analysis.
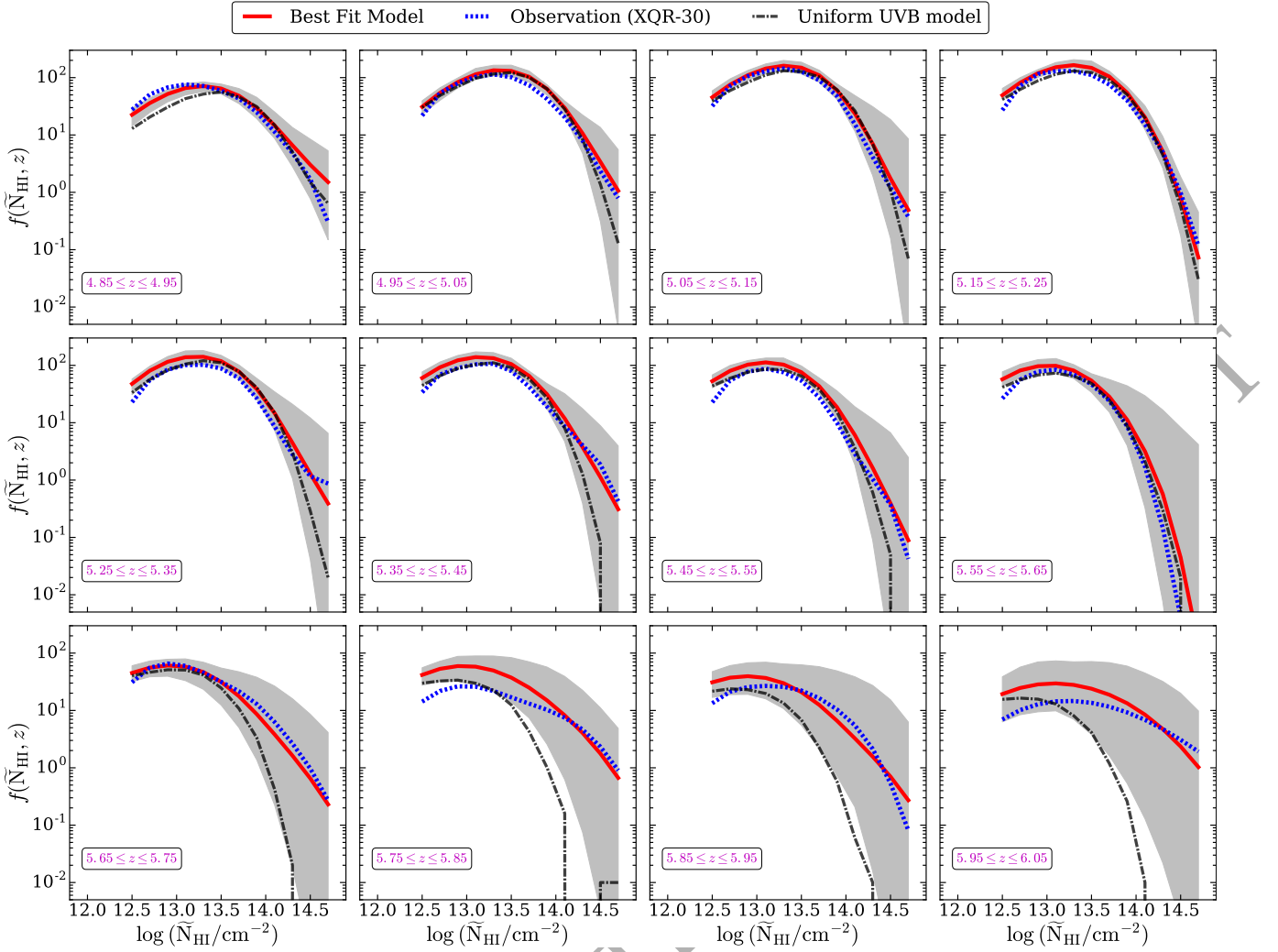
Figure 14. The figure compares the pseudo-column density distribution function (pCDDF) from observations (blue dashed curve) with that from a uniform UVB Haardt & Madau (2012) model (black dashed curve) and our best fit model (red solid curve). The transmission spikes are fitted with inverted Voigt profiles to obtain pseudo-column densities, $\log\widetilde{N}_{\mathrm{HI}}$. The turnover in the pCDDF at lower $\log\widetilde{N}_{\mathrm{HI}}$ values is due to the incompleteness of the sample that has not been accounted for in the simulations or observations. The normalization of the pCDDF is sensitive to $\langle\Gamma_{\mathrm{HI}}\rangle$ while the shape is sensitive to $\lambda_{\mathrm{mfp,HI}}$. The 1σ uncertainty (gray shaded) in the pCDDF is calculated from 1000 mocks. The uncertainty in the pCDDF is larger at higher redshift because of the less frequent occurrence of transmission spikes. The uniform UVB model fails to reproduce the pCDDF at $z > 5.7$ because of the larger mean free path. However, the agreement between uniform UVB model and observations is reasonably good at $z < 5.7$ although the shape of pCDDF at the high $\log\widetilde{N}_{\mathrm{HI}}$ end is somewhat steeper in uniform UVB models. Our best fit model (that includes fluctuations in ionizing background) is in good agreement with observations at all redshifts.

For a fair comparison, we calculate the dark gap lengths from simulations as well as observations because our redshift bins are slightly different than those used in Zhu et al. (2021). For this analysis, we chose a slightly larger redshift bin size $\Delta z = 0.2$, because the size of observed dark gaps is as long as $\sim 110\,h^{-1}$ cMpc. Our box size $L \sim 160\,h^{-1}$ cMpc is sufficiently large to model the longest observed trough. Fig. 13 shows the comparison of the cumulative distribution function of dark gap lengths in our best fit models and that from observations. The observed dark gap distribution is well reproduced by the best fit model in all the redshift bins. The scatter in dark gap lengths increases with redshift such that larger dark gaps are more frequent at higher redshift. Note that the cumulative distribution function from all the sightlines in a simulation box (red curve) is in good agreement

with that of that obtained from the observed spectra at all redshifts.

The dark gaps in spectra correspond to relatively neutral regions along the sightlines whereas the location of transmission spikes are sensitive to ionized (or under-dense) regions along the sightlines. Thus, the statistics derived from transmission spikes is complementary to the dark gap statistics. The transmission spikes are usually described by height, width and location parameters. The height of the transmission spikes is mainly sensitive to the mean and fluctuations in the ionizing radiation field. The width of the transmission spikes is sensitive to the temperature of the IGM. The resolution of the observed and simulated spectra used in this work is not sufficient to resolve the transmission spike widths (see Gaikwad et al. 2020, for a detailed discussion

on resolving transmission spikes). Hence in this work we mainly focus on comparing the height distribution from simulations with observations. Following Gaikwad et al. (2020), we decompose the transmission spikes into multi-component Voigt profiles. We fit the inverted flux with Voigt profiles using VIPER (Gaikwad et al. 2017b). We calculate the pseudo Column Density Distribution Function (pCDDF) from the observations and simulations as explained in §5.5. We use only XShooter spectra for comparison of the pCDDF from simulations and models. Fig. 14 compares the pCDDF of our best fit model with that from the observations. The turnover seen in the observed pCDDF at low $\log \widetilde{N}_{HI}$ is due to the incompleteness that has not been accounted for in the simulations and observations. The uncertainty in the pCDDF is calculated from 1000 mocks. As before, each mock has the same redshift path length as the observations in each redshift bin. Fig. 14 shows that the uncertainty in the pCDDF is larger at higher redshift. This is because the transmission spikes occur less frequently at high redshift and the uncertainty is dominated by Poisson statistics. Our best fit models are consistent with observations within 1σ in all the redshift bins. The shape of the pCDDF is significantly different for uniform UVB models and fails to match the observations at $z \geq 5.7$. At lower redshift $z < 5.7$, the pCDDF from uniform UVB is in relatively good agreement with the observations. However, we can not conclude from this that reionization is completed by $z \sim 5.7$. The transmission spikes probe only the ionized (or under-dense) regions. Our definition of the end of reionization based on the $\tau_{\mathrm{eff,HI}}$ distribution uses the information of transmission spikes as well as dark gaps and is more robust.

The best fit $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{HI}\rangle$, $\langle f_{HI}\rangle$ parameter model that matches the observed $\tau_{\mathrm{eff,HI}}$ distribution, also reproduces the observed dark gap statistics and the pseudo-Column Density Distribution Function (pCDDF). This consistency of our best fit model with the two additional statistics suggest that our $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{HI}\rangle$ and $\langle f_{HI}\rangle$ measurements are robust and reasonably accurate.

7 SUMMARY

We have measured the mean free path $\lambda_{\mathrm{mfp,HI}}$ of H I ionizing photons, the spatially averaged photo-ionization rate $\langle\Gamma_{HI}\rangle$ of H I and the spatially averaged neutral fraction $\langle f_{HI}\rangle$ of hydrogen at $4.9 \leq z \leq 6.0$ from a new sample of 67 XShooter and ESI QSO absorption spectra of unprecedented quality. $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{HI}\rangle$ and $\langle f_{HI}\rangle$ are measured by comparing the statistics of the Lyα forest at $4.9 \leq z \leq 6.0$ with state-of-the-art cosmological simulations post-processed with our new code EX-CITE that captures the fluctuation in the ionizing radiation field. Our main results are as follows

• We have developed the EX-CITE code based on the Code for Ionization and Temperature Evolution (CITE) to capture the fluctuations in the ionizing radiation field that are important during H I reionization. EX-CITE generates $\Gamma_{HI}/\langle\Gamma_{HI}\rangle$ maps using an iterative method to account for a spatially fluctuating mean free path. The octree summation implemented in EX-CITE requires $\mathcal{O}(N \log N)$ operations to generate $\Gamma_{HI}/\langle\Gamma_{HI}\rangle$ maps at a given redshift allowing us to probe a large parameter space of $\lambda_{\mathrm{mfp,HI}} -\langle\Gamma_{HI}\rangle$ combinations. Using EX-CITE, we have generated 648 $\Gamma_{HI}/\langle\Gamma_{HI}\rangle$ maps

(resolution $0.32 h^{-1}$cMpc) with varying $\lambda_{\mathrm{mfp,HI}}$ at 6 redshifts bins for simulations from the Sherwood suite with a box size $L = 160\ h^{-1}$ cMpc and particle number $N_{\mathrm{particle}} = 2048^3$. The resolution and number of models generated in this work using EX-CITE are factors 10 and 35 larger than those in previous studies, respectively. We show good consistency between our simulations with EX-CITE and the state-of-the-art radiative transfer code *Aton*. The $\Gamma_{HI}/\langle\Gamma_{HI}\rangle$ maps show remarkable similarity on large scale while on small scales EX-CITE shows more fluctuations compared to *Aton*. We demonstrate that the $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{HI}\rangle$ in *Aton* simulations can be recovered within 1σ using EX-CITE models at $4.9 \leq z \leq 6$.

• We extract random skewers from the simulations and generate Lyα forest spectra. We forward model the simulated Lyα forest spectra to match properties (e.g., S/N, LSF etc) of our observed sample. We derive three statistics from the simulations and observations namely the CDF of the effective optical depth ($\tau_{\mathrm{eff,HI}}$), the dark gap length CDF and the pseudo-Column Density Distribution Function (pCDDF). We demonstrate the sensitivity of the three statistics to $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{HI}\rangle$. The median and scatter of $\tau_{\mathrm{eff,HI}}$ are sensitive to $\langle\Gamma_{HI}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$, respectively. The normalization and shape of pCDDF are likewise sensitive to $\langle\Gamma_{HI}\rangle$ and $\lambda_{\mathrm{mfp,HI}}$, respectively.

• We have measured $\lambda_{\mathrm{mfp,HI}} -\langle\Gamma_{HI}\rangle$ by comparing the $\tau_{\mathrm{eff,HI}}$ CDF from simulations with observations using a non-parametric two-sample Anderson-Darlington test. Our final $\lambda_{\mathrm{mfp,HI}} -\langle\Gamma_{HI}\rangle$ measurements account for the thermal parameter uncertainty, modeling parameter uncertainty, cosmic variance and observational uncertainties. Using the best fit $\lambda_{\mathrm{mfp,HI}} -\langle\Gamma_{HI}\rangle$ parameters along with the total 1σ uncertainty, we further measure the spatially averaged neutral fraction $\langle f_{HI}\rangle$. A given $\lambda_{\mathrm{mfp,HI}} -\langle\Gamma_{HI}\rangle$ parameter combination leads to a unique $\langle f_{HI}\rangle$ measurement with our EX-CITE models. Using the absorption limited approximation, we also measure the emissivity at 912 Å (ϵ_{912}) and the number density of ionizing photons per unit time ($\dot{n}$).

• We have performed a detailed comparison between our measurements and those in the literature. Our best fit $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{HI}\rangle$ and $\langle f_{HI}\rangle$ measurements are in good agreement with the measurements (or limits) from the literature. Our best fit $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{HI}\rangle$ and $\langle f_{HI}\rangle$ show significant evolution with redshift such that the $\lambda_{\mathrm{mfp,HI}}$ decreases from $z = 4.9$ to $z = 5.9$ by a factor ~ 6. The best fit $\langle\Gamma_{HI}\rangle$ drops from a value of $5.5 \times 10^{-13}\ \mathrm{s}^{-1}$ at $z = 4.9$ to $2 \times 10^{-13}\ \mathrm{s}^{-1}$ at $z = 5.9$. The best fit $\langle f_{HI}\rangle$ evolves from 0.1 at $z = 5.9$ to 2×10^{-5} at $z = 4.9$ with a sudden drop at $z \sim 5.2$. With our measurements of $\lambda_{\mathrm{mfp,HI}}$ and $\langle\Gamma_{HI}\rangle$ we then estimate the ionizing photon emission rate $\dot{n}$. Our $\dot{n}$ measurements at $4.9 \leq z \leq 6.0$ show a fairly constant value of $\log \dot{n} \approx 50.8$ ($\mathrm{s}^{-1}\ \mathrm{cMpc}^{-3}$) that is consistent with lower redshift measurements at $z < 5$ and high redshift measurements at $z > 6$ in the literature. For the models with the best fit $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{HI}\rangle$ and $\langle f_{HI}\rangle$ values we have also compared simulated and observed dark gap length CDF and pCDDF statistics. Both observed statistics are well reproduced by the best fit EX-CITE models and are consistent within 1σ uncertainty at $4.9 \leq z \leq 6.0$.

• We have compared our best fit $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{HI}\rangle$, $\langle f_{HI}\rangle$ and $\dot{n}$ evolution with those from four state-of-the-art cosmological radiative transfer simulations of H I reionization namely our *Aton* simulations from Keating et al. (2020b), the *The-*

san, *CoDa-III* and the simulations by Cain et al. (2021). All four simulations show reasonable agreement with our measured parameters at $4.9 \leq z \leq 6.0$ with some distinct differences. Reionization in our *Aton* model occurs slightly late than that suggested by our measurements. The reionization in the *CoDa-III* simulation is rapid and completes at $z \leq 5.6$ whereas our measurements suggest it to be completed only by $z \leq 5.2$. The simulations by Cain et al. (2021) show a somewhat later end of reionization with $\lambda_{\mathrm{mfp,HI}}$ systematically lower than our best fit measurements. The *Thesan* simulation shows mostly good agreement for all three parameters, but it is unclear when reionization completes in *Thesan*. We have also performed a new *Aton* simulation with a slight change in emissivity evolution and spectral energy distribution of sources. The evolution of $\lambda_{\mathrm{mfp,HI}}$, $\langle\Gamma_{\mathrm{HI}}\rangle$, $\langle f_{\mathrm{HI}}\rangle$ and $\dot{n}$ in the new *Aton* reionization model seems to be in very good agreement with our best fit evolution of these parameters. This illustrates that the observed evolution of the four parameters can be well reproduced in state-of-the-art cosmological radiative transfer simulations with appropriately calibrated ionizing emissivity and energy.

Our $\langle\Gamma_{\mathrm{HI}}\rangle$, $\lambda_{\mathrm{mfp,HI}}$ and $\langle f_{\mathrm{HI}}\rangle$ measurements further corroborate a picture in which reionization is largely completed by $z \sim 5.2$. The observed evolution of these parameters can be well reproduced by state-of-the art radiative transfer simulations by careful calibration of emissivities and spectral energy distribution of sources. The uncertainty in thermal parameters is one of the main source of uncertainty in the current measurement of the reionization history from Lyα forest data. High-resolution, high-S/N spectra are needed to accurately measure the thermal parameters. In future, the high-resolution spectrograph ANDES that is being developed for the Extremely Large Telescope (ELT) and similar high resolution spectrographs on the Thirty Meter Telescope (TMT) and the Giant Magellan Telescope (GMT) will provide unprecedented quality high-redshift QSO absorption spectra that will allow us to extend accurate measurements of the thermal and reionization history of the Universe to higher redshift.

DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material. Any other data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Alvarez M. A., Abel T., 2012, ApJ, 747, 126
Aubert D., Teyssier R., 2008, MNRAS, 387, 295
Barnes J., Hut P., 1986, Nature, 324, 446
Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Becker G. D., Davies F. B., Furlanetto S. R., Malkan M. A., Boera E., Douglass C., 2018, ApJ, 863, 92
Becker G. D., et al., 2019, ApJ, 883, 163
Becker G. D., D'Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, MNRAS, 508, 1853
Bischetti M., et al., 2022, Nature, 605, 244
Bolton J. S., Haehnelt M. G., 2007, MNRAS, 382, 325
Bolton J. S., Viel M., Kim T. S., Haehnelt M. G., Carswell R. F., 2008, MNRAS, 386, 1131
Bolton J. S., Becker G. D., Raskutti S., Wyithe J. S. B., Haehnelt M. G., Sargent W. L. W., 2012, MNRAS, 419, 2880
Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T.-S., Meiksin A., Regan J. A., Viel M., 2017, MNRAS, 464, 897
Bosman S. E. I., 2021, arXiv e-prints, p. arXiv:2108.12446
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055
Bosman S. E. I., et al., 2022, MNRAS, 514, 55
Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015, ApJ, 811, 140
Cain C., D'Aloisio A., Gangolli N., Becker G. D., 2021, ApJ, 917, L37
Cain C., D'Aloisio A., Gangolli N., McQuinn M., 2022, arXiv e-prints, p. arXiv:2207.11266
Cain C., D'Aloisio A., Iršič V., Gangolli N., Dhami S., 2023, J. Cosmology Astropart. Phys., 2023, 002
Calverley A. P., Becker G. D., Haehnelt M. G., Bolton J. S., 2011, MNRAS, 412, 2543
Chardin J., Haehnelt M. G., Aubert D., Puchwein E., 2015, MNRAS, 453, 2943
Chardin J., Puchwein E., Haehnelt M. G., 2017, MNRAS, 465, 3429
Chardin J., Haehnelt M. G., Bosman S. E. I., Puchwein E., 2018a, MNRAS, 473, 765
Chardin J., Kulkarni G., Haehnelt M. G., 2018b, MNRAS, 478, 1065
Chen H., et al., 2022, ApJ, 931, 29
Choudhury T. R., Paranjape A., Bosman S. E. I., 2021, MNRAS, 501, 5782
D'Aloisio A., McQuinn M., Davies F. B., Furlanetto S. R., 2018, MNRAS, 473, 560

D'Aloisio A., McQuinn M., Trac H., Cain C., Mesinger A., 2020, ApJ, 898, 149

D'Odorico V., et al., 2023, MNRAS, 523, 1399

Davies F. B., Furlanetto S. R., 2014, MNRAS, 437, 1141

Davies F. B., Furlanetto S. R., 2016, MNRAS, 460, 1328

Davies F. B., Furlanetto S. R., Dixon K. L., 2017, MNRAS, 465, 2886

Davies F. B., Hennawi J. F., Eilers A.-C., Lukić Z., 2018, ApJ, 855, 106

Davies R. L., et al., 2022, arXiv e-prints, p. arXiv:2211.15816

Eilers A.-C., Davies F. B., Hennawi J. F., 2018, ApJ, 864, 53

Fan X., et al., 2001, AJ, 122, 2833

Fan X., et al., 2006, AJ, 132, 117

Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761

Furlanetto S. R., Oh S. P., 2005, MNRAS, 363, 1031

Gaikwad P., Khaire V., Choudhury T. R., Srianand R., 2017a, MNRAS, 466, 838

Gaikwad P., Srianand R., Choudhury T. R., Khaire V., 2017b, MNRAS, 467, 3172

Gaikwad P., Choudhury T. R., Srianand R., Khaire V., 2018, MNRAS, 474, 2233

Gaikwad P., Srianand R., Khaire V., Choudhury T. R., 2019, MNRAS, 490, 1588

Gaikwad P., et al., 2020, MNRAS, 494, 5091

Gaikwad P., Srianand R., Haehnelt M. G., Choudhury T. R., 2021, MNRAS, 506, 4389

Gallerani S., Choudhury T. R., Ferrara A., 2006, MNRAS, 370, 1401

Garaldi E., Gnedin N. Y., Madau P., 2019, ApJ, 876, 31

Garaldi E., Kannan R., Smith A., Springel V., Pakmor R., Vogelsberger M., Hernquist L., 2022, MNRAS, 512, 4909

Gnedin N. Y., 2014, ApJ, 793, 29

Gnedin N. Y., Madau P., 2022, Living Reviews in Computational Astrophysics, 8, 3

Gnedin N. Y., Becker G. D., Fan X., 2017, ApJ, 841, 26

Haardt F., Madau P., 1996, ApJ, 461, 20

Haardt F., Madau P., 2012, ApJ, 746, 125

Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27

Iliev I. T., Shapiro P. R., Raga A. C., 2005a, MNRAS, 361, 405

Iliev I. T., Scannapieco E., Shapiro P. R., 2005b, ApJ, 624, 491

Jiang L., et al., 2022, Nature Astronomy,

Jin X., et al., 2023, ApJ, 942, 59

Kannan R., Garaldi E., Smith A., Pakmor R., Springel V., Vogelsberger M., Hernquist L., 2022, MNRAS, 511, 4005

Keating L. C., Puchwein E., Haehnelt M. G., 2018, MNRAS, 477, 5501

Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020a, MNRAS, 491, 1736

Keating L. C., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020b, MNRAS, 497, 906

Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24

Lai S., et al., 2022, MNRAS, 513, 1801

Lewis J. S. W., et al., 2022, MNRAS, 516, 3389

Maity B., Choudhury T. R., 2022a, MNRAS, 511, 2239

Maity B., Choudhury T. R., 2022b, MNRAS, 515, 617

Mason C. A., et al., 2019, MNRAS, 485, 3947

McGreer I. D., Mesinger A., Fan X., 2011, MNRAS, 415, 3237

McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499

McQuinn M., 2016, ARA&A, 54, 313

McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043

Meiksin A. A., 2009, Reviews of Modern Physics, 81, 1405

Meiksin A., 2020, MNRAS, 491, 4884

Meiksin A., White M., 2003, MNRAS, 342, 1205

Mesinger A., 2010, MNRAS, 407, 1328

Mesinger A., Furlanetto S., 2009, MNRAS, 400, 1461

Mesinger A., Ewall-Wice A., Hewitt J., 2014, MNRAS, 439, 3262

Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1

Molaro M., et al., 2022, MNRAS, 509, 6119

Muñoz J. A., Oh S. P., Davies F. B., Furlanetto S. R., 2016, MNRAS, 455, 1385

Nasir F., D'Aloisio A., 2020, MNRAS, 494, 3080

Nasir F., Cain C., D'Aloisio A., Gangolli N., McQuinn M., 2021, ApJ, 923, 161

Ocvirk P., et al., 2020, MNRAS, 496, 4087

Ocvirk P., Lewis J. S. W., Gillet N., Chardin J., Aubert D., Deparis N., Thélie É., 2021, MNRAS, 507, 6108

Park H., Shapiro P. R., Choi J.-h., Yoshida N., Hirano S., Ahn K., 2016, ApJ, 831, 86

Park H., Shapiro P. R., Ahn K., Yoshida N., Hirano S., 2021, ApJ, 908, 96

Planck Collaboration et al., 2014, A&A, 571, A16

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical recipes in FORTRAN. The art of scientific computing

Puchwein E., Bolton J. S., Haehnelt M. G., Madau P., Becker G. D., Haardt F., 2015, MNRAS, 450, 4081

Puchwein E., Haardt F., Haehnelt M. G., Madau P., 2019, MNRAS, 485, 47

Puchwein E., et al., 2023, MNRAS, 519, 6162

Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021, MNRAS, 506, 2390

Rauch M., 1998, ARA&A, 36, 267

Rollinde E., Theuns T., Schaye J., Pâris I., Petitjean P., 2013, MNRAS, 428, 540

Rosdahl J., Blaizot J., Aubert D., Stranex T., Teyssier R., 2013, MNRAS, 436, 2188

Rosdahl J., et al., 2018, MNRAS, 479, 994

Shapiro P. R., Iliev I. T., Raga A. C., 2004, MNRAS, 348, 753

Shapiro P. R., Iliev I. T., Alvarez M. A., Scannapieco E., 2006, ApJ, 648, 922

Sheinis A. I., Bolte M., Epps H. W., Kibrick R. I., Miller J. S., Radovan M. V., Bigelow B. C., Sutin B. M., 2002, PASP, 114, 851

Springel V., 2005, MNRAS, 364, 1105

Suzuki N., Tytler D., Kirkman D., O'Meara J. M., Lubin D., 2005, ApJ, 618, 592

Verner D. A., Ferland G., Korista K., 1994, in American Astronomical Society Meeting Abstracts. p. 41.05

Vernet J., et al., 2011, A&A, 536, A105

Viel M., Haehnelt M. G., Springel V., 2004, MNRAS, 354, 684

Weinberg D. H., Katz N., Hernquist L., 1998, in Woodward C. E., Shull J. M., Thronson Harley A. J., eds, Astronomical Society of the Pacific Conference Series Vol. 148, Origins. p. 21 (arXiv:astro-ph/9708213)

Worseck G., et al., 2014, MNRAS, 445, 1745

Worseck G., Davies F. B., Hennawi J. F., Prochaska J. X., 2019, ApJ, 875, 111

Wu X., McQuinn M., Eisenstein D., 2021, J. Cosmology Astropart. Phys., 2021, 042

Wyithe J. S. B., Bolton J. S., 2011, MNRAS, 412, 1926

Yang J., et al., 2020, ApJ, 904, 26

Zhu Y., et al., 2021, ApJ, 923, 223

Zhu Y., et al., 2022, ApJ, 932, 76

ORIGINAL UNEDITED MANUSCRIPT