

# Studying the Lyman $\alpha$ optical depth fluctuations at $z \sim 5.5$ using fast semi-numerical methods

T. Roy Choudhury<sup>1</sup>,<sup>\*</sup> Aseem Paranjape<sup>2</sup> and Sarah E. I. Bosman<sup>3</sup>

<sup>1</sup>National Centre for Radio Astrophysics, TIFR, Post Bag 3, Ganeshkhind, Pune 411007, India

<sup>2</sup>Inter-University Center for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411007, India

<sup>3</sup>Department of Physics and Astronomy, University College London, London WC1E 6BT, UK

Accepted 2020 December 16. Received 2020 November 27; in original form 2020 March 20

## ABSTRACT

We present a computationally efficient and fast semi-numerical technique for simulating the Lyman  $\alpha$  ( $\text{Ly } \alpha$ ) absorption optical depth in presence of neutral hydrogen ‘islands’ left over from reionization at redshifts  $5 \lesssim z \lesssim 6$ . The main inputs to the analysis are (i) a semi-numerical photon-conserving model of ionized regions during reionization (named `script`) along with a prescription for simulating the shadowing by neutral islands and (ii) the fluctuating Gunn–Peterson approximation to model the  $\text{Ly } \alpha$  absorption. Our model is then used for simulating the large-scale fluctuations in the effective optical depth as observed along sightlines towards high- $z$  quasars. Our model is fully described by five parameters. By setting two of them to default values and varying the other three, we obtain the constraints on reionization history at  $5 \lesssim z \lesssim 6$  as allowed by the data. We confirm that reionization is *not* complete before  $z \sim 5.6$  at  $\gtrsim 2\sigma$  confidence, with the exact confidence limits depending on how the non-detections of the flux in the data are treated. We also confirm that the completion of reionization can be as late as  $z \sim 5.2$ . With further improvements in the model and with more sightlines at  $z \sim 6$ , we can take advantage of the computational efficiency of our analysis to obtain more stringent constraints on the ionization fraction at the tail end of reionization.

**Key words:** galaxies: high-redshift – intergalactic medium – quasars: absorption lines – dark ages, reionization, first stars.

## 1 INTRODUCTION

The detection of quasars at high redshifts  $z \sim 6$  enabled a novel way of studying the end stages of reionization of neutral hydrogen (HI) by the early star formation (Fan et al. 2000, 2001, 2002, 2003, 2004; Songaila 2004; Fan et al. 2006a,b). The Lyman- $\alpha$  ( $\text{Ly } \alpha$ ) absorption spectra of these quasars are expected to contain information on the distribution of HI in the intergalactic medium (IGM) along the lines of sight. These observations of the  $\text{Ly } \alpha$  optical depth, combined with high-quality numerical simulations, allowed one to estimate the HI photoionization rate  $\Gamma_{\text{HI}}$  and thus the number of ionizing photons available in the IGM (Fan et al. 2006b; Bolton & Haehnelt 2007; Calverley et al. 2011; Wyithe & Bolton 2011). More detailed studies of these spectra based on, e.g. the damping wings and near zones (Wyithe & Loeb 2004; Bolton & Haehnelt 2007; Maselli et al. 2007; Bolton et al. 2011; Wyithe & Bolton 2011; Eilers et al. 2017; Greig et al. 2017; Eilers, Hennawi & Davies 2018b; Davies, Hennawi & Eilers 2020; Āurovčíková et al. 2020), evolution of the IGM temperature (Raskutti et al. 2012; Boera et al. 2019), fraction of ‘dark’ pixels in the spectra (McGreer, Mesinger & Fan 2011; McGreer, Mesinger & D’Odorico 2015), dark gap statistics (Songaila & Cowie 2002; Gallerani, Choudhury & Ferrara 2006; Gallerani et al. 2008), have revealed a wealth of information on reionization.

The constraints obtained on the global HI fraction from these studies were relatively straightforward to interpret and to implement

in semi-analytical models (Wyithe & Loeb 2003; Choudhury & Ferrara 2005; Pritchard, Loeb & Wyithe 2010; Mitra, Choudhury & Ferrara 2011, 2012). When combined with other observations, e.g. the Thomson scattering optical depth of the cosmic microwave background (CMB) photons (Hinshaw et al. 2013; Planck Collaboration V 2019), these models were able to constrain the reionization history to a significant extent. Being analytical or semi-analytical in nature, probing a wide range of parameter space was natural for these models and hence they could be coupled to advanced statistical techniques, e.g. Markov chain Monte Carlo (MCMC; Mitra et al. 2011, 2012; Greig & Mesinger 2015; Greig, Mesinger & Bañados 2019). Overall, the data seemed to be consistent with a picture wherein the reionization was completed by  $z \sim 5.8$ .

More recently, the  $\text{Ly } \alpha$  effective optical depth  $\tau_{\text{eff}}$  of the quasar absorption spectra at  $5.5 < z < 6$ , when averaged over large scales ( $50 h^{-1}$  cMpc), showed significant fluctuations (Becker et al. 2015; Bosman et al. 2018; Eilers, Davies & Hennawi 2018a; Eilers et al. 2019). These fluctuations could not be explained by simple models of uniform  $\Gamma_{\text{HI}}$  and thus led to various extensions to the existing picture of that time. These included, e.g. temperature fluctuations in the IGM (D’Aloisio, McQuinn & Trac 2015), presence of an undetected population of faint quasars (Chardin et al. 2015; Chardin, Puchwein & Haehnelt 2017), fluctuations in the mean-free path  $\lambda_{\text{mfp}}$  of ionizing photons (Davies & Furlanetto 2016), presence of HI islands left over from reionization (Nasir & D’Aloisio 2020), shot noise in the placement of bright sources like quasars (Meiksin 2020). The state of the IGM at these redshifts has been modelled extensively through radiative transfer in the numerical simulations of Kulkarni

\* E-mail: tirth@ncra.tifr.res.in

et al. (2019) and Keating et al. (2019), where the essential features of many of the other models have been incorporated in a self-consistent manner (e.g. the mean-free path and temperature fluctuations, and the presence of left over neutral islands).

In case the  $\tau_{\text{eff}}$  fluctuations are indeed due to the neutral islands, then these fluctuations are directly probing the tail end of reionization. Hence, these observations need to be taken into account while attempting to constrain the reionization history. As mentioned above, Kulkarni et al. (2019) have modelled these neutral patches using high-resolution SPH simulations and a cosmological radiative transfer code. Such simulations are usually computationally expensive and hence are not suited for probing the parameter space. Using semi-analytical or semi-numerical models to constrain reionization has the advantage that one is able to identify all possible histories allowed by the data by varying the free parameters, and subsequently study the state of the IGM along with the properties of the ionizing sources (e.g. cooling, feedback, escape of ionizing photons; Mitra, Ferrara & Choudhury 2013; Mitra, Choudhury & Ferrara 2018). In case one wants to include the  $\tau_{\text{eff}}$  fluctuation data in such statistical analyses, it becomes imperative to devise ways to model the HI islands in a computationally efficient manner.

The main aim of this work is to build a model of reionization and Ly $\alpha$  forest at  $z \sim 5.5$ , which is computationally efficient and hence can be used for parameter space exploration. To achieve this, we use a previously developed semi-numerical method to generate ionized regions, driven by Lyman-continuum photons from galaxies, within relatively low-resolution but large simulation volumes (Choudhury & Paranjape 2018). Once the distribution of the ionized (and neutral) regions is generated, we then model the Ly $\alpha$  optical depth of neutral hydrogen (as would be imprinted on spectra of background point sources such as quasars) using the so-called fluctuating Gunn–Peterson approximation (Croft et al. 1998). The resulting realizations of the quasar absorption spectra are then used for calculating the  $\tau_{\text{eff}}$  along large lines of sight to allow proper comparison with observations. Because of the simplifications employed, we need to introduce a few free parameters in the model, thus decreasing its predictive power as compared to radiative transfer simulations. The free parameters are constrained by comparing the model predictions with the observational data, the analysis being possible due to the computational efficiency of our algorithm. The end result of the analysis is that we obtain the range in reionization histories at  $5 \lesssim z \lesssim 6$  that are statistically allowed by the data.

This paper is organized as follows: We discuss our method of calculating the Ly $\alpha$  optical depth in Section 2. In Section 3, we present the main results of our analysis before summarizing and discussing the future outlook in Section 4. The Appendices are devoted to exploring the model parameters beyond their default values and testing the convergence of our results with respect to the resolution. The cosmological parameters used in this work are  $\Omega_m = 0.308$ ,  $\Omega_\Lambda = 1 - \Omega_m$ ,  $\Omega_b = 0.0482$ ,  $h = 0.678$ ,  $n_s = 0.961$ , and  $\sigma_8 = 0.829$  (Planck Collaboration XVI 2014).

## 2 METHOD

### 2.1 Generation of ionization maps using SCRIPT

The ionization maps needed for this work are generated using the semi-numerical method introduced in Choudhury & Paranjape (2018). The method consists of two steps. In the first, we use a collisionless  $N$ -body simulation to generate the large-scale smoothed density fields, which are then used for generating the large-scale distribution of the collapsed haloes through a subgrid prescription. In

the second step, the density and the halo fields are used as input to an explicitly photon-conserving semi-numerical formalism to generate the distribution of ionized regions.

For the  $N$ -body simulation, we use the publicly available code GADGET-2<sup>1</sup> (Springel 2005) and simulate a box of length  $256 h^{-1}$  cMpc with  $512^3$  particles. The initial conditions for the simulation are generated using the N-GenIC code.<sup>2</sup> At redshifts of interest, the simulation outputs in the form of the particle positions are smoothed using a Cloud-in-Cell (CIC) algorithm to generate the matter overdensity field  $\Delta_i = \rho_i / \bar{\rho}$  in a uniform grid with cells labelled by  $i$ .

Since the particle resolution of our box is not sufficient to identify the collapsed haloes of interest, we employ a subgrid scheme to compute the large-scale halo distribution from the density field. Given the overdensity field, we use the conditional mass function from ellipsoidal collapse (Sheth & Tormen 2002), with parameters calibrated to match simulation results, to generate the fraction of mass  $f_{\text{coll},i}$  in collapsed haloes above mass  $M_{\text{min}}$  inside every grid cell. Note that this approximate way of computing the collapsed mass works only for relatively larger grid volumes, hence we do not use grids finer than  $2 h^{-1}$  cMpc. Our method not only produces the halo mass function consistent with  $N$ -body simulations (Jenkins et al. 2001), but also the large-scale clustering of haloes.

The generation of ionization maps requires computing two numbers in every cell in the box. The first is the number of hydrogen atoms, which is assumed to follow the dark matter at scales of our interest:

$$N_{\text{H},i} = \bar{n}_{\text{H}} V_{\text{cell}} \Delta_i, \quad (1)$$

where  $\bar{n}_{\text{H}}$  is the mean comoving hydrogen number density and  $V_{\text{cell}}$  is the comoving volume of the grid cells. Secondly, we need the cumulative number of ionizing photons produced, which can be assumed to be proportional to the mass within collapsed haloes above a mass  $M_{\text{min}}$  and is given by

$$N_{\text{ion},i} = \zeta N_{\text{H},i} f_{\text{coll},i} = \zeta \bar{n}_{\text{H}} V_{\text{cell}} \Delta_i f_{\text{coll},i}, \quad (2)$$

where  $\zeta$  is the ionizing efficiency.

The ionization maps are generated using the photon-conserving semi-numerical scheme introduced in Choudhury & Paranjape (2018), which is named `script` (Semi-numerical Code for ReIonization with PhoTon-conservation). The algorithm consists of two steps: in the first step, we generate ionized ‘bubbles’ around individual grid cells (or sources, as the case may be) allowing the cells where multiple bubbles overlap to be ‘overionized’. In the second step, we deal with the overionized cells in the overlapped bubbles by distributing the excess photons to nearby cells. The code is developed primarily for studying the redshifted 21-cm signal from HI during reionization and includes several improvements over existing semi-numerical codes based on the excursion-set formalism, namely (i) it conserves the number of ionizing photons thus fixing a known shortcoming of earlier models (see, e.g. Zahn et al. 2007; Paranjape, Choudhury & Padmanabhan 2016), and (ii) consequently ensures the numerical convergence of large-scale properties of the ionization field with respect to the resolution at which the maps are made. The code has recently been optimized so that it takes 2–4 s to complete on a single processor for a  $128^3$  grid and  $\sim 0.01$  s for a  $32^3$  grid.

<sup>1</sup><https://wwwmpa.mpa-garching.mpg.de/gadget/>.

<sup>2</sup><https://wwwmpa.mpa-garching.mpg.de/gadget/right.html#ICcode>.

## 2.2 Calculation of the photoionization rate within ionized regions

The output of the semi-numerical method described above provides the neutral hydrogen fraction  $x_{\text{HI}}$  for each cell, which can, in principle, be zero in regions that are completely ionized. In reality, however, the recombinations will ensure that there exist some residual neutral atoms even in these cells. The fraction of such neutral atoms would be exceedingly small (unless they are in high-density self-shielded regions) and hence would hardly affect the large-scale properties of the 21-cm signal (which was the original motivation for developing SCRIPT). On the other hand, for studying the quasar spectra, this residual neutral hydrogen in the low-density IGM would cause the Ly  $\alpha$  absorption observed in quasar spectra and hence needs to be modelled properly.

The main ingredient in modelling the residual neutral fraction is the photoionization rate  $\Gamma_{\text{HI}}$ . To do so, let us start with the flux incident on the  $i$ th cell,

$$J_i(\nu) = \sum_{j \neq i} \frac{L_j(\nu)}{(4\pi)^2 (a x_{ij})^2} e^{-\tau_{ij}(\nu)}, \quad (3)$$

where  $L_j(\nu)$  is the luminosity of the  $j$ th cell,  $x_{ij}$  is the *comoving* distance between the  $i$ th and the  $j$ th cells and  $\tau_{ij}$  is the optical depth between the two cells. The summation extends over all cells other than the cell under consideration. The luminosity  $L_i(\nu)$  is related to the quantity  $N_{\text{ion},i}$  used for generating the ionization maps in Section 2.1. However, since  $N_{\text{ion},i}$  is the cumulative number of photons produced, we need to introduce a characteristic time-scale  $t_*$  to relate it to the instantaneous luminosity. We can then write

$$L_i(\nu) = \dot{N}_i(\nu) h\nu \equiv \frac{N_i(\nu)}{t_*} h\nu, \quad (4)$$

where  $N_i(\nu)$  is the cumulative number of photons produced per unit frequency range and is related to  $N_{\text{ion},i}$  by

$$N_{\text{ion},i} = \int_{\nu_{\text{HI}}}^{\infty} d\nu N_i(\nu), \quad (5)$$

where  $\nu_{\text{HI}}$  is the Lyman-limit frequency.

The optical depth between the two cells  $i$  and  $j$  can be calculated by integrating along the sightline joining the cell (Davies & Furlanetto 2016),

$$\tau_{ij}(\nu) = \int_{x_i}^{x_j} \frac{dx}{\lambda_{\text{mfp}}(\nu, x)}. \quad (6)$$

As shown in Davies & Furlanetto (2016), computing the mean-free path self-consistently requires iterative solutions and can be computationally expensive (Hutter 2018). To start with, for ionized regions, we can make the simplifying assumption that the mean-free path takes just one value for the whole box and is determined by the typical distance between the self-shielded regions. Let us denote this mean-free path by  $\lambda_{\text{ss}}$ . This approximation is believed to be adequate in the post-reionization universe. This approximation is probably acceptable also for cells within an ionized region as long as its size is significantly larger than  $\lambda_{\text{ss}}$ . However, the assumption breaks down at early stages of reionization where the mean-free path is pre-dominantly determined by the bubble size.

With this assumption, we can write a simplified expression for the optical depth,

$$\tau_{ij} = \frac{x_{ij}}{\lambda_{\text{ss}}(\nu)}. \quad (7)$$

The calculation then follows the usual approach outlined in Davies & Furlanetto (2016) and Hutter (2018), and the photoionization rate

can be shown to be given by

$$\Gamma_{\text{HI},i} = \frac{1}{a^2} \frac{\alpha}{\alpha + \beta} \frac{\sigma_{\text{HI}}(\nu_{\text{HI}})}{t_*} \sum_{j \neq i} N_{\text{ion},j} \frac{e^{-x_{ij}/\lambda_{\text{ss}}}}{4\pi x_{ij}^2}, \quad (8)$$

where  $\alpha$  is the spectral index of the ionizing sources,  $\beta$  is the spectral index of the hydrogen ionization cross-section, and  $\sigma_{\text{HI}}$  is the cross-section at  $\nu = \nu_{\text{HI}}$ . While deriving the above equation, we have assumed that the mean-free path is independent of  $\nu$  (which is reasonable because the hydrogen ionization cross-section is a steeply declining function of  $\nu$ ).

Note that the summation on the right-hand side depends only on  $x_{ij}$ , hence, it can be expressed as a sum over contributions from spherical shells around the  $i$ th cell. Let us write it as

$$\Gamma_{\text{HI},i} = \frac{1}{a^2} \frac{\alpha}{\alpha + \beta} \frac{\sigma_{\text{HI}}(\nu_{\text{HI}})}{t_*} \sum_J N_{\text{ion}}^{(i)}(r_J) \frac{e^{-r_J/\lambda_{\text{ss}}}}{4\pi r_J^2}, \quad (9)$$

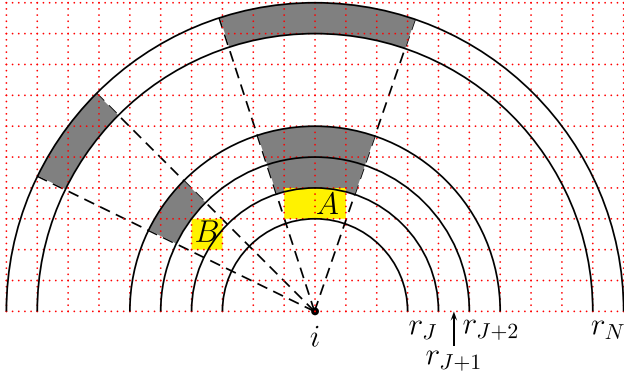
where the summation index  $J$  is over all spherical shells. The  $J$ th shell has a comoving radius  $r_J$ . The quantity  $N_{\text{ion}}^{(i)}(r_J)$  is the number of photons contributed by all the grid cells within the  $J$ th spherical shell and the superscript  $(i)$  signifies that the spheres are constructed around the  $i$ th grid cell. The summation over shells in the above equation can be computed using spherical filters and allows the calculation to be extremely efficient computationally.

A major shortcoming of the above formalism is that it does not account for ‘shadows’ arising because of the neutral islands. We expect that the cells close to the boundaries of the ionized regions will not receive contribution from sources in the direction of the neutral regions; hence, the photoionization rate in these cells should be less than what is given by equation (9). Nasir & D’Aloisio (2020) implement this effect in a direct way by removing contributions from sources whose lines of sight pass through neutral islands. However, such a method based on ray-tracing turns out to be computationally time-consuming in our simulations. Ideally, we would prefer to retain the summation over spherical shells as this allows the method to be computationally efficient.

To this end, for every pair of grid cell  $i$  and spherical shell  $J$ , we introduce a correction factor for the photon number  $N_{\text{ion}}^{(i)}(r_J)$  to account for the shadows arising from the neutral islands within radius  $r_J$  around the  $i$ th cell. Our method for computing the correction factor is best explained by the illustration in Fig. 1. To obtain the photoionization rate  $\Gamma_{\text{HI},i}$ , we construct spherical shells of different radii around the  $i$ th cell and sum over the contribution from each of them as given in equation (9). We show a few such shells in the figure. For convenience of implementing the algorithm, we assume all spherical shells to have the same thickness  $\Delta r$ .

Let us imagine that there are two neutral regions  $A$  and  $B$  in the volume of interest, the corresponding neutral grid cells being marked in yellow. Although the shapes of the regions look somewhat contrived in the chosen example, our final expression would be written in terms of the fraction of volume occupied by the neutral regions and hence would work even for arbitrary shaped regions.

Clearly, the neutral cells would *not* allow photons to reach  $i$  from regions on the other side. For example, the neutral region  $A$ , which is situated in shell  $J$ , would block photons from the outer shells  $J+1$ ,  $J+2$ ,  $\dots$ ,  $N$ ,  $\dots$ , the affected regions are marked in grey in the figure. Let the number of neutral cells in shell  $J$  be denoted as  $N_{\text{cell, HI}}(r_J)$ , which in our example would be the number of cells occupied by  $A$ . The cells in shell  $J+1$  that would be affected by these neutral cells would be the ones within the solid angle denoted by the dashed lines joining  $i$  and the edges of region  $A$ . Since all the shells have the same thickness  $\Delta r$ , the effective number of cells in the  $(J+1)$ th shell that



**Figure 1.** Illustration of our method of implementing ‘shadows’ arising from neutral islands. The simulation box is divided into (cubic) grid cells that are assigned a neutral fraction using semi-numerical method `script`. In this example, the grids in yellow are neutral and all others are assumed to be ionized. The photoionization rate at the grid cell under consideration  $i$  is computed by summing over contributions from ionizing sources within spherical shells  $J$  having radii  $r_J$ . The neutral regions will block radiation from sources within the regions marked in grey (showed only for a few representative shells).

are affected by the neutral cells in  $J$  is approximately given by

$$N_{\text{cell,shadow}}^{(i)}(r_{J+1}) = N_{\text{cell,H I}}^{(i)}(r_J) \frac{r_{J+1}^2}{r_J^2}, \quad (10)$$

as is obvious from the figure. The above relation is approximate due to edge effects of fitting cubes inside spherical shells. We can now work out, under the same approximation, the effective number of cells affected by the neutral cells in shell  $J+2$ ,

$$N_{\text{cell,shadow}}^{(i)}(r_{J+2}) = N_{\text{cell,shadow}}^{(i)}(r_{J+1}) \frac{r_{J+2}^2}{r_{J+1}^2} + N_{\text{cell,H I}}^{(i)}(r_{J+1}) \frac{r_{J+2}^2}{r_{J+1}^2}, \quad (11)$$

where the first term on the right-hand side is arising from region A and the second term from B.

It is straightforward to generalize the relation and write down the equivalent formula for any given shell at  $r_N$ :

$$N_{\text{cell,shadow}}^{(i)}(r_N) = \left[ N_{\text{cell,H I}}^{(i)}(r_{N-1}) + N_{\text{cell,shadow}}^{(i)}(r_{N-1}) \right] \frac{r_N^2}{r_{N-1}^2}. \quad (12)$$

Since the total number of cells in a spherical shell  $N_{\text{cell,tot}}(r_N) \propto r_N^2 \Delta r$ , and since  $\Delta r$  is chosen to be the same for all shells, the fraction of cells in the  $N$ th shell that are affected by the neutral islands is given by

$$Q_{\text{shadow}}^{(i)}(r_N) = \min \left[ Q_{\text{H I}}^{(i)}(r_{N-1}) + Q_{\text{shadow}}^{(i)}(r_{N-1}), 1 \right], \quad (13)$$

where  $Q_{\text{H I}}^{(i)}(r_N)$  is the fraction of neutral cells in the  $N$ th shell. As we consider contributions from shells farther away from the  $i$ th cell, the fraction  $Q_{\text{shadow}}^{(i)}(r_N)$  keeps on increasing (or remains the same) depending on how the neutral islands are distributed. This allows us to account for the shadows by simply decreasing the photon contribution from the  $N$ th shell by a factor  $N_{\text{ion}}^{(i)}(r_N) \rightarrow \left[ 1 - Q_{\text{shadow}}^{(i)}(r_N) \right] N_{\text{ion}}^{(i)}(r_N)$ . Since the equation is written solely in terms of neutral volume fractions  $Q_{\text{H I}}^{(i)}(r_N)$ , it automatically allows for partially ionized cells (which may arise because the boundaries of the neutral regions can be of arbitrary shapes).

With the above modification, our new relation for the photoionization rate is given by

$$\Gamma_{\text{H I},i} = \frac{1}{a^2} \frac{\alpha}{\alpha + \beta} \frac{\sigma_{\text{H I}}(\nu_{\text{H I}})}{t_*} \times \sum_J \left[ 1 - Q_{\text{shadow}}^{(i)}(r_J) \right] N_{\text{ion}}^{(i)}(r_J) \frac{e^{-r_J/\lambda_{\text{ss}}}}{4\pi r_J^2}. \quad (14)$$

Since the algorithm operates at the level of spherical shells (instead of individual cells), it effectively penalizes all sources that happen to lie in a shell. This is clearly only an approximation, since it ignores all direction dependence in the placement of neutral islands and sources around the cell in question. For example, the method tends to double-count the contribution from the neutral regions that lie in the shadow of another neutral region. Nevertheless, we expect the spherical averaging to level some of these discrepancies. Our method still gives lower fluxes at cells that are close to the neutral islands and hence are likely to be affected by shadowing. It also introduces additional fluctuations in the photoionization field within the ionized regions. The main advantage of the method is that it is computationally much faster than any method that depends on lines of sight.

A final point to note is that the summation in equation (14) accounts only for the cells other than the cell under consideration. For the local contribution, we assume that the sources within  $r_0$  (the radius of the sphere corresponding to the grid volume) are distributed uniformly; hence, the photoionization rate is given by (Davies & Furlanetto 2016)

$$\Gamma_{\text{H I},i}^{\text{local}} = \frac{1}{a^2} \frac{\alpha}{\alpha + \beta} \frac{\sigma_{\text{H I}}(\nu_{\text{H I}})}{t_*} N_{\text{ion},i} \left( 1 - e^{-r_0/\lambda_{\text{ss}}} \right) \frac{3\lambda_{\text{ss}}}{4\pi r_0^3}. \quad (15)$$

The final photoionization rate is calculated by adding this local contribution to the one computed using equation (14).

We can write the expression for  $\Gamma_{\text{H I},i}$  in a more useful form as

$$\frac{\Gamma_{\text{H I},i}}{10^{-12} \text{ s}^{-1}} = A_\Gamma \left( \frac{1+z}{6.5} \right)^2 \frac{S_i}{2.37 \times 10^{18} \text{ cm}^{-2}}, \quad (16)$$

where

$$A_\Gamma \equiv \left( \frac{\alpha}{3} \right) \left( \frac{6}{\alpha + \beta} \right) \left( \frac{10^7 \text{ yr}}{t_*} \right) \quad (17)$$

is a normalization factor and

$$S_i \equiv \sum_J \left[ 1 - Q_{\text{shadow}}^{(i)}(r_J) \right] N_{\text{ion}}^{(i)}(r_J) \frac{e^{-r_J/\lambda_{\text{ss}}}}{4\pi r_J^2} + N_{\text{ion},i} \left( 1 - e^{-r_0/\lambda_{\text{ss}}} \right) \frac{3\lambda_{\text{ss}}}{4\pi r_0^3} \quad (18)$$

is the cumulative photon flux at cell  $i$ . Our method of calculating  $\Gamma_{\text{H I},i}$  thus depends on two parameters, namely  $\lambda_{\text{ss}}$  and  $A_\Gamma$  (which, in turn, depends on several other parameters  $\alpha$ ,  $\beta$ , and  $t_*$ ). We shall return to discuss these parameters in Section 2.4.

### 2.3 The Ly $\alpha$ optical depth

Having calculated the distribution of H I in the IGM, as caused by the Lyman-continuum photons from galaxies, we now compute the Ly  $\alpha$  optical depth  $\tau_\alpha$  arising from the H I field. This  $\tau_\alpha$  field would get imprinted on the spectra of distant quasars that act as background sources.

For a cell that is identified as completely ionized by our semi-numerical model of reionization, the residual neutral hydrogen

fraction is obtained assuming photoionization equilibrium,

$$x_{\text{H},i} \Gamma_{\text{H},i} = \frac{\chi_{\text{He}}}{a^3} \alpha_B(T_i) n_{\text{H},i} (1 - x_{\text{H},i})^2, \quad (19)$$

where  $\chi_{\text{He}} \approx 1.08$  accounts for the excess electron produced by singly ionized helium,  $\alpha_B(T)$  is the case-B recombination rate, and the factor  $a^3$  accounts for the fact that the number densities used are in comoving units. The solution to the above quadratic equation is straightforward provided we assume a relation between the temperature  $T_i$  of the cell and the density, which is usually taken to be a power-law  $T_i = T_0 \Delta_i^{\gamma-1}$ , where  $T_0$  is the temperature of the cell at the mean density and  $\gamma$  is the slope. In numerical simulations where reionization is assumed to be instantaneous and uniform, the value of  $\gamma$  is found to be around unity right after the reionization is completed and subsequently approaches a value  $\sim 1.5$  (Puchwein et al. 2015; Gaikwad et al. 2018). In reality, the temperature distribution could be more complicated given that different points in the IGM get reionized at different times and hence the  $T_i-\Delta_i$  relation is not necessarily one-to-one. Note that the value of  $\gamma$  used in the above applies to pixels that are ionized. If we make a simplifying assumption that a substantial fraction of them got ionized sufficiently early, we can take a one-to-one  $T_i-\Delta_i$  relation in such regions. One should also realize that the relation is defined for the  $\Delta_i$  field smoothed over some grid size, hence we expect  $\gamma$  to depend on the resolution used for carrying out the analysis.

For most cases of interest, applying the above photoionization equilibrium equation to the completely ionized cells yields neutral fractions much smaller than unity, which turn out to be

$$x_{\text{H},i} \approx \frac{\chi_{\text{He}}}{a^3} \frac{\alpha_B(T_i) n_{\text{H},i}}{\Gamma_{\text{H},i}}. \quad (20)$$

The above relation is applied only to those cells that are identified as completely ionized by the semi-numerical method, whereas for cells that are partially or completely neutral, we assign the neutral fraction as obtained from the semi-numerical calculation itself.

Under the fluctuating Gunn–Peterson approximation, the Ly  $\alpha$  optical depth is given by

$$\tau_{\alpha,i} = \kappa_{\text{res}} \frac{\pi e^2}{m_e c} f_\alpha \lambda_\alpha H^{-1}(z) x_{\text{H},i} \frac{n_{\text{H},i}}{a^3}, \quad (21)$$

where  $\kappa_{\text{res}}$  is a normalization factor to account for the small-scale fluctuations in the density and velocity fields that are not resolved in our coarse-resolution simulations (Dixon & Furlanetto 2009; Davies & Furlanetto 2016),  $f_\alpha$  is the Ly  $\alpha$  oscillator strength, and all other symbols have their usual meanings. A straightforward calculation shows that the optical depth can be written as

$$\begin{aligned} \tau_{\alpha,i} = & 5.01 \kappa_{\text{res}} \left( \frac{\chi_{\text{He}}}{1.08} \right) \left( \frac{1-Y}{0.76} \right)^2 \left( \frac{\Omega_b h^{3/2}}{0.0269} \right)^2 \\ & \times \left( \frac{1+z}{6.5} \right)^6 \left( \frac{9.23}{H(z)/H_0} \right) \\ & \times \left( \frac{T_0}{10^4 \text{ K}} \right)^{-0.7} \left( \frac{10^{-12} \text{ s}^{-1}}{\Gamma_{\text{H},i}} \right) \Delta_i^{2.7-0.7\gamma}, \end{aligned} \quad (22)$$

where we have assumed  $\alpha_B(T) \propto T^{-0.7}$ . A more useful form can be obtained by substituting  $\Gamma_{\text{H},i}$  from equation (16):

$$\begin{aligned} \tau_{\alpha,i} = & 5.01 A_\tau \left( \frac{\chi_{\text{He}}}{1.08} \right) \left( \frac{1-Y}{0.76} \right)^2 \left( \frac{\Omega_b h^{3/2}}{0.0269} \right)^2 \\ & \times \left( \frac{1+z}{6.5} \right)^4 \left( \frac{9.23}{H(z)/H_0} \right) \\ & \times \frac{2.37 \times 10^{18} \text{ cm}^{-2}}{S_i} \Delta_i^{2.7-0.7\gamma}, \end{aligned} \quad (23)$$

where we define a new normalization factor,

$$\begin{aligned} A_\tau \equiv & \kappa_{\text{res}} \left( \frac{T_0}{10^4 \text{ K}} \right)^{-0.7} A_\Gamma^{-1} \\ = & \kappa_{\text{res}} \left( \frac{T_0}{10^4 \text{ K}} \right)^{-0.7} \left( \frac{3}{\alpha} \right) \left( \frac{\alpha + \beta}{6} \right) \left( \frac{t_*}{10^7 \text{ yr}} \right). \end{aligned} \quad (24)$$

The transmitted flux for the cell is given by  $e^{-\tau_{\alpha,i}}$ ; hence, the effective optical depth averaged over  $N$  pixels is given by

$$\tau_{\text{eff}} = -\ln \left[ \frac{1}{N} \sum_i e^{-\tau_{\alpha,i}} \right]. \quad (25)$$

This is the main observable in our work that will be compared with the observations.

It is obvious from equation (23) that it is *not* possible to constrain the physical quantities  $\alpha$ ,  $\beta$ ,  $t_*$ , and  $T_0$  individually from observations of the optical depth, we can only hope to constrain the combination  $A_\tau$ . Additionally to be noted is that the value of  $A_\Gamma$  is completely degenerate with  $T_0$  and  $\kappa_{\text{res}}$ ; therefore, we cannot measure the amplitude of  $\Gamma_{\text{H},i}$  from our low-resolution simulations. The fluctuations in the rate, however, should be correctly captured in our model.

## 2.4 Model parameters

Our model has *five* parameters (which, in general, can be functions of  $z$ ):

(i) The first parameter is the ionizing efficiency  $\zeta$ , which is used for generating the ionization maps. We treat this as a free parameter.

(ii) The other parameter that is required for generating the ionization maps is the minimum mass  $M_{\text{min}}$  of haloes that are capable of producing ionizing photons. In this work, we choose  $M_{\text{min}} = 10^9 M_\odot$ , which is appropriate for late stages of reionization. At these redshifts, most regions are photoheated and hence the star-formation threshold is set by the radiative feedback. We have also varied  $M_{\text{min}}$  in the range  $10^8-10^{10} M_\odot$  and found that our constraints on the reionization history are insensitive to the value of  $M_{\text{min}}$ ; see Appendix A1.

(iii) The next parameter is the mean-free path  $\lambda_{\text{ss}}$  of ionizing photons as determined by the distance between the self-shielded regions. We choose its value as extrapolated from  $z \lesssim 5$  observations (Worseck et al. 2014) having the empirical power-law fitting form

$$\lambda_{\text{ss}}(z) = 175 \text{ cMpc} \left( \frac{1+z}{5.0} \right)^{-4.4}. \quad (26)$$

We study the effect of  $\lambda_{\text{ss}}$  values different from the above default choice in Appendix A2.

(iv) The fourth parameter in our list is the slope  $\gamma$  of the temperature–density relation in the IGM. We leave it as a free parameter.

(v) The final parameter is the normalization factor  $A_\tau$  used for computing the Ly  $\alpha$  optical depth; see equation (24). This too is kept

free. Since we would want to vary this parameter over orders of magnitude, we prefer  $\log A_\tau$  as the free parameter while carrying out the statistical analysis.

To summarize, we treat  $\zeta$ ,  $\gamma$ , and  $\log A_\tau$  as free parameters and constrain them by comparing with Ly $\alpha$  optical depth data. We fix the values of  $M_{\min}$  and  $\lambda_{\text{ss}}$  to default values (and check the effect of varying them in Appendix A). We also emphasize here that the constraints on  $\gamma$  and  $A_\tau$  are expected to depend on the resolution used for the analysis; hence, it is important to check whether our conclusions regarding the reionization history remain unchanged with respect to the resolution. We investigate this aspect in Appendix B.

## 2.5 Observational data

The main observational data used in this work are from Bosman et al. (2018), who have measured the Ly $\alpha$  effective optical depth  $\tau_{\text{eff}}$  averaged over  $50 h^{-1}$  cMpc chunks in the redshift range  $5 \lesssim z \lesssim 6$ . We use their ‘GOLD’ sample that consists of spectra with various quality-cuts. Their results are presented as ‘optimistic’ and ‘pessimistic’ limits on the cumulative distribution function (CDF)  $P(< \tau_{\text{eff}})$  depending on how they treat the non-detections of the transmitted flux. In the optimistic case, the lower limits on  $\tau_{\text{eff}}$  are treated as measurements just below the detection sensitivity, while in the pessimistic case, these are assumed to have  $\tau_{\text{eff}} \rightarrow \infty$ .

## 3 RESULTS

### 3.1 Constraints on the reionization history

We now compare the predictions of our model with the observational data on  $\tau_{\text{eff}}$  fluctuations to constrain the reionization history. We treat each redshift bin as independent and constrain the three free parameters of our model using a Bayesian likelihood method. The main steps followed in the analysis can be summarized as follows:

(i) We first convert the observational data on the CDF  $P(< \tau_{\text{eff}})$  into the differential PDF  $p(\tau_{\text{eff}}) \equiv dP(< \tau_{\text{eff}})/d\tau_{\text{eff}}$ . Using the differential distribution for our analysis ensures that each measurement of  $\tau_{\text{eff}}$  from the observed spectra contributes to only one bin thus reducing correlations across different bins.

(ii) The pessimistic and optimistic limits of Bosman et al. (2018) data differ in the way the non-detections of the transmitted flux are treated. In principle, one can take the forward modelling approach and use the noise characteristics of the telescopes to contaminate the simulated spectra appropriately. This will allow a fair comparison with the data without making any assumptions about the value of  $\tau_{\text{eff}}$  in case of non-detections. However, the features most affected while adding the noise are narrow transmission spikes ( $\lesssim$  cMpc across; see Chardin et al. 2018; Gaikwad et al. 2020) that are not resolved by the low-resolution pixels of our model. Hence, we take a different approach where we treat the optimistic and pessimistic bounds as two independent data sets and compare them with the simulated spectra without adding any noise. Since the bounds provide reliable extrema for the recovery of the underlying distribution, the two sets of constraints thus obtained on the reionization history should bracket the full range of allowed histories.

(iii) The likelihood analysis requires computing the  $\chi^2$  defined as

$$\chi^2 = \sum_{\alpha, \beta=1}^{N_{\text{bins}}} \Delta p(\tau_{\text{eff}, \alpha}) [C^{-1}]_{\alpha\beta} \Delta p(\tau_{\text{eff}, \beta}), \quad (27)$$

where  $\tau_{\text{eff}, \alpha}$  is the value corresponding to  $\alpha$ -bin,  $C^{-1}$  is the inverse of the error covariance matrix, and

$$\Delta p(\tau_{\text{eff}, \alpha}) \equiv p_{\text{model}}(\tau_{\text{eff}, \alpha}; \boldsymbol{\theta}) - p(\tau_{\text{eff}, \alpha}). \quad (28)$$

In the above definition,  $p(\tau_{\text{eff}, \alpha})$  is the binned PDF computed from the observational data and  $p_{\text{model}}(\tau_{\text{eff}, \alpha}; \boldsymbol{\theta})$  is the theoretical PDF for the parameter set  $\boldsymbol{\theta} \equiv \{\zeta, \gamma, \log A_\tau\}$ . For a given redshift, we compute  $p_{\text{model}}(\tau_{\text{eff}, \alpha}; \boldsymbol{\theta})$  from our simulations by first drawing as many random lines of sight of length  $50 h^{-1}$  cMpc as there are in the observational data and then computing the mean over 5000 independent realizations of the sightlines.

(iv) We assume that, for a given data set, the errors on the distribution are dominated by the variations across different sightlines. We thus estimate the covariance matrix elements  $C_{\alpha\beta}$  from 5000 independent realizations of the PDF from the simulation, accounting for any possible correlations between different  $\tau_{\text{eff}}$  bins. The number of realizations is chosen to ensure that the  $\chi^2$  is numerically converged. Clearly, the elements  $C_{\alpha\beta}$  depend on the parameter values  $\boldsymbol{\theta}$  and ideally one should compute them for every point  $\boldsymbol{\theta}$  in the parameter space during the Bayesian statistical analysis. This, however, substantially increases the time taken for exploring the parameter space, hence, we compute the covariance matrix only for a fiducial parameter set  $\tilde{\boldsymbol{\theta}}$  (for each redshift and each data set) and use it throughout the analysis. The fiducial parameter values  $\tilde{\boldsymbol{\theta}}$  are obtained by minimizing the  $\chi^2$  in equation (27). We use the BOBYQA-bounded minimization routine of Powell (2009) for this purpose. During the minimization, we compute  $C_{\alpha\beta}$  individually for every point  $\boldsymbol{\theta}$  in the parameter space, which is manageable only because the minimization requires substantially less number of evaluations of the  $\chi^2$  compared to the full Bayesian analysis.

(v) We use the publicly available affine-invariant ensemble sampler for MCMC called emcee (Foreman-Mackey et al. 2013) to obtain the posterior distribution of the parameters. The sampler requires evaluation of a likelihood for every  $\boldsymbol{\theta}$ , which we define as  $\mathcal{L} = e^{-\chi^2/2}$ . We use flat priors on all the parameters in the ranges given below:

(a)  $\zeta$  is assumed to have a flat prior in the range  $[0, \zeta_{\text{max}}]$ , where  $\zeta_{\text{max}}$  is the value corresponding to complete ionization of the IGM.

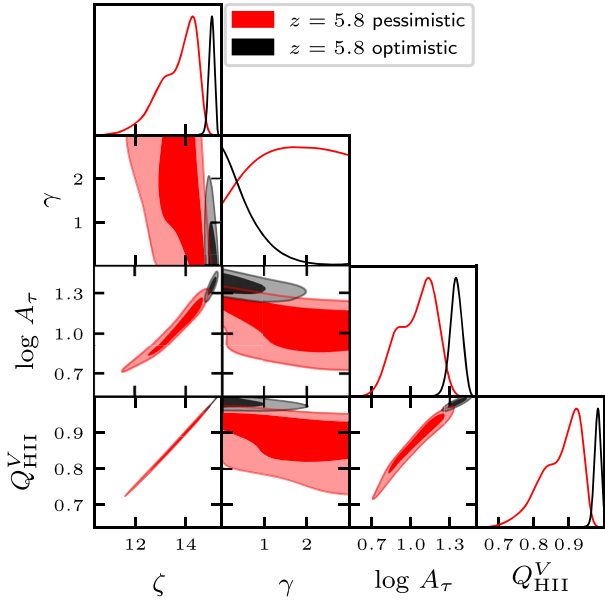
(b)  $\gamma$  is assumed to have a flat prior in the range  $[0, 3]$ . This prior is quite conservative and allows for a wide range in thermal states of the IGM (including ‘inverted’ temperature–density relations).

(c)  $\log A_\tau$  is assumed to have a flat prior in the range  $[-3, 3]$ .

We use 20 walkers and run the chains long enough so that they converge, which is assessed through the autocorrelation analysis of Goodman & Weare (2010).

(vi) Our default runs are carried out at grid resolution  $\Delta x = 8 h^{-1}$  cMpc. Given our box, this leads to  $32^3$  grid cells, thus making the code extremely fast. The sensitivity of our results to the grid size is investigated in Appendix B.

The result of our analysis, for one representative redshift  $z = 5.8$ , is shown in Fig. 2 where we plot the marginalized posterior distributions for the three free parameters along with a derived parameter  $Q_{\text{H II}}^V$ . The distributions are obtained by comparing the model with the optimistic (black) and pessimistic (red) data sets. We can see that the allowed values of  $\zeta$ , and hence of  $Q_{\text{H II}}^V$ , are higher for the optimistic data set than the pessimistic one. This is along the expected lines as the inferred values of  $\tau_{\text{eff}}$  for sightlines with no detections are higher

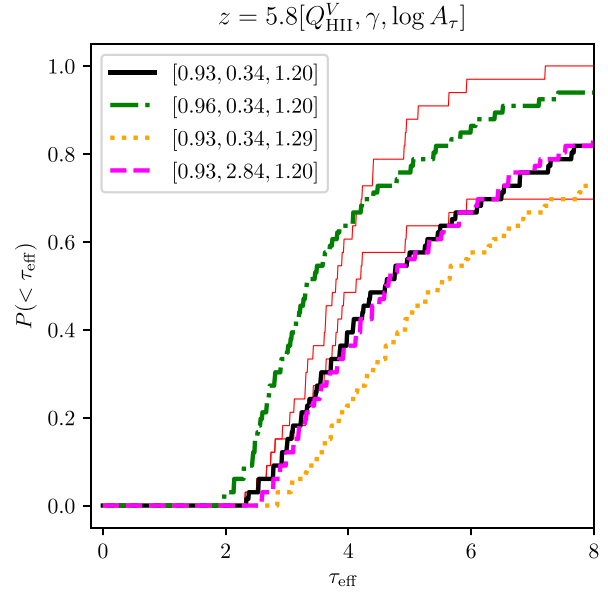


**Figure 2.** The marginalized posterior distributions for the free parameters  $\zeta$ ,  $\gamma$ , and  $\log A_\tau$  and one derived parameter  $Q_{\text{HII}}^V$  obtained from the MCMC analysis. The results are shown at  $z = 5.8$  for the pessimistic (red) and optimistic (black) data sets. The contours enclose 68 and 95 per cent of the points.

in the pessimistic case; hence, matching this data set requires more neutral patches in the IGM. The statistical errors on the parameters are typically larger in the pessimistic case. This is because the best-fitting values of  $Q_{\text{HII}}^V$  for the pessimistic case are usually smaller and the presence of more neutral islands introduces larger variations across lines of sight. This naturally leads to larger errors on the  $\tau_{\text{eff}}$  distribution (i.e. larger values of the covariance matrix elements  $C_{\alpha\beta}$ ).

To further understand the constraints on different parameters, we show in Fig. 3 the dependence of the observable  $P(< \tau_{\text{eff}})$  on the three parameters  $Q_{\text{HII}}^V$ ,  $\log A_\tau$ , and  $\gamma$ .<sup>3</sup> It is clear from the figure that increasing  $Q_{\text{HII}}^V$  leads to less number of high-opacity sightlines (compare the black solid curve with the green dot-dashed), thus decreasing the scatter in the distribution. This is expected since it is the presence of neutral islands that leads to the high-opacity regions. The effect of increasing the normalization  $A_\tau$  is to increase the value of  $\tau_{\text{eff}}$ , hence the whole distribution shifts to the right (compare the black solid curve with the orange dotted). Interestingly, the effect of  $\gamma$  on the  $\tau_{\text{eff}}$  distribution is quite minimal. One can see that in spite of increasing  $\gamma$  by  $\sim 2.5$ , the change in  $P(< \tau_{\text{eff}})$  is negligible (compare the black solid curve with the magenta dashed). A closer look at the curves reveals that a higher  $\gamma$  tends to increase the value of  $\tau_{\text{eff}}$  along lines of sight of relatively lower opacities and vice versa, and hence the distribution becomes slightly narrower. If we ignore the neutral islands for the moment, this can be understood as follows: High-density regions tend to remain more neutral because they recombine more efficiently. Increasing  $\gamma$ , on the other hand, has the opposite effect where it leads to a higher temperature in the high-density regions, thus making them more ionized. So the

<sup>3</sup>The dependence of  $P(< \tau_{\text{eff}})$  on  $\zeta$  is very similar to that on  $Q_{\text{HII}}^V$  as these two parameters are perfectly correlated; hence, we do not show the effect of  $\zeta$  separately.



**Figure 3.** Dependence of the cumulative  $\tau_{\text{eff}}$  distribution on model parameters. The two thin red curves show the observational data for the optimistic (left/upper) and pessimistic (right/lower) cases of Bosman et al. (2018). The black solid curve is the ‘default’ model for the purpose of this plot, which is essentially the best fit to the pessimistic data at  $z = 5.8$ . The green dot-dashed, orange dotted and magenta dashed curves show the variation of the  $\tau_{\text{eff}}$  distribution with respect of  $Q_{\text{HII}}^V$ ,  $\log A_\tau$ , and  $\gamma$ , respectively.

main effect of increasing  $\gamma$  is to reduce the scatter in the neutral hydrogen distribution and hence in the optical depth distribution. This is essentially the reason a higher  $\gamma$  leads to a narrower  $\tau_{\text{eff}}$  distribution.

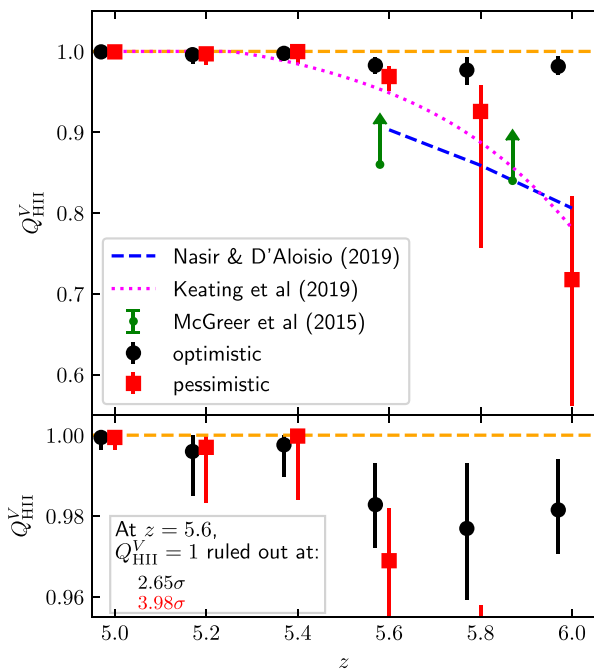
Returning to the posterior distribution in Fig. 2, we find that the constraints on  $\gamma$  are extremely weak, which follows from the fact that  $\tau_{\text{eff}}$  distributions are relatively insensitive to  $\gamma$  in the probed range. In fact, there is no constraint for the pessimistic case within the prior range chosen. For the optimistic case, the statistical errors on the  $\tau_{\text{eff}}$  distribution are smaller and hence the data are able to put some constraints on  $\gamma$ . In this case, we can rule out  $\gamma > 1.6$  (at  $2\sigma$ ). This is because for higher values of  $\gamma$  the distribution becomes too narrow to be allowed by the data. In general, we found that fixing  $\gamma = 1.5$  and carrying out a two-parameter MCMC analysis leave the posteriors of the other parameters unaffected. There is a strong positive correlation between  $\zeta$  and  $\log A_\tau$  and, as a consequence, between  $Q_{\text{HII}}^V$  and  $A_\tau$ . Higher  $\zeta$ -values correspond to more ionized IGM, which lead to less neutral islands and hence lower  $\tau_{\alpha}$ . This can be compensated by increasing  $A_\tau$  appropriately that thus gives rise to the positive correlation. Note that the main goal of our work is to constrain the reionization history that is achieved essentially from the marginalized posterior of the derived parameter  $Q_{\text{HII}}^V$ . We find that the correlations between different parameters and other properties of the posteriors are similar for other redshifts; hence, we do not show the detailed contour plots separately. Instead, we provide the best-fitting values of the parameters  $\gamma$ ,  $\log A_\tau$ , and  $Q_{\text{HII}}^V$  along with their  $2\sigma$  allowed range in Table 1.

The best-fitting values of  $Q_{\text{HII}}^V$  along with the  $2\sigma$  errors obtained from our likelihood analysis for the pessimistic (red squares with error bars) and optimistic (black circles with error bars, shifted slightly along the horizontal axis for clarity) cases are shown in

**Table 1.** Best-fitting values of the model parameters along with the  $2\sigma$  confidence limits (in parentheses) obtained from the likelihood analysis.

Parameters: data type	$\gamma$		$\log A_\tau$		$Q_{\text{HII}}^V$	
	Optimistic	Pessimistic	Optimistic	Pessimistic	Optimistic	Pessimistic
$z = 5.0$	0.80 [0.18, 1.82]	0.80 [0.18, 1.82]	1.49 [1.31, 1.66]	1.49 [1.31, 1.66]	0.999 [ $>0.996$ ]	0.999 [ $>0.996$ ]
$z = 5.2$	0.94 [0.09, 1.68]	0.89 [0.15, 2.13]	1.44 [1.39, 1.60]	1.44 [1.36, 1.47]	0.996 [ $>0.985$ ]	0.997 [ $>0.983$ ]
$z = 5.4$	0.52 [ $<1.67$ ]	0.69 [ $<2.55$ ]	1.42 [1.36, 1.45]	1.43 [1.33, 1.44]	0.998 [ $>0.990$ ]	1.000 [ $>0.984$ ]
$z = 5.6$	0.25 [ $<1.62$ ]	0.45	1.37 [1.30, 1.43]	1.30 [1.24, 1.37]	0.983 [0.972, 0.993]	0.969 [0.952, 0.982]
$z = 5.8$	0.14 [ $<1.61$ ]	0.34	1.35 [1.26, 1.44]	1.20 [0.78, 1.29]	0.974 [0.959, 0.993]	0.926 [0.757, 0.958]
$z = 6.0$	0.08	0.70	1.33 [1.31, 1.38]	0.67 [0.36, 1.00]	0.982 [0.971, 0.994]	0.718 [0.561, 0.820]

Notes. Cases for which the  $2\sigma$  range is not mentioned imply that there were no constraints within the prior range considered.  $\gamma$  and  $\log A_\tau$  are the free parameters of the model, while  $Q_{\text{HII}}^V$  is a derived parameter that is perfectly correlated with the third free parameter  $\zeta$ .



**Figure 4.** Constraints on the ionized volume fraction: The top panel shows the constraints on  $Q_{\text{HII}}^V$  obtained from the MCMC analysis for the optimistic (black circles with error bars, shifted slightly along the redshift axis for clarity) and pessimistic (red squares with error bars) data sets of Bosman et al. (2018). The horizontal orange dashed line corresponds to  $Q_{\text{HII}}^V = 1$ . The blue dashed curve corresponds to the late-reion-long-mfp reionization history of Nasir & D’Aloisio (2020), while the magenta dotted curve is the default reionization model of Keating et al. (2019). The green points with error bars reflect the  $2\sigma$  lower limits on  $Q_{\text{HII}}^V$  obtained using the dark pixel fraction (McGreer et al. 2015). The bottom panel shows our constraints on  $Q_{\text{HII}}^V$  (same as in the top panel), but the vertical axis scaled such that the behaviour around  $Q_{\text{HII}}^V \sim 1$  can be clearly visualized. It is clear that  $Q_{\text{HII}}^V = 1$  is ruled out at  $z \geq 5.6$  with a significance  $>2\sigma$  for either of the data sets.

Fig. 4. The horizontal orange dashed line corresponds to  $Q_{\text{HII}}^V = 1$ . For reference, we also show the late-reion-long-mfp model of Nasir & D’Aloisio (2020) (blue dashed curve) and the default reionization model of Keating et al. (2019) (magenta dotted curve, almost identical to the very late model of Kulkarni et al. 2019 in the redshift range of our interest). The  $2\sigma$  lower limits on  $Q_{\text{HII}}^V$  obtained using the dark pixel fraction (McGreer et al. 2015) are shown by green points with error bars. Our  $Q_{\text{HII}}^V$  limits for the optimistic data set are higher than the simulations of Keating et al. (2019) and Nasir & D’Aloisio (2020). The constraints for the pessimistic data set, on

the other hand, are consistent with the reionization history of Keating et al. (2019) within the  $2\sigma$  error bars (except that our  $2\sigma$  lower limit is marginally higher than the Keating et al. 2019 value). However, at  $z = 5.6$ , the allowed values of  $Q_{\text{HII}}^V$  from our analysis are significantly higher than that of Nasir & D’Aloisio (2020). This indicates that our method, for the same value of neutral fraction, produces more  $\tau_{\text{eff}}$  fluctuations than the model of Nasir & D’Aloisio (2020). All our constraints are consistent with the lower limits on  $Q_{\text{HII}}^V$  from the model-independent dark pixel fraction (McGreer et al. 2015).

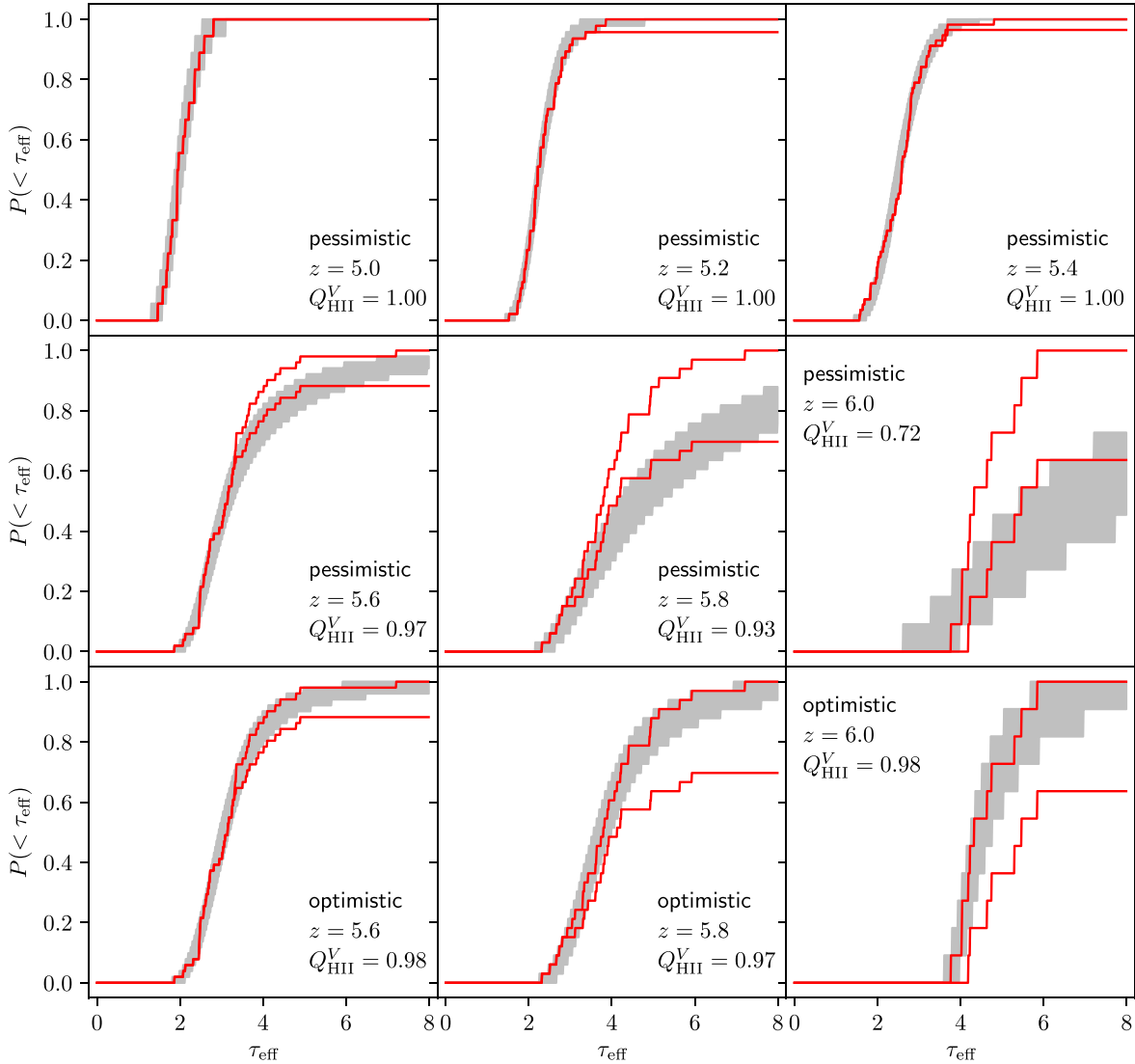
As expected, the allowed values of  $Q_{\text{HII}}^V$  are higher for the optimistic data set than the pessimistic one at  $z \geq 5.2$ . The differences between the two data sets decrease at smaller  $z$  because of fewer non-detections. At  $z = 5.6$ , the  $2\sigma$  upper limit on  $Q_{\text{HII}}^V$  is 0.99 for the optimistic data set, while it is 0.98 for the pessimistic one. In fact, we can rule out  $Q_{\text{HII}}^V = 1$  at  $3.98\sigma$  (99.993 per cent confidence) for the pessimistic data and at  $2.65\sigma$  (99.2 per cent confidence) for the optimistic data. This indicates that, in order to match the data, the completion of reionization must be delayed until  $z \sim 5.6$  (independent of how the non-detections are treated), i.e. the data are not consistent with complete reionization at  $z > 5.6$ . Interestingly, the  $2\sigma$  lower limit on  $Q_{\text{HII}}^V$  at  $z = 5.2$  is  $\approx 0.98$  for both the data sets, which would imply a rather significantly late completion of reionization. We find that our analysis requires the IGM to be completely ionized at  $z = 5$  and is able to limit the range of  $Q_{\text{HII}}^V$  to within a tight limit of  $\sim 0.02$  in the redshift range  $5.2 \leq z \leq 5.4$  (irrespective of which data set is used). The constraints at  $z = 5.8$  and 6 for the pessimistic data set are rather weak because of the larger errors on the  $\tau_{\text{eff}}$  distribution. One can see that  $Q_{\text{HII}}^V$  can be as small as 0.76 at  $z = 5.8$  and 0.56 at  $z = 6$  if the pessimistic data set represents reality, although the lower limits may be underestimated by  $\sim 0.06$  because of resolution effects (see Appendix B).

This statistical analysis shows the main benefit of our model. Given its computational efficiency, we can probe the parameter space in a reasonable amount of time and hence determine the range of histories allowed by the data. It is important to keep in mind here that, although our constraints on  $Q_{\text{HII}}^V$  are reasonably robust, our model cannot constrain other interesting physical quantities like the temperature and the amplitude of the photoionization rate. With improved data sets in the future, we can expect to test some of the model assumptions more critically and put more stringent constraints on reionization.

### 3.2 Aspects of the best-fitting reionization history

Having obtained the allowed ranges in reionization history from our model, we now examine in some detail the properties of the best-fitting model (red squares and black circles in Fig. 4). We first show the CDF  $P(< \tau_{\text{eff}})$  for the best-fitting model and the match with the data in Fig. 5. In each panel, the grey shaded regions denote





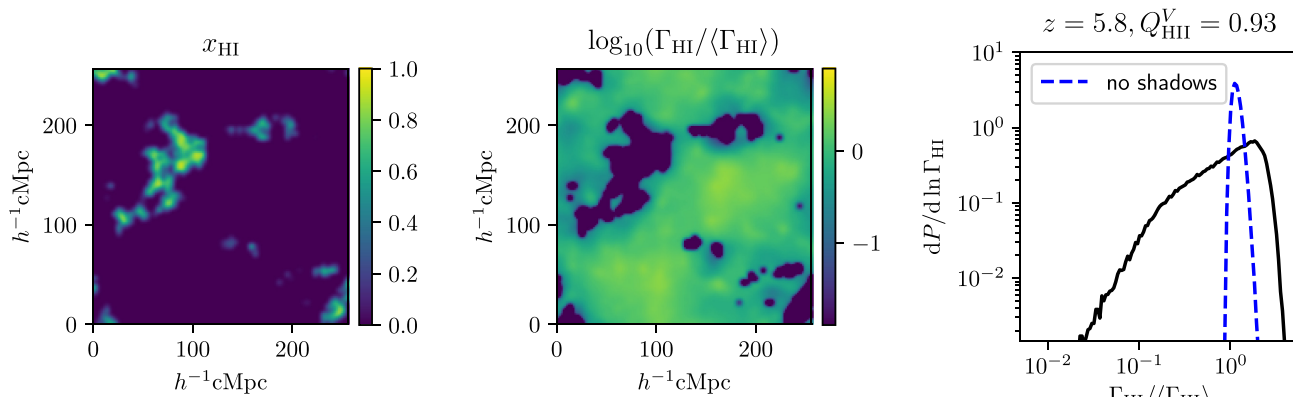
**Figure 5.** The cumulative  $\tau_{\text{eff}}$  distribution for the best-fitting reionization model obtained from the MCMC analysis in Section 3.1. The two red curves show the observational data for the optimistic (left/upper) and pessimistic (right/lower) cases of Bosman et al. (2018), with the legend mentioning which of the two data sets is used for model comparison. The grey shaded regions correspond to the model predictions accounting for statistical fluctuations along different sightlines. The redshifts and the ionized volume fractions  $Q_{\text{HII}}^V$  are mentioned in the respective panels. Note that we do *not* show the results for the optimistic data set at  $z = 5.0, 5.2,$  and  $5.4$  as the best-fitting models in these cases are almost identical to those for the pessimistic data.

the predictions of our model accounting for statistical fluctuations along different sightlines while the red curves denote  $P(< \tau_{\text{eff}})$  for the optimistic (left/upper) and pessimistic (right/lower) cases. The legend specifies which of the two cases is used for the model comparison.

The top panels show the predictions for the pessimistic data set at  $z = 5$  (left-hand panels),  $5.2$  (middle panels), and  $5.4$  (right-hand panels), while the middle panels are at  $z = 5.6$  (left-hand panels),  $5.8$  (middle panels), and  $6$  (right-hand panels). The bottom panels show the match for the optimistic data set at  $z = 5.6$  (left-hand panels),  $5.8$  (middle panels), and  $6$  (right-hand panels). Since the best-fitting models for the optimistic cases at  $z = 5, 5.2,$  and  $5.4$  are almost identical to the pessimistic cases (see Table 1), we do not show them separately in the figure. Visual comparisons between model and data can sometimes be misleading when the data points are highly correlated (which is the case here as different  $\tau_{\text{eff}}$  bins are

indeed correlated), however, at least in this case, one can see that the best-fitting model is a good description of the data. In particular, the model is able to reproduce the high values of  $\tau_{\text{eff}}$  that are required to match the data. The data at  $z \leq 5.4$  are consistent with a completely ionized medium and hence a relatively uniform  $\Gamma_{\text{HII}}$ , which was also noted by Becker et al. (2015) and Bosman et al. (2018). However, at higher redshifts, the best-fitting model contains some neutral patches in the IGM (characterized by a ionized volume fraction  $Q_{\text{HII}}^V < 1$ ). The introduction of neutral islands in the model allows for sightlines with large  $\tau_{\text{eff}}$ , which captures the high- $\tau_{\text{eff}}$  tail of the distribution.

Let us now understand the characteristics of the ionization and the radiation fields for our model. We choose the best-fitting model corresponding to the pessimistic data set at  $z = 5.8$  for which the effect of patchy reionization is quite prominent. The results are shown in Fig. 6. The left-hand panel shows the neutral fraction  $x_{\text{HII}}$  map for

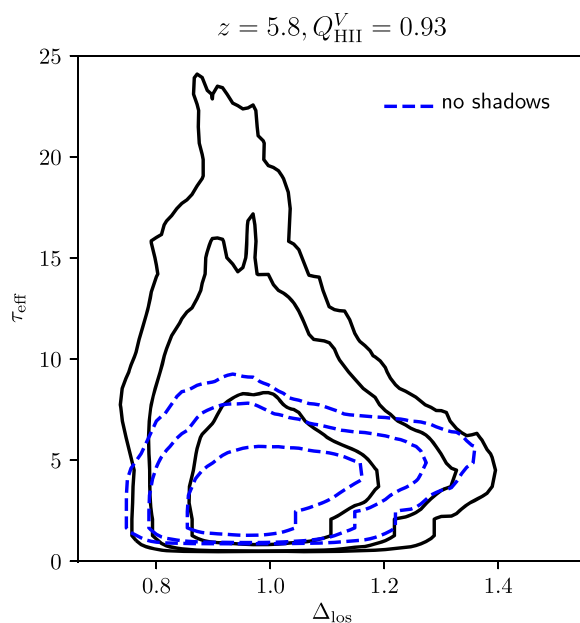


**Figure 6.** Various quantities obtained from our semi-numerical simulation at  $z = 5.8$  where the ionized volume fraction  $Q_{\text{HII}}^V = 0.93$ . The left-hand panel shows the neutral hydrogen fraction  $x_{\text{HI}}$  field for a slice of thickness  $4h^{-1}$  cMpc, while the middle panel shows the  $\Gamma_{\text{HI}}$  fluctuations for the same slice. The black curve in the right-hand panel shows the PDF of  $\Gamma_{\text{HI}}$  for only points in the ionized regions, while the blue dashed curve is for the case where the effect of shadows is turned off.

a two-dimensional slice of thickness  $4h^{-1}$  cMpc, while the middle panel shows the fluctuations in the photoionization rate  $\Gamma_{\text{HI}}$ . The ionization field resembles an almost ionized universe with patches of neutral islands in between. As expected,  $\Gamma_{\text{HI}}$  is non-zero in the ionized regions and (almost close to) zero in the neutral regions, thus tracing the overall topology of the ionization map. What is interesting to note is that even in the ionized regions there are pixels close to the neutral islands where the photoionization rate is quite small. These regions are the ‘shadows’ that arise because they do not receive photons from sources beyond the neutral islands. In our model, this effect is captured via the inclusion of the shadowing algorithm outlined in Section 2.2.

To see the effect of these shadows on the distribution of  $\Gamma_{\text{HI}}$  in the ionized regions, we plot the PDF  $dP/d\ln \Gamma_{\text{HI}}$  (computed using only pixels in the ionized regions) in the right-hand panel of Fig. 6 (the black line). One can see that the distribution peaks around the mean  $\langle \Gamma_{\text{HI}} \rangle$  and there exists a long tail for small values of  $\Gamma_{\text{HI}}$ . This tail arises from the shadows near the neutral islands. To confirm this point, we have plotted in blue dashed curve the  $\Gamma_{\text{HI}}$  distribution for the case where we turn-off the shadows, i.e. we simply use equation (9) instead of equation (14) to compute the photoionization rate. Clearly, the absence of the shadows leads to a very sharply peaked distribution of  $\Gamma_{\text{HI}}$ . Our  $\Gamma_{\text{HI}}$  distribution can be compared with those of Davies & Furlanetto (2016) and Nasir & D’Aloisio (2020), who also find a low- $\Gamma_{\text{HI}}$  tail. In fact, Nasir & D’Aloisio (2020) have explicitly checked that such a tail arises from shadowing of the neutral islands (see their fig. 5). In this sense, our findings are qualitatively similar to theirs, although the implementations are very different.

We further compare our model with that of Davies & Furlanetto (2016) by computing the mean overdensity  $\Delta_{\text{los}}$  and the mean effective optical depth  $\tau_{\text{eff}}$  along different sightlines of length  $50h^{-1}$  cMpc and plotting the resulting two-dimensional distribution in Fig. 7. The black contours correspond to the distribution for the best-fitting model, while the blue dashed contours are for the case where the shadows have been tuned off. The three contours in each case enclose 68, 95, and 99 per cent sightlines from the sample. It is clear that there is a mild negative correlation between  $\Delta_{\text{los}}$  and  $\tau_{\text{eff}}$  for the best-fitting model, entirely different from the positive correlation that is expected for an IGM with uniform  $\Gamma_{\text{HI}}$ ; see equation (22). Our model is thus qualitatively consistent with Davies & Furlanetto (2016). We further conclude from the figure that the presence of



**Figure 7.** The black curves show the distribution of sightlines in the  $\tau_{\text{eff}} - \Delta_{\text{los}}$  plane, where  $\tau_{\text{eff}}$  is the effective optical depth and  $\Delta_{\text{los}}$  is the average overdensity along each line of sight. The distribution is calculated for the best-fitting model obtained using the pessimistic data at  $z = 5.8$ . The contours contain 68, 95, and 99 per cent of the sightlines. For comparison, we also show the results when the shadows are turned off (blue dashed contours).

shadows increases the fluctuations in  $\tau_{\text{eff}}$  allowing for a better fit to the data for the pessimistic case.

#### 4 SUMMARY AND DISCUSSION

Recent observations of the effective optical depth  $\tau_{\text{eff}}$  of Ly $\alpha$  absorption at  $5 \lesssim z \lesssim 6$  show significant fluctuations when averaged over reasonably large-scale  $50h^{-1}$  cMpc scales. One possible interpretation of these observations is that the fluctuations arise because of left-over HI islands and that HI reionization is complete only at  $z \sim 5.2$  (Keating et al. 2019; Kulkarni et al. 2019; Nasir & D’Aloisio 2020). If this interpretation of the data is indeed true, it becomes

imperative to include these observations in any parameter constraints related to the reionization history (in addition to, e.g. the existing CMB observations of electron scattering optical depth). Obtaining constraints, in turn, requires efficient methods of computing the relevant observables (in this case, the Ly $\alpha$  optical depth) to probe the space of unknown parameters.

To achieve this goal, we have developed a semi-numerical technique to constrain the reionization history at  $5 \lesssim z \lesssim 6$ . Our method is appropriate for probing large-scale properties of the Ly $\alpha$  absorption in relatively low-resolution simulation boxes and relies on two main inputs: (i) the modelling of ionized regions using a photon-conserving semi-numerical code of reionization (`script`; Choudhury & Paranjape 2018) combined with a prescription for blocking photons from sources along sightlines passing through neutral regions and (ii) modelling the Ly $\alpha$  optical depth using the fluctuating Gunn–Peterson approximation. To our knowledge, this is the least computationally expensive model to study the Ly $\alpha$  opacity fluctuations.

We find that the model is able to capture the essential properties of the HI field as observed in the Ly $\alpha$  absorption, similar to those found in other semi-numerical models (Davies & Furlanetto 2016; Nasir & D’Aloisio 2020) and more detailed simulations (Keating et al. 2019; Kulkarni et al. 2019). Since the method is computationally fast, it allows us to probe the parameter space quite efficiently and thus obtain the range of histories consistent with the data (keeping in mind that other physical quantities like the temperature and the amplitude of the photoionization rate cannot be constrained by our model). We find that the inferred reionization history is delayed when we use the data set where non-detections of the flux are treated as having infinite optical depth (the so-called ‘pessimistic’ case of Bosman et al. 2018) compared to the case where non-detections are assumed to have optical depths just below the detection limit (the ‘optimistic’ case). The data are *inconsistent* with reionization being complete at  $z > 5.6$  (independent of which data set is used). The completion can be as late as  $z \sim 5.2$ , corresponding to the  $2\sigma$  lower limits on the ionized fraction. We also find that the ionized volume fraction can be as low as  $\sim 60$  per cent at  $z \sim 6$  for the pessimistic data set. The analysis thus indicates the potential of our technique in constraining the reionization history with more number of quasar sightlines at  $z \sim 6$ .

The number of known  $z > 5$  quasars will increase dramatically in the next decade with the upcoming quasar searches that will be performed by *Euclid* (Euclid Collaboration et al. 2019; Griffin et al. 2020), the Vera Rubin Observatory (formerly LSST; Ivezić et al. 2019), and the *Nancy Grace Roman Space Telescope* (formerly *WFIRST*; Spergel et al. 2015). Coupled with more efficient spectroscopic observations owing to the ELT (Gilmozzi & Spyromilio 2007), new high- $z$  quasars will significantly increase the amount and quality of Ly $\alpha$  opacity information at  $z > 6$  in the next decade. The discovery of bright quasars beyond  $z = 7.5$  (Bañados et al. 2018) ensures we will be able to map the large-scale evolution of opacity until  $z \sim 6.5$ , where Gunn–Peterson absorption is expected to saturate fully.

Indeed, fast semi-numerical models like ours often are unable to track all the physical processes self-consistently. We have also seen in Fig. 4 that the reionization history inferred from our analysis can be different from that predicted by other simulations (Keating et al. 2019; Kulkarni et al. 2019; Nasir & D’Aloisio 2020). One possible source of uncertainty in our model arises from the treatment of the regions that do not receive photons from sources whose lines of sight pass through the HI islands, thus creating shadows and suppressing the photoionization rate. Another possibility could be

that our assumed value of mean-free path  $\lambda_{\text{ss}}$  is different than that in reality. In the future, we plan to make a detailed comparison of our model with simulations to check the validity of our method of producing shadows as well as test various other assumptions of the model.

We also plan to further improve the analysis by including various effects that have been ignored in this work. The first is to model the Ly $\beta$  absorption from the same sightlines and compare with the available data (Eilers et al. 2019), thus obtaining more stringent constraints on the reionization history. The second is to forward model the noise in the observations and include it in the model, instead of considering the two extreme cases as done in this work. The third and perhaps most important improvement would be to self-consistently model the temperature evolution in each cell across redshifts. This would relax the assumption of the power-law relation between the temperature and density of the cell and should be able to account for the dependence of the temperature field on the reionization history. In addition, we also plan to expand the scope of the analysis to include other data sets and hence constrain the reionization history accounting for causal correlations between different redshift bins. Such improvements, combined with the fact that the model is computationally inexpensive, would then allow for comparing with a wide variety of observations simultaneously and hence obtain constraints on reionization.

## ACKNOWLEDGEMENTS

TRC acknowledges support of the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.02-0700 and the Associateship Scheme of ICTP. The research of AP is supported by the Associateship Scheme of ICTP, Trieste, and the Ramanujan Fellowship awarded by the Department of Science and Technology, Government of India. SEIB acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 669253). We thank the anonymous referee for constructive comments that helped improve the content of this paper.

## DATA AVAILABILITY

The code for generating the ionized regions, `script`, is available at <https://bitbucket.org/rctirthankar/script>. The rest of the data underlying this paper will be shared on reasonable request to the authors.

## REFERENCES

- Bañados E. et al., 2018, *Nature*, 553, 473  
 Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, *MNRAS*, 447, 3402  
 Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, *ApJ*, 872, 101  
 Bolton J. S., Haehnelt M. G., 2007, *MNRAS*, 382, 325  
 Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, *MNRAS*, 416, L70  
 Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M. I., 2018, *MNRAS*, 479, 1055  
 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., Haehnelt M. G., Bosman S. E. I., Puchwein E. I., 2018, *MNRAS*, 473, 765  
 Chardin J., Puchwein E., Haehnelt M. G., 2017, *MNRAS*, 465, 3429

- Choudhury T. R., Ferrara A., 2005, *MNRAS*, 361, 577
- Choudhury T. R., Paranjape A., 2018, *MNRAS*, 481, 3821
- Croft R. A. C., Weinberg D. H., Katz N., Hernquist L., 1998, *ApJ*, 495, 44
- Davies F. B., Furlanetto S. R., 2016, *MNRAS*, 460, 1328
- Davies F. B., Hennawi J. F., Eilers A.-C., 2020, *MNRAS*, 493, 1330
- Dixon K. L., Furlanetto S. R., 2009, *ApJ*, 706, 970
- Ď urovč íková D., Katz H., Bosman S. E. I., Davies F. B., Devriendt J., Slyz A. I., 2020, *MNRAS*, 493, 4256
- D'Aloisio A., McQuinn M., Trac H., 2015, *ApJ*, 813, L38
- Eilers A.-C., Davies F. B., Hennawi J. F., 2018a, *ApJ*, 864, 53
- Eilers A.-C., Davies F. B., Hennawi J. F., Prochaska J. X., Lukić Z., Mazzucchelli C., 2017, *ApJ*, 840, 24
- Eilers A.-C., Hennawi J. F., Davies F. B., 2018b, *ApJ*, 867, 30
- Eilers A.-C., Hennawi J. F., Davies F. B., Oñorbe J., 2019, *ApJ*, 881, 23
- Euclid Collaboration et al., 2019, *A&A*, 631, A85
- Fan X., Narayanan V. K., Strauss M. A., White R. L., Becker R. H., Pentericci L., Rix H.-W., 2002, *AJ*, 123, 1247
- Fan X. et al., 2000, *AJ*, 120, 1167
- Fan X. et al., 2001, *AJ*, 122, 2833
- Fan X. et al., 2003, *AJ*, 125, 1649
- Fan X. et al., 2004, *AJ*, 128, 515
- Fan X. et al., 2006a, *AJ*, 131, 1203
- Fan X. et al., 2006b, *AJ*, 132, 117
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Gaikwad P., Choudhury T. R., Srianand R., Khaire V., 2018, *MNRAS*, 474, 2233
- Gaikwad P. et al., 2020, *MNRAS*, 494, 5091
- Gallerani S., Choudhury T. R., Ferrara A., 2006, *MNRAS*, 370, 1401
- Gallerani S., Ferrara A., Fan X., Choudhury T. R., 2008, *MNRAS*, 386, 359
- Gilmozzi R., Spyromilio J., 2007, *The Messenger*, 127, 11
- Goodman J., Weare J., 2010, *Commun. Appl. Math. Comput. Sci.*, 5, 65
- Greig B., Mesinger A., 2015, *MNRAS*, 449, 4246
- Greig B., Mesinger A., Bañados E., 2019, *MNRAS*, 484, 5094
- Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, *MNRAS*, 466, 4239
- Griffin A. J., Lacey C. G., Gonzalez-Perez V., Lagos C. d. P., Baugh C. M., Fanidakis N., 2020, *MNRAS*, 492, 2535
- Hinshaw G. et al., 2013, *ApJS*, 208, 19
- Hutter A., 2018, *MNRAS*, 477, 1549
- I vezić Ž. et al., 2019, *ApJ*, 873, 111
- Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, *MNRAS*, 321, 372
- Keating L. C., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2019, *MNRAS*, 497, 906
- Kulkarni G., Choudhury T. R., Puchwein E., Haehnelt M. G., 2017, *MNRAS*, 469, 4283
- Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D. I., 2019, *MNRAS*, 485, L24
- Maselli A., Gallerani S., Ferrara A., Choudhury T. R., 2007, *MNRAS*, 376, L34
- McGreer I. D., Mesinger A., D'Odorico V. I., 2015, *MNRAS*, 447, 499
- McGreer I. D., Mesinger A., Fan X. I., 2011, *MNRAS*, 415, 3237
- Meiksin A., 2020, *MNRAS*, 491, 4884
- Mitra S., Choudhury T. R., Ferrara A., 2011, *MNRAS*, 413, 1569
- Mitra S., Choudhury T. R., Ferrara A., 2012, *MNRAS*, 419, 1480
- Mitra S., Choudhury T. R., Ferrara A., 2018, *MNRAS*, 473, 1416
- Mitra S., Ferrara A., Choudhury T. R., 2013, *MNRAS*, 428, L1
- Nasir F., D'Aloisio A., 2020, *MNRAS*, 494, 3080
- Paranjape A., Choudhury T. R., Padmanabhan H., 2016, *MNRAS*, 460, 1801
- Planck Collaboration et al., 2019, *A&A*, 641, 92
- Planck Collaboration et al., 2014, *A&A*, 571, A16
- Powell M., 2009, Technical Report, Department of Applied Mathematics and Theoretical Physics
- Pritchard J. R., Loeb A., Wyithe J. S. B., 2010, *MNRAS*, 408, 57
- Puchwein E., Bolton J. S., Haehnelt M. G., Madau P., Becker G. D., Haardt F., 2015, *MNRAS*, 450, 4081
- Raskutti S., Bolton J. S., Wyithe J. S. B., Becker G. D., 2012, *MNRAS*, 421, 1969
- Sheth R. K., Tormen G., 2002, *MNRAS*, 329, 61
- Songaila A., 2004, *AJ*, 127, 2598
- Songaila A., Cowie L. L., 2002, *AJ*, 123, 2183
- Spergel D. et al., 2015, preprint ([arXiv:1503.03757](https://arxiv.org/abs/1503.03757))
- Springel V., 2005, *MNRAS*, 364, 1105
- Worseck G. et al., 2014, *MNRAS*, 445, 1745
- Wyithe J. S. B., Bolton J. S., 2011, *MNRAS*, 412, 1926
- Wyithe J. S. B., Loeb A., 2003, *ApJ*, 586, 693
- Wyithe J. S. B., Loeb A., 2004, *Nature*, 427, 815
- Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, *ApJ*, 654, 12

## APPENDIX A: DEPENDENCE OF THE RESULTS ON THE PARAMETER CHOICES

In this Appendix, we discuss the dependence of our model on different parameters once we vary them beyond their default values.

### A1 The minimum halo mass $M_{\min}$

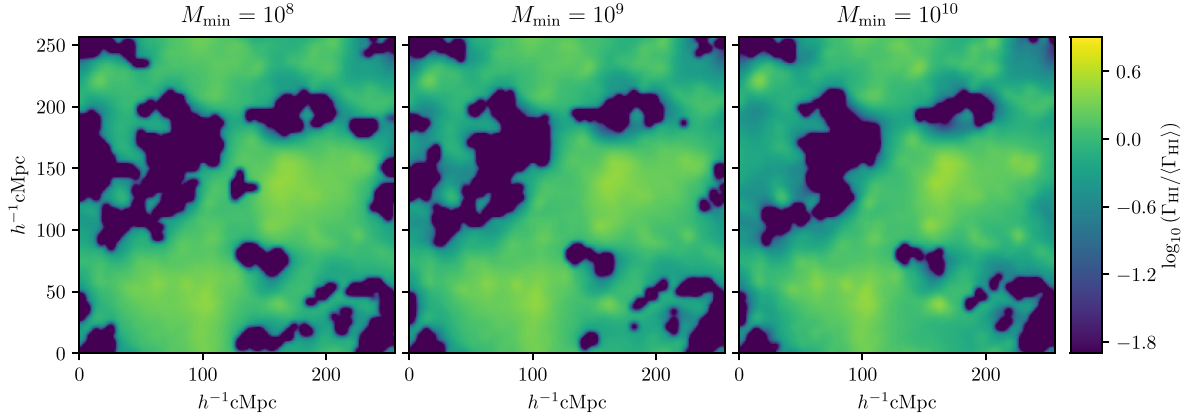
We have fixed the value of the minimum mass  $M_{\min}$  of haloes that can produce ionizing photons to  $10^9 M_{\odot}$  for the default analysis in Section 3.1. In this section, we investigate if the choice affects our constraints on the reionization history. Let us first see the  $\Gamma_{\text{HI}}$  fluctuations for a two-dimensional slice for three values of  $M_{\min}$  at a representative redshift, chosen to be  $z = 5.8$ . The maps are shown in Fig. A1. We have fixed the value of  $Q_{\text{HII}}^V$  for the three cases (by changing the  $\zeta$  values appropriately). We can see that although the large-scale properties look very similar, there are some obvious differences. As  $M_{\min}$  increases, the neutral regions (shown by the dark points, corresponding to  $\Gamma_{\text{HI}}$  values almost zero), which are relatively smaller in size, tend to disappear. Since the amount of neutral volume is the same, the existing neutral regions have larger sizes for higher  $M_{\min}$ .

However, these changes, which are visually obvious, seem to affect the maps at scales smaller than  $50 h^{-1}$  cMpc. As a result, we expect that these may not affect the  $\tau_{\text{eff}}$  distributions when averaged over large path lengths. To confirm this, we plot the CDF  $P(< \tau_{\text{eff}})$  for the three cases in Fig. A2. For the default case  $M_{\min} = 10^9 M_{\odot}$ , we choose parameters corresponding to the best-fitting model obtained in Section 3.1. For the other two cases, we keep the same value of  $Q_{\text{HII}}^V$  and adjust  $A_{\tau}$  to match the mean  $\tau_{\text{eff}}$  for the default case. As is clear from the figure, the three models produce identical  $P(< \tau_{\text{eff}})$ , thus confirming our expectation that the  $Q_{\text{HII}}^V$  constraints are insensitive to  $M_{\min}$  (although the values of  $A_{\tau}$  could vary by  $\sim 10$  per cent as we change  $M_{\min}$  by an order of magnitude). It is possible that models with different  $M_{\min}$  can be distinguished when the transmitted flux is averaged over sightlines of smaller length; this is a possible avenue that we plan to explore in the future.

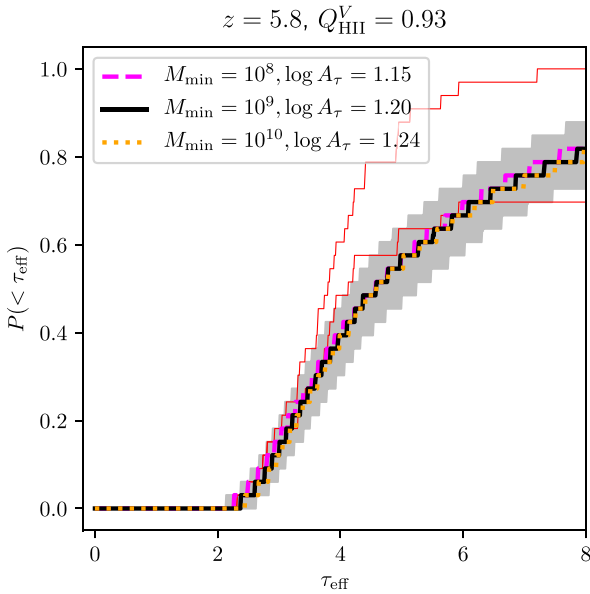
We did not investigate whether the  $\tau_{\text{eff}}$  distributions change more substantially for a reionization model dominated by much rarer sources, e.g. AGNs (Kulkarni et al. 2017). Such models would require higher values of  $M_{\min}$  than what is considered here. In such cases, our subgrid based method of computing the collapsed fraction becomes less accurate and hence a more careful analysis is warranted. We plan to study such cases separately in the future.

### A2 The mean-free path $\lambda_{\text{ss}}$

The default value of the mean-free path  $\lambda_{\text{ss}}$ , determined by the distance between the self-shielded regions, has been chosen as the value extrapolated from lower redshift observations. In this section,



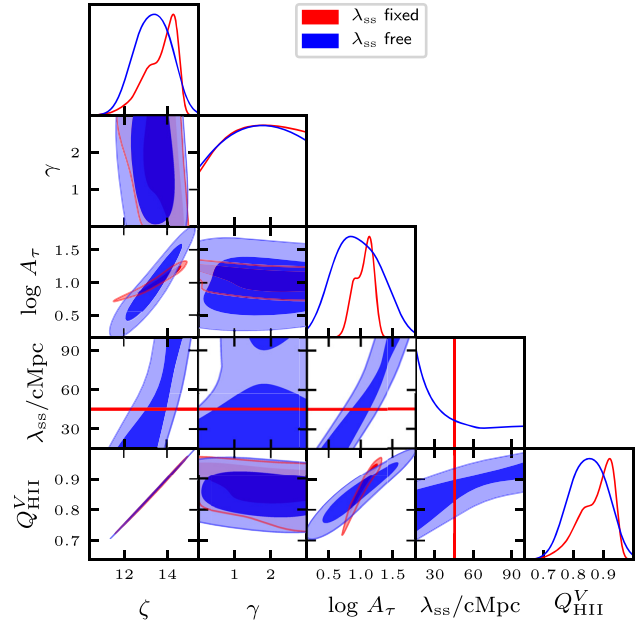
**Figure A1.** The two-dimensional maps of the photoionization rate fluctuations for a slice of thickness  $4h^{-1}$  cMpc at  $z = 5.8$  for different values of  $M_{\min}$  (in units of  $M_{\odot}$ ) as mentioned above the respective panels.



**Figure A2.** The  $\tau_{\text{eff}}$  CDF for three different values of  $M_{\min}$  at  $z = 5.8$  (with red curves showing the optimistic and pessimistic data sets used in this paper). The default runs in the paper are for  $M_{\min} = 10^9 M_{\odot}$ . All the three cases have the same value of  $Q_{\text{HII}}^V$  (corresponding to the best-fitting value in the default case), while the  $A_{\tau}$  values are adjusted so as to match the mean  $\tau_{\text{eff}}$  for the default case. All other model parameters are identical for the three cases.

we investigate if the uncertainty in the value of  $\lambda_{\text{ss}}$  affects our constraints on  $Q_{\text{HII}}^V$  as discussed in Section 3.1.

We first perform a full MCMC analysis treating  $\lambda_{\text{ss}}$  as a free parameter. We choose it to have a flat prior in the range 15–100 cMpc. These limits are set by the resolution and the box size of our simulation, however, they cover the most interesting range of values for the parameter. The posterior distribution of the parameters for this case, along with those of our three-parameter default run with  $\lambda_{\text{ss}}$  fixed, are shown in Fig. A3. The constraints are obtained using the pessimistic data set at  $z = 5.8$ . We can immediately see that the posterior distribution of  $Q_{\text{HII}}^V$  is different when  $\lambda_{\text{ss}}$  is left free. The best-fitting value is smaller than the default case. However, interestingly, the  $2\sigma$  limits on  $Q_{\text{HII}}^V$  remain relatively unaffected (the range is [0.76–0.96] for the default case, while it is modified to [0.73–

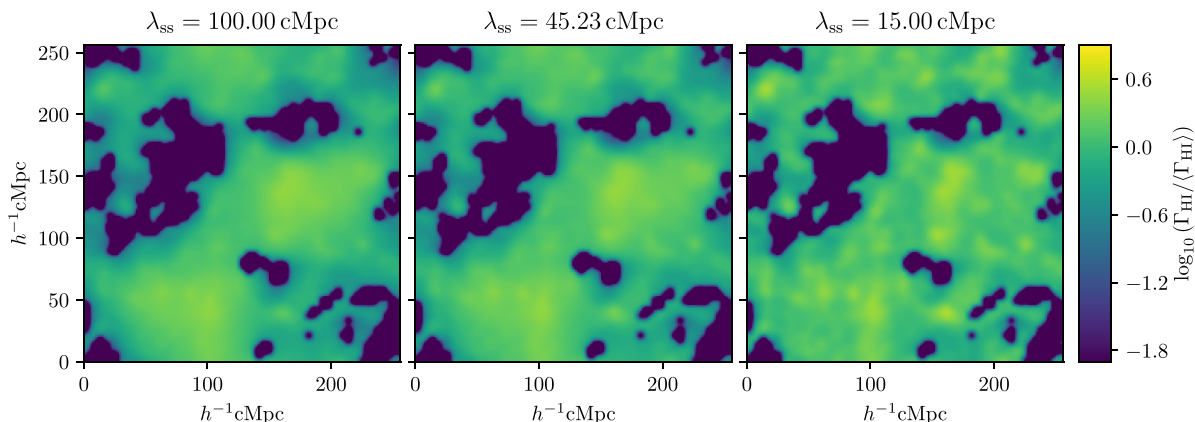


**Figure A3.** The posterior distribution of parameters obtained using MCMC analysis when  $\lambda_{\text{ss}}$  is kept fixed to the default value (red) and when  $\lambda_{\text{ss}}$  is left free (blue). The results are shown for the pessimistic data set at  $z = 5.8$ . For the case where  $\lambda_{\text{ss}}$  is fixed, the red straight lines parallel to the axes denote the fixed value of  $\lambda_{\text{ss}}$ .

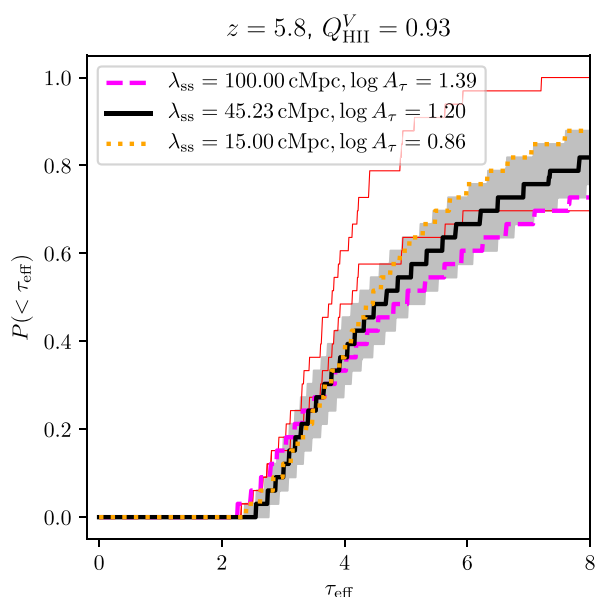
0.97] when  $\lambda_{\text{ss}}$  is free). Hence, we can conclude that our constraints on the reionization history remain relatively unaffected even when  $\lambda_{\text{ss}}$  is allowed to vary.

We find that it is *not* possible to constrain  $\lambda_{\text{ss}}$  from our analysis. The marginalized posterior is almost flat with marginal increase at lower values. In our approach, it is perhaps more convenient to make use of constraints on  $\lambda_{\text{ss}}$  from other studies, either observations or simulations.

We can see from the figure that there is a positive correlation between  $\lambda_{\text{ss}}$  and  $\log A_{\tau}$ . The reason is that a larger mean-free path allows sources from larger distances to contribute to the photon flux  $S_i$ , thus decreasing the optical depth  $\tau_{\alpha, i}$ . One requires a higher value of  $A_{\tau}$  to compensate for this increase in the flux and increase  $\tau_{\alpha, i}$ ;



**Figure A4.** Same as Fig. A1 but for different values of  $\lambda_{\text{ss}}$ .



**Figure A5.** Same as Fig. A2 but for different values of  $\lambda_{\text{ss}}$ .

see equation (23). This correlation between  $\lambda_{\text{ss}}$  and  $\log A_\tau$  leads to a wider range of  $A_\tau$  values to be allowed than that in the default case.

There is also a positive correlation between  $\lambda_{\text{ss}}$  and  $Q_{\text{HII}}^V$  (or  $\zeta$ ). To understand this, we compare the  $\Gamma_{\text{HII}}$  maps for different values of  $\lambda_{\text{ss}}$  in Fig. A4. For this plot, we have fixed the value of  $Q_{\text{HII}}^V = 0.93$  for all the three cases. In the left-hand panel, we choose a high value  $\lambda_{\text{ss}} = 100$  cMpc, which essentially implies that points in the ionized regions can see sources at large distances (unless blocked by a neutral region). Thus, for large  $\lambda_{\text{ss}}$ , only points that are considerably far away from the islands (say, with distance  $\gtrsim \lambda_{\text{ss}}$ ) can receive photons from all directions without being obstructed by the islands. This implies that the number of points that are affected by these islands is relatively larger, which then leads to more fluctuations. Hence, we end up with a somewhat counter-intuitive result where increasing  $\lambda_{\text{ss}}$  leads to more large-scale fluctuations in  $\Gamma_{\text{HII}}$ . It follows that, to produce the same amount of fluctuations as the default  $\lambda_{\text{ss}}$ , we require less neutral islands and thus larger  $Q_{\text{HII}}^V$  for the case of a higher  $\lambda_{\text{ss}}$ . This is the cause of the positive correlation found in the MCMC analysis.

In the right-hand panel of Fig. A4, we show the  $\Gamma_{\text{HII}}$  map for a smaller value  $\lambda_{\text{ss}} = 15$  cMpc. As expected, we find a lot of small-scale fluctuations in  $\Gamma_{\text{HII}}$  in this case. Moreover, only points that are very close to the neutral islands ( $\lesssim 15$  cMpc) are affected by them, leading to a much smaller effect of the shadows.

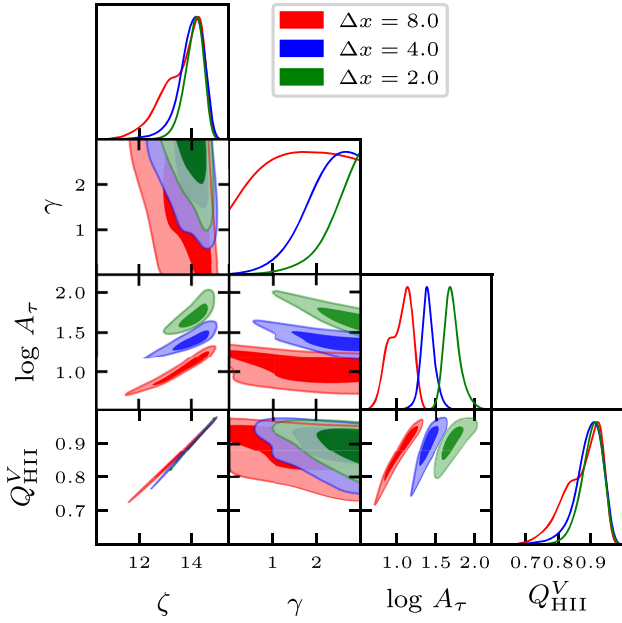
The effect of  $\lambda_{\text{ss}}$  on the  $\tau_{\text{eff}}$  CDF is shown in Fig. A5. Note that we have kept the value of  $Q_{\text{HII}}^V = 0.93$  same for the three cases plotted, while the value of  $A_\tau$  is chosen to obtain the same mean  $\tau_{\text{eff}}$ . As expected, the  $\tau_{\text{eff}}$  CDF is wider for larger  $\lambda_{\text{ss}}$  because of more large-scale fluctuations, and similarly narrower for smaller  $\lambda_{\text{ss}}$ . However, the variation even in such extreme cases is within the cosmic variance of the observable, thus not affecting the constraints on  $Q_{\text{HII}}^V$  that severely.

To summarize the results of this section, we find that the constraints on the reionization history are, in principle, affected by the chosen value of  $\lambda_{\text{ss}}$ . In addition, we also find that our model is unable to put any constraints on  $\lambda_{\text{ss}}$ .

## APPENDIX B: CONVERGENCE OF THE RESULTS WITH RESPECT TO RESOLUTION

In this Appendix, we study the dependence of our results on the grid size chosen for generating the ionized bubbles and computing the Ly $\alpha$  optical depth. This is particularly important as some of the model parameters depend on the resolution. For this study, we have chosen the pessimistic data set at  $z = 5.8$  and carried out the MCMC analysis for grid sizes of  $\Delta x = 4$  ( $64^3$  grids) and  $2h^{-1}$  cMpc ( $128^3$  grids). Recall that our default runs are for a coarser resolution  $\Delta x = 8h^{-1}$  cMpc ( $32^3$  grids). While running for the different resolutions, we choose the same priors as the default case and fix  $M_{\text{min}}$  and  $\lambda_{\text{ss}}$  to their default values as mentioned in Section 2.4. The posterior distributions of the various parameters for the three cases are shown in Fig. B1.

The first point to note from the figure is that the allowed values of  $A_\tau$  decrease with coarsening resolution (i.e. increasing  $\Delta x$ ). This is not surprising as  $A_\tau$  is directly dependent of  $\kappa_{\text{res}}$ , a parameter that itself depends on the resolution. The fact that the constraints on  $A_\tau$  are resolution-dependent is a direct consequence of using the fluctuation Gunn–Peterson approximation and our inability to capture the small-scale fluctuations in the low-resolution simulations. We also find that the other parameter that is affected by resolution, namely,  $\gamma$ , is also different for the different resolutions. However, the constraints on  $\gamma$  are weak no matter which resolution

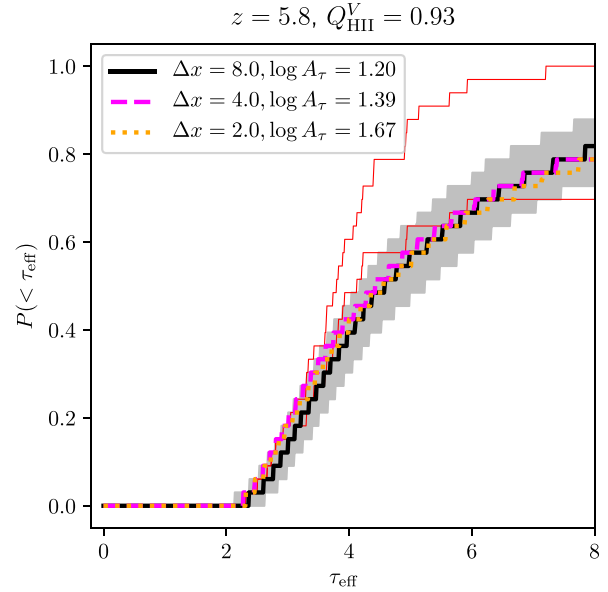


**Figure B1.** The posterior distribution of parameters obtained using MCMC analysis for three different grid resolutions used for generating the ionization maps and the Ly $\alpha$  optical depth. The results are shown at  $z = 5.8$  for the pessimistic data set. The default run in the paper corresponds to  $\Delta x = 8 h^{-1}$  cMpc (red) whose results are identical to those shown by red contours/curves in Fig. 2.

is chosen and  $\gamma \approx 1.5$ –2 seems to be a good choice for all the cases.

What is important for our analysis is that the constraints of  $Q_{\text{HII}}^V$  remain similar for the three resolutions. The best-fitting values are almost the same (they differ by only  $\lesssim 0.01$ ). The  $2\sigma$  upper limits are also within  $\sim 0.005$  of each other. There is some difference in the three cases at the lower tail of the  $Q_{\text{HII}}^V$  posterior distribution. The  $2\sigma$  lower limit for the default case  $\Delta x = 8 h^{-1}$  cMpc is 0.76, while they are 0.80 and 0.82 for  $\Delta x = 4$  and  $2 h^{-1}$  cMpc, respectively. This implies that our default runs underestimate the lower  $2\sigma$  values by  $\sim 0.06$  compared to the highest resolution probed here. The reason for  $Q_{\text{HII}}^V$  lower limits to be more stringent for higher resolution is as follows: More neutral IGM leads to more fluctuations in  $\tau_{\text{eff}}$ , whereas coarser resolution simulations tend to smooth the fluctuations. Hence, coarser resolution simulations allow agreement with the data even for lower values of  $Q_{\text{HII}}^V$  (values that produce more fluctuations in the finer resolution runs and thus tend to get ruled out). The resolution-dependence of the constraints is expected to be less prominent for cases where the allowed  $Q_{\text{HII}}^V$  values are higher, e.g. for the optimistic data sets and for lower redshifts.

We choose  $\Delta x = 8 h^{-1}$  cMpc as our default grid size because the MCMC runs take much less time to complete. For the three cases plotted in Fig. B1, it takes  $\sim 6$  h,  $\sim 36$  h, and  $\sim 15$  d for the runs to complete for  $\Delta x = 8, 4$ , and  $2 h^{-1}$  cMpc, respectively. Hence, using a coarse resolution allows us to perform many more MCMC runs in



**Figure B2.** The  $\tau_{\text{eff}}$  CDF for three different grid resolutions used while generating the H I field and the Ly $\alpha$  optical depth at  $z = 5.8$  (with red curves showing the optimistic and pessimistic data sets used in the paper). The model parameters correspond to the best-fitting values found from the MCMC run shown in Fig. B1. Note that all the three cases have the same value of  $Q_{\text{HII}}^V$ , while the  $A_\tau$  values are mentioned in the legend.

a reasonable amount of time and study the different features of the model in more detail. The downside of using the coarse resolution is that the lower limits quoted are somewhat conservative.

For completeness, we also show the CDF  $P(< \tau_{\text{eff}})$  for the three resolutions in Fig. B2. We choose the best-fitting  $Q_{\text{HII}}^V$  and  $A_\tau$  values for the three resolutions as found from the MCMC runs and fix  $\gamma = 1.5$  for all the cases. It is clear that the models produce almost identical  $\tau_{\text{eff}}$  distributions for the same value of  $Q_{\text{HII}}^V$  although the  $A_\tau$  values are different. This is consistent with our findings that the reionization constraints are insensitive to the grid size.

Overall, we can safely conclude that our results are not sensitive to the resolution, except for the lower limits being slightly underestimated for the coarser resolution. We use  $\Delta x = 8 h^{-1}$  cMpc for the MCMC analysis in the paper.

We have also checked whether our results are sensitive to the box size by comparing with a smaller box of length  $128 h^{-1}$  cMpc. We find that the  $\tau_{\text{eff}}$  distributions remain unchanged for the smaller box. This is perhaps not surprising as all the relevant scales in the problem (the size of the neutral regions, the mean-free path, and the length of the sightlines over which the optical depth is calculated) are smaller than the boxes used.

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