

Efficient modeling of Lyman- α opacity fluctuations during the late reionization epoch

Barun Maity^{1,*}, Frederick Davies¹, and Prakash Gaikwad²

¹ Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

² Department of Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology Indore, Simrol, MP 453552, India

Received 14 May 2025 / Accepted 27 October 2025

ABSTRACT

The Lyman- α forest opacity fluctuations observed from high-redshift quasar spectra have been proven to be extremely successful in probing the late phase of the reionization epoch. For ideal modeling of these opacity fluctuations, one of the main challenges is to satisfy the extremely high dynamic range requirements of the simulation box, resolving the Lyman- α forest while probing the large cosmological scales. In this study, we adopted an efficient approach to model Lyman- α opacity fluctuations in a coarse simulation volume, utilizing the semi-numerical reionization model SCRIPT (including inhomogeneous recombination and radiative feedback) integrated with a realistic photoionization background fluctuation generating model. Our model crucially incorporates ionization and temperature fluctuations, which are consistent with the reionization model. After calibrating our method with respect to high-resolution full hydrodynamic simulation, Nyx, we compared the models with available observational data at the redshift range, $z = 5.0-6.1$. With a fiducial reionization model (reionization end at $z = 5.8$), we demonstrated that the observed scatter in the effective optical depth can be matched reasonably well by tuning the free parameters of our model, (i.e., the effective ionizing photon mean free path and the mean photoionization rate). We further pursued an MCMC-based parameter space exploration, utilizing the available data to put constraints on the above free parameters. Our estimation prefers a slightly higher photoionization rate and slightly lower mean free path than the previous studies, which is also a consequence of temperature fluctuations. This study holds significant promise for efficiently extracting important physical information about the Epoch of Reionization, utilizing the wealth of available and upcoming observational data.

Key words. cosmology: observations – cosmology: theory – large-scale structure of Universe – dark ages, reionization, first stars

1. Introduction

The epoch of reionization (EoR) represents an important phase in the history of the Universe, marking the time when the first stars and galaxies formed and ionized neutral hydrogen atoms (HI) in the intergalactic medium (IGM). The precise timeline and morphology of this epoch can be crucial to understand the structure formation and evolution of the universe at those early times (see reviews for details, Barkana & Loeb 2001; Choudhury 2009; McQuinn 2016; Dayal & Ferrara 2018; Gnedin & Madau 2022; Choudhury 2022). There exists a handful of observational probes at various wavelengths to trace the IGM during the EoR (Zaroubi 2013; Mesinger 2016). Among these, spectra obtained from high-redshift quasars have been proven to be a treasure trove for extracting critical information about the nature of the IGM.

Specifically, Lyman- α (Ly- α) forest absorption of quasar spectra contains information about the line-of-sight distribution of neutral hydrogen atoms (HI) in the IGM, which has been widely used in the literature to study various physical aspects. For example, these include exploiting realistic numerical simulations to constrain the photoionization rate (Bolton & Haehnelt 2007; Wyithe & Bolton 2011; Gaikwad et al. 2023; Davies et al. 2024b) and tracking the temperature evolution of the IGM (Walther et al. 2019; Gaikwad et al. 2020). The high quality of existing quasar spectra also enables us to directly estimate the

neutral fraction of the universe with damping wing analysis (Spina et al. 2024) and with dark pixel analysis (McGreer et al. 2015; Jin et al. 2023). Recently, a popular scenario of late reionization end ($z_{\text{end}} \sim 5.3$) has come out based on the findings of large Gunn-Peterson (GP) troughs, giving rise to excess Ly- α forest opacity fluctuations at $z \geq 5.3$ (Zhu et al. 2021, 2022; Bosman et al. 2022).

Several traditional models have been found to be inconsistent to explain the large GP trough and the excess Ly- α opacity fluctuations. There exist a few suggestions to match the data, keeping the reionization end at $z \geq 6$. One of these includes a hidden abundance of faint quasar populations at $z \geq 5$ (Meiksin 2020); however, it is not yet completely understood whether they have the required photon budget to make a significant contribution. An extended and hot reionization scenario can also produce significant fluctuations via the temperature dependence of recombination rate (D'Aloisio et al. 2015). But, this model associates the opaque trough with large-scale galaxy overdensity (Davies et al. 2018a), which is disfavoured by the galaxy survey results (Becker et al. 2018; Christenson et al. 2021). It is also possible that the excess ionizing background fluctuations are amplified due to very short mean free path of the ionizing photons in the IGM (Davies & Furlanetto 2016; D'Aloisio et al. 2018). However, this is not in agreement with the extrapolation from existing lower redshift measurements (Worseck et al. 2014). Another emerging explanation involves a late reionization end ($z_{\text{end}} \leq 5.5$) where the shadowing effect on the large-scale ionization fluctuations by the residual neutral islands is

* Corresponding author: maity@mpia.de

responsible for the excess opacity fluctuations (Kulkarni et al. 2019; Keating et al. 2020; Nasir & D’Aloisio 2020). However, these models are finely tuned to match the ionizing photon emissivity according to the demand of both Ly- α forest fluctuations and other high-redshift observables such as CMB scattering optical depth. In recent times, this explanation has also been adopted by semi-numerical models, which are further able to provide constraints on the global ionization fraction utilizing Ly- α forest observations (Choudhury et al. 2021; Qin et al. 2021, 2025). However, these models lack a realistic description of photoionization background fluctuations and temperature fluctuations. Despite recent improvements along these lines (Choudhury & Chakraborty 2025), those models are still devoid of proper calibration for individual components of opacity fluctuations against a realistic forest model. For instance, Qin et al. (2025) used a single simulation suite at a fixed redshift ($z = 5.0$) to map out an analytic relation between true opacity and FGPA approximation via probability density function. However, as we discuss in our study, the FGPA relation itself may need to be modified while dealing with coarse resolution simulation, accounting for the calibration of temperature, UVB fluctuations, and redshift dependence of Ly- α optical depth.

Nevertheless, it is crucial to determine the strength of photoionization background fluctuations as it provides an important boundary limit to the reionization scenario. It has been shown that the fluctuations rise rapidly when the ionization bubbles start to overlap with each other, eliminating the neutral hydrogen patches (Bolton & Haehnelt 2007; Wyithe & Bolton 2011; Davies et al. 2018d). The measurements of the photoionization rate (Γ_{HI}) also show an increment in the redshift range $z \sim 5-6$ supporting the above statement. However, the constraints are not very stringent due to a lack of transmission flux at these high redshifts and a degeneracy with the thermal state of the IGM. In recent studies, the constraints have become more stringent with the abundance of high-quality data and robust modeling (Gaikwad et al. 2023; Davies et al. 2024b).

In addition to the photoionization rate, one needs the information on the photon mean free path in order to estimate the reionization source emissivities. At relatively lower redshifts, this can be measured by identifying the absorption lines corresponding to the neutral hydrogen structures (Prochaska et al. 2014). However, this identification becomes difficult towards higher redshifts, which demands the stacking of the quasar spectra. Recently, relatively short values for the mean free path at $z \sim 6$ have been reported by utilizing this stacking method (Becker et al. 2021; Zhu et al. 2023), which again suggests a scenario of late reionization or excess ionization background fluctuations.

From the discussion thus far, it is apparent that we need a reasonably accurate and efficient model for estimating the Ly- α opacity fluctuations during later stages of reionization to properly interpret the observations. However, the modeling faces the challenge of high dynamic range requirements for the simulation box. On the one hand, it has to resolve small scales ($\lesssim 100$ Kpc) for a converged description of Ly- α forest spectra; whereas, on the other hand, it should have a large enough scale to access the observed fluctuation spanning a wide cosmological scale. For example, Davies et al. (2024b) utilizes a very high resolution hydrodynamic simulation, sufficient to resolve small-scale structures while their large-scale fields are not fully self-consistent. In another similar study by Gaikwad et al. (2023), an efficient algorithm was exploited with a relatively larger simulation box ($160 h^{-1}$ cMpc) and coarser resolution. In principle, there could be fluctuation features characterized by an even greater length scale.

In this work, we presented an efficient semi-numerical model for estimating the large-scale Ly- α opacity fluctuations at late stages of reionization. The motivation behind this model is to provide a reasonably accurate description of forest fluctuations within a large cosmological simulation volume while keeping the small-scale information intact. We took into account these small-scale contributions by properly calibrating the model with respect to a full hydrodynamical simulation. As we discuss in the following, the model incorporated other relevant ingredients for large-scale opacity fluctuations in a self-consistent way. Furthermore, the model also allowed us to provide constraints on background mean photoionization rate and an effective mean free path after comparing with the recent Ly- α opacity measurements (Bosman et al. 2022). The method laid out here is a proof-of-concept study with the aim of efficient EoR inference in the future.

The layout of the paper is organized as follows. In Section 2, we outline the formalism of our modeling along with the calibration procedure. Next, we discuss the basic properties of our model with different variants of model parameters in Section 3. This is followed by a brief summary of available observables and likelihood analysis techniques in Section 4. Then, in Section 5, we lay out our main findings and corresponding interpretation from the parameter space exploration study. Finally, we present our conclusions in Section 6. In this paper, the default assumed cosmological parameters are $\Omega_M = 0.308$, $\Omega_\Lambda = 0.691$, $\Omega_b = 0.0482$, $h = 0.678$, $\sigma_8 = 0.829$ and $n_s = 0.961$ (Planck Collaboration XIII 2016). We primarily assumed h^{-1} cMpc as the distance unit unless otherwise stated.

2. Formalism

In this section, we discuss the methodology for simulating the Ly- α forest optical depth distributions. This includes the calibration of the model and description of the modeling ingredients.

2.1. Computing Ly- α optical depth

In a semi-numerical set up (for examples, Qin et al. 2025; Choudhury & Chakraborty 2025), the optical depth for Ly- α absorption is generally obtained by Fluctuating Gunn Peterson Approximation (FGPA; Weinberg et al. 1998). The approximation provides a relation between the Ly- α optical depth (τ_α) and the underlying density field (Δ) of the IGM i.e.

$$\tau_\alpha \propto \Delta^2 T^{-0.724} \Gamma_{\text{HI}}^{-1} (1+z)^6 H^{-1}(z) \quad (1)$$

where T and Γ_{HI} are the temperature and photoionization rate with $H(z)$ being the Hubble parameter at a redshift represented by z . However, this relation is applicable when we can access reasonably small scales ($\sim 10-20$ Kpc) structures with a very high-resolution simulation. To apply this on coarse resolution semi-numerical simulations, we need to modify the relation after properly calibrating with respect to the high-resolution full hydrodynamical simulation. With the aim of calibration, we assumed that for each coarse cell (i , with $4 h^{-1}$ cMpc resolution in this study), the Ly- α optical depth, τ_i , follows power law relations with different physical quantities at the scale of pixel resolution. These power law indices are then calibrated with the post-processing estimates of high-resolution hydrodynamical simulations. Hence, the modified optical depth for a cell in a fully ionized state (i.e., obeying ionizing equilibrium criterion)

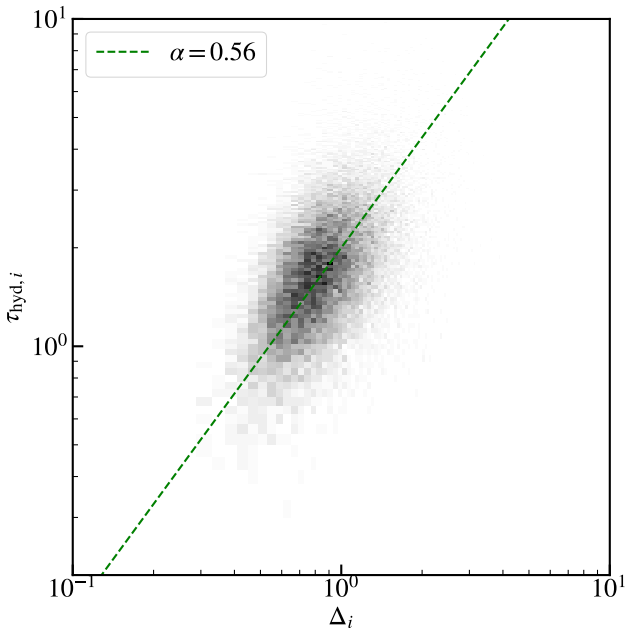


Fig. 1. 2D density plot of optical depth ($\tau_{\text{hyd},i}$) vs density (Δ_i) averaged at the pixel scale ($4 h^{-1} \text{cMpc}$) from the NyX hydro simulation suite at redshift $z = 5.0$. The slope of the correlation matches with $\alpha = 0.56$, which is used to calibrate our semi-analytic model.

can be written as

$$\tau_i = 5.8\kappa \times \Delta_i^{2\alpha} \left(\frac{T_i}{7500} \right)^{-0.724\beta} \left(\frac{2.5 \times 10^{-13}}{\Gamma_{\text{HI},i}} \right)^\delta \left(\frac{1+z}{7} \right)^{6\xi-1.5} \quad (2)$$

where Δ_i , T_i and $\Gamma_{\text{HI},i}$ are the overdensity, temperature and photoionization rate for the i -th cell respectively, with k as a normalization factor. We utilized the full hydrodynamical simulation, Nyx, to estimate the different indices, as we shall discuss in the next paragraph. On the other side, the optical depth for a neutral region can be assumed to be very large, providing essentially zero transmitted flux. Hence, the effective transmitted flux in each cell is given by

$$F_i = x_{\text{HI},i} e^{-\tau_i} + x_{\text{HI},i} e^{-\tau_i^n} \simeq x_{\text{HI},i} e^{-\tau_i} \quad (3)$$

where τ_i^n is the optical depth for the neutral region, which is sufficiently large and can be approximated to ∞ . The ionization fraction of i -th cell is denoted as $x_{\text{HI},i} = 1 - x_{\text{HI},i}$. The effective optical depth of the box was then computed as

$$\tau_{\text{eff}} = \left\langle -\ln \left(\frac{\sum_{i=1}^n F_i}{n} \right) \right\rangle \quad (4)$$

where n is the number of pixels in a single sightline to be averaged over according to the data, and $\langle \dots \rangle$ represents the averaging over all the sightlines. In this study, the averaging length scale was chosen as $\Delta z = 0.1$ to be consistent with Ly- α forest data (Bosman et al. 2022) prepared from the extended XQR-30 sample.

2.2. Calibration with Nyx simulation suite

As mentioned earlier, we utilized a cosmological hydrodynamical simulation, Nyx, to extract the skewers and calibrate our

technique. The simulation description closely follows the previous work by Davies et al. (2024b), which provides sufficient box size ($100 h^{-1} \text{Mpc}$) and resolution (4096^3 dark matter and baryon grids) for converged Ly- α forest statistics. Furthermore, the Ly- α forest opacities are deposited in redshift space using an approximate form for the Voigt profile (Tepper-García 2006) in 2 km/s grid resolution, given the neutral hydrogen field that was estimated assuming ionization equilibrium. We used the skewers from this simulation suite at three redshifts ($z = 5.0, 5.5$ & 6.0) to calibrate the redshift dependence of our empirical model. For each of these redshifts, we extracted 10 000 random skewers of overdensity and temperature fields from the simulation box. Similarly, we utilized different temperature variants of the simulation suites. The optical depth skewers also include peculiar velocity information, although the effect is supposed to be small, and we did not explicitly take into account that effect in the semi-numerical setup (described later). However, due to the large-scale coherence of the velocity field, this effect can be worthwhile to explore in a future study. As this simulation assumes instantaneous reionization, these temperature variations can be obtained by adjusting the reionization end. We utilized the models with sudden reionization end at $z = 6.2, 6.7, 7.7$ & 15 ; providing a range of temperature variations where earlier reionization end corresponds to lower temperature and vice versa. We also created a few different variants of effective mean photoionization rate as well by scaling the optical depth skewers in post-processing, where the opacity is inversely related to the photoionization rate.

We then binned the skewers in 3D at the scale of $\Delta x = 4 h^{-1} \text{cMpc}$, which is suitable for calibrating our empirical model, applicable to implement on efficient semi-numerical techniques. In Figure 1, we show the distribution of optical depth skewers (τ_i) with density (Δ_i) at redshift, $z = 5.0$ for the lowest temperature model. We found a power law correlation between them, which is consistent with $\alpha = 0.56$, which corresponds to a slope of $2\alpha = 1.12$ in the modified FGPA relation (Eq. (2)). We also checked that this slope remains consistent across different redshifts and temperatures. Hence, we fixed this slope in the modified FGPA relation for further calibration. In Figure 2, we show the variations of globally averaged optical depth with respect to the mean photoionization rate ($\langle \Gamma_{\text{HI}} \rangle$, left) and mean temperature ($\langle T \rangle$, right) for the variants of our simulation suite. Here, the points represent the mean differences in optical depth distribution ($\Delta \log \tau$) over all the binned skewers corresponding to the differences in variants of mean photoionization rate ($\Delta \log \langle \Gamma_{\text{HI}} \rangle$) and mean temperature ($\Delta \log \langle T \rangle$) while the error-bars signify 1σ variation in the differences of optical depth distribution among skewers. Again, we found that the trends are consistent with the power law relation. For $\langle \Gamma_{\text{HI}} \rangle$, the relation has a typical slope of $\delta = 0.67$, while the temperature variations provide a typical slope of $\beta = 0.51$ (excluding the point marked in red, corresponding to extreme reionization model ending at $z = 15$, as the temperature relation may not be straightforward for those). As our simulation doesn't include any heating effects (such as X-ray, Ly α) relevant for the cosmic dawn, the temperature estimates from $z_{\text{end}} = 15$ model may be unrealistic. Further, this would set adiabatic cooling as the dominant mechanism at a much earlier redshift, providing an unusually cold IGM at $z < 10$. This cold scenario may already be in tension with the lower redshift IGM temperature estimates (at $z = 5 - 6$) using Ly α forest spike statistics (Gaikwad et al. 2020). The model also provides a high power law index (γ) in temperature-density ($T = T_0 \Delta^{\gamma-1}$) relation (around $\gamma \sim 1.6$), which is again strongly disfavoured by inferred estimates using observations (see Figure 5

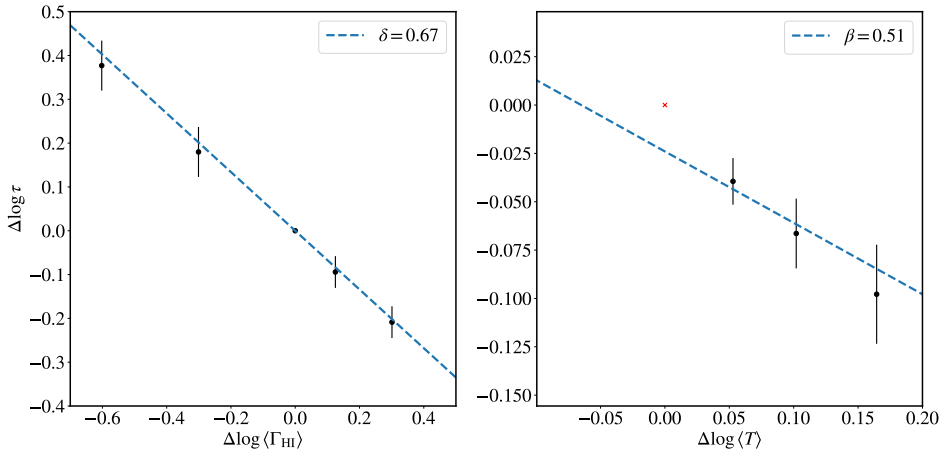


Fig. 2. Global variation of optical depth ($\Delta \log \tau$) at the resolution scale ($4 h^{-1} \text{cMpc}$) for variations in mean photoionization rates ($\Delta \log \langle \Gamma_{\text{HI}} \rangle$, *left*) and for variations in temperatures ($\Delta \log \langle T \rangle$, *right*).

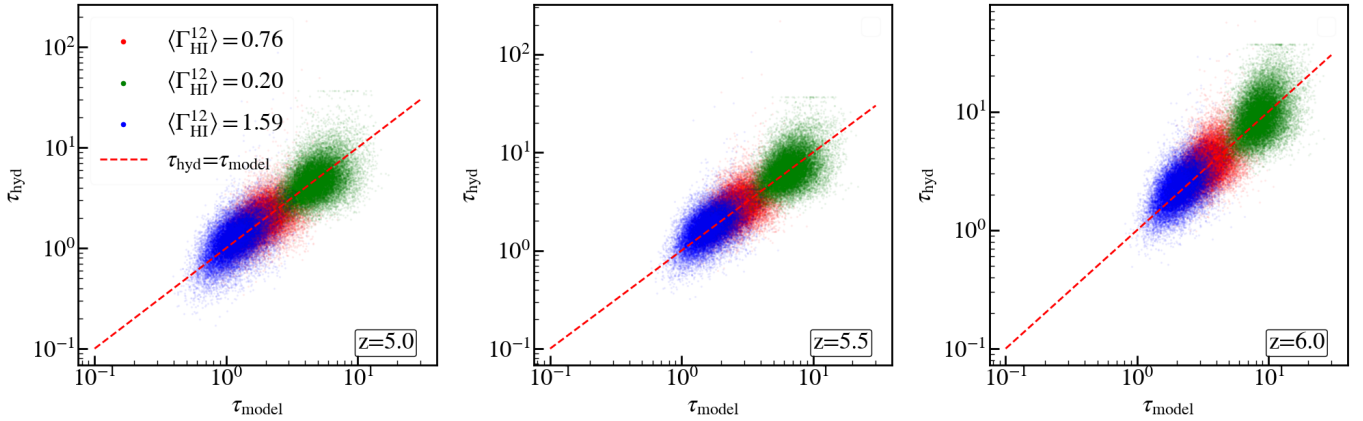


Fig. 3. Comparison of optical depths from hydro simulation (τ_{hyd}) and calibrated empirical model (τ_{model}) for three different redshifts ($z = 5.0, 5.5$ & 6.0) used for calibration. The different colors indicate the set of different values of mean photoionization rates.

for reference). Hence, we excluded that point to keep our conclusion robust against the current observational state. We fixed the β index to the above-mentioned value for further calibration at the pixel level. Next, we estimated the other three parameters (κ , δ & ξ) by fitting the optical depth (τ_i) relation for each pixel using all the simulation suites with variants of redshifts and mean photoionization rate. The bestfit slope δ is again consistent with 0.67, while the bestfit κ and ξ values are 1.28 and 0.86, respectively. In Figure 3, we show the comparison of optical depth distribution from the Nyx simulation and the calibrated model. We found that those are nicely correlated with each other for different redshifts (each panel) and photoionization rates (each color). This gave the confidence to use the relation within a semi-numerical setup. Notably, there is a significant scatter in the relation, which could be included (e.g., Qin et al. 2025). However, we obtained satisfactory results even without that extra scatter, as we discuss in the next section.

2.3. Semi-numerical setup

In this subsection, we briefly describe semi-numerical model setup for computing the different ingredient fields, i.e., density, temperature, ultraviolet background (UVB)/photoionization, for estimating the effective optical depths, which can be further compared against the observational data.

2.3.1. Generating density and collapsed halo field

To generate the cosmological density field, we used the approach adopted in the publicly available semi-numerical reionization model 21cmFAST (Mesinger et al. 2011). The initial conditions on density and velocity fields are generated in Lagrangian space following the gravitational evolution of the fields via Zel'Dovich Approximation (Zel'dovich 1970). This approximation has been proven to be successful in closely resembling the full N-body simulations with less computational cost. We used a simulation of box size $256 h^{-1} \text{cMpc}$, which is sufficiently large to serve the purpose of this study. After having the particle fields, we gridded those at a high resolution ($0.5 h^{-1} \text{cMpc}$) and then averaged them to $\Delta x = 4 h^{-1} \text{cMpc}$, which is the resolution we adopted throughout the study.

Besides, to construct the collapsed halo mass fraction, we adopted the excursion set formalism (ESF-L) in Lagrangian space, following another semi-numerical reionization model AMBER (Trac et al. 2022). To be specific, the formalism uses extended Press Schechter theory, which computes the collapsed mass fraction (f_{coll}) above a minimum threshold halo mass (M_{min}) given the filtered overdensity at a cosmological scale. Here, we used a spherical tophat filter for filtering the overdensity at a certain scale (2cMpc), which provides a good match with the full N-body simulation. At each redshift, we

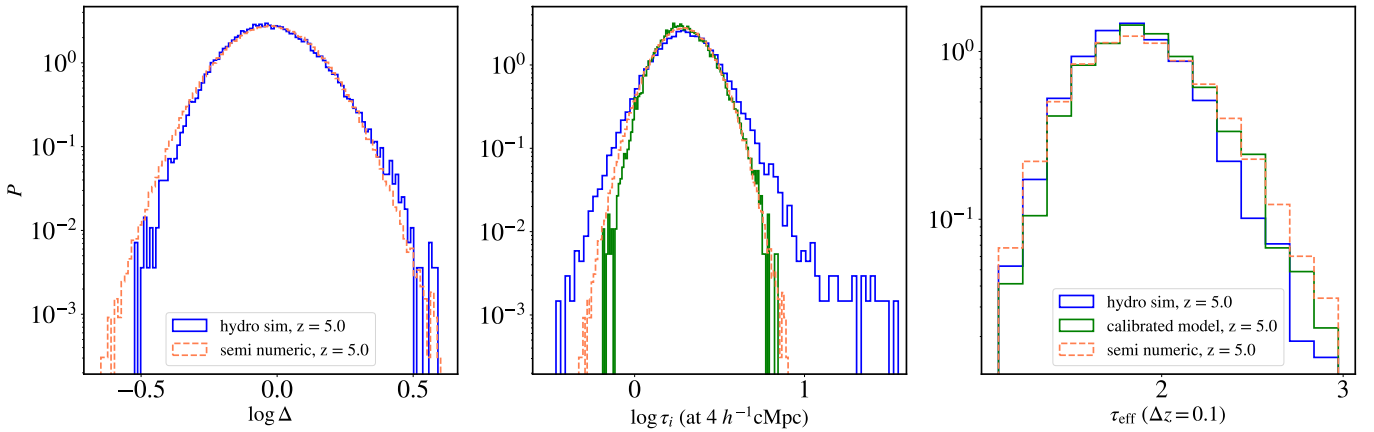


Fig. 4. *Left Panel:* Distribution of densities ($\log \Delta$) from the hydro simulation (blue) and the semi-numerical setup (dashed orange). *Middle Panel:* Distribution of optical depths averaged at $4 h^{-1} \text{cMpc}$ scale (τ_i) for hydro simulation (blue), calibrated empirical relation (green), and calibrated semi-numerical setup (dashed orange). *Right Panel:* Distributions of effective optical depth skewers (τ_{eff}) for the previous three cases.

generated a collapsed fraction field for a logarithmic minimum mass ($\log M_{\text{min}}$) interval of 0.1 in the range 7–12.

In the left panel of Figure 4, we show the overdensity distribution of the Nyx simulation (blue) and the generated overdensities using semi-numerical method (dashed orange) as discussed above. We checked that the natures of these two distributions are in agreement with each other. This resemblance is useful to avoid any discrepancies in optical depth due to the background density field.

2.3.2. Semi-numerical reionization model

We utilized an explicitly photon-conserving semi-numerical reionization model, SCRIPT, which provides the ionization and thermal state of the universe in a cosmologically representative simulation volume (Choudhury & Paranjape 2018; Maity & Choudhury 2022a). To initiate the model, we provided inputs comprising the density field and the distribution of collapsed halos capable of emitting ionizing radiation as generated by the method discussed earlier. For our model to compute the complete reionization history, we utilized comoving simulation boxes spanning from $z = 15$ to $z = 5$, with an interval of $\Delta z = 0.1$. However, for the purpose of this study, we essentially focused on the redshift range $z = 5 - 6.1$.

Our model utilizes the photon-conserving algorithm to construct the reionization topology within the simulation box. The ionization field relies on the ionization efficiency parameter $\zeta(M_h, z)$, which estimates the available ionizing photons per hydrogen atom. This parameter may vary based on halo mass (M_h) and redshift (z). To address inhomogeneous recombinations, we adjusted the ionization criteria to compensate for excess neutral atoms, while introducing small-scale fluctuations through a globally averaged clumping factor, C_{HII} . This parameter remains highly uncertain due to the absence of direct observational measurements. Recently, a claim of a relatively high clumping factor (~ 12) was made via an analytical argument applied to photoionization rate and mean free path measurements (Davies et al. 2024a), although most simulation studies suggest a value around 2–5 (e.g., D’Aloisio et al. 2020; Asthana et al. 2025). In this study, we consistently fixed this parameter at a value of 3. To calculate the recombination number density, it is necessary to consider the thermal

history of the medium. Therefore, alongside tracking the ionization process, we also simulated the temperature evolution within each grid cell of our computational volume. The code is designed to automatically factor in the effects of spatially inhomogeneous reionization on temperature evolution. Specifically, when a region becomes ionized for the first time, its temperature is assumed to increase by a specific amount, denoted as T_{re} (Hui & Gnedin 1997; Furlanetto & Oh 2009; Keating et al. 2018; Maity & Choudhury 2022a).

Additionally, our model accounts for radiative feedback, which inhibits the production of ionizing photons in halos that experience gas heating. We followed the ‘step feedback’ framework introduced in Maity & Choudhury (2022a), where halos below a certain mass threshold, defined as $M_{\text{min}} = \text{Max}[M_{\text{cool}}, M_J]$, are assumed to retain no gas, while those above retain all their gas. Here, M_{cool} is the atomic cooling threshold mass, and M_J is the Jeans mass at virial overdensity. This simplified step-function approach to radiative feedback is both efficient and effective for modeling reionization. Importantly, M_{min} varies with redshift and across spatial locations, especially in ionized zones where temperature-driven feedback becomes significant. The Jeans mass, being temperature-dependent, contributes to this variability. In contrast, the neutral regions are not impacted by feedback, where M_{min} aligns with M_{cool} , which is roughly $\sim 3 \times 10^7 M_{\odot}$ at $z = 20$ and $\sim 2 \times 10^8 M_{\odot}$ at $z = 5$. In feedback-affected zones, M_{min} typically surpasses $10^9 M_{\odot}$.

2.3.3. Model for UVB fluctuations

The fluctuations in HI photoionization rate are a crucial component in determining the Lyman- α optical depth during the later stages of the reionization era. These mainly arise due to the discrete nature and clustering of ionizing sources, which remains even after ionization equilibrium is established in the post-ionization phase (Mesinger & Furlanetto 2009). However, this is hard to accurately model within the standard semi-numerical reionization framework due to the lack of small-scale information.

In this work, we deployed the model developed by Davies & Furlanetto (2016), which incorporates the effect of fluctuating mean free path on the photoionization rate. The photoionization rate in each cell (i) is computed by integrating the

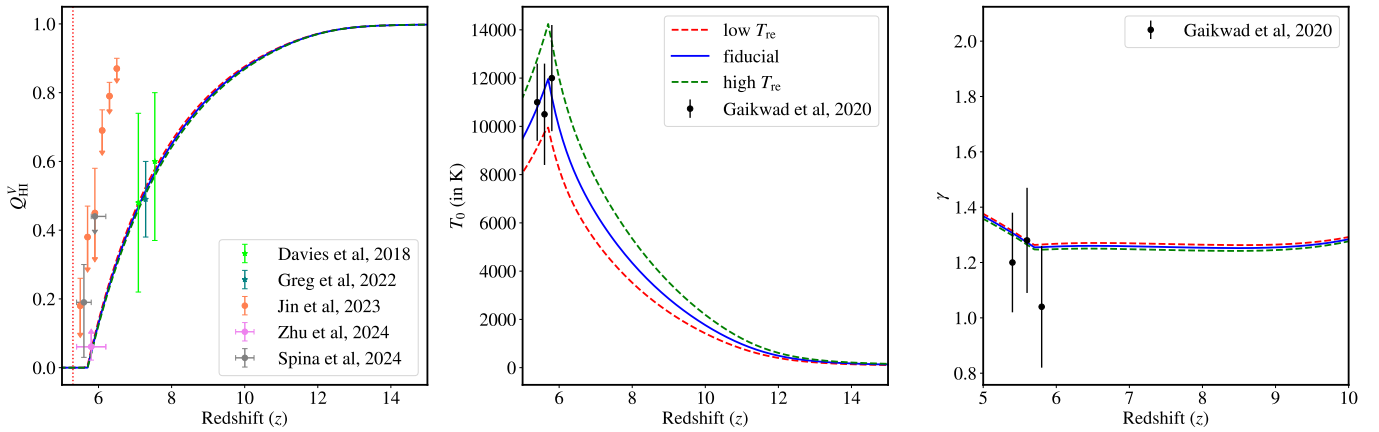


Fig. 5. Neutral fraction (Q_{HI}^V), mean IGM temperature (T_0), and $T-\Delta$ power law index (γ) evolution with redshift for our fiducial reionization model (blue solid lines) along with other two variants using different reionization temperature increment, T_{re} (high: green dashed; low: red dashed). We also show the recent observational constraints on the neutral fractions (Davies et al. 2018b; Greig et al. 2022; Zhu et al. 2024; Spina et al. 2024) and IGM temperatures (Gaikwad et al. 2020).

product of specific ionizing intensity (including local, $J_{\nu,i}^L$ and non local, $J_{\nu,i}^{\text{NL}}$ contributions) and interaction cross section (σ_{HI}) across UVB photon frequency (ν), i.e.

$$\Gamma_{\text{HI},i} = 4\pi \int_{\nu_{\text{HI}}}^{4\nu_{\text{HI}}} \left[\frac{J_{\nu,i}^L + J_{\nu,i}^{\text{NL}}}{h\nu} \right] \sigma_{\text{HI}}(\nu) d\nu. \quad (5)$$

This integration over the frequency makes the computation slightly inefficient, although it has been checked that a single frequency ($\nu \sim 1.32\nu_{\text{HI}}$) in the middle provides a good approximation (Davies & Furlanetto 2016), which is sufficient for this study. To compute this relation, we need to estimate the non-local contribution, which is given as

$$J_{\nu,i}^{\text{NL}} = \sum_{j \neq i}^N \frac{L_{\nu,j}}{4\pi r_{ij}^2} \exp \left[-\int_{r_j}^{r_i} \frac{dx}{\lambda(x)} \right] \quad (6)$$

where r_{ij} is the distance between i -th and j -th cells; $\lambda(x)$ is the mean free path at each position along the sightlines, and $L_{j,\nu}$ is the specific luminosity on the corresponding (j -th) cell. This requires the mean free path variation for each cell that is computed as

$$\lambda_{\nu}(x) \propto \lambda_0 \Delta^{-1}(x) \Gamma_{\text{HI}}^{2/3}(x) \nu^{0.9}. \quad (7)$$

Here, λ_0 gives a proxy of the averaged mean free path, which can be treated as a free parameter in the model, and Δ is the overdensity distribution. The dependencies on the different physical parameters via the above-mentioned power law indices come from realistic assumptions about absorbing gas density distribution; however, small variations in these indices do not change the fluctuations significantly (Gaikwad et al. 2023). On the other hand, the local contribution takes into account the intensity budget within a cell, which is estimated using an approximation (Davies & Furlanetto 2016) to get a resolution-independent average photoionization rate, i.e.

$$J_{\nu,i}^L \approx \frac{\epsilon_{\nu,i} \lambda_{\nu,i}}{4\pi} \left(1 - e^{-0.72\Delta x / \lambda_{\nu,i}} \right) \quad (8)$$

with Δx is the resolution of the cell and $\epsilon_{\nu,i}$ is the specific emissivity. Equations (5) and (6) can be solved iteratively in each cell to get the UVB fluctuations field.

2.3.4. Comparison with hydro skewers

In the middle and right panels of Figure 4, we show the optical depth distribution (τ_i) at pixel scale ($4 h^{-1} \text{cMpc}$) and effective optical depth (τ_{eff} , averaged at $\Delta z = 0.1$), respectively, at a fiducial redshift, $z = 5.0$. The distribution of opacities from hydro skewers is shown in blue, the optical depths estimated using our calibrated empirical relation on hydro density and temperature field are shown in green, while the estimates from the semi-numerical methods (the fiducial model has been discussed separately in the next section) are in dashed orange. We found that the fluctuations at the pixel scale are slightly lower for the semi-numerical setup than the full hydro simulation, which is not very surprising due to the involvement of complex astrophysical processes in those simulations and the fact that we do not introduce the additional scatter in the modified FGPA relation. However, the τ_{eff} distributions are very similar to each other, providing a check on the robustness of the calibration procedure. One possible reason behind these findings may be the contribution of large-scale correlations in peculiar velocities, which are present in the hydro simulation, which we neglect in the semi-numerical model. The peculiar velocity component is ideally a part of the FGPA relation that can suppress the scatter in optical depth while averaging along the line of sight. However, this needs to be investigated in a separate study later, following a proper comparison of velocity fields between the semi-numerical model and hydrodynamical simulation.

3. Demonstration with the fiducial model

In this section, we demonstrate the features of our model assuming a fiducial reionization model. We chose a fiducial model from the allowed posterior parameter spaces satisfying a variety of available observational constraints (Maity & Choudhury 2022b), including CMB scattering optical depth, UV luminosity functions, and low-density IGM temperature estimates. In Figure 5, we show the evolution of neutral fractions (reionization history), mean temperature evolution, and power law index (γ) evolution of $T-\Delta$ relation for our fiducial model. We show the recent observational constraints on these quantities as well, i.e., neutral fractions (Davies et al. 2018b; Greig et al. 2022; Jin et al. 2023; Zhu et al. 2024; Spina et al. 2024) and IGM temperature

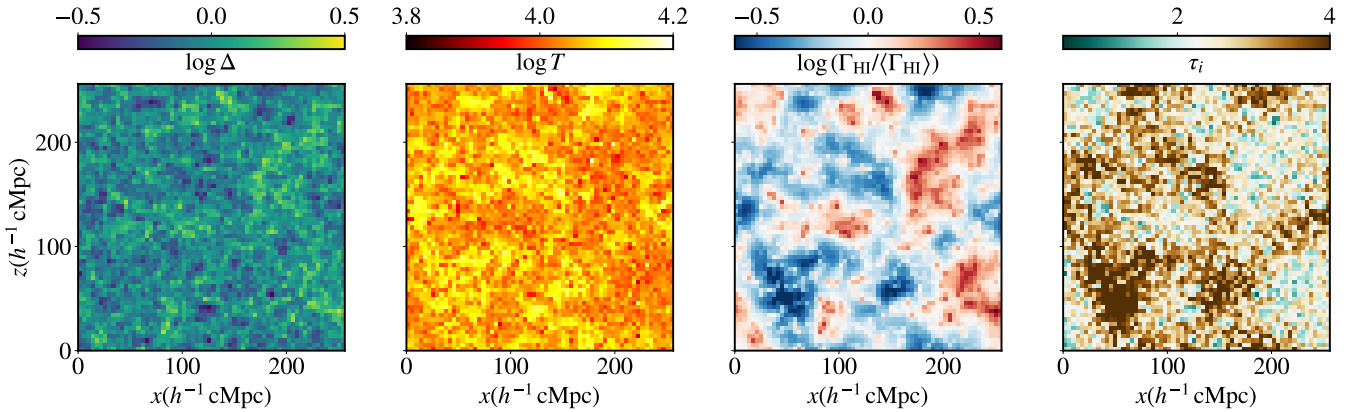


Fig. 6. Snapshots of density (Δ), temperature (T), UVB fluctuations ($\Gamma_{\text{HI}}/\langle\Gamma_{\text{HI}}\rangle$) and Ly- α optical depth (τ_i), gradually from left to right using a fiducial setup of the calibrated semi-numerical model at redshift, $z = 5.4$.

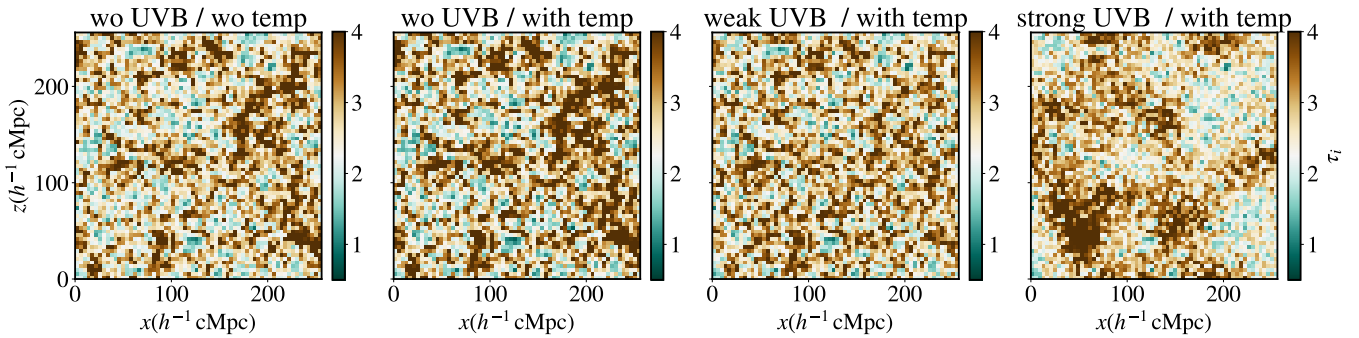


Fig. 7. Snapshots of Ly- α optical depth for different combinations of temperature and UVB fluctuation inputs. From left to right, we show the cases gradually where we include (i) no UVB/temperature fluctuation; (ii) only temperature fluctuation but no UVB fluctuations; (iii) weak UVB fluctuations with temperature fluctuations; and (iv) strong UVB fluctuations with temperature fluctuations.

(Gaikwad et al. 2020). It is apparent that the fiducial model obeys all the constraints reasonably well. We also show two variants of the model using different temperature increments (T_{re}) values representing two extremes. We chose these models such that they just become consistent with the observational constraints on mean IGM temperature from the two ends. Later, we used these models to marginalize over different allowed temperatures while estimating the photoionization rate and mean free path.

Next, we studied the effects of varying the parameters that are responsible for generating the UVB fluctuations. In Figure 6, we show the snapshots of various physical quantities for the fiducial model at redshift $z = 5.4$. These are density (Δ), temperature (T), UVB fluctuations ($\Gamma_{\text{HI}}/\langle\Gamma_{\text{HI}}\rangle$) and the Ly- α optical depth (τ) from left to right respectively. The correlation between different fields is quite apparent from the snapshots. The overdense regions are ionized at a relatively earlier time, allowing more time for cooling. Hence, the temperatures in those regions are comparatively lower. The photoionization rate, on the other hand, is higher for high-density regions. This is expected as high-density regions naturally host galaxies and stars, which are supposed to be dominant sources of ionizing photons. In this fiducial scenario, we have used $\lambda_0 = 10 \text{ h}^{-1} \text{ cMpc}$ for the UVB field. For this small mean free path, the opacity fluctuations are mainly dominated by the ionization background fluctuation. However, the strength of this fluctuation can vary depending on the mean free path parameter, which we discuss later in this paper.

In Figure 7, we demonstrate the effect of UVB and temperature fluctuations on the Ly- α opacities. In the left panel, we show the case where no UVB or temperature fluctuations are included, and the opacities are driven solely by the density field. We include temperature fluctuations in the next panel, which increases the fluctuations as overdense regions become more opaque due to lower temperature and vice versa. In the third panel, we incorporate a weakly fluctuating UVB field, which slightly neutralises the fluctuations. This happens as the high-density regions have higher photoionization rates, contributing to lower opacities, which further results in smoothing the optical depth field. At the rightmost panel, the fluctuations are again increased due to a strong UVB field. In Figure 8, we show the cumulative probability distributions (CDF) of the effective optical depth (τ_{eff}) for different mean photoionization rate (in $10^{-12}/s$ i.e. $\langle\Gamma_{12}^{\text{HI}}\rangle$, at left) and effective mean free path (λ_0 , at right). It can be seen that the mean photoionization rate mainly shifts the mean value of τ_{eff} while λ_0 contributes to the amount of fluctuation in the distribution. A lower value of the mean photoionization rate corresponds to a higher effective optical depth distribution and vice versa. This is also expected from equation 2, where these two quantities are inversely related to each other. Physically, the photoionization rate also controls the transmission flux; namely, a higher photoionization rate provides a higher transmission, which corresponds to a lower optical depth. On the other hand, the mean free path essentially provides an estimate of how far a photon can travel before getting absorbed by

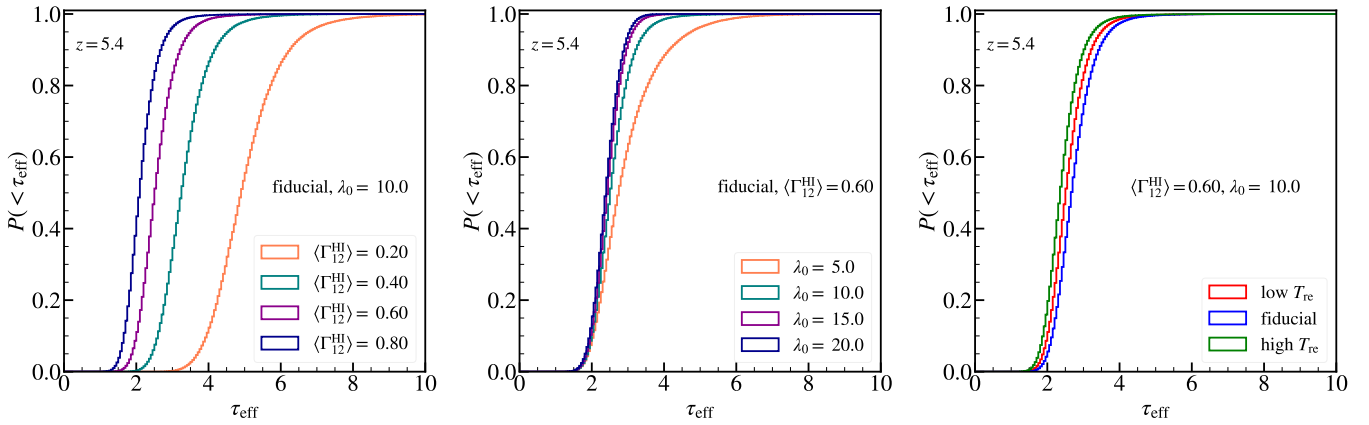


Fig. 8. Cumulative probability distributions (CDFs) of τ_{eff} at redshift, $z = 5.4$ for a variety of parameter values. The left panel shows the variation of CDF for different mean photoionization rate (in $10^{-12}/s$, i.e., $\langle \Gamma_{12}^{\text{HII}} \rangle$). The middle panel shows the CDF variation with effective mean free path (λ_0 in $h^{-1} \text{cMpc}$). The right panel shows the CDF variation with different reionization temperature increment (T_{re}).

the residual neutral gas. Hence, a higher value of λ_0 results in a smoother transmission with a lower amount of fluctuations due to the absorption, which further narrows down the optical depth distribution.

4. Observational data and likelihood analysis

We utilized the high quality Ly- α forest data from 67 high-redshift quasars following [Bosman et al. \(2022\)](#). This sample includes 51 optical quasar spectra from VLT X-shooter spectrograph, among which 25 come from the XQR-30 program ([D’Odorico et al. 2023](#)) and 26 come from archival X-shooter data. The remaining 16 spectra were taken with the Keck ESI spectrograph. All these high redshift (at a redshift $z \geq 5.6$) quasars spectra with a signal-to-noise ratio of $S/N \geq 10$ per pixel provide an ideal dataset to probe the late phase of the reionization epoch. However, the quasar continuum is very uncertain at these redshifts due to the lack of enough pixels to recover the continuum flux. Hence, to compute the effective optical depth, a novel principal component analysis (PCA) method was followed in recent studies ([Davies et al. 2018c](#); [Bosman et al. 2021](#)). Then the effective optical depths were computed with the bins of $\Delta z = 0.1$ ([Bosman et al. 2022](#)).

In our analysis, the observational uncertainties and wavelength-dependent continuum uncertainties are taken into account via forward modeling. To be specific, the measurement uncertainties are added with each realization of the model skewers according to the data, while the continuum uncertainties are multiplicative. The sightlines for which no significantly detected transmission is available, are treated as optimistic limits, i.e., assumed to be 2σ where σ is the data uncertainty corresponding to that sightline. In a pessimistic scenario, the flux could be assumed to be zero, producing an infinitely opaque sightline (following [Bosman et al. 2022](#)).

Given the above dataset, we used the standard Bayesian approach for parameter space exploration. Specifically, we computed the conditional probability distribution or the posterior $\mathcal{P}(\theta|\mathcal{D})$ of the model parameters θ , provided the observational data sets \mathcal{D} as mentioned in the previous paragraph. This can be estimated using Bayes’ theorem as

$$\mathcal{P}(\theta|\mathcal{D}) = \frac{\mathcal{L}(\mathcal{D}|\theta) \pi(\theta)}{\mathcal{P}(\mathcal{D})}, \quad (9)$$

where $\mathcal{L}(\mathcal{D}|\theta)$ is the conditional probability distribution of data given the parameters or the likelihood, $\pi(\theta)$ is the prior and $\mathcal{P}(\mathcal{D})$ is the evidence (which can be treated as the normalization parameter and does not play any role in our analysis).

However, the raw dataset can not be used for likelihood analysis due to the dominance of Poisson noise arising from its discrete nature. To this end, different summary statistics, such as the cumulative distribution function (CDF), mean fluctuations, and so on, can be used for inference studies. Such studies would be based on the assumption that the posterior $\mathcal{P}(\theta|\mathcal{D}) \approx \mathcal{P}(\theta|\mathcal{D}_S)$, where \mathcal{D}_S is the corresponding statistics derived from data. Among these, probability distribution functions (PDFs) fitted with Gaussian kernel density estimator (KDE) are commonly used ([Gaikwad et al. 2021](#); [Choudhury et al. 2021](#)) for robust likelihood inference. Here, we adopted a similar approach and evaluated the KDE distribution into 15 bins to compute the likelihood. We did the exact same procedure with both the observational data and our models. We kept the number of sightlines same as the data. Once we had the binned probability distribution, we computed the logarithm of the likelihood (\mathcal{L}) distribution as

$$\begin{aligned} -2 \ln \mathcal{L} &= \chi^2 \\ &= \sum_{i,j} [p_{\mathcal{D}}(\tau_{\text{eff}}) - p_{\mathcal{M}}(\tau_{\text{eff}})]_i \Sigma_{ij}^{-1}(\mathcal{M}) [p_{\mathcal{D}}(\tau_{\text{eff}}) - p_{\mathcal{M}}(\tau_{\text{eff}})]_j \end{aligned} \quad (10)$$

where $p_{\mathcal{D}}(\tau_{\text{eff}})$ is the observed KDE PDF of effective optical depth and $p_{\mathcal{M}}(\tau_{\text{eff}})$ is the corresponding model KDE PDF. We also took into account the correlations between the bins by computing the covariance matrix $\Sigma_{ij}(\mathcal{M})$, which is model-dependent. This was computed using 1000 random independent realizations of the sightlines.

In the left panel of Figure 9, we show the probability distributions of the data along with the fiducial model at redshift, $z = 5.4$. The blue histogram represents the effective optical depth distribution according to the original data, while the dashed black line is achieved after applying the Gaussian KDE on it. The estimation from the model is shown in red, which seems to match the data KDE reasonably well. In the right panel, we show the normalized covariance matrix for this model. Notably, our covariance matrices are dependent upon the model, which then goes into the computation

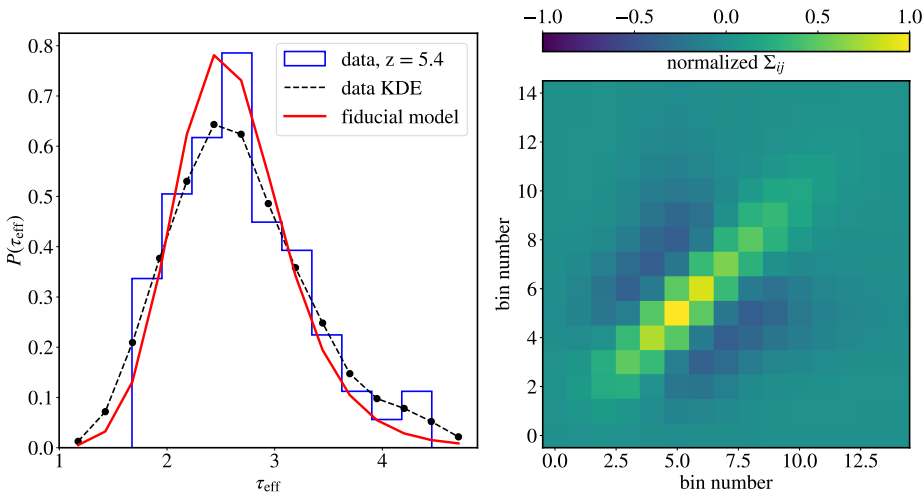


Fig. 9. *Left Panel:* Example of probability distribution functions (PDFs) from the data and the fiducial model at redshift $z = 5.4$. The blue line shows the histogram obtained from the data, while the dashed black line shows the fit using Gaussian KDE. Similarly, the red line shows the distribution obtained from our fiducial model. For the likelihood analysis, we use 15 equally spaced bins as marked by the black solid points. *Right Panel:* Normalized covariance matrix at $z = 5.4$.

of the likelihood for the Markov chain Monte Carlo (MCMC) sampling.

Next, we sampled the posterior distribution using the MCMC method; more specifically, the Metropolis-Hastings algorithm (Metropolis et al. 1953). We utilized the publicly available package *cobaya* (Torrado & Lewis 2021)¹ to run the MCMC chains. We checked the convergence of the chains following the Gelman-Rubin $R-1$ statistic (Gelman & Rubin 1992). The chain was assumed to be converged if the $R-1$ value became less than a threshold 0.01.

5. Results and interpretations

In this section, we discuss the main results from our parameter space exploration studies using the mean photoionization rate ($\langle \Gamma_{\text{HI}} \rangle$) and the effective mean free path (λ_0) as free parameters.

5.1. Cumulative probability distribution of τ_{eff}

We display the cumulative probability distributions (CDFs) of the effective optical depth (τ_{eff}) in Figure 10, for the redshift range of our interest ($z = 5.0 - 6.1$) in this study. The red lines show the bestfit CDF obtained by combining 500 random realizations of the sightlines, while the number of sightlines is kept the same as provided by the observational data. The cyan curves in the background represent the individual realizations of the bestfit model. For the data (shown as dashed black lines), we show both optimistic and pessimistic limits, although we only used the optimistic case for the parameter space exploration. We checked that the bestfit evolution model can nicely match the observed scatter in the optical depth distribution as well as the mean transmission flux in the majority of the cases with mild deviation at $z = 5.6$ and 5.9 . The deviation is mainly due to the dominance of τ_{eff} values at the distribution of the edges in the data for those redshifts. Furthermore, the width of the distribution increases as we go towards higher redshifts, suggesting increasing fluctuations in the opacity fields.

5.2. Joint parameter posteriors

Next, we quantified the posterior distribution of the parameters extracted from three different temperature models allowed by the

current constraints on the IGM temperatures, as shown earlier in Figure 5. We performed this analysis for all the redshifts of interests and show the combined posteriors in Figure 11 while in Figure 12, we show the 2D posterior contours (at 68% and 95% confidence level) for the individual temperature models along with the combined contour of all the three samples again at the fiducial redshift, $z = 5.4$. It is apparent that the combined contour nicely represents the whole posterior space allowed by the temperature constraints. Now, moving the discussion to Figure 11 for different redshifts, it can be seen that the data are consistent even with almost uniform UVB models towards low redshifts ($z \leq 5.3$), demanding only a small amount of UVB fluctuations. This interpretation is driven by the fact that, in our definition, post-reionization UVB fluctuations are solely quantified by the mean free path parameter λ_0 . The large parameter values correspond to a relatively uniform UVB field and vice versa. Since, at those redshifts, all the posteriors hit the upper end of the prior values on λ_0 , we can conclude that almost uniform UVB models are also allowed by the data. An interesting feature can be noticed at $z = 5.3$ where a bimodality in mean free path distribution arises disfavoring some of the moderate UVB fluctuation models ($\lambda_0 \sim 20 - 30 h^{-1} \text{cMpc}$). This happens because those models are not able to provide enough fluctuations to match the data distribution. For a large mean free path, the opacity fluctuations are driven by the underlying density field and temperature fluctuations. As the mean free path decreases, it initially neutralizes the opacity fluctuations as demonstrated in Figure 7, which corresponds to the disfavored region in the posterior. This happens because high-density regions are ionized earlier and have relatively lower temperatures at later times due to the availability of more time for cooling (D'Aloisio et al. 2015). This enhances the opacity fluctuations correlating with density fluctuations. However, the correlation between density and UVB field is opposite, providing high photoionization rate at high-density regions and vice versa, which further drives the fluctuations in the other direction (see also; Davies et al. 2018a). Then, with a reasonably short mean free path, the opacity fluctuations are dominated by the UVB field itself. Hence, the bimodal feature portrays the interplay among these competing fluctuations, generating the opacity field.

As we move towards higher redshifts (up to $z = 5.8$, where reionization ends in our model), the required UVB fluctuations also increase, allowing for shorter mean free paths. Before reionization ends ($z \geq 5.9$), the ionization fluctuations (driven

¹ <https://cobaya.readthedocs.io/en/latest/>

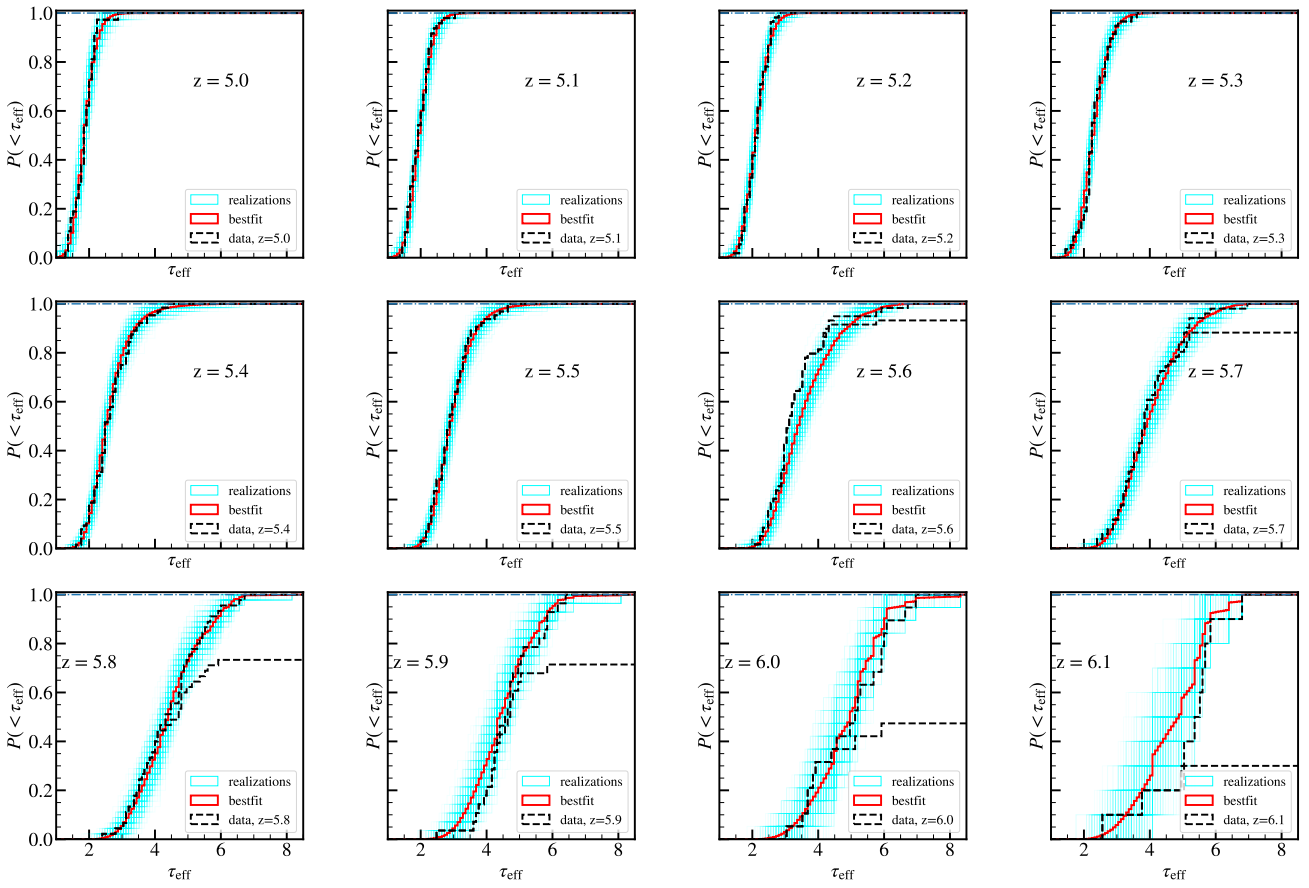


Fig. 10. CDF plots of τ_{eff} for different redshifts ($z = 5.0 - 6.1$). Each panel shows the combined CDFs of the bestfit model using 500 random realizations (red) along with the individual realization (cyan) and observational data (black). We show both the optimistic and pessimistic scenarios for the data.

by leftover neutral islands) are strong enough to produce the required large-scale UVB fluctuations matching the observed opacity fluctuations, hence requiring little or almost no fluctuations driven by post-reionization λ_0 . This further keeps the mean free path parameter unconstrained. This explanation can be clearer from Figure 13, where we show two snapshots of UVB fluctuations. Among these two snapshots, one is at redshift, $z = 5.8$ (after reionization completion) with a very short mean free path ($\lambda_0 = 4 h^{-1} \text{cMpc}$) and the other one is at redshift, $z = 5.9$ (before reionization completion) with a relatively high mean free path parameter ($\lambda_0 = 40 h^{-1} \text{cMpc}$), as suggested by the posteriors. It is apparent that the presence of neutral regions can produce strong fluctuations in the UVB field even if λ_0 is larger (see also, Choudhury et al. 2021). This supports our above argument for an unconstrained mean free path at higher redshifts where reionization is still not ended. Therefore, despite the fact that the λ_0 posteriors look very similar to the cases at $z \leq 5.3$, the UVB fluctuations are not uniform here (at $z = 5.9 - 6.1$), and, rather, they are driven by the leftover neutral islands. This analysis provides us with the confidence that our simplistic model can be utilized to pursue parameter space statistics study and interpret the relevant physics from the observational data.

5.3. constraints on $\langle \Gamma_{12}^{\text{HI}} \rangle$ and λ_0

In Figure 14, we show the marginalized constraints on the mean photoionization rate ($\langle \Gamma_{12}^{\text{HI}} \rangle$, at the left panel) and effective mean

free path (λ_0 , at the right panel) at 95% confidence level along with the points from recent literature. Our estimates are denoted with red circles, where solid ones indicate the redshifts where reionization has completed, and hollow ones indicate the redshifts with incomplete reionization. We found that our photoionization rate estimates are consistent with the findings of Davies et al. (2024b, shown in blue stars with displaced redshift for better visualization). This is not very surprising, as the UVB fluctuation generation model is the same for both cases. However, our points are slightly higher as we have a realistic reionization history allowing sufficient time for cooling in the early ionized high-density regions. On the other hand, Davies et al. (2024b) assumed a sudden reionization for all the cells in the simulation box, which results in larger temperature and hence, a slightly shorter photoionization rate to match the mean transmission flux. Additionally, our estimates are also in agreement with the corresponding constraints of Gaikwad et al. (2023, shown in green triangle, again shifted in redshift).

In the case of the mean free path, we are able to derive the constraints from posteriors at redshift range $z = 5.4 - 5.7$. For the redshifts providing constraints from only one side, we quote the value where the posterior value falls down to $1/e^2$ from the maximum (Davies et al. 2024b). We do not get any constraints at $z = 6.0$. Our constraints prefer a shorter mean free path distribution than Davies et al. (2024b) values, although these are consistent within the error bars, particularly given that the lower end of the Davies et al. (2024b) uncertainties corresponds to an

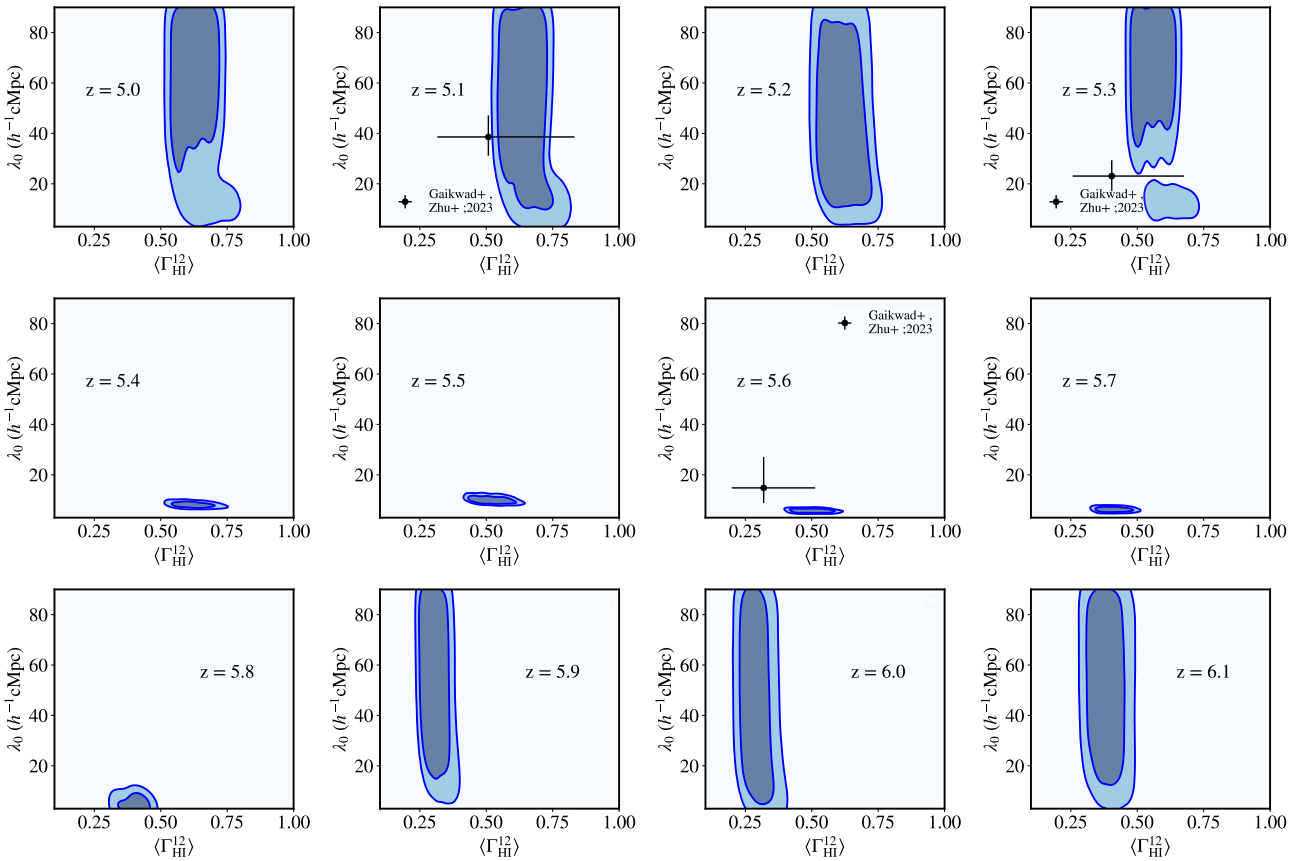


Fig. 11. Joint 2D posterior distribution of effective mean free path (λ_0) and mean photoionization rate ($\langle \Gamma_{\text{HI}}^{12} \rangle$) for different redshifts (from $z = 5.0$ in the top-left to $z = 6.1$ in the bottom-right).

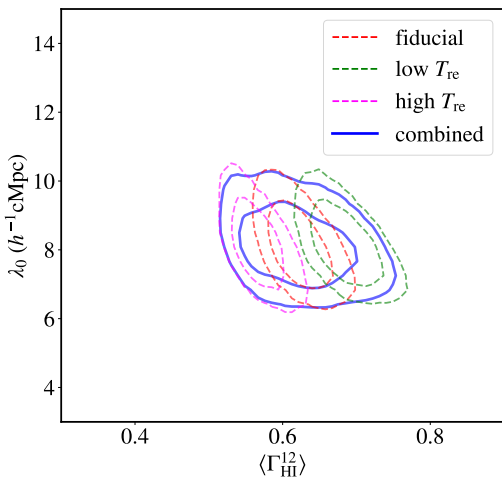


Fig. 12. 2D posterior distribution between λ_0 and Γ_{HI} for different temperature models individually, along with the combined contours. The contours are shown at 68% and 95% levels.

approximate correction to account for their lack of model self-consistency. The main reason for this difference is again traced back to the assumed sudden reionization and the lack of post-reionization large-scale temperature fluctuations in their model. It is expected that the large-scale temperature field is anticorrelated with the density field due to the variations in reionization

timing at different regions. As a result of that, a stronger UVB fluctuation is required to overcome the correlated fluctuations of thermal and density field, which in turn demands a lower λ_0 value. Intriguingly, this more physical treatment with a gradual and patchy thermal evolution puts our model constraints within a similar ballpark of recent high-redshift mean free path measurements (Becker et al. 2021; Zhu et al. 2023). At this point, it is important to highlight that the definition of our mean free path is similar to that of Davies et al. (2024b), which is solely a subgrid parameter quantifying clumping of the unresolved gas. Hence, our inferred mean free path only characterizes the excess UVB fluctuations over the uniform case without including density field fluctuations. As a consequence, we do not get any upper bound on λ_0 for $z \leq 5.3$ as an almost uniform UVB field is consistent with fluctuations required by the data. This is unlike the approach adopted by Gaikwad et al. (2023), where the mean free path is computed along with their density field after including fluctuating UVB contribution. That further enables them to put constraints on the mean free path even without requiring explicit UVB fluctuations. At redshift $z \geq 5.9$, reionization is incomplete in our model, which further allows us to provide only lower limits on the mean free path, as demonstrated earlier in this paper (Section 5.2). The constraints used in the plots are tabulated in Table 1.

In continuation with the above discussion, it is worthwhile mentioning that most of the semi-numerical models give a sharp transition to the reionization end, while detailed numerical simulations (Kannan et al. 2022; Asthana et al. 2025) prefer relatively smooth transitions along with slight inflection in

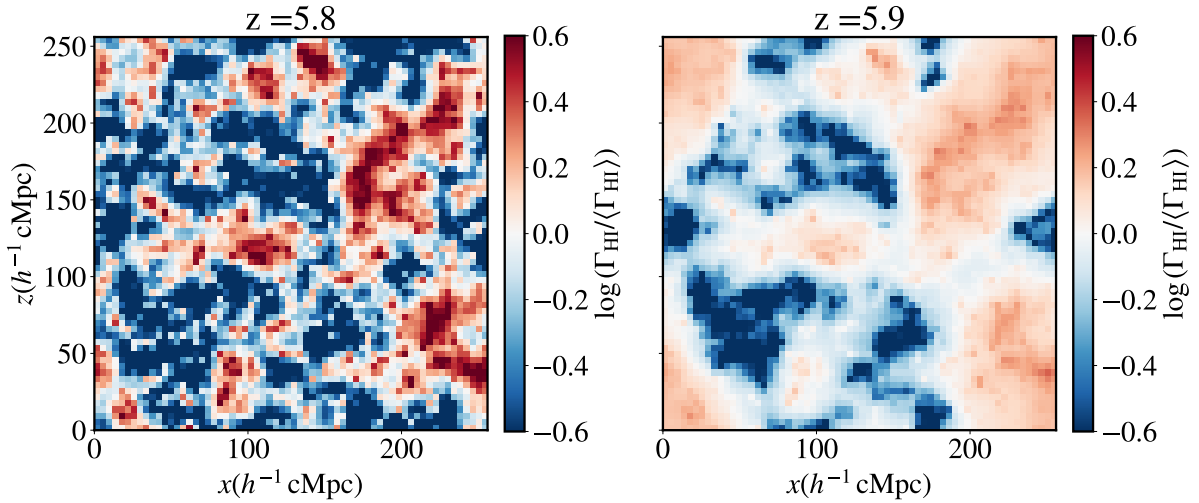


Fig. 13. Comparison of UVB fluctuation fields allowed by the posteriors at redshifts, $z = 5.8$ (reionization has ended) and $z = 5.9$ (reionization is incomplete). The mean free path requirement for $z = 5.8$ ($\lambda_0 = 4 h^{-1} \text{cMpc}$) is much smaller than $z = 5.9$ ($\lambda_0 = 40 h^{-1} \text{cMpc}$), showing the effect of incomplete ionization on the UVB field.

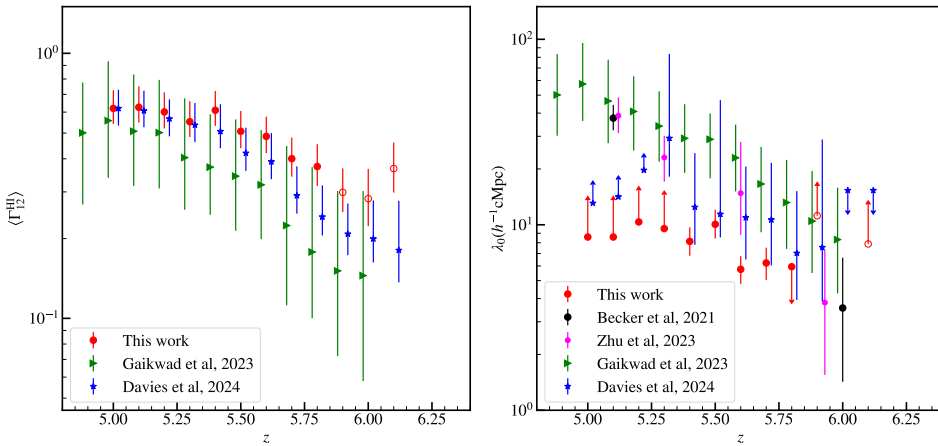


Fig. 14. Marginalized 1D posterior constraints for the mean photoionization rate ($\langle \Gamma_{\text{HI}}^{\text{HI}} \rangle$) and the effective mean free path (λ_0) at different redshifts. The constraints correspond to 95% confidence limits where the constraints are available from both sides. For the limits, we quote the value where the posterior falls down to $1/e^2$ from the maximum. The red circles denote the constraints derived from this work (hollow circles represent the redshift where reionization is not complete). The redshifts for green and blue points have been shifted by 0.02 dex for visual clarity.

neutral fraction evolution. A gradual transition will translate to a smoother evolution in mean free path, as hinted from our findings. A smooth transition towards reionization end can be achieved within a semi-numerical setup with a more realistic modeling of photon sinks (e.g. [Davies & Furlanetto 2022](#); [Qin et al. 2025](#)). This could provide a useful direction for potential improvements to our model in the future.

6. Summary and conclusions

The large-scale fluctuations of Ly- α forest spectra observed in high-redshift quasars contain crucial information about the physical properties of the IGM. The modeling of these fluctuations becomes challenging due to the effects of various physical quantities at different scales. Ideally, we would require an extremely high dynamic range simulation to accurately characterize the fluctuations. However, this becomes computationally intractable given a wide range of uncertain astrophysical parameters. To this end, fast semi-numerical approaches can offer an alternate way to achieve some degree of optimization between speed and accuracy.

In this work, we developed an efficient semi-numerical model to estimate the Ly- α opacity fluctuations during the late and post-EoR phase. We further compared our model with the available observational data, aiming to constrain the model parameters. Below, we summarize the key points of this work.

- We calibrated our model with respect to high-resolution full hydrodynamic Nyx simulation. This essentially modified the fluctuating GP relation accordingly to be applicable on coarse resolution grid cell (in our case $4 h^{-1} \text{cMpc}$). Our model self-consistently includes the relevant ingredients (i.e., density, temperature, and UVB fluctuations) for generating the opacity fluctuations. We checked that the opacity distribution using the calibrated model can provide a reasonable match with the full hydro skewers. The small differences may arise due to the lack of velocity field correlation in our model. However, this effect is unlikely to be strong and can be explored in a separate future project.
- Next, we demonstrated the capabilities of our model assuming fiducial ionization and thermal histories that are consistent with the available constraints. We found that, along with the underlying density field, large-scale temperature

Table 1. Parameter constraints obtained from the MCMC-based analysis after combining different temperature models. For each redshift, we show the mean value of combined samples along with 95% confidence limits on different parameters where constraints are achieved from both sides. We quote the value where the posterior value falls down to $1/e^2$ from the maximum (Davies et al. 2024b) for the one-sided limit.

Redshifts (z)	$\langle \Gamma_{\text{HI}}^{12} \rangle$		$\lambda_0 (h^{-1} \text{cMpc})$	
	Mean	95% Limits	Mean	95% Limits
5.0	0.62	[0.54, 0.73]	–	>8.59
5.1	0.63	[0.55, 0.75]	–	>9.12
5.2	0.60	[0.52, 0.71]	–	>10.36
5.3	0.55	[0.48, 0.66]	–	>9.54
5.4	0.61	[0.53, 0.72]	8.14	[6.81, 9.69]
5.5	0.51	[0.43, 0.66]	10.05	[8.44, 12.05]
5.6	0.49	[0.42, 0.58]	5.75	[4.80, 6.76]
5.7	0.40	[0.34, 0.48]	6.22	[5.04, 7.53]
5.8	0.37	[0.31, 0.45]	–	<5.94
5.9	0.30	[0.25, 0.37]	–	>11.19
6.0	0.28	[0.22, 0.37]	–	–
6.1	0.37	[0.30, 0.46]	–	>7.88

fluctuations can also play a role in shaping the opacity fluctuations. Furthermore, the photoionization rate/UVB fluctuations act oppositely to the above two. Hence, we need a very short effective mean free path (λ_0) to produce sufficient opacity fluctuations after overcoming the density and temperature contribution. We also found that the mean photoionization rate ($\langle \Gamma_{\text{HI}}^{12} \rangle$) mainly controls the mean transmission flux by shifting the opacity distribution.

- We then compared our model predictions with the opacity fluctuations measured by the extended XQR-30 dataset. The efficiency of our model also allowed us to perform a parameter space exploration study by varying λ_0 and $\langle \Gamma_{\text{HI}}^{12} \rangle$. We found that the bestfit model, along with the scatter due to realization, can provide a reasonable match to the data distribution. This further enabled us to derive constraints on these parameters. The constraints remained consistent within the error bars with the previous measurements.

Given the prospective wealth of high-quality spectroscopic data at high redshifts in the near future, it is crucial to have efficient and reasonably accurate models for gleaning relevant physical information without major computational bottlenecks. To this end, our model can be useful to study Ly- α opacity fluctuations in a large cosmological simulation volume within a tractable computational cost. Hence, this model can complement the full-scale hydrodynamic simulations when there is a requirement for uncertain parameter space exploration. With the availability of upcoming large field-of-view telescope surveys, this will be crucial from a theoretical perspective. Furthermore, this can be useful for the synergetic multi-wavelength study of EoR along with other probes such as the 21 cm signal from neutral hydrogen distribution. We plan to exploit the capabilities of our model in future projects.

Data availability

The data presented in this article will be shared on reasonable request to the corresponding author (BM).

References

- Asthana, S., Kulkarni, G., Haehnelt, M. G., et al. 2025, *MNRAS*, 539, L18
- Barkana, R., & Loeb, A. 2001, *Phys. Rep.*, 349, 125
- Becker, G. D., Davies, F. B., Furlanetto, S. R., et al. 2018, *ApJ*, 863, 92
- Becker, G. D., D’Aloisio, A., Christenson, H. M., et al. 2021, *MNRAS*, 508, 1853
- Bolton, J. S., & Haehnelt, M. G. 2007, *MNRAS*, 382, 325
- Bosman, S. E. I., Ďurovčiková, D., Davies, F. B., & Eilers, A.-C. 2021, *MNRAS*, 503, 2077
- Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, *MNRAS*, 514, 55
- Choudhury, T. R. 2009, *Current Sci.*, 97, 841
- Choudhury, T. R. 2022, *General Relativ. Gravitation*, 54, 102
- Choudhury, T. R., & Chakraborty, A. 2025, arXiv e-prints [arXiv:2504.03384]
- Choudhury, T. R., & Paranjape, A. 2018, *MNRAS*, 481, 3821
- Choudhury, T. R., Paranjape, A., & Bosman, S. E. I. 2021, *MNRAS*, 501, 5782
- Christenson, H. M., Becker, G. D., Furlanetto, S. R., et al. 2021, *ApJ*, 923, 87
- D’Aloisio, A., McQuinn, M., & Trac, H. 2015, *ApJ*, 813, L38
- D’Aloisio, A., McQuinn, M., Davies, F. B., & Furlanetto, S. R. 2018, *MNRAS*, 473, 560
- D’Aloisio, A., McQuinn, M., Trac, H., Cain, C., & Mesinger, A. 2020, *ApJ*, 898, 149
- Davies, F. B., & Furlanetto, S. R. 2016, *MNRAS*, 460, 1328
- Davies, F. B., & Furlanetto, S. R. 2022, *MNRAS*, 514, 1302
- Davies, F. B., Becker, G. D., & Furlanetto, S. R. 2018a, *ApJ*, 860, 155
- Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018b, *ApJ*, 864, 142
- Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018c, *ApJ*, 864, 143
- Davies, F. B., Hennawi, J. F., Eilers, A.-C., & Lukić, Z. 2018d, *ApJ*, 855, 106
- Davies, F. B., Bosman, S. E. I., & Furlanetto, S. R. 2024a, arXiv e-prints [arXiv:2406.18186]
- Davies, F. B., Bosman, S. E. I., Gaikwad, P., et al. 2024b, *ApJ*, 965, 134
- Dayal, P., & Ferrara, A. 2018, *Phys. Rep.*, 780, 1
- D’Odorico, V., Bañados, E., Becker, G. D., et al. 2023, *MNRAS*, 523, 1399
- Furlanetto, S. R., & Oh, S. P. 2009, *ApJ*, 701, 94
- Gaikwad, P., Rauch, M., Haehnelt, M. G., et al. 2020, *MNRAS*, 494, 5091
- Gaikwad, P., Srianand, R., Haehnelt, M. G., & Choudhury, T. R. 2021, *MNRAS*, 506, 4389
- Gaikwad, P., Haehnelt, M. G., Davies, F. B., et al. 2023, *MNRAS*, 525, 4093
- Gelman, A., & Rubin, D. B. 1992, *Stat. Sci.*, 7, 457
- Gnedin, N. Y., & Madau, P. 2022, arXiv e-prints [arXiv:2208.02260]
- Greig, B., Mesinger, A., Davies, F. B., et al. 2022, *MNRAS*, 512, 5390
- Hui, L., & Gnedin, N. Y. 1997, *MNRAS*, 292, 27
- Jin, X., Yang, J., Fan, X., et al. 2023, *ApJ*, 942, 59
- Kannan, R., Garaldi, E., Smith, A., et al. 2022, *MNRAS*, 511, 4005
- Keating, L. C., Puchwein, E., & Haehnelt, M. G. 2018, *MNRAS*, 477, 5501
- Keating, L. C., Weinberger, L. H., Kulkarni, G., et al. 2020, *MNRAS*, 491, 1736
- Kulkarni, G., Keating, L. C., Haehnelt, M. G., et al. 2019, *MNRAS*, 485, L24
- Maity, B., & Choudhury, T. R. 2022a, *MNRAS*, 511, 2239
- Maity, B., & Choudhury, T. R. 2022b, *MNRAS*, 515, 617
- McGreer, I. D., Mesinger, A., & D’Odorico, V. 2015, *MNRAS*, 447, 499
- McQuinn, M. 2016, *ARA&A*, 54, 313
- Meiksin, A. 2020, *MNRAS*, 491, 4884
- Mesinger, A. 2016, *Astrophys. Space Sci. Libr.*, 423
- Mesinger, A., & Furlanetto, S. 2009, *MNRAS*, 400, 1461
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, 411, 955
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *J. Chem. Phys.*, 21, 1087
- Nasir, F., & D’Aloisio, A. 2020, *MNRAS*, 494, 3080
- Planck Collaboration XIII. 2016, *A&A*, 594, A13
- Prochaska, J. X., Madau, P., O’Meara, J. M., & Fumagalli, M. 2014, *MNRAS*, 438, 476
- Qin, Y., Mesinger, A., Greig, B., & Park, J. 2021, *MNRAS*, 501, 4748
- Qin, Y., Mesinger, A., Prelogović, D., et al. 2025, *PASA*, 42
- Spina, B., Bosman, S. E. I., Davies, F. B., Gaikwad, P., & Zhu, Y. 2024, *A&A*, 688, L26
- Tepper-García, T. 2006, *MNRAS*, 369, 2025
- Torrado, J., & Lewis, A. 2021, *JCAP*, 2021, 057
- Trac, H., Chen, N., Holst, I., Alvarez, M. A., & Cen, R. 2022, *ApJ*, 927, 186
- Walther, M., Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2019, *ApJ*, 872, 13
- Weinberg, D. H., Katz, N., & Hernquist, L. 1998, *ASP Conf. Ser.*, 148, 21
- Worseck, G., Prochaska, J. X., O’Meara, J. M., et al. 2014, *MNRAS*, 445, 1745
- Wyithe, J. S. B., & Bolton, J. S. 2011, *MNRAS*, 412, 1926
- Zaroubi, S. 2013, *Astrophys. Space Sci. Libr.*, 396, 45
- Zel’dovich, Y. B. 1970, *A&A*, 5, 84
- Zhu, Y., Becker, G. D., Bosman, S. E. I., et al. 2021, *ApJ*, 923, 223
- Zhu, Y., Becker, G. D., Bosman, S. E. I., et al. 2022, *ApJ*, 932, 76
- Zhu, Y., Becker, G. D., Christenson, H. M., et al. 2023, *ApJ*, 955, 115
- Zhu, Y., Becker, G. D., Bosman, S. E. I., et al. 2024, *MNRAS*, 533, L49

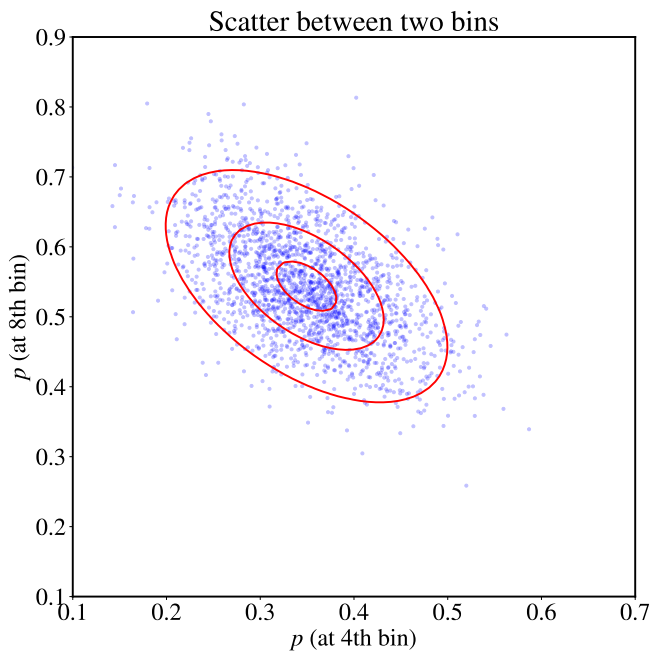


Fig. A.1. Scatter between two bins (specifically the fourth and the eighth).

Appendix A: Justification for Gaussian likelihood

Here, we show a plot justifying the choice of our Gaussian likelihood for the parameter exploration studies. The plot shows the scatter between two random bins due to different realizations of the sightlines, along with a few Gaussian contours. The nature of the scatter is reasonably close to being Gaussian, which gives us the confidence to use a standard multivariate Gaussian likelihood in this study.