





# Robustness of direct measurements of the mean free path of ionizing photons in the epoch of reionization

Sindhu Satyavolu <sup>1</sup>★, Girish Kulkarni <sup>1</sup>, Laura C. Keating <sup>2</sup> and Martin G. Haehnelt <sup>3,4</sup>

<sup>1</sup>Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India

<sup>2</sup>Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK

<sup>3</sup>Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

<sup>4</sup>Kavli Institute of Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

Accepted 2024 July 9. Received 2024 July 9; in original form 2023 November 10

## ABSTRACT

Measurements of the mean free path of Lyman-continuum photons in the intergalactic medium during the epoch of reionization can help constrain the nature of the sources as well as sinks of hydrogen-ionizing radiation. A recent approach to this measurement has been to utilize composite spectra of multiple quasars at  $z \sim 6$ , and infer the mean free path after correcting the spectra for the presence of quasar proximity zones. This has revealed not only a steep drop in the mean free path from  $z = 5$  to  $z = 6$ , but also potentially a mild tension with reionization simulations. We critically examine such direct measurements of the mean free path for biases due to quasar environment, incomplete reionization, and quasar proximity zones. Using cosmological radiative transfer simulations of reionization combined with one-dimensional radiative transfer calculations of quasar proximity zones, we find that the bias in the mean free path due to overdensities around quasars is minimal at  $z \sim 6$ . Patchiness of reionization at this redshift also does not affect the measurements significantly. Fitting our model to the data results in a mean free path of  $\lambda_{\text{mfp}} = 1.49^{+0.47}_{-0.52}$  pMpc at  $z = 6$ , which is consistent with the recent measurements in the literature, indicating robustness with respect to the modelling of quasar proximity zones. We also compare various ways in which the mean free path has been defined in simulations before the end of reionization. Overall, our finding is that recent measurements of the mean free path appear to be robust relative to several sources of potential bias.

**Key words:** radiative transfer – methods: numerical – intergalactic medium – quasars: absorption lines – cosmology: theory.

## 1 INTRODUCTION

The epoch of reionization marks the era during which the neutral hydrogen in the Universe became largely ionized, most likely due to ultraviolet (UV) photons from stars in galaxies across the cosmic volume, with potentially an additional minor contribution from quasars. To understand how reionization happened, we need to know how the photons that escaped from the various sources of radiation interacted with the intergalactic medium (IGM) to ionize the Universe. An important characteristic of the propagation of these ionizing photons is determined by their mean free path (MFP), defined as the average distance a photon travels before getting absorbed (Rybicki & Lightman 1985). In a homogeneous IGM, the MFP increases with the background UV photoionization rate, as this lowers the IGM opacity. More generally, the MFP  $\lambda_{\text{MFP}}$  varies with the background photoionization rate  $\Gamma_{\text{bg}}$  as a power law (Miralda-Escudé, Haehnelt & Rees 2000), and, furthermore, both of these quantities can be expressed as functions of the cosmological emissivity  $\epsilon$  of ionizing radiation. Knowing the MFP therefore puts constraints on the ionizing sources as well as the sinks of ionizing photons. At higher redshifts, the rate of increase of the

MFP with redshift indicates how rapid the progress of reionization is. Measurement of the MFP of hydrogen-ionizing photons is thus important for characterizing reionization.

During reionization, the sources of radiation carve out regions of ionized hydrogen in the IGM around themselves. These ionized regions later coalesce, at which point a background of ionizing radiation gets established throughout the Universe (Gnedin & Kaurov 2014; Bauer et al. 2015). This is considered to be the end of reionization and is thought to occur at  $z \sim 5.3$  (Kulkarni et al. 2019; Bosman et al. 2022; Zhu et al. 2022). At lower redshifts, in the post-reionization Universe, the MFP is dominated by the residual neutral hydrogen systems in the otherwise fully ionized IGM. These systems retain neutral hydrogen due to self-shielding thanks to their high density. These appear as high-column-density absorbers of the ionizing radiation background. The corresponding MFP is then set by the average spacing between such absorbers, obtained by counting the number of optically thick absorbers or Lyman-limit systems (LLS) along the lines of sight in quasar spectra (Storrie-Lombardi et al. 1994; O’Meara et al. 2007, 2013). At higher redshifts, before reionization has ended, however, the boundaries of the ionized regions also play an important role in setting the MFP. At these times, the MFP is the average distance between LLSs or the typical size of ionized regions, depending on which is smaller (Shukla et al. 2016; Madau 2017).

\* E-mail: [sindhu@theory.tifr.res.in](mailto:sindhu@theory.tifr.res.in)

Observationally, measurements of the MFP have been inferred using direct and indirect means. One way to obtain a direct measurement of the MFP is to use stacked quasar spectra bluewards of rest-frame 912 Å. The MFP is then computed as the distance at which the effective Lyman-continuum optical depth becomes unity (Prochaska, Worseck & O’Meara 2009; Fumagalli et al. 2013; O’Meara et al. 2013; Worseck et al. 2014). This has resulted in a measurement of  $\lambda_{\text{mfp}} = 10.3 \pm 1.6$  pMpc at  $z = 5.16$ , with the MFP increasing as  $(1+z)^{-5.4}$  down to  $z = 4.56$ . At higher redshifts, the MFP can however become comparable or smaller than the size of the proximity zone of the quasars whose spectra are used in the stacks. This biases the MFP towards lower values (Worseck et al. 2014; D’Aloisio et al. 2018). Becker et al. (2021, B21) reported a measurement of the MFP in which they sought to correct the bias due to the quasar proximity zones by modifying the direct measurement method to account for the excess flux due to the quasar. The resultant value of the MFP inferred by B21 is  $0.75^{+0.65}_{-0.45}$  pMpc at  $z = 6$ . The MFP thus shows a sharp drop at  $z > 5$  relative to the  $(1+z)^{-5.4}$  behaviour seen at redshifts lower than 5. Moreover, the value of the MFP at  $z = 6$  is smaller than the value in several of the latest reionization simulations by a factor of 2 or more (Keating et al. 2020a). Davies et al. (2021) discussed the implications of this tension for the ionizing budget of galaxies to argue that a shorter MFP requires an ionizing emissivity that is up to six times larger than the typically assumed values. Cain et al. (2021) found that the short MFP reported by B21 is consistent with a late reionization scenario powered by fainter galaxies with a high ionizing photon production efficiency (Lewis et al. 2022; Garaldi et al. 2022). More recently, Zhu et al. (2023) have measured the MFP for quasars at redshifts between  $5.1 < z < 6$  using the B21 method and found the value at redshift 6 to agree with that measured by B21. They also find the decrease in the MFP to be steeper with increasing redshift, with a sharp drop in MFP by nearly 75 per cent between  $z \sim 5.6$  and 5.9. Such a sharp decline in the MFP points towards a rapid end to reionization. Another direct measurement technique is to define the MFP by averaging free paths measured from the distribution of absorbers along individual quasar sightlines (Songaila & Cowie 2010; Romano et al. 2019). This approach suggested a milder evolution of the MFP with redshift, between  $3 < z < 6$ , when compared to the direct stacking method. Bosman (2021) have extended this approach to detect absorption due to LLS in the six lowest order Lyman-series transitions and put a lower limit based on the average of individual free paths defined this way to be  $\lambda_{\text{MFP}} > 0.31$  pMpc at redshift 6, consistent with the measurement of B21.

Indirect measurements of the MFP have been obtained by comparing the observed Ly  $\alpha$  opacity of the IGM with that in numerical simulations. The mean and the scatter of the cumulative distribution function (CDF) of the effective Ly  $\alpha$  opacity has been used to simultaneously constrain both the MFP and the photoionization rate for a given emissivity (e.g. Gaikwad et al. 2023; Wolfson et al. 2023; Davies et al. 2024). Overall, the indirect MFP measurements based on the Ly  $\alpha$  opacity (Gaikwad et al. 2023) are consistent with the direct MFP measured from the quasar stack beyond the Lyman limit, within the error bars (Becker et al. 2021; Zhu et al. 2023). However, the slope of evolution of the MFP with redshift differs beyond redshift  $z \sim 5.5$ , with the direct measurements yielding a steeper slope than the indirect measurements. A gradual slope agrees better with the reionization simulations of D’Aloisio et al. (2018), Keating et al. (2020a), and Cain et al. (2021), as compared to the steeper slope seen in the models of Lewis et al. (2022) and Garaldi et al. (2022). Possible reasons for the mismatch between the indirect and direct measurements of the MFP could be the assumptions made

by the two methods. B21 assume an analytic model to measure the effective optical depth, where the opacity  $\kappa$  is proportional to the photoionization rate as  $\Gamma^{-\xi}$ . In the QSO proximity zone, the total  $\Gamma$  is a sum of both  $\Gamma_{\text{bg}}$  and  $\Gamma_{\text{qso}}$ , the latter predominantly varying with distance from the quasar due to geometric dilution. B21 solve for  $\Gamma_{\text{qso}}$  numerically, assigning average parameters for their QSO stack. They thereby keep the background value of the photoionization rate,  $\Gamma_{\text{bg}}$  fixed while keeping the MFP as a free parameter. In reality, the photoionization rate and MFP will both depend on the ionizing emissivity and will co-evolve. B21 also fix the opacity due to higher order Lyman series absorption to the values obtained using an optically thin simulation while fitting their model. Roth et al. 2024 discuss the effect of this on the inference of B21. The slope of the variation of the MFP with redshift is also different between the counting LLS method, indirect inference method and the direct stacking method, the latter showing the steepest evolution. These differences reflect the dependence of the MFP on the nature of absorbing sources, each sensitive to a different measurement technique (Inoue et al. 2014).

In this work, we critically examine the direct measurement method of B21 for possible biases due to (a) higher cosmological densities around high-redshift quasars, (b) incomplete reionization at  $z > 5.3$ , and (c) differences between the structure of quasar proximity zones as computed using the analytical model of B21 and that obtained via radiative transfer (RT) calculations. We also investigate the challenge of defining the MFP during the epoch of reionization, when a cosmological UV radiation background is not yet established uniformly. This paper is structured as follows. We discuss our data and models in Section 2. We discuss the definition of the MFP before the end of reionization, and the effect of overdense quasar environments on the MFP in Section 3. Section 4 presents a discussion on the effect of incomplete reionization on the B21 measurements by discussing mock measurements of the MFP in our simulations using the B21 method. Finally, we present a new measurement of the MFP using the B21 data in Section 5, incorporating an RT model of quasar proximity zones. We end with a discussion and a summary of our conclusions in Section 6. Our  $\Lambda$ CDM cosmological model has  $\Omega_b = 0.0482$ ,  $\Omega_m = 0.308$ ,  $\Omega_\Lambda = 0.692$ ,  $h = 0.678$ ,  $n_s = 0.961$ ,  $\sigma_8 = 0.829$ , and  $Y_{\text{He}} = 0.24$  (Planck Collaboration XVI 2014). All distances are in proper (physical) units unless specified otherwise.

## 2 DATA AND MODELS

Our strategy in this paper is to use a cosmological RT simulation of reionization. Such a simulation provides us with a realistic model for a partially ionized IGM at  $z > 5.3$ , on top of which we model quasar proximity zones using one-dimensional RT. We then use the resultant one-dimensional spectra to examine the MFP. The cosmological RT simulation used here is described by Kulkarni et al. (2019). It consists of a cosmological hydrodynamical simulation performed using P-GADGET-3 (which is a modified version of GADGET-2, described by Springel 2005), that is post-processed with three-dimensional RT using the ATON code (Aubert & Teyssier 2008, 2010). The box size of the simulation is  $160 \text{ cMpc}/h$  with  $2048^3$  gas and dark matter particles. This model is calibrated to reproduce the observed mean Ly  $\alpha$  transmission at  $z > 5$ , and agrees with several high-redshift observations such as the spatial fluctuations in the Ly  $\alpha$  effective opacity (Becker et al. 2015; Bosman et al. 2022), the Thomson scattering optical depth to the last scattering surface (Planck Collaboration VI 2020), the large-scale radial distribution of galaxies around opaque Ly  $\alpha$  troughs (Keating et al. 2020a; Becker et al. 2015), quasar damping wings (Greig et al. 2017; Greig,

Mesinger & Bañados 2019; Davies et al. 2018; Wang et al. 2020), measurements of the IGM temperature (Gaikwad et al. 2020), and the luminosity function and clustering of Ly  $\alpha$  emitters (Weinberger et al. 2018; Weinberger, Haehnelt & Kulkarni 2019). The mid-point of reionization in this simulation is at  $z = 7$ , and reionization ends at  $z = 5.3$ . At  $z = 5.95$ , the range of halo masses resolved in the simulation is in between  $2.32 \times 10^8 M_{\odot}$  and  $4.59 \times 10^{12} M_{\odot}$ . We direct the reader to Kulkarni et al. (2019) and Keating et al. (2020a) for further details.

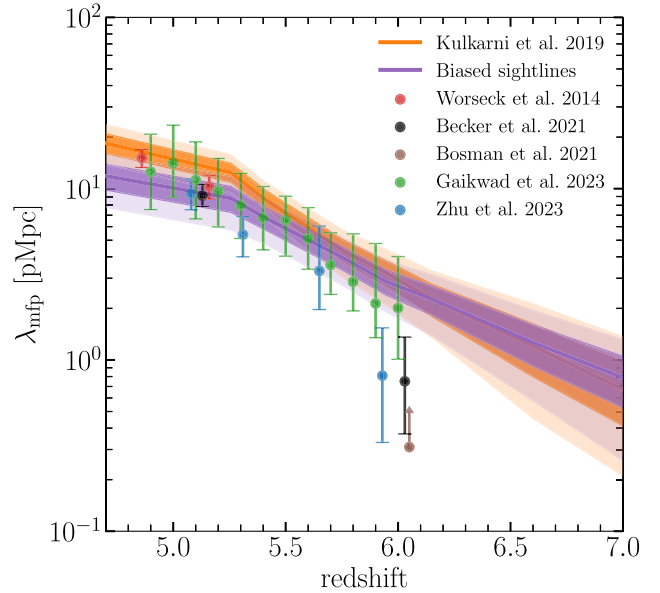
To construct mock spectra including the proximity zone of the QSO, we post-process our simulations with a one-dimensional RT code that is described by Satyavolu et al. (2023). All quasars are assigned a magnitude of  $M_{1450} = -27.0$ , corresponding to the mean magnitude of the quasars in the B21 stack. In our fiducial model, the quasars have a ‘lightbulb’ light curve with a lifetime of  $10^6$  yr and are placed in haloes with masses between  $10^{11} M_{\odot} \lesssim M_{\text{halo}} \lesssim 10^{12} M_{\odot}$ . We assume a broken power-law SED, given by Lusso et al. (2015), such that the spectral slopes above and below  $912 \text{ \AA}$  are  $-0.61$  and  $-1.7$ , respectively. We then solve the thermochemistry equations for the hydrogen and helium ionized fractions along a sightline for a given quasar magnitude and lifetime at each redshift. We also simultaneously solve for the gas temperature, which is regulated by photoionization heating, adiabatic cooling, radiative cooling due to recombinations, collisional excitations, Compton cooling due to CMB photons, and Bremsstrahlung.

For our measurement of the MFP, we use the observed data from B21. This is in the form of a composite of quasar spectra with mean redshift  $z_{\text{qso}} = 5.97$ , constructed with 13 quasar spectra obtained using the Keck/ESI and VLT/XSHOOTER spectrographs. The redshifts used for the quasars in the stack were obtained using the [C II] line, from the Ly  $\alpha$  halo emission, or from the apparent start of the Ly  $\alpha$  forest absorption. While selecting quasars in the stack, a minimum cut-off of  $S/N \gtrsim 20$  per  $30 \text{ km s}^{-1}$  was applied for the signal-to-noise ratio near the rest-frame  $1285 \text{ \AA}$  in the continuum. The mean brightness of the quasars in the sample is  $M_{1450} = -27$ , with a range of  $-27.8 < M_{1450} < -25.7$ . The typical resolution for the quasars in the stack is full width at half-maximum (FWHM)  $\sim 45 \text{ km s}^{-1}$  for spectra obtained using ESI and FWHM  $\sim 25 \text{ km s}^{-1}$  for spectra obtained using XSHOOTER. While stacking, the quasar flux is normalized by the median flux between  $1270$  and  $1380 \text{ \AA}$ , although the normalization with respect to the intrinsic continuum is kept a free parameter for our analysis, similar to B21. In order to study the Lyman-continuum MFP, we focus on the wavelength range of  $912$ – $825 \text{ \AA}$ .

### 3 MFP OF HYDROGEN-IONIZING PHOTONS BEFORE THE END OF REIONIZATION

The post-reionization ( $z \lesssim 5.5$ ) MFP measurements are in good agreement with several reionization simulations (Keating et al. 2020b; D’Aloisio et al. 2020; Cain et al. 2021; Lewis et al. 2022). At higher redshifts  $z \gtrsim 5.5$ , when reionization is still believed to be ongoing, both observations and simulations differ in the measured MFP values. While some simulations agree better with the MFP measured by B21 (Lewis et al. 2022; Garaldi et al. 2022), some models (e.g. Cain et al. 2021), including ours (Kulkarni et al. 2019), agree better with the MFP measured by Gaikwad et al. (2023). Keeping this in mind, we look at possible biases in the computation of the MFP from the simulations.

We use the standard definition for the MFP (Kulkarni et al. 2016) where the average Lyman-continuum transmission across random sightlines in the comoving frame is fit by an exponential with an



**Figure 1.** Effect of higher overdensity around high-redshift quasars on the MFP in simulations. The orange curve and shaded regions show the median value at the  $1\sigma$  and  $2\sigma$  scatter of the MFP in our reionization model, when the MFP is defined using equation (1). The purple curve and shaded regions show the same but when only sightlines originating in the highest mass haloes in the simulation volume are used. The data points and error bars show various measurements of the MFP. Measurements by Becker et al. (2021) and Bosman (2021) have been displaced in redshift by  $\delta z = 0.03$  and  $0.05$ , respectively, for legibility.

$e$ -folding length scale of  $\lambda_{\text{MFP}}$ ,

$$(F) = F_0 \exp\left(-\frac{x}{\lambda_{\text{MFP}}}\right). \quad (1)$$

Fig. 1 shows the evolution in our model of the MFP defined in this manner. The orange curve shows the MFP obtained by fitting equation (1) to composite spectra of 13 randomly drawn sightlines from our simulation to match the number of sightlines used in B21. In order to obtain the sample variance, we repeat this computation for 10 000 randomly drawn samples. This method for measuring the MFP is analogous to the approach of Prochaska et al. (2009).<sup>1</sup> The solid curves in Fig. 1 show the median of the resultant distribution of MFP values, with the shaded regions showing the  $1\sigma$  and  $2\sigma$  scatter. We see that the disagreement between the MFP measured in this manner from the measured value of B21 and Zhu et al. (2023) is more than  $2\sigma$  at  $z = 6$ . Also missing in the simulations is the steep decline in the MFP observed by B21 and Zhu et al. (2023) between  $z \sim 5.5$  and  $z \sim 6$ . We also see that the evolution of the MFP in our model becomes even more gradual once reionization is complete at  $z \sim 5.3$ . At these post-reionization redshifts, the optical depth is controlled by the overdense regions responsible for LLSs.

Direct measurements infer MFP from sightlines towards QSOs, which tend to reside in overdensities (e.g. Wang et al. 2023). To check for the bias due to large-scale structure, we draw random sightlines originating only from haloes with the highest masses sampled by our simulation volume, between  $10^{12}$  and  $10^{13} M_{\odot}$ , otherwise applying equation (1) as above. The resulting median MFP with the  $1\sigma$  and  $2\sigma$

<sup>1</sup>We have verified this by fitting our mock quasar stack discussed in Section 4 with the model of Prochaska et al. (2009).

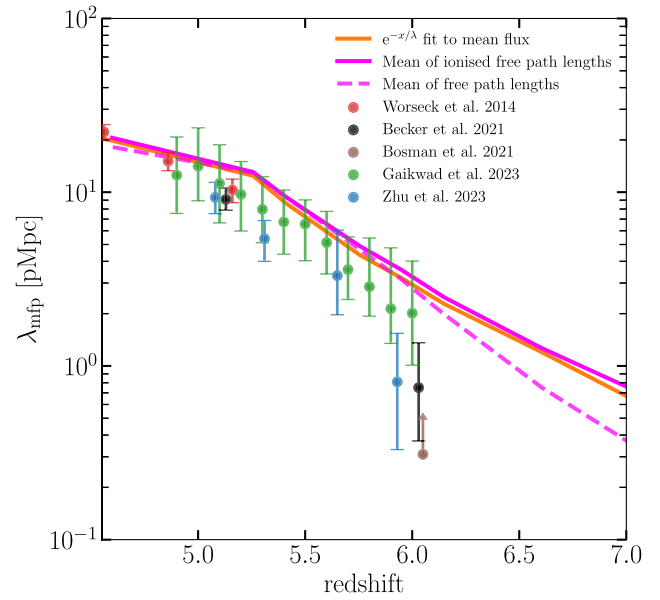
scatter is shown in Fig. 1 in purple. The MFP along biased sightlines shows qualitatively similar behaviour with redshift when compared to the MFP along random sightlines, but with a slightly shallower slope. The difference between the two MFPs is insignificant at  $z = 6$ . The random sightlines include sightlines that start from any location, including haloes. Any difference seen between the two cases in Fig. 1 is because sightlines originating from massive haloes are ionized earlier than other regions in the IGM. Once reionization is complete, the overdensity along sightlines originating from haloes will also lead to a higher optical depth or lower MFP when compared to the measurement of random sightlines. We see in Fig. 1 that this biased MFP is in better agreement with the measurement by Zhu et al. (2023) and B21 than that by Worseck et al. (2014) at  $z \lesssim 5.5$ . We discuss this further in Section 4.

Another potential source of bias in the MFP as formally defined by equation (1) is that this exponential attenuation assumes a spatially constant opacity. In reality, reionization is not yet complete at  $z = 6$ . Consequently, the Lyman-continuum opacity of the IGM has large spatial variations. An alternative definition of the MFP that addresses this complexity is one that obtains a distribution of the free-path lengths of photons in a simulation box and computes its mean (Rahmati & Schaye 2018). We implement this in our simulation by following Rahmati & Schaye (2018) and measuring the free path as the distance at which the Lyman-continuum optical depth along the sightline becomes unity. The distribution of such free-path lengths is bimodal (see Appendix A). One part of the distribution describes the mean free path in neutral regions that are on average not overdense. The sightlines with larger free paths on the other hand also include those which encounter neutral islands along the path. The fraction of sightlines with small free paths becomes smaller than those with larger free paths at lower redshifts, until the free path distribution becomes unimodal Gaussian. We measure the MFP as the average of the dominant part of the bimodal distribution, which represents the ionized IGM. The MFP defined in this manner is shown by the solid pink curve in Fig. 2. We see that this results in a value of the MFP that is very close to the value obtained by using equation (1), shown by the orange curve in Fig. 2.

In their CoDa III simulation, Lewis et al. (2022) compute the MFP using a similar free-path length method. The free-path lengths are defined in this case by doing an exponential fit to the flux along individual sightlines, which they find to be similar to the free paths measured following Rahmati & Schaye (2018). The dashed pink curve in Fig. 2 shows the result from our simulation when we average over all free paths, similar to one of the approaches used in Lewis et al. (2022) to measure the MFP. As expected from the shape of the distribution of the free-path lengths, the MFP is now biased towards lower values. We find that this bias is small at  $z \lesssim 6$ . This could be a reflection of the inadequate spatial resolution of our simulation, due to which a large number of free paths become smaller than our cell size.

#### 4 EFFECT OF INCOMPLETE REIONIZATION ON THE B21 MEASUREMENT

We have seen above that neither the large-scale structure bias nor the variations in the definition of the MFP cause a significant change in the value of the MFP in our simulation at  $z = 6$ . We now investigate the effect of the residual neutral hydrogen ‘islands’ in the IGM at this redshift on the B21 measurement of the MFP. Given the good agreement of our simulations with the measurements of the MFP by Worseck et al. (2014), Gaikwad et al. (2023), and Bosman (2021), it is important to consider if the absence of patchy reionization in



**Figure 2.** A comparison of three definitions of the MFP in an incompletely reionized IGM. The solid orange curve shows the MFP obtained from our simulation using equation (1). This curve is also shown in Fig. 1. The solid pink curve shows the value of the MFP computed as the mean of the higher value subpopulation of the bimodal distribution of free-path lengths in the simulation box. The dashed pink curve shows the MFP when computed as the mean of the full bimodal distribution of the free-path lengths.

the models of B21 and Zhu et al. (2023) could potentially bias their measurements. In order to do this, we construct mock data from our simulation and apply the B21 method to it. We then compare the resultant measurement of the MFP with the ‘true’ value of the MFP in the simulation given by equation (1). A similar test was also performed by B21 themselves. However, the mock spectra used in their test were created from numerical simulations where reionization is assumed as instantaneous from a uniform ionizing background and post-processed to include fluctuations following the approach of Davies & Furlanetto (2016).

#### 4.1 The B21 method

We begin by briefly reviewing the B21 method for measuring the MFP. The effective Lyman-continuum optical depth for a photon that is emitted at  $z_{\text{qso}}$  and redshifts to the Lyman limit at  $z_{912}$  is given by

$$\tau_{\text{eff}}^{\text{LL}}(z_{912}) = \frac{c}{H_0 \Omega_m^{1/2}} (1+z)^{2.75} \int_{z_{912}}^{z_{\text{qso}}} \kappa_{912}(z') (1+z')^{-5.25} dz', \quad (2)$$

where  $\kappa_{912}$  is the opacity to  $912 \text{ \AA}$  photons. The dependence on redshift is as follows: the exponent 2.75 is a result of the dependence of cross-section on frequency as  $\sigma_v^{\text{LL}} \propto \nu^{-2.75}$ , while the exponent  $-5.25$  comes from the conversion of comoving distance to redshift as  $r \propto aH \propto (1+z)^{-1} \times (1+z)^{-3/2}$  in the matter-dominated era. The opacity  $\kappa_{912}$  is assumed to scale with the photoionization rate as a power law, so that at a distance  $r$  from the quasar,

$$\kappa_{912}(r) = \kappa_{912}^{\text{bg}} \left( \frac{\Gamma_{\text{tot}}(r)}{\Gamma_{\text{bg}}} \right)^{-\xi}, \quad (3)$$

where  $\kappa_{912}^{\text{bg}}$  is the opacity to the ionized background and  $\Gamma_{\text{tot}}(r) = \Gamma_{\text{qso}}(r) + \Gamma_{\text{bg}}(r)$  is the sum of the photoionization rate due to the

QSO and the local background photoionization rate. The opacity and photoionization rate are related to each other through their mutual dependence on the shape and number density of neutral gas absorbers. This information is parametrized using the power-law index  $\xi$ , which has been studied using analytic arguments as well as numerical simulations (McQuinn, Oh & Faucher-Giguère 2011; Muñoz et al. 2016). We discuss this parameter in detail in Section 4. Equation (4) is applied to a stack of QSO spectra, so the terms in the equation are all averaged quantities, and the average of the local background photoionization is assumed to be uniform and equal to  $\Gamma_{\text{bg}}$ . The average photoionization rate due to all quasars at a location  $r$  is computed by iteratively solving for

$$\Gamma_{\text{qso}}(r + \delta r) = \Gamma_{\text{qso}}(r) \left( \frac{r + \delta r}{r} \right)^{-2} e^{-\kappa_{912}^{\text{bg}} \delta r}, \quad (4)$$

with the initial condition

$$\Gamma_{\text{qso}}(\delta r) = \Gamma_{\text{bg}} \left( \frac{\delta r}{R_{\text{eq}}} \right)^{-2}. \quad (5)$$

Here,  $R_{\text{eq}}$  is the distance at which the photoionization rate due to the quasar's radiation is equal to the background photoionization rate in the absence of any absorption.  $R_{\text{eq}}$  depends on the QSO magnitude and spectrum in addition to the background photoionization rate. The free parameters in this model are therefore,  $\kappa_{912}^{\text{bg}}$ ,  $R_{\text{eq}}$ ,  $\xi$ , and  $\Gamma_{\text{bg}}$ . The mean flux of the stack is fit by the above model of opacity and the MFP  $\lambda_{\text{MFP}}$  is inferred as the distance from the quasar at which  $\tau_{\text{eff}}^{\text{LL}}(z_{912})$  becomes equal to 1.

#### 4.2 Mock data

To construct stacks of mock spectra at redshift 6, we consider QSOs with magnitude of  $M_{1450} = -27$  with a broken power-law spectral profile for the specific luminosity as in Lusso et al. (2015),

$$L_{\nu} \propto \begin{cases} \nu^{-0.61} & \text{if } \lambda \geq 912 \text{ \AA}, \\ \nu^{-1.70} & \text{if } \lambda < 912 \text{ \AA}. \end{cases} \quad (6)$$

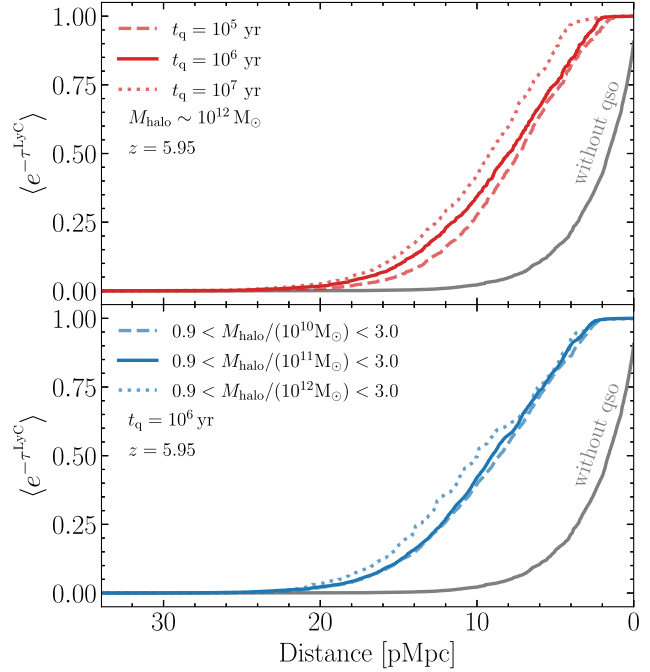
Although it is not a deciding factor, we choose the magnitude of our mock stack to be of the value same as the mean magnitude of quasars in the B21 stack at redshift 6. The power-law index values are slightly different from those assumed in B21 ( $-0.5$  and  $-1.5$  for wavelengths greater than and less than  $912 \text{ \AA}$ , respectively).

After placing QSOs in haloes, we post-process the sightlines with our 1D RT code as described in Section 2 to obtain the distribution of the neutral hydrogen density. Using this, we compute the ionizing optical depth along a sightline as

$$\tau^{\text{LL}}(r) = \sum_r n_{\text{HI}} \sigma_{\text{HI}}^{\text{LL}} dr, \quad (7)$$

where  $\sigma_{\text{HI}}^{\text{LL}} = 6.3 \times 10^{-22} \text{ m}^2$  is the hydrogen ionization cross-section. Since equation (4) does not explicitly assume the QSO lifetime or the host halo mass, we randomize over them in our stack, with halo masses between  $10^{10} M_{\odot} \lesssim M_{\text{halo}} \lesssim 10^{12} M_{\odot}$  and lifetimes between  $10^5 < t_{\text{q}} < 10^7$  Myr. Each of our composite stacks comprises 13 sightlines.

Fig. 3 shows composite stacks obtained in this manner with and without quasars for the above range of quasar lifetimes and host halo masses. The top panel shows the variation of the mean continuum flux along quasar sightlines originating from  $\sim 10^{12} M_{\odot}$  haloes and having lifetimes between  $10^5 < t_{\text{q}} < 10^7$  Myr. The mean flux computed along sightlines drawn at random directions is shown in grey. Once the quasar is on, the opacity near the quasar decreases as the quasar emits ionizing photons that reduce the neutral hydrogen



**Figure 3.** Effect of quasar lifetime (top panel) and host halo mass (bottom panel) on the Lyman-continuum composite spectrum of 1000 sightlines in our simulation. In both cases, the grey curve shows the transmission without the effect of the quasar proximity zone. The volume averaged neutral hydrogen fraction in our simulation at  $z = 5.95$  is  $\langle x_{\text{HI}} \rangle = 0.13$ .

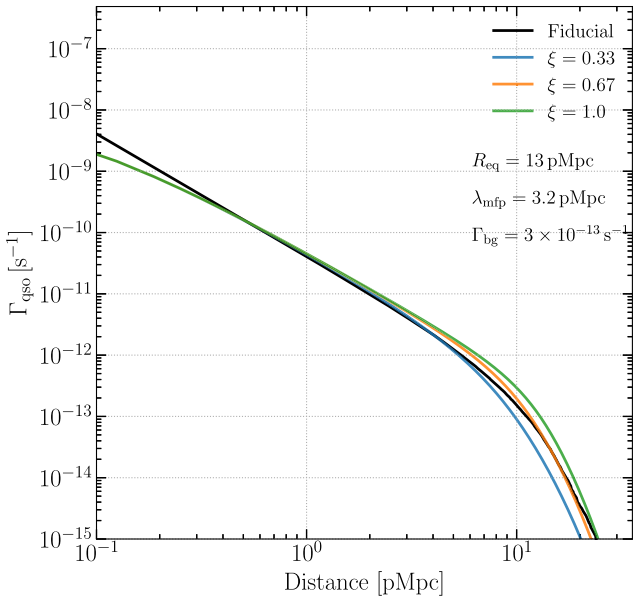
density. The mean flux thus increases with increase in the quasar lifetime. While the decrease in flux is gradual when the quasar turns on in a uniformly ionized medium, we notice that the flux fall off is steeper in our patchy model. The bottom panel shows the variation of the mean continuum flux when the quasar lifetime is fixed to be  $t_{\text{q}} = 10^6$  yr and the quasar host halo masses are varied to be in three ranges between  $10^{10} M_{\odot} \lesssim M_{\text{halo}} \lesssim 10^{12} M_{\odot}$ . In our simulation, more massive haloes are ionized earlier and therefore the mean flux is slightly higher along these sightlines. Fig. 3 shows that the mean flux changes almost doubles in the presence of the quasar, while changing the lifetime or host halo masses of the quasars by an order of magnitude changes the mean flux by less than around 10 per cent. For the purpose of recovery of the MFP using the B21 method, we do not include the absorption due to higher order Lyman series transitions in the mock spectra. We correspondingly fit the mock spectra without including the contribution from the Lyman series opacity, as shown in equation (8). We also do not add a zero-point correction to our mock stack and hence do not include this parameter while fitting using equation (8).

#### 4.3 Choice of $R_{\text{eq}}$ , $\Gamma_{\text{bg}}$ , and $\xi$

Following B21, we fit the mock spectra by an exponential

$$\langle F \rangle = F_0 \exp(-\tau_{\text{eff}}^{\text{LL}}), \quad (8)$$

with  $\kappa_{912}^{\text{bg}}$  as free parameter. The other parameters are  $R_{\text{eq}}$ ,  $\Gamma_{\text{bg}}$ , and  $\xi$ , which we discuss below. We chose to keep these parameters fixed, as B21 did for their nominal measurement. We have inspected the results while keeping  $\xi$  as a free parameter, and found them to be consistent with what was reported in Zhu et al. (2023). We leave self-consistent parametrization of  $\Gamma_{\text{bg}}$  with  $\kappa_{912}^{\text{bg}}$  for future work.



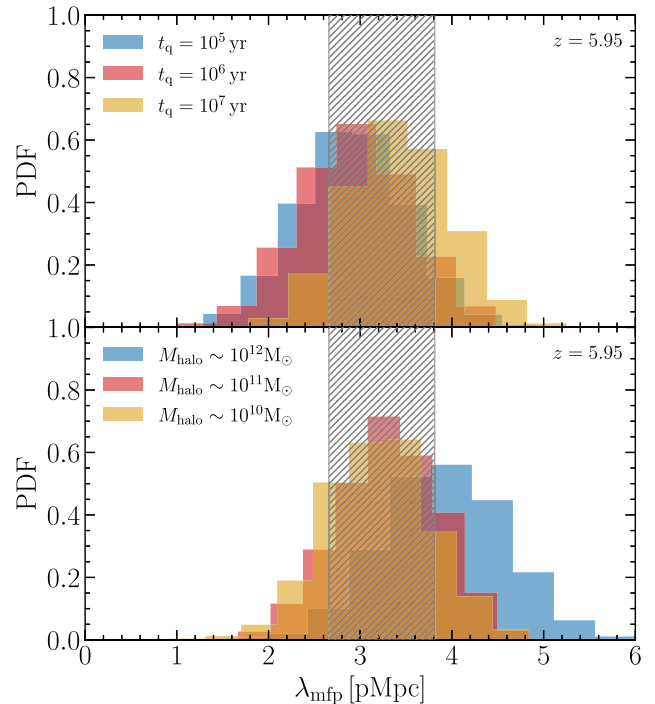
**Figure 4.** H I photoionization rate due to QSO. The black curve shows the average  $\Gamma_{\text{qso}}$  for QSOs with lifetimes ranging between  $10^5 < t_q < 10^7$  yr with magnitude  $M_{1450} = -27$  at redshift  $z = 5.95$ , computed using our 1D RT simulation. The green, orange, and blue curves show the average  $\Gamma_{\text{qso}}$  computed analytically for the same QSO magnitude at redshift 6, for  $\xi = 0.33, 0.67$ , and  $1.0$ , respectively. The value of  $R_{\text{eq}}$  is computed using  $\Gamma_{\text{bg}} = 3 \times 10^{-13} \text{ s}^{-1}$ .

The distance  $R_{\text{eq}}$  is given by Calverley et al. (2011)

$$R_{\text{eq}} = \left[ \frac{L_{912} \sigma_0}{8\pi h^2 \Gamma_{\text{bg}} (\alpha_v^{\text{ion}} + 2.75)} \right]^{1/2}. \quad (9)$$

The analytic expression for  $R_{\text{eq}}$  is computed under the approximation that the quasar photoionization rate falls as  $1/r^2$ , where  $r$  is the distance from the source. This equation can be used to further compute  $\Gamma_{\text{qso}}$  in the presence of absorption by iterating over equation (4). For the mock stacks, we chose a value of  $R_{\text{eq}} = 13$  pMpc, for  $L_{912}$  corresponding to magnitude of  $M_{1450} = -27.0$  and  $\Gamma_{\text{bg}}$  of  $3 \times 10^{-13} \text{ s}^{-1}$  as discussed below.

B21 chose a value of  $3 \times 10^{-13} \text{ s}^{-1}$  for  $\Gamma_{\text{bg}}$  in their analysis. The background photoionization rate in our simulations is inhomogeneous. In ionized regions, we find the average  $\Gamma_{\text{bg}}$  to be around  $3 \times 10^{-13} \text{ s}^{-1}$  at redshift  $z \sim 6$ . However, the average value of  $\Gamma_{\text{bg}}$  is  $10^{-13} \text{ s}^{-1}$ , almost 3 times smaller than the optically thin value assumed by B21. Given that at redshift  $z \sim 6$ , the volume-averaged ionized hydrogen fraction is close to 80 per cent, it would suffice to use this value of background photoionization rate to compute  $\Gamma_{\text{qso}}$ . We note that Zhu et al. (2023) use a different  $\Gamma_{\text{bg}}$  of  $1.5 \times 10^{-13} \text{ s}^{-1}$  at  $z \sim 6$  from Gaikwad et al. (2023). They find that at  $z \sim 5.93$ , assuming a value of  $3 \times 10^{-13} \text{ s}^{-1}$  for  $\Gamma_{\text{bg}}$  increases their mean free path from the measured value of  $0.81$  to  $\sim 1$  pMpc. The parameter  $\xi$  encodes the nature of the density of the absorbers that set the local MFP (Furlanetto & Oh 2005). For an isothermal absorber, the theoretical prediction is  $\xi = 2/3$ , while the scaling inferred from simulations is dependent on the self-shielding systems and can range between  $0.33$  and  $1.0$  (McQuinn et al. 2011; D’Aloisio et al. 2020). For  $\xi$ , we use the nominal parameter from B21 where it is assumed to be  $0.67$  based on the arguments presented in Miralda-Escudé et al. (2000). In Fig. 4, we compute the photoionization rate using equation (4) for a quasar at redshift 6 and magnitude



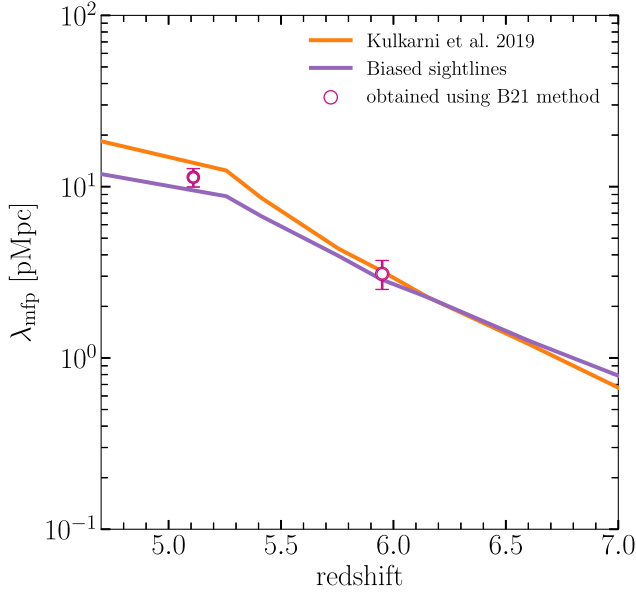
**Figure 5.** Distribution of the MFP obtained using 10 000 draws of 13-sightline stacks from our simulation for different quasar ages (top panel), and different quasar host halo mass (bottom panel). In each case, the MFP is determined using the B21 method, where  $R_{\text{eq}}$  and  $\Gamma_{\text{bg}}$  are chosen from our simulation and  $\xi = 0.67$ . As a comparison, the hatched region shows the 68 per cent scatter in the MFP in the simulation, obtained using equation (1).

$M_{1450} = -27$ .  $R_{\text{eq}}$  was computed using equation (9), assuming  $\Gamma_{\text{bg}}$  of  $\sim 3 \times 10^{-13} \text{ s}^{-1}$ . The resulting  $\Gamma_{\text{qso}}$  is compared to the  $\Gamma_{\text{qso}}$  from our simulations averaged over 1000 lines of sight to the QSO. The analytic  $\Gamma_{\text{qso}}$  is in agreement with our simulated value within 25 per cent for a  $\xi$  value of  $0.67$ . The analytic method, however, underestimates  $\Gamma_{\text{qso}}$  by almost two orders of magnitude close to the quasar for all values of  $\xi$ . Keeping  $\xi$  as a free parameter can change the MFP measurement by up to a factor of 2, as has been discussed in Zhu et al. (2023). Roth et al. 2024 discuss that keeping  $\xi$  as a free parameter yields a better fit to their mock spectra in the presence of neutral islands. They however find that such a fit might result in an MFP that is biased high with respect to the true value in their simulation.

#### 4.4 Recovery of the ‘true’ MFP using the B21 method

We use the B21 method to measure the MFP in our simulation using the mock data generated using the methods discussed in Section 4. We find that the resulting MFP agrees with the MFP of our simulation measured in Kulkarni et al. (2019). This shows that the MFP measured using the B21 method is not biased due to the residual neutral hydrogen islands, quasar lifetimes or due to the overdensities around quasars.

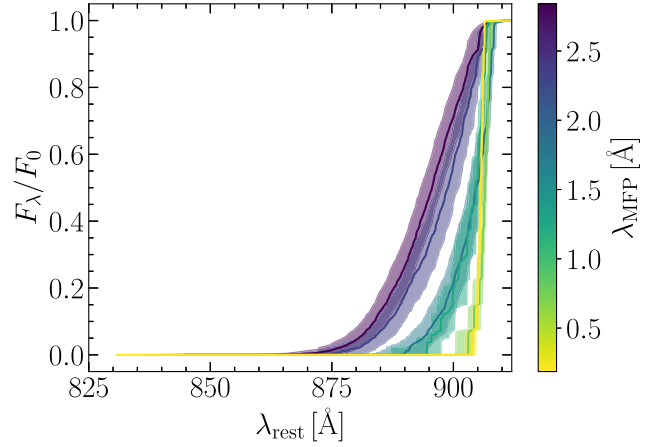
In Fig. 5, we show the distribution of MFP measured in our simulations along 10 000 stacks of 13 QSOs each. In the fiducial model, the MFP is measured along random sightlines without a QSO by simply fitting an exponential to the average flux as discussed in Section 3. To compute MFP from mock stacks including a QSO, we use the B21 method with parameters discussed in the previous



**Figure 6.** Open circles show the MFP inferred using the B21 method from a mock composite spectrum drawn from our simulation. The curves show the MFP in the simulation, obtained using equation (1), using sightlines starting from randomly chosen halos (orange curve), and sightlines starting from only the most massive haloes (purple curve). When we use the B21 method for measuring the MFP, we use  $\xi = 0.67$  and use values of  $R_{\text{eq}}$  and  $\Gamma_{\text{bg}}$  derived from the ionized regions in the simulation.

subsection. We consider mock stacks with different QSO properties, varying QSO lifetimes and varying host halo masses. The resulting distribution of MFP for 10 000 stacks each of 13 QSOs for different QSO lifetimes and host halo masses is shown in the top and bottom panels of Fig. 5, respectively. We find that irrespective of the quasar lifetime, we are able to recover the MFP from the mock stacks using the B21 method reasonably well, up to within  $\sim 10$  per cent of our fiducial value. Similarly, the MFP computed from mock stacks with lower mass haloes is in good agreement with the fiducial MFP. If we use a quasar stack constructed along sightlines originating in the heaviest mass haloes with  $10^{11} M_{\odot} \lesssim M_{\text{halo}} \lesssim 10^{12} M_{\odot}$ , the measured MFP distribution is offset from the fiducial value by  $\sim 25$  per cent.

We show the MFP measured from the 10 000 mock stacks for nominal values of quasar lifetime ( $10^6$  yr) and host halo mass ( $\sim 10^{12} M_{\odot}$ ) at redshifts  $z = 5.1$  and 6 in Fig. 6. Also shown are the MFPs along random sightlines and biased sightlines originating in haloes but without a QSO, as discussed in Section 3. The MFP recovered from our mock stacks using the B21 method matches better with the MFP measured along biased sightlines, which is particularly evident at post-reionization redshifts. This would suggest that the analytic computation of  $\Gamma_{\text{qso}}$  is robust in accounting for the QSO ionization flux, but is biased to measure the MFP along sightlines originating in overdensities. Note however that at higher redshifts, this bias becomes negligible. Another concern with the analytic model is the uncertainties introduced by the assumptions about the free parameters in the model  $R_{\text{eq}}$ ,  $\xi$ , and  $\Gamma_{\text{bg}}$ . Zhu et al. (2023) have shown that increasing or decreasing the value of  $\xi$  can half or double the MFP estimates relative to the estimates with nominal values. This raises a need for a better constrained value of the MFP independent of the choice of  $\xi$ . For this purpose, we use simulated models as discussed in the following section.



**Figure 7.** Stacked quasar spectra bluewards of rest-frame 912 Å in our model for a range of values of the MFP at  $z = 6$ . In each case, the solid curve shows the composite of 13 randomly chosen sightlines from the simulation box. The shaded region shows the 68 per cent scatter across 10 000 such samples.

## 5 A DIRECT MEASUREMENT OF THE MFP USING RT MODELS

We now proceed to analyse the composite spectrum of B21 using our RT models of patchy reionization and quasar proximity zones. Fig. 7 shows the composite spectra from 13 randomly drawn sightlines along a QSO with magnitude  $M_{1450} = -27.0$  and lifetime of 1 Myr in our simulation for a range of values of the mean free path. As our reionization simulation is calibrated to the observed Ly  $\alpha$  mean transmission, there is no freedom to change the mean free path at a given redshift. Instead, to construct Fig. 7, we use the values of the ionized hydrogen fraction, the gas density, and temperature from different redshifts, in each case scaling the densities by  $(1+z)^3$  to  $z = 6$ .

In order to fit the model stacks in Fig. 7 to the data, two modifications need to be made. First, we need to account for the absorption due to higher order Lyman series transitions that will cause absorption bluewards of the rest-frame quasar Lyman limit. Secondly, to match the observed flux, we need to consider the absorption due to the Lyman-series and continuum photons of the intrinsic QSO spectrum. Following B21, we assume the intrinsic quasar spectrum is a power law with  $f_{\lambda}^{\text{SED}} = f_{912}(\lambda/912 \text{ \AA})^{-\alpha_{\text{ion}}}$ , and keep  $f_{912}$  as a free parameter.

We include the foreground absorption due to higher order Lyman-series transitions as follows. At a given  $\lambda_{\text{obs}}$ , the absorption due to transition of the  $j$ th Lyman series line will happen due to an absorber at a redshift  $z_j = \lambda_{\text{obs}}/\lambda_j - 1$ . The effective Lyman series optical depth due to all transitions is then,

$$\tau_{\text{eff}}^{\text{Lyman}}(\lambda_{\text{obs}}) = \sum_j \tau_{\text{eff}}^j(z_j). \quad (10)$$

To obtain the optical depth due to the  $j$ th transition, we rescale the corresponding Ly  $\alpha$  optical depths at  $z_j$  by the product of their oscillator strength and rest-frame wavelength such that  $\tau_j = \tau_{\alpha} f_j \lambda_j / f_{\alpha} \lambda_{\alpha}$ . This scaling is valid in the optically thin regime corresponding to most of the redshifts of our interest, while in the damping wings the optical depths scale roughly as square of oscillator strengths (Malloy & Lidz 2015). We use the post-processed sightlines to compute the Ly  $\alpha$  optical depths in the presence of a quasar. The optical depths outside the proximity zone were assumed to be from the 40–2048 Sherwood simulation (Bolton et al. 2017), rescaled to match the

observed mean flux at low  $z$  as discussed in B21. From the mean flux at  $z_j$ , computed as negative exponential of  $\tau_j$ , the corresponding effective optical depth is obtained as  $\tau_{\text{eff}}^j = -\log(F)$ . Including the absorption from Lyman-limit and 38 higher order Lyman-series terms, the observed flux is

$$F_\lambda = f_{912} \left( \frac{\lambda}{912 \text{ \AA}} \right)^{-\alpha_{\text{ion}}} \exp(-\tau_{\text{eff}}^{\text{Lyman}}) \exp(-\tau_{\text{eff}}^{\text{LyC}}) + f_0, \quad (11)$$

where  $f_0$  is the zero-point correction factor that accounts for the uncertainty in the estimated zero-point of the data in B21. We fit our models to the data by sampling the posterior distribution in a Bayesian manner. We use the likelihood

$$\mathcal{L}(\mathbf{d}|\theta) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{C}}} \exp \left[ -\frac{1}{2} (\mathbf{m} - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{m} - \mathbf{d}) \right], \quad (12)$$

where the column matrices  $\mathbf{m}$  and  $\mathbf{d}$  denote the model and data vectors, respectively, both with  $n$  elements corresponding to the number of wavelength bins used. The covariance matrix  $\mathbf{C}$  is obtained as follows. The data covariance matrix  $\mathbf{C}_{\text{data}}$  over the wavelength range of our interest (912–850 Å) is computed from the data by B21, using bootstrap realisations of the mean flux. Due to the limited data set of 13 quasars, this matrix is noisy and may underestimate the sample variance. We therefore compute a separate bootstrapped model covariance matrix for each of the model parameters using 30 000 simulated stacks and obtain a model correlation matrix  $r$ . We then regularize the data covariance matrix  $\mathbf{C}_{\text{data}}$  using the model correlation matrix  $r$  to obtain the covariance matrix  $\mathbf{C}$  used in equation (12) at each point in the parameter space as

$$\mathbf{C}^{ij} = \frac{r_{\text{sim}}^{ij}}{\mathbf{C}_{\text{data}}^{ii} \mathbf{C}_{\text{data}}^{jj}}, \quad (13)$$

where

$$r_{\text{sim}}^{ij} = \frac{\mathbf{C}_{\text{sim}}^{ij}}{\mathbf{C}_{\text{sim}}^{ii} \mathbf{C}_{\text{sim}}^{jj}}. \quad (14)$$

See applications by Lidz et al. (2016), Viel et al. (2013), and Iršič et al. (2017) for a discussion of this approach.

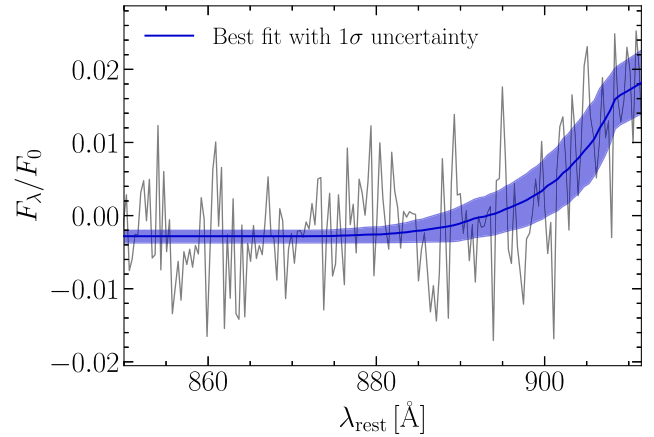
We assumed uniform priors in the ranges  $0.3 < \lambda_{\text{MFP}} < 2$ ,  $0 < f_{912} < 1$ , and  $-0.001 < f_0 < 0.01$ . The model stacks and the likelihood function were linearly interpolated over a grid of models wherever necessary. We used the EMCEE package (Foreman-Mackey et al. 2013) to perform the inference by sampling the posterior distribution using Markov chain Monte Carlo via Bayes' theorem,

$$p(\theta|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\theta)p(\theta). \quad (15)$$

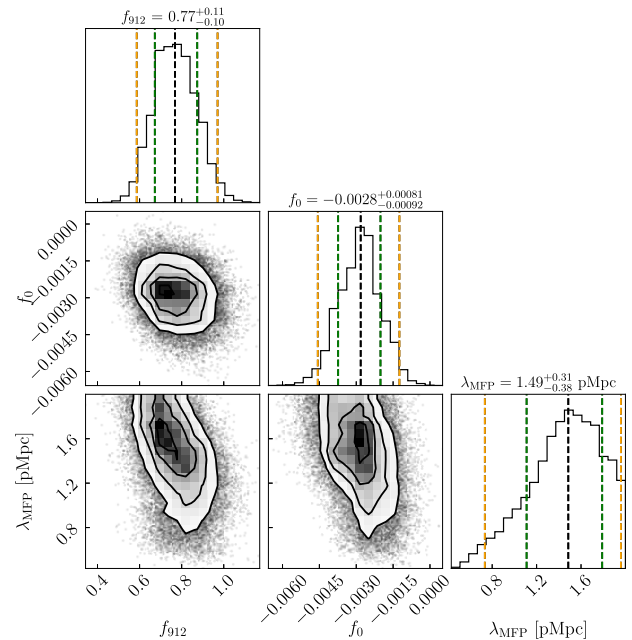
The resultant best-fitting curve, given by the mean of the posterior, is shown in Fig. 8, along with the  $1\sigma$  uncertainty. The marginalized posterior distributions of our model parameters is shown in Fig. 9.

We expect our assumptions for quasar lifetimes and host halo masses to introduce additional uncertainty on these parameters. We approximate these uncertainties by adding them in quadrature to the uncertainty derived from the posterior distributions. We saw in Fig. 5 that varying quasar lifetimes and host halo masses leads to a  $\sim 10$  per cent and  $\sim 25$  per cent change, respectively, in the mean free path. With this additional error included, our inferred value of MFP is  $\lambda_{\text{MFP}} = 1.49^{+0.47}_{-0.52}$  pMpc.

Fig. 10 shows the comparison of our new measurement with previous measurements and our simulation. While our new measurement is slightly higher than the MFP measured by B21, it is consistent with their result. Our measurement is also consistent with the other estimates by Gaikwad et al. (2023) and Zhu et al. (2023). Our new



**Figure 8.** Our best-fitting model for the B21 composite spectrum, with the  $1\sigma$  uncertainty indicated by the shaded region.



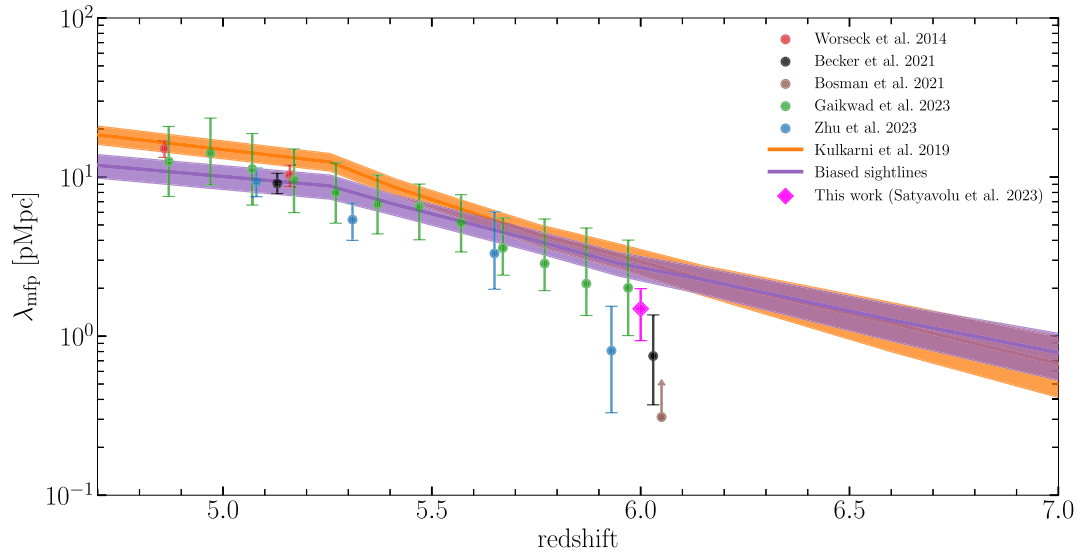
**Figure 9.** Corner plot showing the posterior distributions of the normalization parameters ( $f_{912}$ ,  $f_0$ ) and the mean free path ( $\lambda_{\text{MFP}}$ ). The rightmost panels in each row indicate the marginalized posteriors. The black, green, and yellow dotted lines represent the median,  $1\sigma$ , and  $2\sigma$  levels of the distribution, respectively.

measurement remains lower than the MFP of our simulations by more than  $2\sigma$ . None the less, the consistency between our inference and that of B21 suggests that the B21 method is robust with respect to the quasar proximity zone modelling.

## 6 CONCLUSIONS

In this paper, we have taken a closer look at the mean free path of hydrogen-ionizing photons at  $z = 6$  in order to examine potential sources of bias in recent measurements. Our findings are as follows:

- (i) At least for  $z \lesssim 6.5$ , there is no significant difference in the value of the mean free path obtained by using any of the multiple



**Figure 10.** The magenta point shows our measurement of the MFP using the B21 data with our model. The error bar on this point shows the  $1\sigma$  uncertainty. We see that this measurement is consistent with that of B21 which is shown by the black point. This suggests that the B21 measurement is robust against the sources of potential bias that we have considered in this paper. The orange curve with shaded regions shows the mean free path in the reionization simulations of Kulkarni et al. (2019). We see that our new measurement prefers a factor of 2 smaller MFP than this simulation. We also show other measurements of the MFP, by Worsack et al. (2014), Bosman (2021), Gaikwad et al. (2023), and Zhu et al. (2023). Measurements by Becker et al. (2021), Gaikwad et al. (2023), and Bosman (2021) have been displaced in redshift by  $\delta z = \pm 0.03$  and  $0.05$ , respectively, for legibility.

definitions of this quantity that have been used in the recent literature for redshifts at which reionization is still incomplete.

(ii) The bias in the mean free path due to the overdensities around quasars is also minimal at  $z \sim 6$ . At lower redshifts, with  $z \lesssim 5$ , the overdensities can bias the mean free path towards lower values by about 50 per cent.

(iii) Due to the short mean free path at  $z \sim 6$ , the Lyman-continuum composite spectra used in the literature for a direct MFP measurement are affected by the quasar proximity zones. The dependence of the inferred mean free path on the variations in quasar lifetime and host halo masses appears, however, only to be small and less than 25 per cent.

(iv) The patchiness of reionization also has minimal effect on the direct measurement of the MFP reported by B21.

(v) Using RT modelling of patchy reionization and quasar proximity zones, we reanalyse the data obtained by B21 to find an MFP of  $\lambda_{\text{MFP}} = 1.49^{+0.47}_{-0.52}$  pMpc at  $z = 6$ . This is nearly two times larger than the B21 measurement and smaller than the MFP in our reionization simulation by less than a factor of  $\sim 2$ .

We fit to the data while self-consistently modelling quasar proximity effect on the MFP in our simulations. We include the quasars through post-processing and hence do not consider the response of gas densities to the photoheating caused by the quasar ionization. However, the dynamical time-scale for the gas to respond ( $\sim 100$  Myr) is usually much larger than the average episodic lifetime of the quasar, which is around  $\sim 1$  Myr (D’Aloisio et al. 2020; Morey et al. 2021; Satyavolu et al. 2023). We also do not include feedback from AGN which could result in gas heating as well as AGN-driven clumpy outflows, although the former effect is shown to be not major (Trebitsch et al. 2018). The response of gas densities to heating from reionization might nonetheless be not captured in our main cosmological RT simulation, which was also run in post-processing. This would require running either hybrid or fully

coupled cosmological hydrodynamical simulations (e.g. Puchwein et al. 2023; Garaldi et al. 2022). We did not check the dependence of our result on the quasar spectral profile due to computational costs. The error on MFP due to this however is expected to be small (B21).

Cain et al. (2021) have modelled sinks in a subgrid fashion and found that the small-scale clumping reduces the MFP. There has also been evidence for presence of excess LLSs along the line of sight to quasars (Hennawi et al. 2006). We furthermore do not resolve the dense gas within mini-haloes in our simulation. Recent works show that the Ly  $\alpha$  flux is suppressed by 10 per cent on average due to mini-haloes, with the suppression being enhanced along lines of sight in the vicinity of large haloes, above redshifts  $z \gtrsim 5.5$  (Park et al. 2023). Inclusion of dense absorbing gas such as LLSs or mini-haloes in our simulations therefore will not only reduce the MFP along the random sightlines shown in orange in Fig. 1, but also further reduce the continuum flux along quasar sightlines. This would push our measured MFP in Fig. 10 to higher than its current value.

In summary, the shortness of the mean free path relative to RT simulation could be due to unresolved excess LLSs along quasar sightlines. Consequently, the overall picture that emerges is consistent with one in which reionization ends much later than redshift 6, with the photon budget dominated by faint galaxies with high ionizing efficiency (Davies et al. 2021; Cain et al. 2021).

## ACKNOWLEDGEMENTS

We thank George Becker and Yongda Zhu for providing data and for valuable discussions. We also thank James Bolton for sharing Sherwood data. We thank Shikhar Asthana, Anson D’Aloisio, Christopher Cain, Prakash Gaikwad, Joe Hennawi, Vid Iršič, and Nabendu Kumar Khan for useful discussions. GK is partly supported by the Department of Atomic Energy (Government of India) research project with Project Identification Number RTI 4002, and by the Max Planck Society through a Max Planck Partner Group. This work

was supported by grants from the Swiss National Supercomputing Centre (CSCS) under project IDs s949 and s1114. This work used the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge ([www.csd3.cam.ac.uk](http://www.csd3.cam.ac.uk)), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council ([www.dirac.ac.uk](http://www.dirac.ac.uk)). This work further used the COSMA Data Centric system operated Durham University on behalf of the STFC DiRAC HPC Facility. This equipment was funded by a BIS National E-infrastructure capital grant ST/K00042X/1, DiRAC Operations grant ST/K003267/1, and Durham University. DiRAC is part of the UK's National E-Infrastructure. Support by ERC Advanced Grant 320596 'Emergence' is gratefully acknowledged. For the purpose of open access, LK has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

## DATA AVAILABILITY

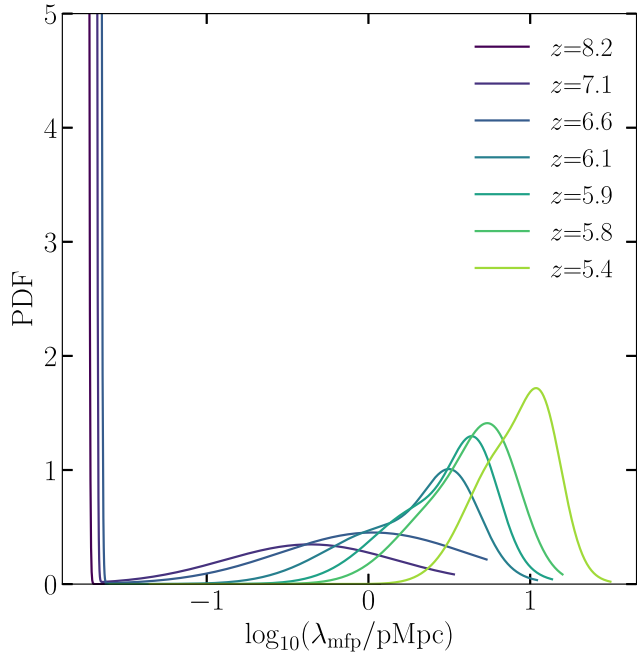
The data and code underlying this article will be shared on reasonable request to the corresponding author.

## REFERENCES

- Aubert D., Teyssier R., 2008, *MNRAS*, 387, 295  
Aubert D., Teyssier R., 2010, *ApJ*, 724, 244  
Bauer A., Springel V., Vogelsberger M., Genel S., Torrey P., Sijacki D., Nelson D., Hernquist L., 2015, *MNRAS*, 453, 3593  
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, *MNRAS*, 447, 3402  
Becker G. D., D'Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, *MNRAS*, 508, 1853  
Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T.-S., Meiksin A., Regan J. A., Viel M., 2017, *MNRAS*, 464, 897  
Bosman S. E. I. et al., 2022, *MNRAS*, 514, 55  
Bosman S. E. I., 2021, preprint ([arXiv:2108.12446](https://arxiv.org/abs/2108.12446))  
Cain C., D'Aloisio A., Gangolli N., Becker G. D., 2021, *ApJ*, 917, L37  
Calverley A. P., Becker G. D., Haehnelt M. G., Bolton J. S., 2011, *MNRAS*, 412, 2543  
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. et al., 2018, *ApJ*, 864, 142  
Davies F. B. et al., 2024, *ApJ*, 965, 134  
Davies F. B., Bosman S. E. I., Furlanetto S. R., Becker G. D., D'Aloisio A., 2021, *ApJ*, 918, L35  
Davies F. B., Furlanetto S. R., 2016, *MNRAS*, 460, 1328  
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306  
Fumagalli M., O'Meara J. M., Prochaska J. X., Worseck G., 2013, *ApJ*, 775, 78  
Furlanetto S. R., Oh S. P., 2005, *MNRAS*, 363, 1031  
Gaikwad P. et al., 2020, *MNRAS*, 494, 5091  
Gaikwad P. et al., 2023, *MNRAS*, 525, 4093  
Garaldi E., Kannan R., Smith A., Springel V., Pakmor R., Vogelsberger M., Hernquist L., 2022, *MNRAS*, 512, 4909  
Gnedin N. Y., Kaurov A. A., 2014, *ApJ*, 793, 30  
Greig B., Mesinger A., Bañados E., 2019, *MNRAS*, 484, 5094  
Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, *MNRAS*, 466, 4239  
Hennawi J. F. et al., 2006, *ApJ*, 651, 61  
Inoue A. K., Shimizu I., Iwata I., Tanaka M., 2014, *MNRAS*, 442, 1805  
Iršič V. et al., 2017, *Phys. Rev. D*, 96, 023522  
Keating L. C., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020b, *MNRAS*, 497, 906  
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020a, *MNRAS*, 491, 1736  
Kulkarni G., Choudhury T. R., Puchwein E., Haehnelt M. G., 2016, *MNRAS*, 463, 2583  
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, *MNRAS*, 485, L24  
Lewis J. S. W. et al., 2022, *MNRAS*, 516, 3389  
Lidz A., Heitmann K., Hui L., Habib S., Rauch M., Sargent W. L. W., 2006, *ApJ*, 638, 27  
Lusso E., Worseck G., Hennawi J. F., Prochaska J. X., Vignali C., Stern J., O'Meara J. M., 2015, *MNRAS*, 449, 4204  
Madau P., 2017, *ApJ*, 851, 50  
Malloy M., Lidz A., 2015, *ApJ*, 799, 179  
McQuinn M., Oh S. P., Faucher-Giguère C.-A., 2011, *ApJ*, 743, 82  
Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, *ApJ*, 530, 1  
Morey K. A., Eilers A.-C., Davies F. B., Hennawi J. F., Simcoe R. A., 2021, *ApJ*, 921, 88  
Muñoz J. A., Oh S. P., Davies F. B., Furlanetto S. R., 2016, *MNRAS*, 455, 1385  
O'Meara J. M., Prochaska J. X., Burles S., Prochter G., Bernstein R. A., Burgess K. M., 2007, *ApJ*, 656, 666  
O'Meara J. M., Prochaska J. X., Worseck G., Chen H.-W., Madau P., 2013, *ApJ*, 765, 137  
Park H., Lukić Z., Sexton J., Alvarez M., 2023, *ApJ*, 969, 46  
Planck Collaboration VI, 2020, *A&A*, 641, A6  
Planck Collaboration XVI, 2014, *A&A*, 571, A16  
Prochaska J. X., Worseck G., O'Meara J. M., 2009, *ApJ*, 705, L113  
Puchwein E. et al., 2023, *MNRAS*, 519, 6162  
Rahmati A., Schaye J., 2018, *MNRAS*, 478, 5123  
Romano M., Grazian A., Giallongo E., Cristiani S., Fontanot F., Boutsia K., Fiore F., Menci N., 2019, *A&A*, 632, A45  
Roth J. T. et al., 2024, *MNRAS*, 530, 5209  
Rybicki G. B., Lightman A. P., 1985, *Radiative Processes in Astrophysics*. Wiley, New Jersey  
Satyavolu S., Kulkarni G., Keating L. C., Haehnelt M. G., 2023, *MNRAS*, 521, 3108  
Shukla H., Mellema G., Iliev I. T., Shapiro P. R., 2016, *MNRAS*, 458, 135  
Songaila A., Cowie L. L., 2010, *ApJ*, 721, 1448  
Springel V., 2005, *MNRAS*, 364, 1105  
Storrie-Lombardi L. J., McMahon R. G., Irwin M. J., Hazard C., 1994, *ApJ*, 427, L13  
Trebitsch M., Volonteri M., Dubois Y., Madau P., 2018, *MNRAS*, 478, 5607  
Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, *Phys. Rev. D*, 88, 043502  
Wang F. et al., 2020, *ApJ*, 896, 23  
Wang F. et al., 2023, *ApJ*, 951, L4  
Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, *MNRAS*, 485, 1350  
Weinberger L. H., Kulkarni G., Haehnelt M. G., Choudhury T. R., Puchwein E., 2018, *MNRAS*, 479, 2564  
Wolfson M., Hennawi J. F., Davies F. B., Oñorbe J., 2023, *MNRAS*, 521, 4056  
Worseck G. et al., 2014, *MNRAS*, 445, 1745  
Zhu Y. et al., 2022, *ApJ*, 932, 76  
Zhu Y. et al., 2023, *ApJ*, 955, 115

## APPENDIX A: DISTRIBUTION OF FREE PATHS

Fig. A1 shows the distribution of free paths in our simulation between redshifts 8 and 5. The divergence seen at redshifts  $z < 6$  are numerical artefacts arising from the short free paths limited by the resolution of our simulation.

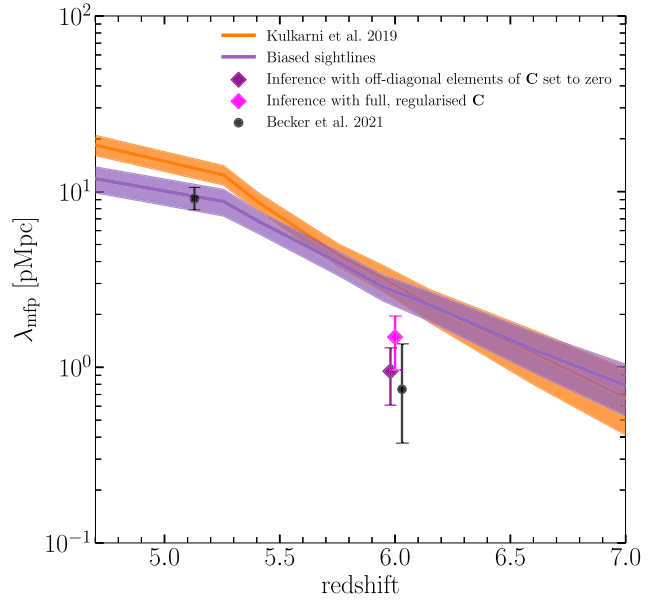


**Figure A1.** Distribution of free paths across redshifts. At redshifts  $z < 6$ , the free path length distribution becomes bimodal with a large number of free paths  $< 0.01$  pMpc, smaller than the resolution of our base simulation at these redshifts.

## APPENDIX B: EFFECT OF OFF-DIAGONAL TERMS IN THE COVARIANCE MATRIX

To understand the effect of the off-diagonal terms in the covariance matrix, we repeat our analysis of Section 5 by fitting our model to the stacked data of B21 using only the diagonal elements of the data covariance matrix provided by B21. The full, unregularized, data covariance matrix is nearly diagonal, but has some non-zero off-diagonal elements, particularly around  $\sim 912 \text{ \AA}$ . Setting all off-diagonal terms to zero results in an inference shown by the purple diamond in Fig. B1. The magenta diamond in this figure, on the other hand, shows the result of the analysis from Section 5. We see that

the off-diagonal elements lead to a larger mean value of the posterior distribution and a narrower posterior distribution, resulting in lower  $1\sigma$  uncertainty.



**Figure B1.** Effect of the off-diagonal terms in the covariance matrix  $\mathbf{C}$  on our inferred value of the mean free path. Magenta diamond shows the mean of the posterior distribution computed using the full, regularized covariance matrix as described in the text. Also shown is the  $1\sigma$  uncertainty. The purple diamond shows the same for the case in which only the diagonal elements of the covariance matrix are used while computing the likelihood. In this case the off-diagonal terms are set to zero. The black points show the B21 measurements. We see that the off-diagonal elements lead to a larger mean value of the posterior distribution and a narrower posterior distribution, resulting in lower  $1\sigma$  uncertainty. The three MFP measurements shown at  $z \sim 6$  are at  $z = 6$ , but we have shown them with small displacements in the redshift directions for better legibility.

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