Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **532,** 2401–2417 (2024) https://doi.org/10.1093/mnras/stae1636
Advance Access publication 2024 July 10

The Lyman-limit photon mean free path at the end of late reionization in the Sherwood–Relics simulations

Jennifer Feron [1]★ Luke Conaboy [1] James S. Bolton [1] Emma Chapman [1] Martin G. Haehnelt [2] Laura C. Keating [3] Girish Kulkarni [4] and Ewald Puchwein [5]

[1] *School of Physics and Astronomy, The University of Nottingham, University Park, Nottingham NG7 2RD, UK*
[2] *Kavli Institute for Cosmology and Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK*
[3] *Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK*
[4] *Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*
[5] *Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany*

Accepted 2024 July 1. Received 2024 June 27; in original form 2024 May 16

ABSTRACT

Recent evidence supporting reionization ending at redshift $z \lesssim 6$ includes the rapid redshift evolution of the mean free path, $\lambda_{\rm mfp}$, for Lyman-limit photons through the intergalactic medium (IGM) at $5 < z < 6$. Here, we investigate $\lambda_{\rm mfp}$ predicted by the Sherwood–Relics suite of hybrid radiation hydrodynamical simulations. Simulations with comoving volumes of $40^3\,h^{-3}$ cMpc^3 ($160^3\,h^{-3}$ cMpc^3), calibrated to match the observed Ly α forest transmission with a late end to reionization at $z < 6$, are consistent with recent $\lambda_{\rm mfp}$ measurements at $z < 5.9$, and are 1.2σ (1.8σ) above the highest redshift $\lambda_{\rm mfp}$ measurement at $z = 5.93$. The majority of the Lyman-limit opacity at the end of reionization is attributable to highly ionized Ly α forest absorbers with neutral hydrogen column densities $N_{\rm HI} \leq 10^{16}$–10^{17} cm^{-2}. Resolving these systems is critical for capturing the redshift evolution of $\lambda_{\rm mfp}$ during the final stages of reionization. After reionization completes, overdense gas will reduce $\lambda_{\rm mfp}$ by up to 20 per cent around haloes with masses $M_{\rm h} \sim 10^9$–$10^{11} h^{-1}$ M$_\odot$, but during reionization ionized bubbles will instead boost $\lambda_{\rm mfp}$ around haloes by up to an order of magnitude when the IGM is as much as 90 per cent neutral by volume. This effect will play an important role in the visibility of Ly α emitting galaxies at $z \gtrsim 10$ discovered with *JWST*.

Key words: methods: numerical – intergalactic medium – quasars: absorption lines – large scale structure of Universe.

1 INTRODUCTION

There are still many unanswered questions about the epoch of reionization, including the precise nature of the ionizing sources (e.g. Finkelstein et al. 2019; Naidu et al. 2020; Grazian et al. 2023; Atek et al. 2024), and the timing and extent of reionization. At present, many observations indicate that reionization has a mid-point around $z \sim 7$–8, with the final neutral islands in the diffuse intergalactic medium (IGM) persisting until at least $z \sim 6$ (e.g. Davies et al. 2018; Mason et al. 2018; Kulkarni et al. 2019; Yang et al. 2020; Qin et al. 2021; Umeda et al. 2023; Ďurovčíková et al. 2024; Nakane et al. 2024). The last stages of reionization may even extend as late as $z \sim 5.3$, based on the observed fluctuations in the Ly α forest transmission (Becker et al. 2015; Bosman et al. 2018; Eilers, Davies & Hennawi 2018; Bosman et al. 2022) and the incidence of dark gaps and pixels in the Ly α and Ly β forests (Zhu et al. 2022; Jin et al. 2023).

Further evidence in support of an end to reionization at $z \lesssim 6$ is provided by the redshift evolution of the mean free path of Lyman-limit photons through the IGM (e.g. Prochaska, Worseck & O'Meara 2009; Fumagalli et al. 2013; O'Meara et al. 2013; Worseck et al. 2014; Lusso et al. 2018; Romano et al. 2019). Recent measurements have demonstrated the mean free path declines by an order of magnitude from $z = 5$ to $z = 6$ (Becker et al. 2021; Bosman 2021; Gaikwad et al. 2023; Zhu et al. 2023; Davies et al. 2024). This evolution is expected if reionization ends below or near to $z = 6$; the mean free path should increase rapidly as ionized hydrogen bubbles merge. The background photoionization rate will then increase and the incidence of Lyman-limit photon sinks will decrease (e.g. Wyithe, Bolton & Haehnelt 2008; Park et al. 2016; Cain et al. 2021; Davies et al. 2021; Nasir et al. 2021; Theuns & Chan 2024).

Intriguingly, recent numerical simulations of inhomogeneous reionization have exhibited mixed success in reproducing the observed rapid decline in the mean free path approaching $z \simeq 6$. Using radiative transfer simulations applied to a hydrodynamical simulation in post-processing, Keating et al. (2020a, b) found a mean free path evolution that was too gradual, matching the data at $z \sim 5.5$ but overshooting at $z \sim 6$. The Keating et al. (2020a, b) model was calibrated to reproduce the large-scale fluctuations observed in the Ly α forest transmission at $z < 6$. In contrast, using the Cosmic Dawn (CoDa) III radiation hydrodynamical simulation, Lewis et al. (2022) found good agreement with the observed mean free path evolution. Notably, the CoDa III simulation was performed at substantially better mass resolution (with a dark matter particle mass $M_{\rm dm} = 5.09 \times 10^4$ M$_\odot$ in a 94.4 cMpc box) compared to the Keating et al. (2020b) model ($M_{\rm dm} = 5.1 \times 10^7$ M$_\odot$ in a 236 cMpc

★ E-mail: jennifer.feron@nottingham.ac.uk

box). However, CoDa III was not calibrated to reproduce the post-reionization Ly α forest transmission. Intermediate in scale between these two cases is the THESAN radiation hydrodynamical simulation (Garaldi et al. 2022), with $M_{\rm dm} = 3.12 \times 10^6\,{\rm M}_\odot$ and a box size of 95.5 cMpc. The redshift evolution of the mean free path in THESAN was found to be in good agreement with the observational data, but with a normalization that lies slightly above the 1σ uncertainty on the Becker et al. (2021) and Zhu et al. (2023) measurements at $z = 6$. In all three of these independent simulations, a late end to reionization at $z < 6$ was assumed. Finally, the CROC radiation hydrodynamical simulations (Gnedin 2014) used a box size of 59 cMpc and a dark matter particle mass $M_{\rm dm} = 7 \times 10^6\,{\rm M}_\odot$. Fan et al. (2024) recently demonstrated that CROC overshoots the mean free path measurements at $z \simeq 6$, most likely because the IGM is already highly ionized (with volume averaged H I fraction $\langle x_{\rm HI}\rangle < 10^{-3}$) by $z \sim 6.4$ in these models.

One possible reason for differences between observation and theory is if the mean free path measurements at $z \simeq 6$ suffer a systematic bias. The measurements presented by Becker et al. (2021) and Zhu et al. (2023) at $z \simeq 6$ are determined from the Lyman continuum opacity in the vicinity of bright quasars. Biases may arise if there is an overcorrection for the quasar proximity effect (D'Aloisio et al. 2018), or an undercorrection for the clustering of ionizing photon sinks around the quasar host haloes (Prochaska et al. 2014; Theuns & Chan 2024). However, Satyavolu et al. (2023) and Roth et al. (2024) have recently investigated a range of potential biases in the mean free path measurements, finding them to be well controlled within the analysis framework used by Becker et al. (2021) and Zhu et al. (2023). Satyavolu et al. (2023) also found no significant difference in mean free path values when comparing the various definitions employed in the recent literature.

Another possibility is the simulations lack the dynamic range necessary for capturing the redshift evolution of the mean free path; the large simulation volumes needed to model the size distribution of ionized bubbles during reionization (e.g. Iliev et al. 2014; Kaur, Gillet & Mesinger 2020) are too low resolution to simultaneously resolve small-scale photon sinks. Cain et al. (2021) suggested that improved agreement with the mean free path measurements may be obtained by invoking a subgrid model for the ionizing photon sinks. These sinks may be associated with unresolved Lyman-limit systems and/or mini-haloes with masses $< 10^8\,{\rm M}_\odot$ (Park et al. 2023). Furthermore, earlier work by Rahmati & Schaye (2018) using the Aurora suite of radiation hydrodyanamical simulations[1] highlighted the importance of H I absorption systems with $N_{\rm H\,I} \sim 10^{16}$–$10^{17}\,{\rm cm}^{-2}$ for setting the mean free path during the final stages of reionization (i.e. strong Ly α forest absorbers that arise from the diffuse, already reionized IGM, rather than dense neutral clumps). Nasir et al. (2021) similarly argued that gas that has dynamically responded to the increased thermal pressure from photoheating will dominate the Lyman-limit opacity by $z \simeq 6$, and this gas will typically have H I column densities $N_{\rm H\,I} < 10^{17.2}\,{\rm cm}^{-2}$. Hence, resolving structure in the diffuse IGM during the end stages of reionization is a key requirement for capturing the evolution of the mean free path in simulations of inhomogeneous reionization.

In this context, we examine the Lyman-limit photon mean free path predicted by a subset of the Sherwood–Relics[2] simulations (Puchwein et al. 2023). These are a large suite of simulations that model the effect of inhomogeneous reionization on the high-redshift IGM and Ly α forest, including the hydrodynamical response of intergalactic gas to patchy photoionization and heating. In contrast to other recent numerical simulation work, the Sherwood–Relics simulations use a hybrid approach that combines radiative transfer calculations performed using ATON (Aubert & Teyssier 2008) with P-GADGET-3 cosmological hydrodynamical simulations (Springel 2005). In addition, because Sherwood–Relics was designed to study the small-scale structure in the Ly α forest (e.g. Molaro et al. 2023; Iršič et al. 2024), the mass resolution of our fiducial model ($M_{\rm dm} = 7.9 \times 10^5\,{\rm M}_\odot$ in a 59 cMpc box, which lies between CoDa III and THESAN) has been selected with this specific goal in mind (cf. Bolton & Becker 2009; Doughty et al. 2023).

This paper is organized as follows. In Section 2, we introduce the simulations used in this work and perform an initial comparison of the Lyman-limit photon mean free path predicted by Sherwood–Relics to the observational data. We investigate the physical properties of the absorption systems that set the mean free path in Section 3, and examine the expected bias in the mean free path around the host haloes of ionizing sources in Section 4. Finally, we conclude in Section 5. A set of appendices contain some numerical tests and a discussion of the analytical scaling relations commonly used for modelling the Lyman-limit opacity.

2 MODELLING THE MEAN FREE PATH WITH SHERWOOD–RELICS

2.1 Hydrodynamical simulations

In this work, we use a subset of the Sherwood−Relics simulations, described in detail in Puchwein et al. (2023). These are a suite of high-resolution cosmological hydrodynamical simulations performed with a modified version of the P-GADGET-3 code (Springel 2005). We use cosmological boxes with size $40\,h^{-1}$ cMpc and $160\,h^{-1}$ cMpc, each with 2×2048^3 dark matter and gas particles. We refer to these simulations as 40–2048 and 160–2048, respectively (see Table 1). The 160–2048 model uses the same mass resolution, box size and initial conditions as the (hydrodynamically decoupled) ATON simulations presented by Keating et al. (2020a, b), although in this work we assume a slightly different reionization history. In addition, we use two smaller simulations of size $40\,h^{-1}$ cMpc with 2×512^3 and 2×1024^3 dark matter and gas particles, which we refer to as 40–512 and 40–1024, respectively. These models are identical to the 40–2048 run except for particle masses that are 64 and 8 times larger. The mass resolution of the 40–512 model matches the mass resolution of 160–2048.

The Sherwood–Relics simulations follow inhomogeneous reionization using a novel hybrid approach. We refer the reader to Puchwein et al. (2023) for further details on the numerical scheme. In brief, the radiative transfer of monochromatic, ultraviolet photons in the simulations is followed using the moment-based M1-closure radiative transfer code ATON (Aubert & Teyssier 2008, 2010). Rather than performing full radiation-hydrodynamical (RHD) simulations, however, we take an intermediate approach, where pre-generated three-dimensional maps of the H I photoionization rates from ATON are applied on-the-fly to the hydrodynamical simulations. This has the advantage of self-consistently following the pressure response of the intergalactic gas to reionization, but without the computational overhead of a full RHD calculation that also incorporates a detailed subgrid model for star formation and feedback.

Gas particles with temperature $T < 10^5\,{\rm K}$ and density $\Delta = \rho/\langle\rho\rangle > 1000$ are immediately converted into star particles in all

[1]The fiducial Aurora simulation used a dark matter particle mass, $M_{\rm dm} = 1 \times 10^7\,{\rm M}_\odot$ and a box size 35.2 cMpc.
[2]https://www.nottingham.ac.uk/astronomy/sherwood-relics/

Table 1. Summary of the Sherwood–Relics simulations (Puchwein et al. 2023) used in this work. The columns, from left to right, list the model name, box size in h^{-1} cMpc, the gas and dark matter particle number, the gravitational softening length in h^{-1} ckpc, the dark matter and gas particle masses in $h^{-1}\mathrm{M}_\odot$, the redshift, z_r, where the volume averaged neutral hydrogen fraction first falls below $\langle x_\mathrm{HI}\rangle = 10^{-3}$, and the mid-point of reionization, z_mid, where $\langle x_\mathrm{HI}\rangle = 0.5$.

Name	L_box (h^{-1} cMpc)	N_part	l_soft (h^{-1} ckpc)	M_dm ($h^{-1}\mathrm{M}_\odot$)	M_gas ($h^{-1}\mathrm{M}_\odot$)	z_r	z_mid
40–2048	40	2×2048^3	0.78	5.37×10^5	9.97×10^4	5.7	7.5
160–2048	160	2×2048^3	3.13	3.44×10^7	6.38×10^6	5.3	7.2
40–1024	40	2×1024^3	1.56	4.30×10^6	7.97×10^5	5.7	7.5
40–512	40	2×512^3	3.13	3.44×10^7	6.38×10^6	5.7	7.5

models (the 'quick Ly α' approximation; Viel, Haehnelt & Springel 2004). The simulations will therefore not correctly capture the incidence of very high density gas in haloes, but this will only affect the mean free path once the Lyman-limit opacity is dominated by gas within the virial radius of haloes (i.e. well after reionization has completed). A test of this 'quick Ly α' approximation and its effect on the mean free path is provided in Appendix A. The mass resolution of our main simulation, 40–2048, has been chosen to adequately resolve small-scale structure in the low-density IGM, while also maintaining a large enough cosmological volume to sample a good range of halo masses. We will use the 160–2048 and 40–512 models to assess the effect of simulation volume and mass resolution on our results. For all simulations, a ΛCDM cosmology with $\Omega_\mathrm{m} = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_\mathrm{b} = 0.0482$, $h = 0.678$, $\sigma_8 = 0.829$, and $n = 0.961$ is assumed (Planck Collaboration XVI 2014).

The luminosity of H I photoionizing sources in the simulations is proportional to the total halo mass, where a minimum host halo mass of $M_\mathrm{h} > 10^9 h^{-1}\mathrm{M}_\odot$ is assumed. The mean energy for the ionizing photons is 18.6 eV, corresponding to a black body spectrum with $T = 40,000$ K. Following the approach used in earlier work by Kulkarni et al. (2019) and Keating et al. (2020a), two of the models we use here have been calibrated to match observational measurements of the mean transmitted flux of the Ly α forest. The 40–2048 simulation was calibrated to match the results from Bosman et al. (2018) and Eilers et al. (2018), whereas the 160–2048 simulation was calibrated to match the more recent Bosman et al. (2022) measurements. Despite the overlap between the different Ly α forest measurements; however, the 40–2048 and 160–2048 simulations have slightly different reionization histories, with mid-points (i.e. where the volume average neutral hydrogen fraction $\langle x_\mathrm{HI}\rangle = 0.5$) at $z_\mathrm{mid} = 7.5$ and $z_\mathrm{mid} = 7.2$, respectively. This difference is attributable to the different simulation volumes (see Section 2.2). The 40–512 and 40–1024 models use the same reionization history as the 40–2048 model.

2.2 Comparison of the simulated mean free path to observations

A variety of approaches have been used to compute the Lyman-limit mean free path in the literature (see Satyavolu et al. 2023 for a recent discussion). In this work, we follow Rahmati & Schaye (2018) and use the mean of the free path distribution obtained from the simulation volume. Satyavolu et al. (2023) demonstrated this yields a mean free path in good agreement with the e-folding scale obtained by fitting the average Lyman continuum transmission with an exponential (e.g. Prochaska et al. 2009).

The free path, λ_fp, is defined as the distance over which a Lyman-limit photon ($\lambda = 912$ Å) will encounter an optical depth, τ_{912}, equal to unity. For a sight-line, i, drawn a proper distance R parallel to the Cartesian axes of the periodic simulation volume, we require

$$\tau_{912,\mathrm{i}} = \sigma_{912} \int_0^{\lambda_\mathrm{fp,i}} n_\mathrm{H\,I}(R)\,\mathrm{d}R = 1, \tag{1}$$

where $\sigma_{912} = 6.348 \times 10^{-18}\,\mathrm{cm}^2$ is the photoionization cross-section (Verner et al. 1996). We choose the starting position of each sight-line in the simulation at random. If the distance required to achieve an optical depth of unity exceeds the box size, we add further randomly selected sight-lines on to the first. The mean free path, λ_mfp, is then the average free-path over an ensemble of N sight-lines

$$\lambda_\mathrm{mfp} \equiv \langle \lambda_\mathrm{fp,i}\rangle = \frac{1}{N}\sum_\mathrm{i=0}^{\mathrm{N}} \lambda_\mathrm{fp,i}. \tag{2}$$

Throughout this work, we use between $N = 10^4$–10^6, which provides a well converged measurement.

The 40–2048 and 160–2048 simulations are compared to observational constraints in Fig. 1. Clockwise from the top panel, these are: the mean free path, λ_mfp, for Lyman limit photons, the ionizing emissivity, $\dot{n}$, the volume averaged H I photoionization rate, $\langle\Gamma_\mathrm{H\,I}\rangle$, the volume averaged ionized hydrogen fraction, $1 - \langle x_\mathrm{H\,I}\rangle$, and the mean transmission in the Ly α forest, $\langle F\rangle$. The good agreement with the observed mean Ly α forest transmission at $z > 5$ in the middle left panel is by design, and is not a prediction of the models. Both simulations have a late end to reionization, at $z < 6$ (see Table 1), and are consistent with observational measurements of the ionized hydrogen fraction and the H I photoionization rate at $z > 5$. The ionizing emissivity drops to $\dot{n} \simeq 3 \times 10^{50}\,\mathrm{s}^{-1}\,\mathrm{cMpc}^{-3}$ at $z < 6$ in both models, where $\dot{n} \propto \Gamma_\mathrm{H\,I}/\lambda_\mathrm{mfp}$. This behaviour is qualitatively similar to other recent reionization models (Kulkarni et al. 2019; Keating et al. 2020a, b; Cain et al. 2021; Gaikwad et al. 2023; Asthana et al. 2024) and is driven by mandating the simulations reproduce the observed Ly α forest transmission. An ionizing emissivity of $\dot{n} \simeq 3 \times 10^{50}\,\mathrm{s}^{-1}\,\mathrm{cMpc}^{-3}$ corresponds to only 1–2 ionizing photons per hydrogen atom emitted over the age of the Universe at $z = 6$ (Miralda-Escudé 2003; Bolton & Haehnelt 2007). Given that the collapsed mass fraction in haloes will monotonically rise towards lower redshift, the apparent need for this rather low ionizing emissivity at $z = 6$ implies some redshift evolution in the escape fraction or efficiency of ionizing photon sources at $z > 6$ (Ocvirk et al. 2021). Alternatively, models of the IGM approaching the end stages of reionization may still be incomplete (see Cain et al. 2024, for a recent discussion).

The redshift evolution of the mean free path in the simulations (see also Table 2) is in good agreement (within 1σ) with the recent (indirect) determination of the mean free path from the Ly α forest transmission at $4.9 < z < 6$ (Gaikwad et al. 2023). Similar λ_mfp measurements have also recently been independently reported by Davies et al. (2024). The models are also in good agreement with the mean free path measured from the Lyman continuum transmission in stacked quasar spectra at $z < 5.7$ (Worseck et al. 2014; Becker

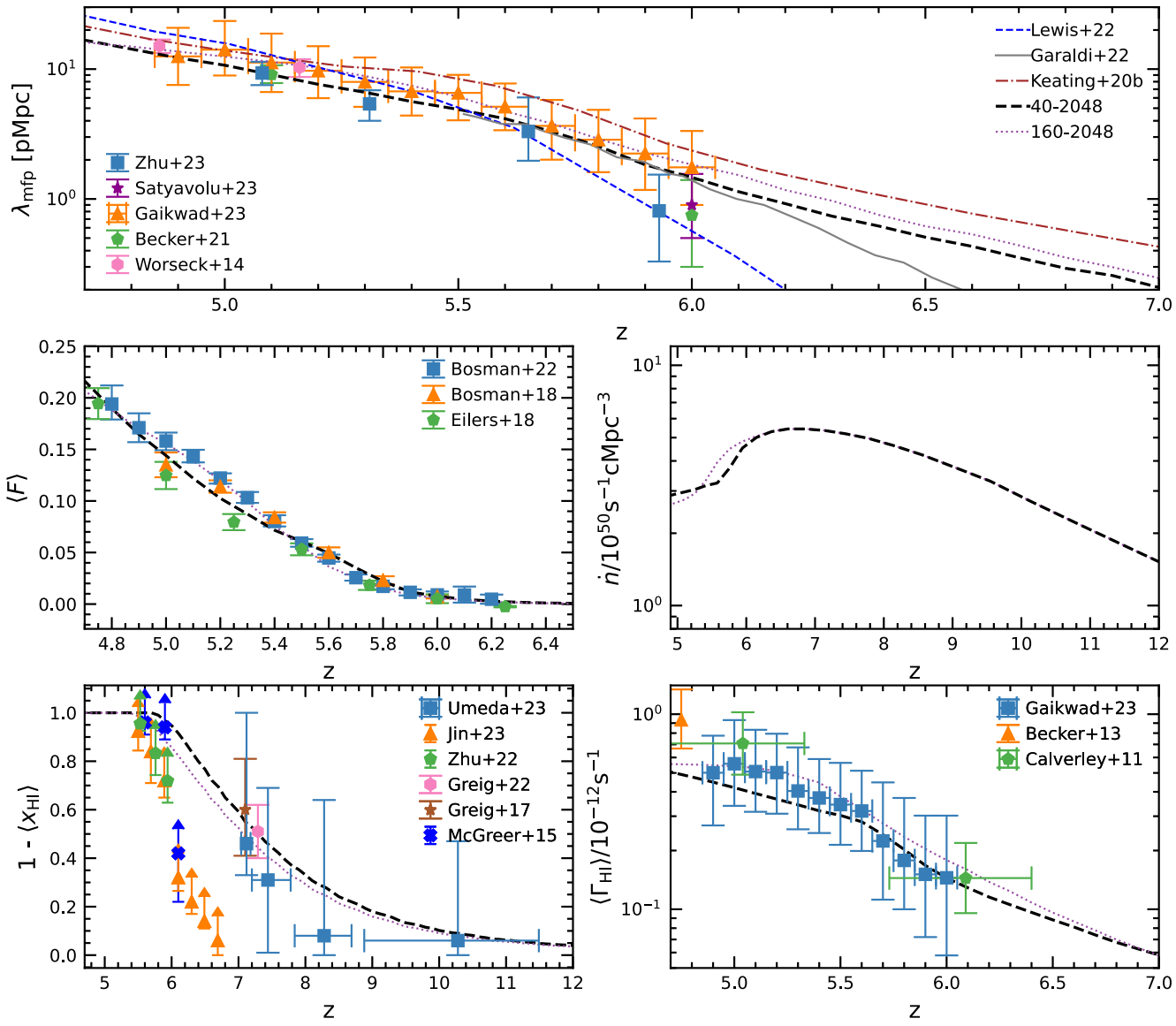

Figure 1. Comparison of two simulations used in this work, 40–2048 (black dashed curves) and 160–2048 (violet dotted curves) to observational data. *Top:* The Lyman-limit photon mean free path. The observational data are obtained from the Lyman-limit transmission in stacked quasar spectra (Prochaska et al. 2009; Worseck et al. 2014; Becker et al. 2021; Zhu et al. 2023) and (indirectly) from the Ly α forest (Gaikwad et al. 2023). All error bars are 1σ. Predictions from other recent simulations are the low τ_{CMB} model in Keating et al. (2020b) (brown dot–dashed curve), THESAN-1 (Garaldi et al. 2022, solid grey curve), and CoDa III (Lewis et al. 2022; blue dashed curve). *Middle left:* The mean transmission of the Ly α forest against redshift (Bosman et al. 2018; Eilers et al. 2018; Bosman et al. 2022). The 40–2048 model was calibrated to match Eilers et al. (2018) and Bosman et al. (2018), whereas the 160–2048 simulation has been calibrated to reproduce Bosman et al. (2022). The good agreement with the data at $z > 5$ is by design and is not a prediction. *Middle right: The ionizing emissivity against redshift. Bottom left:* The volume averaged ionized hydrogen fraction. The observational data include lower limits from dark pixels and gaps in the Ly α and Ly β forest (McGreer, Mesinger & D'Odorico 2015; Zhu et al. 2022; Jin et al. 2023) and the Ly α damping wing in quasar (Greig et al. 2017, 2022) and galaxy (Umeda et al. 2023) spectra. *Bottom right:* The volume averaged H I photoionization rate. Observational constraints are from the quasar proximity effect (Calverley et al. 2011) and the Ly α forest transmission (Becker & Bolton 2013; Gaikwad et al. 2023). *Note the redshift range on the horizontal axis is not the same in all panels.*

et al. 2021; Zhu et al. 2023). For the highest redshift data point at $z = 5.93$, the 40–2048 (160–2048) model is 1.2σ (1.8σ) above the recent Zhu et al. (2023) measurement. At $z < 6$, the mean free path redshift evolution for the 40–2048 model is very similar to THESAN-1 (Garaldi et al. 2022; solid grey curve), but more gradual than CoDa III (Lewis et al. 2022; blue dashed curve). Note also the mean free path from Keating et al. (2020b) is larger than the 160–2048 simulation at $z > 5.2$, despite these models both using P-GADGET-3 and ATON simulations performed at the same mass resolution and box size. This is in part because Keating et al. (2020b) adopted a larger emissivity at $z < 7$ by calibrating to match the Bosman et al. (2018) Ly α forest transmission. However, another difference is that Keating et al. (2020b) used ATON simulations that did not follow the patchy hydrodynamical response of the gas to photoheating during reionization. Instead, they performed post-processed radiative transfer simulations on IGM density fields

Table 2. The volume averaged neutral hydrogen fraction, $\langle x_{\rm H\,I}\rangle$, and mean free path for Lyman-limit photons, $\lambda_{\rm mfp}$, predicted by the 40–2048 and 160–2048 simulations, at selected redshifts, z.

z	$\langle x_{\rm H\,I}\rangle$ (40–2048)	$\lambda_{\rm mfp}$ (pMpc) (40–2048)	$\langle x_{\rm H\,I}\rangle$ (160–2048)	$\lambda_{\rm mfp}$ (pMpc) (160–2048)
5.4	$10^{-4.19}$	5.63	$10^{-2.35}$	7.44
6.0	0.05	1.46	0.14	1.84
7.0	0.41	0.20	0.48	0.25
8.0	0.67	0.05	0.71	0.05
10.0	0.90	$10^{-2.17}$	0.91	$10^{-2.24}$

extracted from a cosmological simulation that was quickly reionized and heated at $z = 15$ by the spatially uniform Haardt & Madau (2012) UV background. This early, uniform heating of the IGM will assist in suppressing small-scale gas clumping and could also increase the mean free path.

Furthermore, at $5 \leq z \leq 7$ – where the mean Ly α forest transmission in the two Sherwood–Relics simulations in Fig. 1 is similar by design – the mean free path in the 160–2048 model is ~ 20 per cent larger than 40–2048. This is despite reionization being slightly delayed (by $\Delta z \sim 0.3$–0.4) in 160–2048 relative to the 40–2048 simulation. This difference is partly because the 160–2048 model contains more massive haloes and larger, rarer ionized bubbles, as a result of the 64 times larger simulation volume (cf. Iliev et al. 2014; Kaur et al. 2020; Lu et al. 2024). However, another difference is mass resolution; the 160–2048 simulation has a factor of 64 poorer mass resolution compared to 40–2048 (see Table 1). The 160–2048 simulation under-resolves the Lyman-limit opacity from small-scale structure in the *already reionized* IGM.

We examine the effect of simulation mass resolution and box size on the mean free path further in Fig. 2. Here, we choose to show the mean free path against the volume averaged H I fraction, $\langle x_{\rm HI}\rangle$, rather than against redshift. This allows us to isolate the effect of box size and mass resolution from the redshift-dependent ionization state of the IGM (see e.g. Rahmati & Schaye 2018). The mean free path in the 40–512 (40–1024) simulation is typically 20 (5) per cent larger than the 40–2048 model, and is as much as 40 (10) per cent larger at $\langle x_{\rm HI}\rangle = 10^{-4}$. This is due to unresolved small-scale structure in the lower resolution models, resulting in an underprediction of the Lyman-limit opacity from the reionized IGM. The mean free path in the 40–2048 simulation should therefore be reasonably well converged (within 5 per cent) with mass resolution. However, the mean free path in the 160–2048 simulation will not be converged with mass resolution and is likely overestimated by *at least* 20 per cent. Hence, as has been discussed elsewhere (e.g. Cain et al. 2021; Park et al. 2023; Satyavolu et al. 2023; Georgiev, Mellema & Giri 2024), part of the reason for previous models failing to fully capture the redshift evolution of the mean free path first reported by Becker et al. (2021) may be unresolved photon sinks. We argue in Section 3.2 these sinks include residual neutral hydrogen in the diffuse, reionized IGM.

In contrast, the smaller box size of $40\,h^{-1}$ cMpc used for the 40–512 simulation results in an underprediction of the mean free path by up to a factor of two, relative to 160–2048. This is because of the larger ionized bubbles and brighter ionizing sources/more massive haloes present in the 160–2048 simulation. This explains why a slightly later end to reionization (by $\Delta z = 0.3$–0.4 when compared to 40–2048) is required for the 160–2048 simulation to match the Ly α forest transmission. For comparison, Iliev et al. (2014) reported a box size of $> 100\,h^{-1}$ cMpc is required to converge on the mean

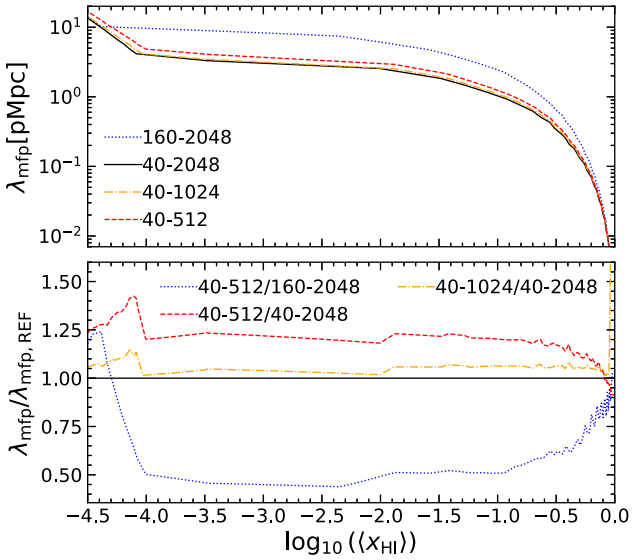

Figure 2. *Top:* The mean free path, $\lambda_{\rm mfp}$, against the volume averaged H I fraction, $\langle x_{\rm HI}\rangle$, for the Sherwood–Relics simulations used in this work. The 40–2048 (black solid curve), 40–1024 (orange dash dotted curve) and 40–512 (red dashed curve) models have the same box size, $40\,h^{-1}$ cMpc, but the gas particle mass in 40–512 (40–1024) is 64 (8) times larger than 40–2048, which has $M_{\rm gas} = 9.97 \times 10^4\,h^{-1}\,{\rm M}_\odot$ (see Table 1). The 160–2048 simulation (blue dotted curve, box size $160\,h^{-1}$ cMpc) has 64 times the volume of 40–2048 but has the same mass resolution as 40–512. *Bottom:* The ratio of the mean free path at fixed box size for a factor of 64 and 8 difference in the gas particle mass (red dashed and orange dash dotted curves, respectively), and at fixed mass resolution for a factor of 64 difference in simulation volume (blue dotted curve). The sharp features at $\log_{10}\langle x_{\rm H\,I}\rangle \sim -4$ arise from differences in the neutral fraction where the IGM transitions to being fully reionized.

reionization history, and a box size of $> 200h^{-1}$ cMpc is needed to adequately capture the large-scale patchiness of reionization. Our 160–2048 model should therefore be marginally converged with box size.

3 THE LYMAN-LIMIT OPACITY IN SHERWOOD–RELICS

3.1 Physical origin of the free-path distribution

We now turn to examine the origin of the free path distribution in the 40–2048 simulation. The free path PDFs for five different volume averaged neutral hydrogen fractions, $\langle x_{\rm H\,I}\rangle = 0.90$, 0.67, 0.41, 0.05, $10^{-4.19}$ (corresponding to redshifts $z = 10$, 8, 7, 6, 5.4, see Table 2) are shown in Fig. 3. The mean free path, $\lambda_{\rm mfp}$, is shown by the cross on each curve.

Prior to the completion of reionization (i.e. for $\langle x_{\rm H\,I}\rangle \gtrsim 10^{-3}$), the free path distribution is bimodal, with a characteristic minimum between the two modes at $\lambda_{\rm fp} \sim 10^{-2}$ pMpc. The mode at smaller (larger) $\lambda_{\rm fp}$ arises from the neutral (ionized) hydrogen in the simulation volume. The relative amplitudes of the modes depend on the average neutral fraction, and are similar for $\langle x_{\rm H\,I}\rangle \sim 0.5$. In all cases, the range of free path values varies over 4–5 orders of magnitude. This range reflects the distribution of gas densities in the simulation, with the smallest free paths often associated with the highest density neutral gas. Following the completion of reionization ($\langle x_{\rm HI}\rangle \lesssim 10^{-3}$), there are no neutral islands and

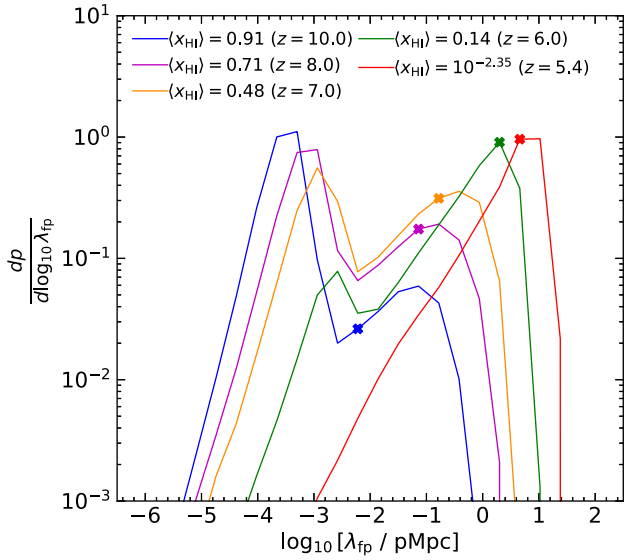

Figure 3. The PDF of the logarithm of the Lyman-limit free path, $\log_{10}(\lambda_{\rm fp}/{\rm pMpc})$, in the 40–2048 model for five different volume weighted neutral hydrogen fractions/redshifts (see Table 2). The mean free path is shown as a cross on each distribution. The bimodality arises from the distribution of ionized and neutral regions in the model and depends on the average H I fraction.

the bimodality is absent. These results are very similar to earlier independent work by Lewis et al. (2022) using the CoDa III radiation hydrodynamical simulation, suggesting a bimodal distribution for $\lambda_{\rm fp}$ is a generic prediction of inhomogeneous reionization models.

The spatial distribution of the free path is illustrated in Fig. 4. This shows a projection of the 40–2048 model at $z = 7$ ($\langle x_{\rm H\,I}\rangle = 0.41$) over a slice of width 64 pkpc – approximately the Jeans length at the mean background density (see equation B1 in Appendix B). The left panel of Fig. 4 shows the logarithm of the H I column density, $\log_{10}(N_{\rm H\,I}/{\rm cm}^{-2})$, while the right panel displays the logarithm of the free path, $\log_{10}(\lambda_{\rm fp}/\,{\rm pMpc})$. Pink ($N_{\rm H\,I} \gtrapprox 10^{18}\,{\rm cm}^{-2}$) and orange ($\lambda_{\rm fp} \lessapprox 10^{-3}$ pMpc) shading corresponds to neutral hydrogen that has yet to be reionized. The origin of the bimodal distribution in Fig. 3 is now evident, with the ionized bubbles around clustered sources (shown as white circles) producing the free path mode at $\lambda_{\rm fp} \sim 1$ pMpc. Note also that $N_{\rm H\,I}$ and $\lambda_{\rm fp}$ vary *within* the ionized and neutral regions, due to the underlying gas density fluctuations in the cosmic web.

We calculate H I column densities, $N_{\rm H\,I}$, by integrating the neutral hydrogen density over windows equal to the Jeans length at the mean background density. The choice of integration window is somewhat arbitrary, but is motivated by the results of Schaye (2001) and Rahmati & Schaye (2018), who found the local Jeans length is close to the typical size of H I absorbers following reionization. We also tried integrating over fixed windows of width 50 km s^{-1} (e.g. Gurvich, Burkhart & Bird 2017), but found this did not qualitatively change our results (see Appendix B for more details).

The resulting H I column density distribution function, $f(\log_{10} N_{\rm H\,I}, R)$, is displayed in Fig. 5. This is defined as the number, $\mathcal{N}$, of integration windows with column density $\log_{10} N_{\rm H\,I}$ per proper distance R, such that

$$f(\log_{10} N_{\rm H\,I}, R) = \frac{\partial^2 \mathcal{N}}{\partial \log_{10} N_{\rm H\,I}\, \partial R}. \quad (3)$$

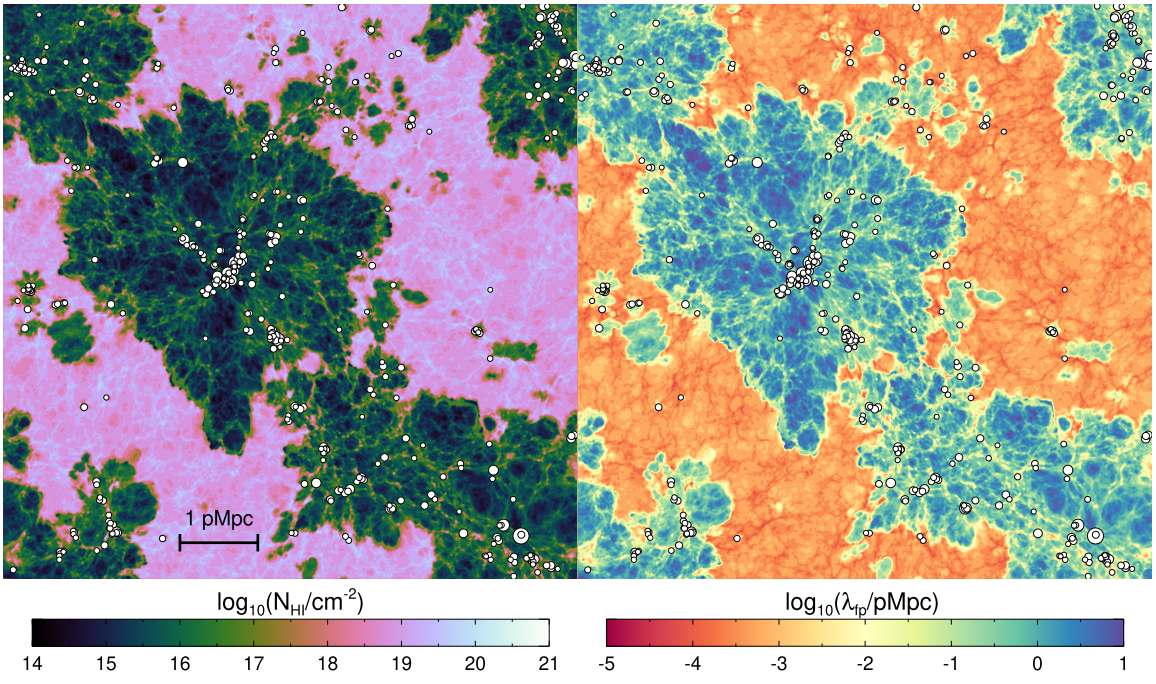

Figure 4. Projection of the logarithm of H I column density, $\log_{10}(N_{\rm H\,I}/{\rm cm}^{-2})$ (left panel), and logarithm of the proper free path for Lyman-limit photons, $\log_{10}(\lambda_{\rm fp}/{\rm pMpc})$ (right panel), in the 40–2048 model at $z = 7$, when $\langle x_{\rm H\,I}\rangle = 0.41$. Note that for this visualization the free path has been estimated using the local H I density in each pixel, rather than using equation (1). The slice width is 64 pkpc and is $40\,h^{-1}$ cMpc on a side. The locations of ionizing source host haloes (i.e. all haloes with total mass $M_{\rm h} \geq 10^9 h^{-1}\,{\rm M}_\odot$) are shown as white circles, with sizes that scale with the logarithm of the halo mass. The most massive halo in the slice – located in the large ionized region slightly left of the box centre – has total mass $M_{\rm h} = 10^{11.25} h^{-1}\,{\rm M}_\odot$.

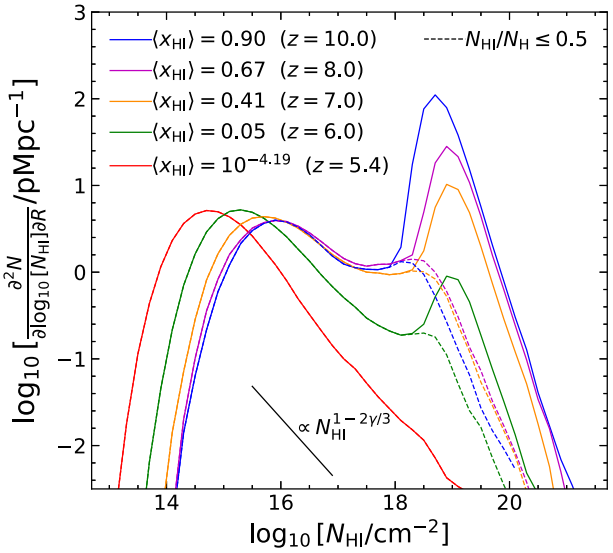

Figure 5. The solid curves show the H I column density distribution for the five different volume weighted H I fractions/redshifts used in Fig. 3 and summarized in Table 2. The dashed curves exclude column densities with $N_{\rm H\,I}/N_{\rm H} > 0.5$, and therefore ignore gas that is yet to be reionized. The diagonal black line, $f(\log_{10} N_{\rm H\,I}, R) \propto N_{\rm H\,I}^{1-2\gamma/3}$, gives the expected slope of the post-reionization H I column density distribution assuming an underlying gas density distribution with a power-law high-density tail, $P(\Delta) \propto \Delta^{-\gamma}$, with $\gamma = 2.6$ (see the text and Appendix C for details).

Similar to the free path distribution, the column density distribution in Fig. 5 is bimodal if $\langle x_{\rm H\,I}\rangle > 10^{-3}$, with modes at $N_{\rm H\,I} \sim 10^{15.5}$ cm^{-2} ($\sim 10^{19}$ cm^{-2}) arising from the predominantly ionized (neutral) hydrogen gas. This is further highlighted by the dashed curves in Fig. 5, where all H I column densities with $N_{\rm H\,I}/N_{\rm H} > 0.5$ (i.e. hydrogen that is more than 50 per cent neutral) have been removed from the distribution. For the largest H I fractions, the dashed curve deviates substantially from the solid curve at $N_{\rm H\,I} > 10^{18}$ cm^{-2}. Following the completion of reionization, however, the dashed and solid curves are almost identical. At column densities $N_{\rm H\,I} > 10^{14}$–10^{15} cm^{-2} (i.e. where the distribution is complete), the shape of the distribution approximately follows a power law for the predominantly ionized gas, although there is a hint of flattening in the slope at $N_{\rm H\,I} \sim 10^{17}$ cm^{-2} due to self-shielded gas (e.g. Altay et al. 2011). When the H I in the post-reionization IGM is in photoionization equilibrium, the shape of $f(\log_{10} N_{\rm H\,I}, R)$ is directly related to the gas density distribution (see Appendix C). For a gas density distribution with a power-law tail, $P(\Delta) \propto \Delta^{-\gamma}$, we expect that $f(\log_{10} N_{\rm H\,I}, R) \propto N_{\rm H\,I}^{1-2\gamma/3}$ (e.g. Furlanetto & Oh 2005; McQuinn, Oh & Faucher-Giguère 2011).

Fig. 6 shows 2D histograms that elucidate the relationship between the H I column density and the gas density, temperature, H I photoionization rate and H I fraction for $\langle x_{\rm H\,I}\rangle = 0.90$, 0.41, and $10^{-4.19}$ (corresponding to $z = 10$, 6, and 5.4, respectively, in the 40–2048 simulation). Here, we denote quantities that have been averaged over the column integration window with $\langle \ldots \rangle_{\rm N_{H\,I}}$. Prior to the completion of reionization at $z < 6$ the temperature and H I fractions are bimodal, although note column densities with $\langle \Gamma_{\rm H\,I}\rangle_{\rm N_{H\,I}} = 0$ (i.e. for $\langle x_{\rm H\,I}\rangle_{\rm N_{H\,I}} = 1$) are not shown on the logarithmic scale used in the lower panels. Column densities $N_{\rm H\,I} > 10^{18}$ cm^{-2} most often correspond to cool, $\langle T\rangle_{\rm N_{H\,I}} \lesssim 10^3$ K neutral gas, with photoionization rates $\langle \Gamma_{\rm H\,I}\rangle_{\rm N_{H\,I}} \lesssim 10^{-14}$ s^{-1}. The largest column densities, $N_{\rm H\,I} > 10^{21}$ cm^{-2}, are almost exclusively associated with the highest density gas with $\langle \Delta\rangle_{\rm N_{HI}} \gtrsim 10$. Gas that has been reionized has $\langle T\rangle_{\rm N_{H\,I}} \sim 10^4$ K, $\langle \Gamma\rangle_{\rm N_{H\,I}} \sim 10^{-12}$–$10^{-13}$ s^{-1}, and $\langle x_{\rm H\,I}\rangle_{\rm N_{H\,I}} \lesssim 10^{-3}$, with typical column densities of $N_{\rm H\,I} \sim 10^{15}$–$10^{16}$ cm^{-2} at the mean background density. The upward scatter at $\langle T\rangle_{\rm N_{H\,I}} > 10^{4.5}$ K, which is particularly evident in the left column of Fig. 6, is associated with gas shocked by gravitational infall (see also fig. 5 of Puchwein et al. 2023).

3.2 Lyman-limit opacity as a function of H I column density

We now quantify the role that different H I column densities play in determining the opacity of the IGM to Lyman-limit photons. Here, we follow Rahmati & Schaye (2018) and Nasir et al. (2021) and compute the Lyman-limit opacity by integrating over the H I column density distribution (see also Prochaska et al. 2009; Becker et al. 2021; Zhu et al. 2023). The differential distribution of the opacity to Lyman-limit photons, $\kappa_{\rm eff}$, is calculated as

$$\frac{{\rm d}\kappa_{\rm eff}}{{\rm d}\log_{10} N_{\rm H\,I}} = \frac{1}{\delta \log_{10} N_{\rm H\,I}} \times \int_{N_{\rm H\,I}}^{N_{\rm H\,I}+\delta N_{\rm H\,I}} f(N_{\rm H\,I}, R)(1 - e^{-N_{\rm H\,I}\sigma_{912}})\,{\rm d}N_{\rm H\,I}, \quad (4)$$

where $f(N_{\rm H\,I}, R) = f(\log_{10} N_{\rm H\,I}, R)/(N_{\rm H\,I} \ln 10)$ is the column density distribution function (i.e. equation 3) and $\delta \log_{10} N_{\rm H\,I} = \log_{10}(1 + \delta N_{\rm H\,I}/N_{\rm H\,I}) = 0.2$ is our adopted bin size. The cumulative distribution of the opacity is then

$$\kappa_{\rm eff}(\leq \log_{10} N_{\rm H\,I}) = \int_0^{\log_{10} N_{\rm H\,I}} \frac{{\rm d}\kappa_{\rm eff}}{{\rm d}\log_{10} N_{\rm H\,I}}\,{\rm d}\log_{10} N_{\rm H\,I}. \quad (5)$$

This quantity is related to the effective optical depth to Lyman-limit photons, $\tau_{\rm eff}$, by $\kappa_{\rm eff} = {\rm d}\tau_{\rm eff}/{\rm d}R$, and the attenuation length $\lambda_{\rm eff}$ by $\kappa_{\rm eff} = \lambda_{\rm eff}^{-1}$. We find the attenuation length is typically within a factor of two of the mean free path given by equation (2) (see e.g. appendix A in Theuns & Chan 2024), but this factor will be sensitive to the choice of column density integration window.

The differential opacity is shown in the left panel of Fig. 7. For $\langle x_{\rm H\,I}\rangle > 10^{-3}$ the differential opacity exhibits a peak at $N_{\rm H\,I} = 10^{18}$–10^{19} cm^{-2} associated with mostly neutral hydrogen[3] with $N_{\rm H\,I}/N_{\rm H} > 0.5$. The peak occurs around 1 dex higher compared to Rahmati & Schaye (2018), but as discussed earlier, this value will depend on our choice of integration window as well as differences in the assumed reionization history, which for 40–2048 ends later than the fiducial model in Rahmati & Schaye (2018). For $\langle x_{\rm H\,I}\rangle < 10^{-3}$ (i.e. at the end of reionization), the contribution of column densities with $N_{\rm H\,I} > 10^{18}$–10^{19} cm^{-2} to the total Lyman-limit opacity is minimal, and instead gas with $N_{\rm H\,I} \sim 10^{16}$–$10^{17}$ cm^{-2} dominates the differential opacity (i.e. at the H I column densities typical of strong Ly α forest absorbers). At $z = 5.4$, these absorbers have average densities $\langle \Delta\rangle_{\rm N_{H\,I}} \sim 3$, temperatures $\langle T\rangle_{\rm N_{H\,I}} \sim 10^{4.1}$ K, and H I fractions $\langle x_{\rm H\,I}\rangle_{\rm N_{H\,I}} \sim 10^{-3.5}$.

Insight into the role the *already reionized* IGM plays in setting the mean free path may be obtained from the cumulative opacity, displayed in the right panel of Fig. 7. Here, we report the fraction of the total opacity that originates from ionized regions with $N_{\rm HI}/N_{\rm H} \leq 0.5$ (the dashed curves in Fig. 7). For the five neutral hydrogen fractions shown at $z = 10$, 8, 7, 6 and 5.4, this represents 3, 13, 27, 61 and > 99 per cent of the total opacity, respectively. Hence, during the

[3]Note that while column densities of $N_{\rm H\,I} = 10^{18}$–10^{19} cm^{-2} are typically associated with Lyman-limit systems, prior to the end of reionization these arise from H I in the diffuse IGM, rather than from high density, self-shielded gas (see Fig. 6).

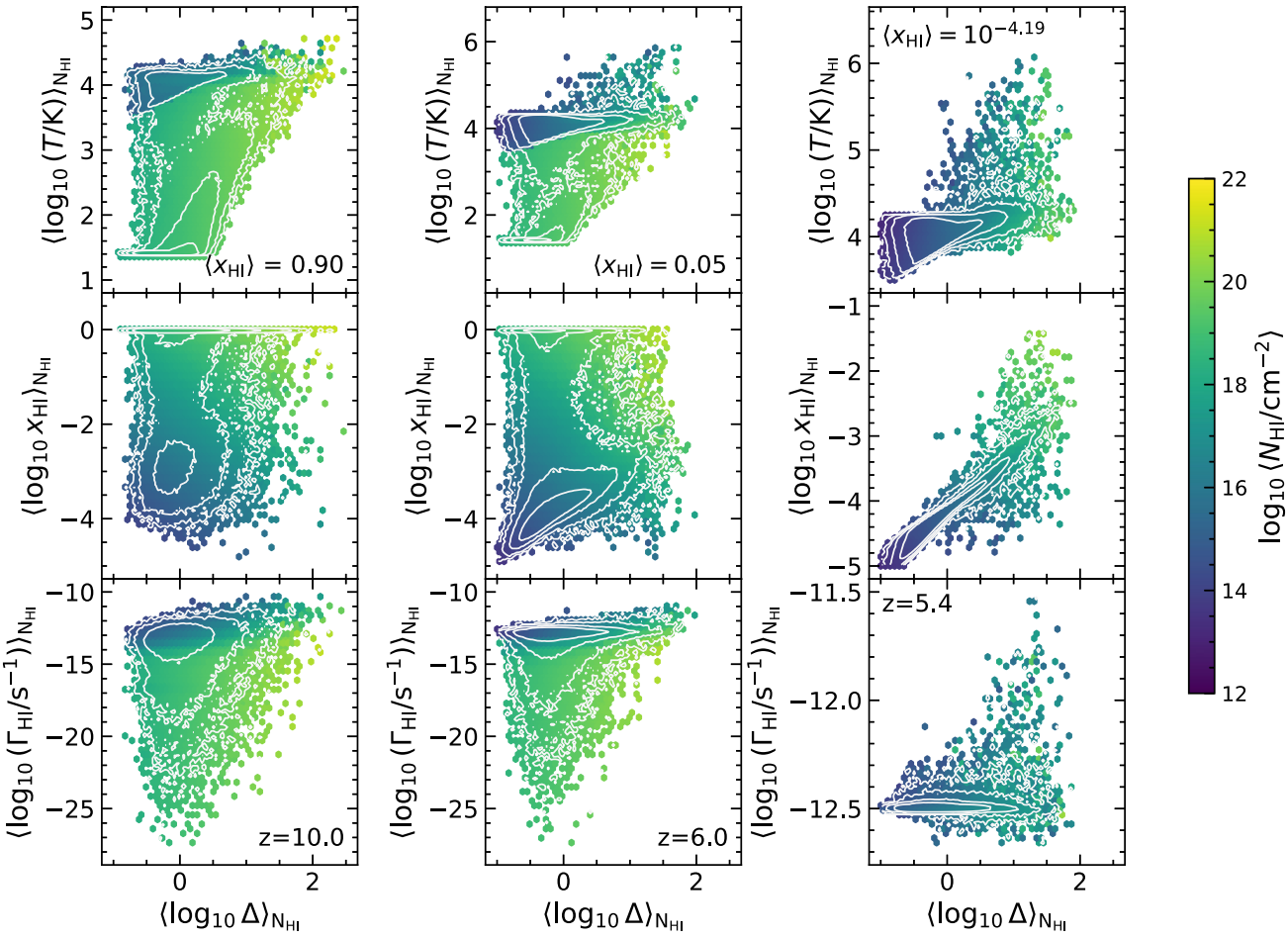

Figure 6. Each column in this figure shows a different volume averaged neutral fraction/redshift, where from left to right, $\langle x_{\rm H\,I}\rangle = 0.90$, 0.05, and $10^{-4.2}$, corresponding to $z = 10$, 6 and 5.4 (see Table 2). *Top row:* A 2D histogram of the logarithm of the average temperature, $\langle \log_{10}(T/{\rm K})\rangle_{\rm N_{H\,I}}$, against the logarithm of the average overdensity, $\langle \log_{10}\Delta\rangle_{\rm N_{H\,I}}$, within the integration windows used to calculate the H I column densities in the 40–2048 simulation. The value of the average H I column density in each bin is displayed in the colour bar on the right of the figure. Contours show the number density of systems in each bin, with the density increasing by 1 dex for each contour level. *Middle row:* As for the top row, but now showing the logarithm of the average H I fraction in each integration window, $\langle \log_{10} x_{\rm H\,I}\rangle_{\rm N_{H\,I}}$ against $\langle \log_{10}\Delta\rangle_{\rm N_{H\,I}}$. *Bottom row:* As for the top row, but now showing the logarithm of the average H I photoionization rate in each integration window, $\langle \log_{10}(\Gamma_{\rm H\,I}/{\rm s}^{-1})\rangle_{\rm N_{H\,I}}$ against $\langle \log_{10}\Delta\rangle_{\rm N_{H\,I}}$. *Note the different range on the vertical axis of each row.*

final stages of late reionization at $z = 6$, *the majority of the Lyman-limit opacity in the IGM arises from gas that is mostly ionized.* This result is consistent with Rahmati & Schaye (2018), who pointed out the importance of the cumulative opacity arising from systems with $N_{\rm H\,I} < 10^{16.5}\,{\rm cm}^{-2}$ (i.e. absorbers below the Lyman-limit system threshold of $N_{\rm H\,I} = 10^{17.2}\,{\rm cm}^{-2}$) for setting the mean free path during the final stages of reionization. These systems will be associated with reionized gas that may also have had time to dynamically respond to photoheating (Nasir et al. 2021; Puchwein et al. 2023). Resolving the small-scale structure of the IGM at moderate overdensity is therefore a necessary requirement for modelling the mean free path during the final stages of reionization.

4 THE MEAN FREE PATH AROUND IONIZING SOURCE HOST HALOES

4.1 Method for calculating the mean free path around haloes

Thus, far we have investigated the Lyman-limit photon mean free path for the 'average' IGM, yet $\lambda_{\rm mfp}$ measurements are typically determined from the Lyman continuum opacity in the vicinity of (stacked) high redshift quasars (Prochaska et al. 2009; Becker et al. 2021; Zhu et al. 2023). The quasar proximity effect (D'Aloisio et al. 2018), neutral islands (Roth et al. 2024) or Lyman-limit systems around the quasar host haloes (Prochaska et al. 2014) could then bias the mean free path measurements. Satyavolu et al. (2023) addressed these biases and concluded that – if correctly accounted for – they should not significantly impact the recent $\lambda_{\rm mfp}$ measurements at $5 < z < 6$. We do not revisit the issue of measurement bias here, therefore. Instead, we focus on understanding the physical origin of variations in the mean free path around haloes in late reionization models.

We first identify haloes in the simulations at the five neutral fractions/redshifts we have investigated previously. At each redshift, we construct mass bins each containing 10^3 unique haloes, where the smallest halo mass in the lower bin, $M = 10^9 h^{-1}\,{\rm M}_\odot$, corresponds to the minimum host halo mass for ionizing sources in Sherwood–Relics (Puchwein et al. 2023). The mean free path in each mass bin is then calculated in the same way as described in Section 2.2, except that (i) the sight-line now begins at a halo centre of mass, drawn in one of six directions (i.e. $\pm x, \pm y, \pm z$) parallel to the Cartesian

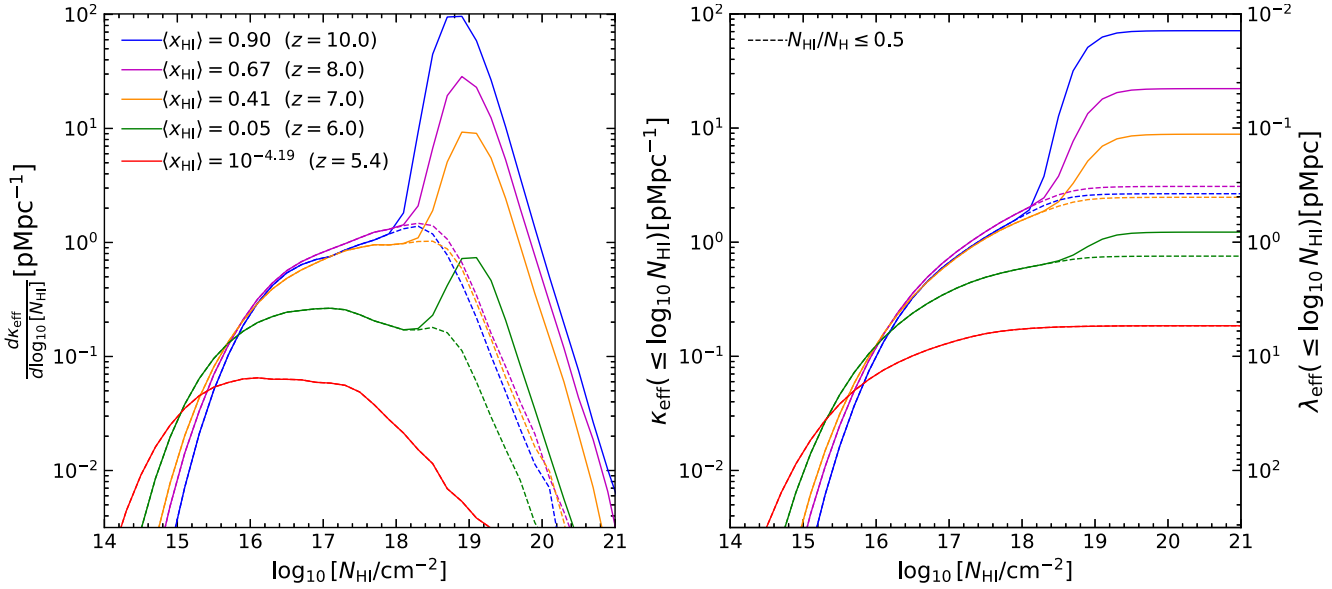

Figure 7. *Left:* The differential distribution of the Lyman-limit opacity, $\kappa_{\rm eff}$, as a function of $\log_{10}(N_{\rm H\,I}/{\rm cm}^{-2})$ from the 40–2048 model. As for Figs 3 and 5, five different volume weighted H I fractions/redshifts are displayed (see Table 2). The dashed curves are computed ignoring all H I column densities with $N_{\rm H\,I}/N_{\rm H} > 0.5$. *Right:* The corresponding cumulative distribution of the Lyman-limit opacity. At $\langle x_{\rm H\,I}\rangle = 0.05$ ($z = 6$, green solid, and dashed curves), 61 per cent of the total Lyman-limit opacity arises from hydrogen that is already ionized (i.e. with $N_{\rm H\,I}/N_{\rm H} \leq 0.5$).

axes of the simulation box, and (ii) the Lyman-limit photon starting position is at one virial radius[4], $R_{\rm vir}$, from the halo centre of mass, where

$$R_{\rm vir} \simeq \left(\frac{2G}{H_0^2\Omega_{\rm m}}\right)^{1/3} M_{\rm h}^{1/3}\Delta_{\rm c}^{-1/3}(1+z)^{-1}, \\ = 8.7\,{\rm pkpc}\left(\frac{M_{\rm h}}{10^{10}\,{\rm M}_\odot}\right)^{1/3}\left(\frac{\Delta_{\rm c}}{18\pi^2}\right)^{-1/3}\left(\frac{1+z}{8}\right)^{-1}, \quad (6)$$

as is appropriate for $\Omega_{\rm m}(1+z)^3 \gg \Omega_\Lambda$. This is because the Lyman-limit opacity within $R_{\rm vir}$ will depend on highly uncertain subgrid physics that we do not model here (e.g. star formation and feedback, the properties of the interstellar medium, and the escape of ionizing radiation).

Finally, to aid with distinguishing the halo mean free path from the mean free path for the average IGM, in the following discussion we introduce the notation, $\Lambda_{\rm mfp}$, for the Lyman-limit photon mean free path starting from a halo.

4.2 Dependence of the mean free path on halo mass

The halo mean free path, $\Lambda_{\rm mfp}$, against halo mass from the 40–2048 (left panel) and 160–2048 (right panel) simulations is displayed in Fig. 8 for five different H I fractions/redshifts. We also report the values for the 40–2048 model in Table 3. Note that, to facilitate comparison, the first three halo mass bins are identical for both simulations. The fourth mass bin in the right panel of Fig. 8 (shown by the filled stars) contains the 10^3 most massive haloes in the 160–2048 volume.

Several general trends are apparent in both models, despite the differences in the simulation mass resolution/box size and reionization history. First, regardless of the underlying H I fraction, the halo mean free path is not equal to the average IGM value, $\Lambda_{\rm mfp} \neq \lambda_{\rm mfp}$, where the average IGM value, $\lambda_{\rm mfp}$, is indicated by the horizontal lines on the left of the panels in Fig. 8. Prior to the completion of reionization the halo mean free path is greater than the average IGM value, $\Lambda_{\rm mfp} > \lambda_{\rm mfp}$, yet once the IGM is almost fully ionized this trend reverses, and $\Lambda_{\rm mfp} < \lambda_{\rm mfp}$. The magnitude of this bias, relative to the average IGM value, $\lambda_{\rm mfp}$, changes with the volume averaged H I fraction; at $z = 5.4$, the halo mean free path, $\Lambda_{\rm mfp}$, is ~ 20 per cent smaller than the average IGM value, while at $z = 10$, for the most massive haloes it is over an order of magnitude larger than the average IGM value. Second, the dependence of $\Lambda_{\rm mfp}$ on halo mass, $M_{\rm h}$, varies with the H I fraction. During reionization, $\Lambda_{\rm mfp}$ is correlated with halo mass, but after reionization the trend is reversed, with halo masses that are (weakly) anticorrelated with $\Lambda_{\rm mfp}$.

The reason for this behaviour is explored further in the distribution of the halo free path, $\Lambda_{\rm fp}$, in Fig. 9. We show results for the 40–2048 model at $\langle x_{\rm HI}\rangle = 0.90$, 0.05 and $10^{-4.19}$ (corresponding to redshifts $z = 10$, 6 and 5.4), for each of the mass bins. For comparison, the dashed green curves show the distribution for the free path, $\lambda_{\rm fp}$, in the average IGM. As before, the left hand column of Fig. 9 excludes all gas within $R_{\rm vir}$ of the halo centre of mass (i.e. as for Fig. 8), but the right column now excludes all gas within a larger region, $10R_{\rm vir}$.

We observe that – at all stages of reionization – the smallest free paths are associated with gas between $1R_{\rm vir}$ and $10R_{\rm vir}$; these small free paths are removed from the distribution that excludes gas within $10R_{\rm vir}$, which converges toward the average IGM distribution. The $\Lambda_{\rm fp}$ distribution is furthermore not as strongly bimodal as the average IGM. At $\langle x_{\rm HI}\rangle = 0.9$ ($z = 10$), for the long mode of the average IGM distribution at $\lambda_{\rm fp} \sim 10^{-1}$ pMpc, the incidence of halo free paths is around an order of magnitude greater than for the average IGM. However, following reionization, when $\langle x_{\rm HI}\rangle = 10^{-4.19}$ ($z = 5.4$), there is instead an excess of halo free paths at the short mode for the average IGM, $\lambda_{\rm fp} \sim 10^{-3}$ pMpc. As discussed in Section 3, these long (short) modes are typically associated with ionized (neutral) gas, indicating that haloes before (after) reionization are preferentially probing the more

[4]At the redshifts we consider here, equation (6) is within 0.3 per cent of the expression given by equation (24) of Barkana & Loeb (2001).

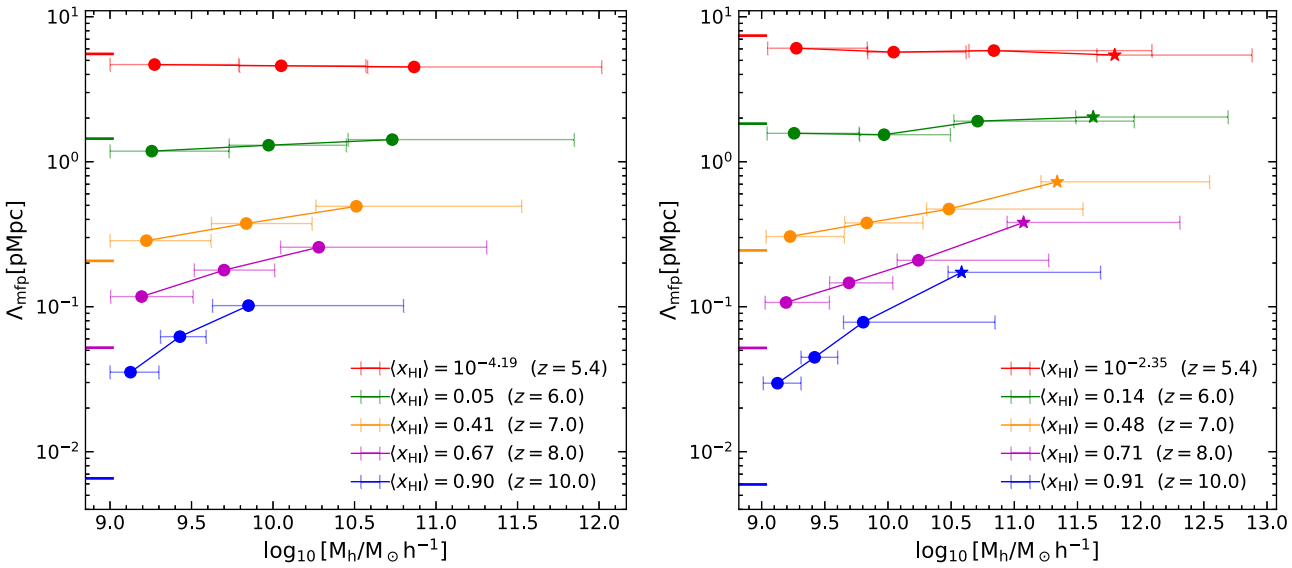

Figure 8. The Lyman-limit mean free path around haloes, $\Lambda_{\rm mfp}$, against the logarithm of the total halo mass, $\log_{10}(M_{\rm h}/h^{-1}\,{\rm M}_\odot)$, for the five different neutral hydrogen fractions/redshifts used in Fig. 3 (see also Table 3). The left panel shows the 40–2048 model, while the right panel shows the 160–2048 simulation. The horizontal error bars give the size of each logarithmic mass bin (see the text for details). Note also the different scales on the horizontal axis of each panel, as well as the additional mass bin in the right panel (filled stars) for the 10^3 most massive haloes in the 160–2048 simulation. The corresponding mean free path for the average IGM, matching the values displayed in Fig. 1, is shown by the horizontal lines on the left of each panel.

Table 3. Tabulation of the halo mean free path, $\Lambda_{\rm mfp}$, in the 40–2048 simulation for the five redshift/neutral fractions displayed in Fig. 8. The columns list, from left to right, the redshift, z, the volume weighted H I fraction, $\langle x_{\rm HI}\rangle$, the average halo mass in each bin, $M_{\rm h}$, the halo mean free path, $\Lambda_{\rm mfp}$ and its ratio with the mean free path in the average IGM, $\lambda_{\rm mfp}$.

z	$\langle x_{\rm H\,I}\rangle$	$\langle M_{\rm h}\rangle$ $(h^{-1}\,{\rm M}_\odot)$	$\Lambda_{\rm mfp}$ (pMpc)	$\Lambda_{\rm mfp}/\lambda_{\rm mfp}$
5.4	$10^{-4.19}$	$10^{10.9}$	4.50	0.81
5.4	$10^{-4.19}$	$10^{10.1}$	4.59	0.83
5.4	$10^{-4.19}$	$10^{9.3}$	4.68	0.85
6.0	0.05	$10^{10.7}$	1.42	0.98
6.0	0.05	$10^{10.0}$	1.30	0.90
6.0	0.05	$10^{9.3}$	1.18	0.82
7.0	0.41	$10^{10.5}$	0.49	2.38
7.0	0.41	$10^{9.8}$	0.37	1.81
7.0	0.41	$10^{9.2}$	0.29	1.38
8.0	0.67	$10^{10.3}$	0.26	4.93
8.0	0.67	$10^{9.7}$	0.18	3.43
8.0	0.67	$10^{9.2}$	0.12	2.25
10.0	0.90	$10^{9.8}$	0.10	15.55
10.0	0.90	$10^{9.4}$	0.06	9.50
10.0	0.90	$10^{9.1}$	0.04	5.40

ionized (neutral) hydrogen. The smallest halo free path values at $z = 5.4$ ($z = 10$) are furthermore associated with the most (least) massive haloes, consistent with the behaviour in Fig. 8 where halo mass anti-correlates (correlates) with $\Lambda_{\rm mfp}$. In other words, post-reionization, differences between the mean free path in the IGM and around haloes are due to large scale structure bias, but when the IGM is more than a few per cent neutral by volume the differences instead arise from ionization bias (cf. Wyithe et al. 2008).

In Fig. 10, the role that the gas density and H I fraction around the haloes play in setting this behaviour is highlighted. The average gas overdensity (left column) and H I fraction (right column) around haloes in the three mass bins used for the 40–2048 model are compared to the average IGM (green curves). Each row shows the same three volume averaged H I fractions/redshifts as Fig. 9. The more massive haloes are always associated with larger overdensities. For a fixed H I fraction, we would expect the mean free path to decrease with increasing halo mass, because $\Lambda_{\rm mfp} \propto (x_{\rm HI}\Delta)^{-1}$. However, the H I fraction around the haloes also varies as reionization progresses. During the early stages of reionization, when $\langle x_{\rm HI}\rangle = 0.9$, the H I fraction around the most massive haloes is substantially *smaller* than for lower mass haloes and the average IGM. This is a consequence of (i) ionizing source luminosity being proportional to halo mass in our model and (ii) the clustering of ionizing sources around the most massive haloes (see Fig. 4). This enhanced ionization increases the mean free path relative to the average IGM, and causes $\Lambda_{\rm mfp}$ to correlate with halo mass, $M_{\rm h}$.

Similar behaviour is observed when the IGM is mostly reionized, with only a few neutral islands remaining (i.e. $\langle x_{\rm H\,I}\rangle = 0.05$ at $z = 6$, in the middle row of Fig. 10). Here the contrast in the H I fraction with halo mass is smaller and the profiles vary less with distance from the halo centre of mass, R. This is a consequence of the larger ionized regions around the haloes compared to higher redshift. However, by $z = 5.4$ the trend has reversed, and the most massive haloes have a larger average H I fraction in their vicinity. Here, the average photoionization rate in the IGM, $\Gamma_{\rm H\,I}$, has increased by a factor of 2–3 from $z = 6$ (see Fig. 1), and the residual H I in the IGM is now approximately in photoionization equilibrium with an increasingly spatially uniform ionizing background. In this case, $x_{\rm H\,I} \propto \Delta$. It is this interplay between gas density and spatial variations in the H I fraction approaching haloes that drives the halo mean free path behaviour in Fig. 8.

Our results are consistent with Rahmati & Schaye (2018), who reported a bimodal distribution for the free path that was halo mass dependent. The halo mean free path was furthermore

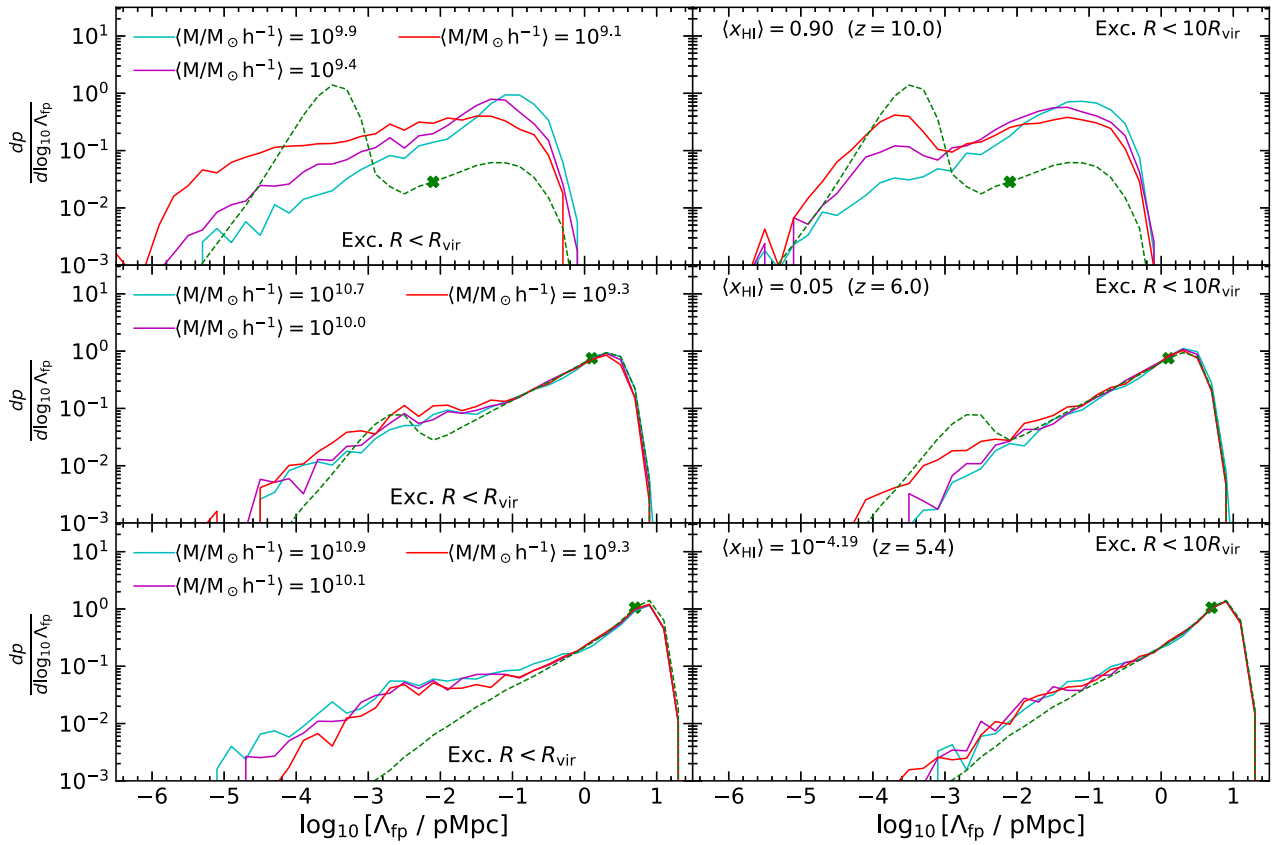

Figure 9. *Left column:* PDFs of the logarithm of the free path for Lyman-limit photons around haloes, $\log_{10}(\Lambda_{\rm fp}/{\rm pMpc})$, in the 40–2048 simulation for three different volume weighted neutral hydrogen fractions/redshifts (the $\langle x_{\rm HI}\rangle$ values are indicated in the right column panels only). Solid curves show $p(\log_{10}[\Lambda_{\rm fp}/{\rm pMpc}])$ for the three different mass bins indicated, while the dashed green curves show the corresponding average IGM distribution, $p(\log_{10}[\lambda_{\rm fp}/{\rm pMpc}])$. The mean free path for the average IGM, $\lambda_{\rm mfp}$, is shown as a cross. All gas within one virial radius, $R_{\rm vir}$, of the centre of mass of the haloes has been excluded when calculating $\Lambda_{\rm fp}$ (see text for details). *Right column:* As for the left column, but for free paths that now start at $10R_{\rm vir}$ from the halo centre of mass. Note these $\Lambda_{\rm fp}$ distributions fall more in line with the average IGM distribution (green dashed curves).

suppressed (enhanced) relative to the average IGM value following (during) reionization, matching our findings. Our results are also broadly similar to Fan et al. (2024), who reported a slightly smaller suppression in the mean free path around $\sim 10^{11}$–$10^{12}\,{\rm M}_\odot$ haloes of ~ 10 per cent at $z \sim 5$. In the analytical halo model for Lyman-limit systems presented by Theuns & Chan (2024) the mean free path was also smaller around the massive haloes hosting quasars at $z \simeq 6$ when compared to the average IGM (see also Prochaska et al. 2014). Theuns & Chan (2024) attributed this difference to the clustering of Lyman-limit systems around the quasar host haloes; the ratio between the attenuation length around haloes and the average IGM in their model was around an order of magnitude for the halo masses, $M_{\rm h} = 10^{12}$–$10^{13}\,{\rm M}_\odot$, thought to be typical of quasar hosts (Pizzati et al. 2024). Our 40–2048 simulation only includes haloes up to $10^{12}\,{\rm M}_\odot$ at $5 < z < 6$, but do we find a small reduction, $0.8 \lesssim \Lambda_{\rm mfp}/\lambda_{\rm mfp} \lesssim 1.0$, at the end of reionization (Table 3). In the 160–2048 model, which includes haloes with masses up to $M_{\rm h} = 10^{12.8}\,{\rm M}_\odot$ at $z = 6$, we instead find $\Lambda_{\rm mfp}/\lambda_{\rm mfp} \simeq 0.7$ for the most massive haloes, although note again that we ignore all gas within the virial radius of the haloes. Our results suggest the Theuns & Chan (2024) model will not hold during reionization, when the mean free path is no longer set by neutral gas in high density, small-scale structures.

In summary, any systematic biases should be accounted for when measuring the mean free path around bright quasars at $z > 5$, but as pointed out in recent work (Satyavolu et al. 2023; Roth et al. 2024) these already appear to be well controlled in the most recent measurements. The biases we discuss here will furthermore be subdominant to effect of the significantly enhanced ionization around bright quasars at $z \simeq 6$, which we do not model in this work. Lastly, as has been noted elsewhere (e.g. Mason et al. 2018; Asthana et al. 2024; Lu et al. 2024; Neyer et al. 2024), a mean free path that is significantly larger than in the average IGM around ionizing source host haloes during reionization will aid the visibility of high redshift Ly α emitting galaxies at $z > 7$. Indeed, multiple objects have now been detected with near-infrared spectroscopic observations (e.g. Larson et al. 2022; Tang et al. 2023; Whitler et al. 2024), with the current highest redshift Ly α emitter recently detected at $z = 10.6$ with *JWST* (Bunker et al. 2023).

5 CONCLUSIONS

In this work, we have examined the mean free path to Lyman-limit photons, $\lambda_{\rm mfp}$, predicted by the Sherwood–Relics simulation suite (Puchwein et al. 2023). Sherwood–Relics has been specifically designed for modelling the Ly α forest at the end stages of inhomogeneous reionization. Here, we consider a sub-set of the

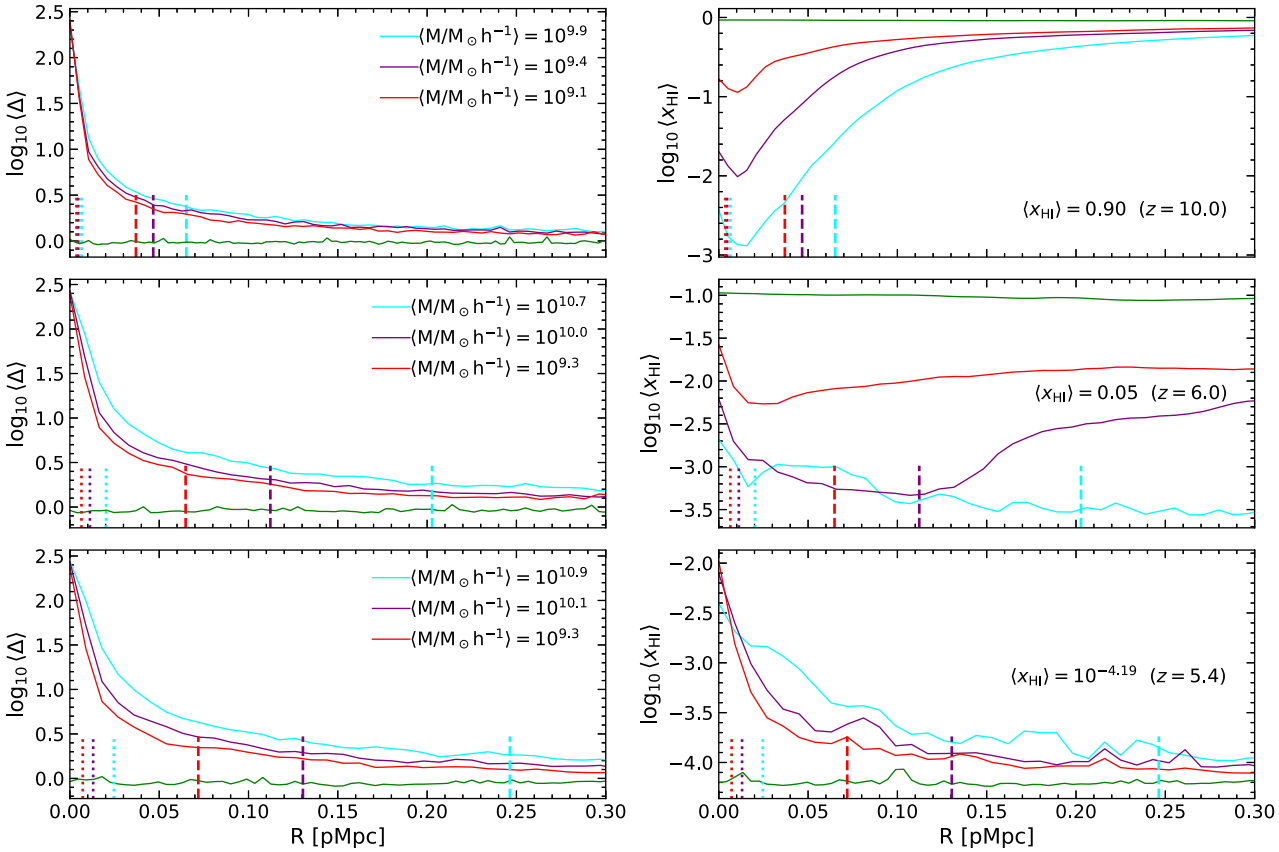

Figure 10. *Left column:* The logarithm of the average gas overdensity, $\log_{10}\langle\Delta\rangle$, against distance from the halo centre of mass in the 40–2048 simulation. Each of the three halo mass bins may be compared to the average IGM (green curves). From top to bottom, the results for three different volume weighted neutral fractions/redshifts are shown, $\langle x_{\rm H\,I}\rangle = 0.90,\ 0.05$ and $10^{-4.19}$ (or $z = 10,\ 6$ and 5.4). The dotted (dashed) vertical lines show the value of $R_{\rm vir}$ ($10R_{\rm vir}$) for the average halo mass in each bin (cf. Fig. 9). Larger overdensities are associated with the more massive haloes, as expected. *Right column*: As for the left column, but now showing the logarithm of the average H I fraction, $\log_{10}\langle x_{\rm H\,I}\rangle$. *Note the scale on the vertical axis is not the same in all panels.*

Sherwood–Relics simulations that have been calibrated to reproduce the observed Ly α forest transmission at $5 < z < 6$, assuming a late end to reionization (i.e. at $z < 6$). The simulations self-consistently follow the hydrodynamical response of the diffuse IGM to photoionization and heating, and our fiducial model has a mass resolution intermediate between the recent THESAN (Garaldi et al. 2022) and CoDa III (Lewis et al. 2022) radiation hydrodynamical simulations. Importantly for this study, while Sherwood–Relics does not explicitly model star formation and dense, cool gas with $\Delta = \rho/\langle\rho\rangle > 1000$ and $T < 10^5$ K, it follows the absorption systems in the Ly α forest that dominate the Lyman-limit opacity during the final stages of late reionization (Rahmati & Schaye 2018; Nasir et al. 2021). Our main findings are as follows:

(i) The Lyman-limit mean free path predicted by our fiducial Sherwood–Relics simulation (40–2048) is reasonably well converged (within 5 per cent) with mass resolution and is in good agreement (typically within 1σ) with recent measurements of the mean free path from direct determinations using the Lyman continuum opacity (Becker et al. 2021; Zhu et al. 2023) and indirect determinations using the observed Ly α forest transmission (Gaikwad et al. 2023). At $z = 5.93$, the 40–2048 simulation furthermore predicts a mean free path 1.2σ above the recent measurement presented by Zhu et al. (2023), although the mean reionization history in this model will not be converged with box size. Our larger simulation volume at lower resolution (160–2048) is 1.8σ above the same measurement and has a reionization history that should be marginally converged with box size (cf. Iliev et al. 2014). However, the lower mass resolution of this model will overpredict the mean free path at fixed H I fraction by at least 20 per cent.

(ii) In order to match the observed Ly α transmission, reionization completes at $z \sim 5.7$ in our fiducial model, and at $z \sim 5.3$ in our larger, lower resolution volume. In both cases, the observed evolution of $\lambda_{\rm mfp}$ is consistent with the emerging consensus that reionization ends at $5 < z < 6$ (see also Keating et al. 2020b; Bosman et al. 2022; Garaldi et al. 2022; Lewis et al. 2022).

(iii) The Lyman-limit photon free path distribution in Sherwood–Relics is bimodal during reionization, with a minimum at $\lambda_{\rm fp} \sim 10^{-2}$ pMpc. The short (long) mode is typically associated with neutral (ionized) hydrogen, and is a generic prediction of inhomogeneous reionization models (cf. Lewis et al. 2022; Satyavolu et al. 2023). This bimodality also manifests in the H I column density distribution, with H I column densities at $N_{\rm H\,I} \sim 10^{19}\,{\rm cm}^{-2}$ ($N_{\rm HI} \sim 10^{15.5}\,{\rm cm}^{-2}$) associated with the neutral (ionized) hydrogen in the diffuse IGM.

(iv) During the final stages of late reionization at $z = 6$, when the volume averaged H I fraction is $\langle x_{\rm HI}\rangle = 0.05$ in our fiducial model, the majority of the total Lyman-limit opacity (61 per cent) arises from the residual H I fraction in the *already reionized IGM*. The residual H I is typically associated with absorber column densities $N_{\rm HI} = 10^{16}$–$10^{17}\,{\rm cm}^{-2}$ with neutral fractions $x_{\rm HI} \sim 10^{-4}$–10^{-2}, in

broad agreement with earlier work by Rahmati & Schaye (2018) using radiation hydrodynamical simulations performed in smaller volumes and at lower mass resolution.

(v) As a consequence of the typical H I absorption systems, $N_{\rm HI} \sim 10^{16}$–$10^{17}\,{\rm cm}^{-2}$, that dominate the Lyman-limit opacity at the tail-end of reionization, we argue that an important requirement for inhomogeneous reionization simulations that capture the redshift evolution of the mean free path at the end of reionization is resolving photon sinks in the diffuse IGM at moderate overdensities. Capturing these systems will also require accurate modelling of the hydrodynamical response of the gas to photo-heating during reionization (e.g. Puchwein et al. 2023).

(vi) The mean free path around dark matter haloes, $\Lambda_{\rm mfp}$, differs from the average IGM throughout and after reionization. We consider halo masses in the range $M_{\rm h} \sim 10^{9}$–$10^{11}\,h^{-1}{\rm M}_{\odot}$ in our fiducial simulation, but we observe the same general trends in a lower resolution simulation with 64 times the volume. Throughout most of reionization the halo mean free path is greater than the average IGM value, but during the final stages of reionization it is smaller. The enhancement during reionization (by as much as an order of magnitude) is because of the highly ionized bubbles centred around (clustered) haloes; these will be relevant for aiding the visibility of high redshift Ly α emitters at $z > 7$ (e.g. Larson et al. 2022; Tang et al. 2023; Asthana et al. 2024; Whitler et al. 2024). The suppression of $\Lambda_{\rm mfp}$ following reionization (typically by ~ 20 per cent relative to the average IGM) is because of the higher density of gas around haloes once the IGM is approximately in photoionization equilibrium, such that $x_{\rm H\,I} \propto \Delta$.

(vii) During reionization, the halo mean free path correlates with halo mass, partly because of ionizing source clustering but also because our modelling assumes the source ionizing luminosity is proportional to halo mass. However, following reionization, the halo mean free path is weakly anticorrelated with halo mass (see also Theuns & Chan 2024). These biases should be accounted for when measuring the mean free path around bright quasars at $z > 5$, but by $z = 6$ they are likely to be subdominant to the significantly enhanced ionization from the proximity effect around bright quasars (see e.g. Satyavolu et al. 2023; Roth et al. 2024).

Further progress with modelling the Lyman-limit mean free path at the end of reionization relies on capturing a large dynamic range. Our results suggest that a $160h^{-1}$ cMpc box with a dark matter particle mass of $5.37 \times 10^{5}\,h^{-1}\,{\rm M}_{\odot}$ (or equivalently 2×8192^3 particles) should yield a mean free path that is reasonably well converged with both box size and mass resolution. The required number of resolution elements is already similar to the CoDa III simulation (Lewis et al. 2022), but at present such models are expensive and have data volumes that are challenging to handle. Accounting for potential biases in mean free path measurements may also be valuable, although recent work has suggested these should already be well controlled in the existing data at $z \simeq 6$ (Satyavolu et al. 2023; Roth et al. 2024). Finally, the rather low emissivity preferred by the Ly α forest and Lyman-limit mean free path at $5 < z < 6$ (where $\dot{n} = 3 \times 10^{50}\,{\rm s}^{-1}\,{\rm cMpc}^{-3}$ corresponds to only 1–2 ionizing photons emitted per hydrogen atom over the age of the Universe at $z = 6$) remains an important constraint on the reionization photon budget, particularly given the apparent abundance of ionizing photons at $z > 6$ now implied by *JWST* (Muñoz et al. 2024). This tension may necessitate substantial redshift evolution in the properties of ionizing photon sources at $z > 6$, or a revision in our understanding of the IGM and structure formation during the reionization era.

ACKNOWLEDGEMENTS

The simulations used in this work were performed using the Joliot Curie supercomputer at the Trés Grand Centre de Calcul (TGCC) and the Cambridge Service for Data Driven Discovery (CSD3), part of which is operated by the University of Cambridge Research Computing on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). We acknowledge the Partnership for Advanced Computing in Europe (PRACE) for awarding us time on Joliot Curie in the 16th call. The DiRAC component of CSD3 was funded by BEIS capital funding via STFC capital grants ST/P002307/1 and ST/R002452/1 and STFC operations grant ST/R00689X/1. This work also used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility. The equipment was funded by BEIS capital funding via STFC capital grants ST/P002293/1 and ST/R002371/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. JSB, LC, and EC are supported by STFC consolidated grant ST/X000982/1. EC also acknowledges the support of a Royal Society Dorothy Hodgkin Fellowship and a Royal Society Enhancement Award. Support by ERC Advanced Grant 320596 'The Emergence of Structure During the Epoch of Reionization' is gratefully acknowledged. MGH has been supported by ⟨0:funding-source 3:href="http://dx.doi.org/10.13039/501100000271"⟩STFC⟨/0:funding-source⟩ consolidated grants ST/N000927/1 and ST/S000623/1. We thank Volker Springel for making P-GADGET-3 available. We also thank Dominique Aubert for sharing the ATON code, and Philip Parry for technical support. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

DATA AVAILABILITY

All data and analysis code used in this work are available from the first author on reasonable request. Further guidance on accessing the publicly available Sherwood–Relics simulation data may also be found at https://www.nottingham.ac.uk/astronomy/sherwood-relics/.

REFERENCES

Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37
Asthana S., Haehnelt M. G., Kulkarni G., Aubert D., Bolton J. S., Keating L. C., 2024, preprint (arXiv:2404.06548)
Atek H. et al., 2024, Nature, 626, 975
Aubert D., Teyssier R., 2008, MNRAS, 387, 295
Aubert D., Teyssier R., 2010, ApJ, 724, 244
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Becker G. D., D'Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, MNRAS, 508, 1853
Bolton J. S., Becker G. D., 2009, MNRAS, 398, L26
Bolton J. S., Haehnelt M. G., 2007, MNRAS, 382, 325
Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T.-S., Meiksin A., Regan J. A., Viel M., 2017, MNRAS, 464, 897
Bolton J. S., Gaikwad P., Haehnelt M. G., Kim T.-S., Nasir F., Puchwein E., Viel M., Wakker B. P., 2022, MNRAS, 513, 864
Bosman S. E. I., 2021, arXiv e-prints,arXiv:2108.12446
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055

Bosman S. E. I. et al., 2022, MNRAS, 514, 55
Bunker A. J. et al., 2023, A&A, 677, A88
Cain C., D'Aloisio A., Gangolli N., Becker G. D., 2021, ApJ, 917, L37
Cain C., D'Aloisio A., Lopez G., Gangolli N., Roth J. T., 2024, MNRAS, 531, 1951
Calverley A. P., Becker G. D., Haehnelt M. G., Bolton J. S., 2011, MNRAS, 412, 2543
Chardin J., Kulkarni G., Haehnelt M. G., 2018, MNRAS, 478, 1065
Crighton N. H. M., Prochaska J. X., Murphy M. T., O'Meara J. M., Worseck G., Smith B. D., 2019, MNRAS, 482, 1456
D'Aloisio A., McQuinn M., Davies F. B., Furlanetto S. R., 2018, MNRAS, 473, 560
Davies F. B., Furlanetto S. R., 2016, MNRAS, 460, 1328
Davies F. B. et al., 2018, ApJ, 864, 142
Davies F. B., Bosman S. E. I., Furlanetto S. R., Becker G. D., D'Aloisio A., 2021, ApJ, 918, L35
Davies F. B. et al., 2024, ApJ, 965, 134
Doughty C. C., Hennawi J. F., Davies F. B., Lukić Z., Oñorbe J., 2023, MNRAS, 525, 3790
Eilers A.-C., Davies F. B., Hennawi J. F., 2018, ApJ, 864, 53
Fan J., Chen H., Avestruz C., Khadir A., 2024, preprint (arXiv:2405.00100)
Finkelstein S. L. et al., 2019, ApJ, 879, 36
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., 2023, MNRAS, 525, 4093
Garaldi E., Kannan R., Smith A., Springel V., Pakmor R., Vogelsberger M., Hernquist L., 2022, MNRAS, 512, 4909
Georgiev I., Mellema G., Giri S. K., 2024, preprint (arXiv:2405.04273)
Gnedin N. Y., 2014, ApJ, 793, 29
Grazian A. et al., 2023, ApJ, 955, 60
Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, MNRAS, 466, 4239
Greig B., Mesinger A., Davies F. B., Wang F., Yang J., Hennawi J. F., 2022, MNRAS, 512, 5390
Gurvich A., Burkhart B., Bird S., 2017, ApJ, 835, 175
Haardt F., Madau P., 2012, ApJ, 746, 125
Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725
Iršič V. et al., 2024, Phys. Rev. D, 109, 043511
Jin X. et al., 2023, ApJ, 942, 59
Kaur H. D., Gillet N., Mesinger A., 2020, MNRAS, 495, 2354
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020a, MNRAS, 491, 1736
Keating L. C., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020b, MNRAS, 497, 906
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Larson R. L. et al., 2022, ApJ, 930, 104
Lewis J. S. W. et al., 2022, MNRAS, 516, 3389
Lu T.-Y., Mason C. A., Hutter A., Mesinger A., Qin Y., Stark D. P., Endsley R., 2024, MNRAS, 528, 4872
Lusso E., Fumagalli M., Rafelski M., Neeleman M., Prochaska J. X., Hennawi J. F., O'Meara J. M., Theuns T., 2018, ApJ, 860, 41
Madau P., Haardt F., Rees M. J., 1999, ApJ, 514, 648
Mason C. A. et al., 2018, ApJ, 857, L11
McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499
McQuinn M., Oh S. P., Faucher-Giguère C.-A., 2011, ApJ, 743, 82
Miller J. S. A., Bolton J. S., Hatch N., 2019, MNRAS, 489, 5381
Miralda-Escudé J., 2003, ApJ, 597, 66
Miralda-Escudé J., Cen R., Ostriker J. P., Rauch M., 1996, ApJ, 471, 582
Miralda-Escude J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1
Molaro M., Iršič V., Bolton J. S., Lieu M., Keating L. C., Puchwein E., Haehnelt M. G., Viel M., 2023, MNRAS, 521, 1489
Muñoz J. B., Mirocha J., Chisholm J., Furlanetto S. R., Mason C., 2024, arXiv e-prints,arXiv:2404.07250
Naidu R. P., Tacchella S., Mason C. A., Bose S., Oesch P. A., Conroy C., 2020, ApJ, 892, 109
Nakane M. et al., 2024, ApJ, 967, 28
Nasir F., Cain C., D'Aloisio A., Gangolli N., McQuinn M., 2021, ApJ, 923, 161
Neyer M. et al., 2024, MNRAS, 531, 2943
O'Meara J. M., Prochaska J. X., Worseck G., Chen H.-W., Madau P., 2013, ApJ, 765, 137
Ocvirk P., Lewis J. S. W., Gillet N., Chardin J., Aubert D., Deparis N., Thélie É., 2021, MNRAS, 507, 6108
Park H., Shapiro P. R., Choi J.-h., Yoshida N., Hirano S., Ahn K., 2016, ApJ, 831, 86
Park H., Lukić Z., Sexton J., Alvarez M., 2023, preprint (arXiv:2309.04129)
Pizzati E. et al., 2024, preprint (arXiv:2403.12140)
Planck CollaborationXVI, 2014, A&A, 571, A16
Prochaska J. X., Worseck G., O'Meara J. M., 2009, ApJ, 705, L113
Prochaska J. X., Madau P., O'Meara J. M., Fumagalli M., 2014, MNRAS, 438, 476
Puchwein E., Springel V., 2013, MNRAS, 428, 2966
Puchwein E. et al., 2023, MNRAS, 519, 6162
Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021, MNRAS, 506, 2390
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., D'Aloisio A., Cain C., Wilson B., Zhu Y., Becker G. D., 2024, MNRAS, 530, 5209
Satyavolu S., Kulkarni G., Keating L. C., Haehnelt M. G., 2023, preprint (arXiv:2311.06344)
Schaye J., 2001, ApJ, 559, 507
Schirber M., Bullock J. S., 2003, ApJ, 584, 110
Songaila A., Cowie L. L., 2010, ApJ, 721, 1448
Springel V., 2005, MNRAS, 364, 1105
Tang M. et al., 2023, MNRAS, 526, 1657
Theuns T., Chan T. K., 2024, MNRAS, 527, 689
Umeda H., Ouchi M., Nakajima K., Harikane Y., Ono Y., Xu Y., Isobe Y., Zhang Y., 2023, arXiv e-prints,arXiv:2306.00487
Verner D. A., Ferland G. J., 1996, ApJS, 103, 467
Verner D. A., Ferland G. J., Korista K. T., Yakovlev D. G., 1996, ApJ, 465, 487
Viel M., Haehnelt M. G., Springel V., 2004, MNRAS, 354, 684
Whitler L., Stark D. P., Endsley R., Chen Z., Mason C., Topping M. W., Charlot S., 2024, MNRAS, 529, 855
Worseck G. et al., 2014, MNRAS, 445, 1745
Wyithe J. S. B., Bolton J. S., Haehnelt M. G., 2008, MNRAS, 383, 691
Yang J. et al., 2020, ApJ, 904, 26
Zhu Y. et al., 2022, ApJ, 932, 76
Zhu Y. et al., 2023, ApJ, 955, 115
Ďurovčíková D. et al., 2024, preprint (arXiv:2401.10328)

APPENDIX A: EFFECT OF THE QUICK LY α APPROXIMATION ON THE MEAN FREE PATH

In this appendix, we test the effect of the quick-Ly α approximation used in Sherwood–Relics on the Lyman-limit mean free path at $z > 5$. Although we do not have a Sherwood-Relics model that includes gas with $\Delta > 1000$ and $T < 10^5$ K, we may gain insight using two simulations from the original Sherwood simulation project (Bolton et al. 2017). Here we make use of the 40–1024 and 40-1024-ps13 models presented in table 1 of Bolton et al. (2017), which we refer to here as Sherwood-qLyα and Sherwood-ps13, respectively. These have the same box size as the 40–2048 model, but use a dark matter particle mass that is 8 times larger. The Sherwood models do not model inhomogeneous reionization. Instead, a spatially uniform UV background with an IGM that remains highly ionized at $z > 6$ is assumed (Haardt & Madau 2012). A correction for the self-shielding of dense hydrogen gas to ionizing photons is also included, following Chardin, Kulkarni & Haehnelt (2018). For the purpose of this comparison, the key difference is that Sherwood-qLyα uses the same quick-Lyα approximation used in this work, whereas Sherwood-

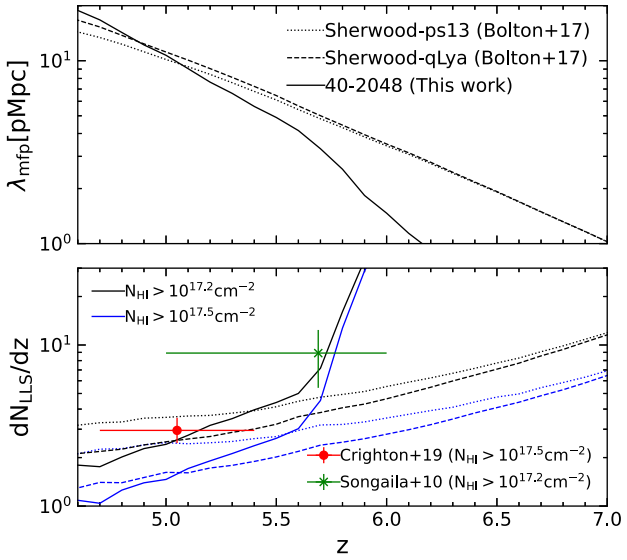

Figure A1. *Top panel:* Comparison of the mean free path in two of the Sherwood simulations (Bolton et al. 2017) to the 40–2048 simulation (solid curve) used in this work. Sherwood-ps13 (dotted curve) uses a sub-grid model for star formation and feedback (Puchwein & Springel 2013) and explicitly follows high density gas in haloes, whereas Sherwood-qLyα (dashed curve) uses the same approximation as 40–2048, where all gas with $\Delta > 1000$ and $T < 10^5$ K is converted into collisionless particles. Both Sherwood models assume a homogeneous UV background, following Haardt & Madau (2012), and the IGM remains highly ionized at $z > 6$. *Bottom panel:* The corresponding number of Lyman limit systems per unit redshift for $N_{\rm H\,I} > 10^{17.2}$ cm^{-2} (black curves) and $N_{\rm H\,I} > 10^{17.5}$ cm^{-2} (blue curves). Observational data are from Songaila & Cowie (2010) (green cross) and Crighton et al. (2019) (red diamond).

ps13 uses the star formation and energy-driven outflow model from Puchwein & Springel (2013). The models are otherwise identical; the difference between them gives an indication of how missing gas with $\Delta > 1000$ and $T < 10^5$ K will alter the predicted mean free path.

The results of the comparison are displayed in Fig. A1. The upper panel shows the mean free path, while the lower panel displays the number of Lyman-limit systems per unit redshift, $\mathrm{d}N_{\rm LSS}/\mathrm{d}z$. The difference in the mean free path predicted by Sherwood-ps13 and Sherwood-qLyα is at most 9 per cent $z \geq 5$, with an average difference of 3 per cent at $5 \leq z \leq 7$. This difference decreases with increasing redshift as the typical overdensity of gas that is optically thick to Lyman-limit photons decreases. A similar trend is observed for $dN_{\rm LSS}/dz$, where the difference between the Sherwood models at $z \simeq 5$ is comparable to the uncertainty on the observational measurements from Crighton et al. (2019). Note, however, that the effect of missing gas will be much larger at lower redshifts as the typical overdensity of Lyman-limit systems increases. At $z < 3$, the quick-Lyα model does not reproduce the observed incidence of Lyman limit systems.[5]

We furthermore note that the 40–2048 model is already in good agreement with the observed Lyman-limit systems with $N_{\rm HI} > 10^{17.2}$ cm^{-2} at $z = 6$, although it underpredicts the number of systems with $N_{\rm HI} > 10^{17.5}$ cm^{-2} at $z = 5$. As discussed in Section 3 and Appendix B below, however, we caution that the number of systems will be sensitive to our choice of integration window size.

[5]See, for example, fig. 1 in Miller, Bolton & Hatch (2019).

APPENDIX B: CHOICE OF INTEGRATION WINDOW FOR CALCULATING COLUMN DENSITIES

As discussed in Section 3, we must select a length scale for computing H I column densities, $N_{\rm HI}$. In this work we have considered windows equal to the (redshift dependent) Jeans length at the mean background density (Schaye 2001; Rahmati & Schaye 2018), and a fixed window of 50 km s^{-1} (e.g. Gurvich et al. 2017). We obtain the proper Jeans length, $\lambda_{\rm J}$, for gas at the mean density using,

$$\lambda_{\rm J} = \left(\frac{40\pi^2 k_{\rm B} \langle T \rangle}{9 \langle \mu \rangle m_{\rm H} H_0^2 \Omega_{\rm m} (1+z)^3} \right)^{1/2}. \tag{B1}$$

Here $\langle T \rangle$ is the volume averaged temperature in the simulation, and $\langle \mu \rangle$ is the volume averaged mean molecular weight,

$$\langle \mu \rangle = \mu_{\rm ionised}(1 - f_{\rm neutral}) + \mu_{\rm neutral} f_{\rm neutral}, \tag{B2}$$

where $\mu_{\rm ionised} = 0.61$ for an H II and He II admixture, $\mu_{\rm neutral} = 1.22$ and $f_{\rm neutral}$ is the fraction of the simulation volume with $x_{\rm HI} > 0.5$. Adding some representative values, equation (B1) then reduces to,

$$\lambda_{\rm J} = 90.5 \text{ pkpc} \left(\frac{\langle T \rangle}{10^4 \text{ K}} \right)^{1/2} \left(\frac{\langle \mu \rangle}{0.61} \right)^{-1/2} \left(\frac{1+z}{8} \right)^{-3/2}. \tag{B3}$$

Fig. 5 displays the column density distribution function obtained using equation (B1). For comparison, Fig. B1 shows the column density distribution for when fixed windows of 50 km s^{-1} are used. The overall shape of the distribution is consistent between the two window sizes; we do not expect qualitative differences to arise from our choice of integration window. It should nevertheless be noted that the absolute values for $N_{\rm HI}$ will always depend on the choice of integration window.

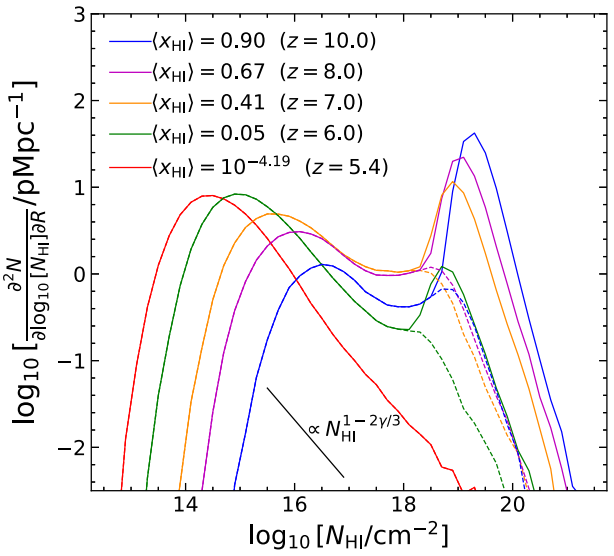

Figure B1. The H I column density distribution, as described in Fig. 5, but now $N_{\rm H\,I}$ has been calculated by integrating over a fixed window of size 50 km s^{-1}. Note that although the peaks of the distribution appear at different column densities, the shape of the distribution is robust to the window choice.

APPENDIX C: ANALYTICAL SCALING RELATIONS

A commonly used approximation for modelling the Lyman-limit photon mean free path and the opacity of the IGM – particularly when correcting for small-scale, unresolved photon sinks in the already reionized IGM – is to use an analytical relation that captures how the opacity varies with the local H I photoionization rate (e.g. McQuinn et al. 2011; Davies & Furlanetto 2016; Becker et al. 2021). The details of this scaling relation derive from the model presented in Miralda-Escude, Haehnelt & Rees (2000, hereafter MHR00), as already discussed elsewhere (Furlanetto & Oh 2005; McQuinn et al. 2011). Here we find it useful to recapitulate this model for a direct comparison to Sherwood–Relics.

In the MHR00 analytical reionization model, all gas up to some critical density, $\Delta_{\rm i}$ is assumed ionized, and gas above this density is fully neutral. The mean free path of ionizing photons is then proportional to the chance of an ionizing photon meeting a neutral photon sink. Given the volume-filling fraction of ionized gas $F_{\rm V}$, MHR00 make the ansatz that the typical extent of a ionizing photon sink is $(1 - F_{\rm V})^{1/3}$; the chance of encountering a photon sink along a random line of sight is then $(1 - F_{\rm V})^{1/3}/(1 - F_{\rm V}) = (1 - F_{\rm V})^{-2/3}$. Hence, the mean free path $\lambda \propto (1 - F_{\rm V})^{-2/3}$.

It is straightforward to compute $F_{\rm V}$ given the differential probability distribution function for the gas density, $P(\Delta)$,

$$F_{\rm V} = \int_0^{\Delta_{\rm i}} P(\Delta){\rm d}\Delta, \tag{C1}$$

and so

$$\lambda \propto \left[\int_{\Delta_i}^{\infty} P(\Delta){\rm d}\Delta\right]^{-2/3}. \tag{C2}$$

MHR00 showed that $P(\Delta)$ from cosmological hydrodynamical simulations (Miralda-Escudé et al. 1996) is well fit over the range $2 \leq z \leq 4$ using a lognormal distribution with a high density, power-law tail,

$$P(\Delta){\rm d}\Delta = A \exp\left[-\frac{\left(\Delta^{-2/3} - C_0\right)^2}{2\,(\delta_0/3)^2}\right] \Delta^{-\gamma}{\rm d}\Delta. \tag{C3}$$

MHR00 extrapolated this fit to higher redshifts, $z > 4$, assuming $\gamma = 5/2$.

Turning now to the production and absorption of ionizing photons, assuming that the mean free path of an ionizing photon is much smaller than the horizon scale, such that cosmological effects can be neglected (sometimes called the 'local source' approximation, e.g. Madau, Haardt & Rees 1999; Schirber & Bullock 2003), then we can write $\Gamma \propto \epsilon\lambda$, where Γ is the hydrogen photoionization rate and ϵ is the ionizing emissivity (i.e. the number of ionizing photons emitted per unit time per unit volume). When the IGM is highly ionized, the emissivity is approximately balanced by the recombination rate, $\epsilon \simeq \mathcal{R}$, where

$$\mathcal{R} = \alpha(T) n_{\rm e} n_{\rm H\,II}. \tag{C4}$$

Here $\alpha(T)$ is the recombination coefficient, $n_{\rm e}$ is the number density of electrons and $n_{\rm H\,II}$ is the number density of ionized hydrogen. If the gas is approximately isothermal, $\alpha(T) \simeq \alpha$, and if it is also highly ionized, $n_{\rm e} \approx n_{\rm H\,II}$. We may then write

$$\mathcal{R} = \alpha \bar{n}_{\rm e}^2 \int_0^{\Delta_i} \Delta^2 P(\Delta){\rm d}\Delta, \tag{C5}$$

where $\bar{n}_{\rm e}$ is the mean number density of electrons.

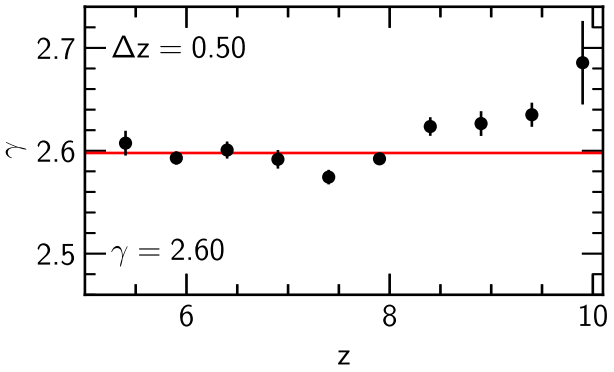

Figure C1. Estimate of the slope of the power-law high density tail, $P(\Delta) \propto \Delta^{-\gamma}$, in equation (C3) from the 40–2048 Sherwood–Relics model used in this work. The values of γ and their associated 1σ standard deviation are obtained by fitting $p(\Delta)$ in the simulation in steps of $\Delta z = 0.1$, before combining into bins of $\Delta z = 0.5$ to reduce noise (black points). The red line indicates a constant value of $\gamma = 2.6$ provides a reasonable fit to the data at $z < 8$.

The final step is to notice that most recombinations happen at high densities, and that equation (C3) tends to a power law $P(\Delta) \propto \Delta^{-\gamma}$ for $\Delta \gg 1$. Plugging this into equation (C2) we find

$$\lambda \propto \Delta_i^{2(\gamma-1)/3}, \tag{C6}$$

and into equation (C5) yields

$$\mathcal{R} \propto \Delta_i^{3-\gamma}. \tag{C7}$$

Assuming $\Gamma \propto \epsilon\lambda \simeq \mathcal{R}\lambda$ and combining equations (C6) and (C7) gives an expression for the photoionization rate

$$\Gamma \propto \Delta_i^{(7-\gamma)/3}. \tag{C8}$$

We can now use equations (C8) in (C6) to remove the explicit dependence on $\Delta_{\rm i}$, giving

$$\lambda \propto \Gamma^{\xi}, \tag{C9}$$

where $\xi = 2(\gamma - 1)/(7 - \gamma)$. For the value of $\gamma = 5/2$ presented in MHR00, this leads to $\lambda \propto \Gamma^{2/3}$. For the opacity, $\kappa \propto \lambda^{-1}$, and so $\kappa \propto \Gamma^{-2/3}$ (e.g. Davies & Furlanetto 2016; Becker et al. 2021).

We first check the expected slope of the high density power-law tail, γ, in equation (C3). In Fig. C1 we show the result of fitting $P(\Delta) \propto \Delta^{-\gamma}$ for $\log_{10}\Delta \geq 1.2$ from $z = 5.2$ to $z = 10.1$ in the 40–2048 model. We group the individual values of γ into bins of width $\Delta z = 0.5$ to reduce noise – within each bin we compute the mean γ and standard deviation. We find that γ is not independent with redshift, but a value of $\gamma = 2.6$ provides a reasonable approximation during the second half of reionization at $z < 8$.

Next, we estimate ξ, the exponent in equation (C9), from our simulation. To do this, we start by computing the Lyman-limit photon attenuation length as

$$\lambda = \left(n_{\rm H\,I}\sigma_{912}\right)^{-1}, \tag{C10}$$

where $n_{\rm H\,I}$ is the number density of neutral hydrogen at a given position in the simulation, and σ_{912} is the photoionization cross-section at the Lyman limit. Note that this is not equivalent to the mean free path of ionizing photons defined in equation (2).

Furthermore, equation (C9) is derived under the assumption of photoionization equilibrium in the reionized IGM, so to estimate ξ we must also select regions where this condition holds in the simulation. To do this, we require that

$$\mathcal{E} = \frac{x_{\rm HI}\Gamma}{n_{\rm H}(1 + \chi_{\rm e})x_{\rm H\,II}^2\alpha(T)} \approx 1, \tag{C11}$$

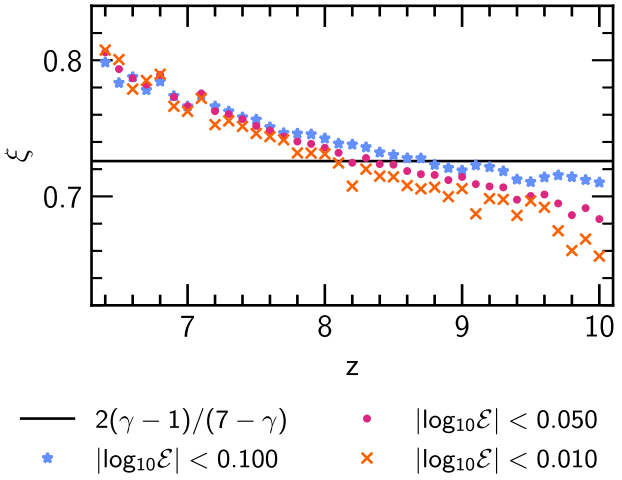

Figure C2. The value of the power-law slope, ξ, obtained by fitting equation (C9) to the distribution of the attenuation length $\lambda = (n_{\rm H\,I}\sigma_{912})^{-1}$ and local photoionization rate Γ in the 40–2048 model as a function of redshift. The coloured data points display the result when applying ionization cuts to the simulation data that select for gas where ionization equilibrium holds (see the text for details). The black horizonal line shows the analytically expected value of $\xi = 2(\gamma - 1)/(7 - \gamma)$ for $\gamma = 2.6$.

is satisfied, where $x_{\rm HI}$ is the neutral hydrogen fraction, $x_{\rm H\,II} = 1 - x_{\rm H\,I}$ is the ionized hydrogen fraction and $\chi_{\rm e} = 0.079$ is a correction to account for electrons due to singly ionized helium (assuming that helium is ionized once at the same time as hydrogen). In practice, when applying equation (C11) we allow for some small deviation around equilibrium by applying a cut on $|\log_{10}\mathcal{E}|$.

Fig. C2 shows the value of ξ obtained by fitting equation (C9) to the distribution of $\lambda = (n_{\rm H\,I}\sigma_{912})^{-1}$ against Γ for (already reionized) gas at each redshift. We find that, for reasonable values of $|\log_{10}\mathcal{E}|$, the choice of ionization cut makes a small difference to the value of ξ and no difference to the redshift evolution. At $z \geq 6.3$, the typical value of ξ we obtain directly from the simulation (data points in Fig. C2) overlaps with the analytical scaling relation (solid black line) derived from the MHR00 model for $\gamma = 2.6$, although there is evidence that a ξ that increases toward lower redshift may be more appropriate. However, this agreement breaks at the point where the ionizing radiation field is close to spatially uniform throughout the simulation box, and it is no longer possible to fit for the λ–Γ correlation in the diffuse IGM. We therefore caution against extrapolation of the redshift dependence of ξ to $z < 6.3$.

Finally, returning to the H I column density distribution function, $f(N_{\rm H\,I}, R)$, Schaye (2001) showed that the extent of overdense absorbers will typically be of order the local Jeans length. Taking now the general expression for the proper Jeans scale (cf. equation B3), we have

$$\lambda_{\rm J} = \left(\frac{40\pi^2 k_{\rm B} T}{9\mu m_{\rm H} H_0^2 \Omega_{\rm m}(1+z)^3 \Delta}\right)^{1/2}. \tag{C12}$$

Dropping any constants and dependence on redshift or cosmology, we can write

$$\lambda_{\rm J} \propto T^{1/2}\mu^{-1/2}\Delta^{-1/2}. \tag{C13}$$

Using this relation in equation (C11) we can express $x_{\rm H\,I}$, for highly ionized gas, as

$$x_{\rm H\,I} \propto T^{-0.72}\Delta\Gamma^{-1}, \tag{C14}$$

where we have used $\alpha(T) \propto T^{-0.72}$ (Verner & Ferland 1996; Bolton et al. 2022). If the typical length of an absorber is $\lambda_{\rm J}$, then its typical column density is $N_{\rm H\,I} \approx \bar{n}_{\rm H}\Delta x_{\rm H\,I}\lambda_{\rm J}$, which we can express as

$$N_{\rm H\,I} \propto T^{-0.22}\mu^{-1/2}\Delta^{3/2}\Gamma^{-1}. \tag{C15}$$

Next, following Furlanetto & Oh (2005), we can relate the H I column density distribution function to the underlying gas density distribution by

$$f(N_{\rm H\,I}, R) \propto \frac{x_{\rm H\,I}}{N_{\rm H\,I}}\Delta P(\Delta)\frac{{\rm d}\Delta}{{\rm d}N_{\rm H\,I}}. \tag{C16}$$

We make one further assumption, namely that, for highly ionized gas T, μ, and Γ are all approximately constant, such that from equation (C15)

$$\frac{{\rm d}N_{\rm H\,I}}{{\rm d}\Delta} \simeq \left(\tfrac{\partial N_{\rm H\,I}}{\partial \Delta}\right)_{T,\mu,\Gamma}, \tag{C17}$$

$$\propto \Delta^{1/2}. \tag{C18}$$

This can be plugged into equation (C16) and the other terms can be reduced to their relevant density dependences (i.e. $x_{\rm H\,I} \propto \Delta$ and $N_{\rm H\,I} \propto \Delta^{3/2}$) to give

$$f(N_{\rm H\,I}, R) \propto P(\Delta). \tag{C19}$$

Finally, using the MHR00 high density power-law fit to $P(\Delta)$, we can write $f(N_{\rm H\,I}, R) \propto P(\Delta) \propto \Delta^{-\gamma}$ and hence

$$f(N_{\rm H\,I}, R) \propto N_{\rm H\,I}^{-2\gamma/3}. \tag{C20}$$

This result is plotted in Fig. 5 for $\gamma = 2.6$; we observe the slope is similar to the high column density tail of $f(N_{\rm H\,I}, R)$ following reionization.

This paper has been typeset from a TEX/LATEX file prepared by the author.