



**HAL**  
open science

## The detectability of strong 21-cm forest absorbers from the diffuse intergalactic medium in late reionization models

Tomáš Šoltinský, James S. Bolton, Nina Hatch, Martin G. Haehnelt, Laura C. Keating, Girish Kulkarni, Ewald Puchwein, Jonathan Chardin, Dominique Aubert

### ► To cite this version:

Tomáš Šoltinský, James S. Bolton, Nina Hatch, Martin G. Haehnelt, Laura C. Keating, et al.. The detectability of strong 21-cm forest absorbers from the diffuse intergalactic medium in late reionization models. *Monthly Notices of the Royal Astronomical Society*, 2021, 506 (4), pp.5818-5835. <10.1093/mnras/stab1830>. <hal-03229347>

**HAL Id: hal-03229347**

**<https://hal.science/hal-03229347v1>**

Submitted on 22 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# The detectability of strong 21-cm forest absorbers from the diffuse intergalactic medium in late reionization models

Tomáš Šoltinský<sup>1</sup>,<sup>\*</sup> James S. Bolton<sup>1</sup>, Nina Hatch<sup>1</sup>, Martin G. Haehnelt<sup>2</sup>, Laura C. Keating<sup>3</sup>, Girish Kulkarni<sup>4</sup>, Ewald Puchwein<sup>3</sup>, Jonathan Chardin<sup>5</sup> and Dominique Aubert<sup>5</sup>

<sup>1</sup>*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

<sup>2</sup>*Kavli Institute for Cosmology and Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK*

<sup>3</sup>*Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany*

<sup>4</sup>*Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*

<sup>5</sup>*Observatoire Astronomique de Strasbourg, 11 rue de l'Université, F-67000 Strasbourg, France*

Accepted 2021 June 24. Received 2021 June 23; in original form 2021 March 17

## ABSTRACT

A late end to reionization at redshift  $z \simeq 5.3$  is consistent with observed spatial variations in the Ly  $\alpha$  forest transmission and the deficit of Ly  $\alpha$  emitting galaxies around extended Ly  $\alpha$  absorption troughs at  $z = 5.5$ . In this model, large islands of neutral hydrogen should persist in the diffuse intergalactic medium (IGM) until  $z \simeq 6$ . We use a novel, hybrid approach that combines high-resolution cosmological hydrodynamical simulations with radiative transfer to predict the incidence of strong 21-cm forest absorbers with optical depths  $\tau_{21} > 10^{-2}$  from the diffuse IGM in these late reionization models. We include the effect of redshift space distortions on the simulated 21-cm forest spectra, and treat the highly uncertain heating of the pre-reionization IGM by soft X-rays as a free parameter. For a model with only modest IGM pre-heating, such that average gas kinetic temperatures in the diffuse IGM remain below  $T_K \simeq 10^2$  K, we find that strong 21-cm forest absorption lines should persist until  $z = 6$ . For a sample of  $\sim 10$  sufficiently radio-loud background sources, a null-detection of 21-cm forest absorbers at  $z \simeq 6$  with SKA1-low or possibly LOFAR should provide an informative lower limit on the still largely unconstrained soft X-ray background at high redshift and the temperature of the pre-reionization IGM.

**Key words:** methods: numerical – intergalactic medium – quasars: absorption lines – dark ages, reionization, first stars.

## 1 INTRODUCTION

At present, the premier technique for examining the small-scale structure of intergalactic neutral hydrogen approaching the reionization era is Lyman series absorption in the spectra of luminous quasars (Becker et al. 2015; Eilers et al. 2017; Bosman et al. 2018; Yang et al. 2020b). However, it is challenging to probe the intergalactic medium (IGM) much beyond redshift  $z \simeq 6$  with this approach. The large cross-section for Ly  $\alpha$  scattering means the IGM becomes opaque to Ly  $\alpha$  photons at neutral hydrogen fractions as low as  $x_{\text{HI}} \simeq 10^{-4}$ . An alternative transition that overcomes this limitation is the hyperfine 21-cm line, which has a cross-section that is a factor of  $\sim 10^7$  smaller than the Ly  $\alpha$  transition.<sup>1</sup> If radio bright sources such as high-redshift quasars (Bañados et al. 2021) or gamma-ray bursts (e.g. Ioka & Mészáros 2005; Ciardi et al. 2015b) can be identified during the reionization era, the intervening neutral IGM may be observed as a 21-cm forest of absorption lines in their spectra. This can be achieved either through the direct identification of individual

absorption features (Carilli, Gnedin & Owen 2002; Furlanetto & Loeb 2002; Meiksin 2011; Xu, Ferrara & Chen 2011; Ciardi et al. 2013; Semelin 2016; Villanueva-Domingo & Ichiki 2021), or by the statistical detection of the average 21-cm forest absorption (Mack & Wyithe 2012; Ewall-Wice et al. 2014; Thyagarajan 2020). This approach is highly complementary to proposed tomographic studies of the redshifted 21-cm line and measurements of the 21-cm power spectrum during reionization (e.g. Mertens et al. 2020; Trott et al. 2020), as it is subject to a different set of systematic uncertainties (Furlanetto, Peng Oh & Briggs 2006; Pritchard & Loeb 2012).

However, any detection of the 21-cm forest relies on the identification of sufficient numbers of radio-loud sources and the existence of cold, neutral gas in the IGM at  $z \gtrsim 6$ . While neither of these criteria are guaranteed, the prospects for both have improved somewhat in the last few years. Approximately  $\sim 10$  radio-loud active galactic nuclei are now known at  $5.5 < z < 6.5$  (e.g. Bañados et al. 2018b, 2021; Ighina et al. 2021; Liu et al. 2021), including the  $z = 6.1$  blazar PSO J0309 + 27 with a flux density  $S_{147\text{MHz}} = 64.2 \pm 6.2$  mJy (Belladitta et al. 2020). The Low Frequency Array (LOFAR) Two-metre Sky Survey (LoTSS; Shimwell et al. 2017; Kondapally, Best & Hardcastle 2021), the Giant Metrewave Radio Telescope (GMRT) all-sky radio survey at 150 MHz (Intema et al. 2017), and the Galactic and Extragalactic All-sky Murchison Widefield Array survey (GLEAM; Wayth et al. 2015) are also projected to detect hundreds of bright  $z > 6$  radio sources, particularly if coupled with large spectroscopic

\* E-mail: [tomas.šoltinský@nottingham.ac.uk](mailto:tomas.šoltinský@nottingham.ac.uk)

<sup>1</sup>Low-ionization metal lines such as O I (Oh 2002; Keating et al. 2014) and Mg II (Hennawi et al. 2021) can also be used to trace neutral intergalactic gas, although the uncertain metallicity of the high-redshift IGM further complicates their interpretation.

follow-up programmes such as the William Herschel Telescope Enhanced Area Velocity Explorer (WEAVE)-LOFAR survey (Smith et al. 2016).

Furthermore, there is now growing evidence that reionization ended rather late, and possibly even extended to redshifts as late as  $z \simeq 5.3$  (Kulkarni et al. 2019; Nasir & D’Aloisio 2020; Qin et al. 2021). This picture is motivated by the large spatial fluctuations observed in the Ly $\alpha$  forest transmission at  $z \simeq 5.5$  (Becker et al. 2015; Eilers, Davies & Hennawi 2018). A late end to reionization is also consistent with the electron scattering optical depth inferred from the cosmic microwave background (CMB; Pagano et al. 2020), the observed deficit of Ly $\alpha$  emitting galaxies around extended Ly $\alpha$  absorption troughs (Becker et al. 2018; Kashino et al. 2020; Keating et al. 2020), the clustering of Ly $\alpha$  emitters (Weinberger, Haehnelt & Kulkarni 2019), the thermal widths of Ly $\alpha$  forest transmission spikes at  $z > 5$  (Gaikwad et al. 2020), and the mean-free path of ionizing photons at  $z = 6$  (Becker et al. 2021). If this interpretation proves to be correct (but see D’Aloisio, McQuinn & Trac 2015; Davies & Furlanetto 2016; Chardin, Puchwein & Haehnelt 2017; Meiksin 2020, for alternative explanations), then there should still be large islands of neutral hydrogen in the IGM as late as  $z = 6$  (e.g. Lidz et al. 2007; Mesinger 2010). If this neutral gas has not already been heated to kinetic temperatures  $T_K \gtrsim 10^3$  K by the soft X-ray background, then it may be possible to detect 21-cm absorbers in the pre-reionization IGM at  $z \simeq 6$ . Alternatively, a null-detection could provide a useful lower limit on the soft X-ray background at high redshift.

The goal of this work is to investigate this possibility further. We use a set of high-resolution hydrodynamical cosmological simulations drawn from the Sherwood–Relics simulation programme (Puchwein et al., in preparation). Using a novel hybrid approach, these are combined with the ATON radiative transfer code (Aubert & Teyssier 2008) to model the small-scale structure of the IGM. Following Kulkarni et al. (2019), we consider a model with late reionization ending at  $z = 5.3$ , and contrast this with a simulation that has an earlier end to reionization at  $z = 6.7$ . We pay particular attention to some of the common assumptions adopted in previous models of the 21-cm forest that affect the absorption signature on small scales. This includes the treatment of gas peculiar motions and thermal broadening, the coupling of the spin temperature to the Ly $\alpha$  background, and the effect of pressure (or Jeans) smoothing on the IGM. Our approach is therefore closest to the earlier work by Semelin (2016), although we do not follow spatial variations in the X-ray and Ly $\alpha$  backgrounds. Instead, we attempt to explore a broader range of parameter space for spatially uniform X-ray heating using hydrodynamical simulations that use several different reionization histories and have an improved mass resolution (by a factor of  $\sim 27$ ). Note, however, that even the high-resolution cosmological simulations considered here will still only capture the 21-cm absorption that arises from the diffuse IGM. We therefore do not model the (uncertain amount of) absorption from neutral gas in haloes below the atomic cooling threshold, or from the cold interstellar medium in much rarer, more massive haloes that host high-redshift galaxies (see e.g. Furlanetto & Loeb 2002; Meiksin 2011).

This paper is structured as follows. We start by describing our numerical model of the IGM in Section 2, and our calculation of the 21-cm optical depths in Section 3. We examine how different modelling assumptions affect the observability of strong 21-cm forest absorbers in Sections 4.1 and 4.2, and estimate how a null-detection of strong 21-cm absorbers at redshift  $z \simeq 6$  with LOFAR or the Square Kilometre Array (SKA) could constrain the high-redshift soft X-ray background in Section 4.3. Finally, we conclude in

Section 5. The appendix contains further technical details regarding our methodology and modelling assumptions.

## 2 NUMERICAL MODEL FOR THE 21-CM FOREST

### 2.1 Hydrodynamical simulations with radiative transfer

We model the 21-cm forest during inhomogeneous reionization using a subset of the high-resolution cosmological hydrodynamical simulations drawn from the Sherwood–Relics simulation programme (see Gaikwad et al. 2020, for an initial application of these models). The Sherwood–Relics simulations were performed with a modified version of the P-GADGET-3 code – which is itself an updated version of the GADGET-2 code described in (Springel 2005) – and uses the same initial conditions as the earlier Sherwood simulation suite (Bolton et al. 2017). In this work, we adopt a flat  $\Lambda$ CDM cosmology with  $\Omega_\Lambda = 0.692$ ,  $\Omega_m = 0.308$ ,  $\Omega_b = 0.0482$ ,  $\sigma_8 = 0.829$ ,  $n_s = 0.961$ ,  $h = 0.678$ , consistent with Planck Collaboration XVI (2014), and a primordial helium fraction by mass of  $Y_p = 0.24$  (Hsyu et al. 2020).

The simulations have a volume  $(40h^{-1}\text{cMpc})^3$  and track  $2 \times 2048^3$  dark matter and gas particles. This yields a dark matter particle mass of  $7.9 \times 10^5 M_\odot$  and resolves dark matter haloes with masses greater than  $\sim 2.5 \times 10^7 M_\odot$ . This high-mass resolution is necessary for capturing the small-scale intergalactic structure probed by the 21-cm forest (cf. Semelin 2016). We furthermore adopt a simple but computationally efficient scheme for converting high-density gas into collisionless particles that robustly predicts the properties of the IGM. If a gas particle has an overdensity  $\Delta = 1 + \delta > 1000$  and kinetic temperature  $T_K < 10^5$  K, it is converted into a collisionless star particle (Viel, Haehnelt & Springel 2004). We have verified this simplified approach is sufficient for modelling the 21-cm forest in the diffuse IGM by direct comparison to a full subgrid star formation model (see Appendix A for further details). The main effect of this approximation is the removal of dense gas from haloes, which slightly reduces the number of strong 21-cm absorbers in models with no X-ray heating.

In order to include the effect of inhomogeneous reionization by UV photons on the IGM, the Sherwood–Relics simulations are combined with the moment-based, M1-closure radiative transfer code ATON (Aubert & Teyssier 2008). We adopt a novel hybrid approach that captures the small-scale hydrodynamical response of the gas in the simulations to patchy heating during reionization (see also Oñorbe et al. 2019, for a related approach). Our hybrid RT/hydrodynamical simulations use inputs in the form of 3D maps of the reionization redshift and H I photoionization rate, produced by ATON simulations performed on the P-GADGET-3 outputs in post-processing. These maps are then fed back into a re-run of the P-GADGET-3 model, where they are called within a non-equilibrium thermo-chemistry solver (Puchwein et al. 2015). Following Kulkarni et al. (2019), the ionizing sources in the ATON simulations have luminosities proportional to the halo mass with a redshift-dependent normalization, and the mean energy of ionizing photons is assumed to be 18.6 eV. Further details can be found in Gaikwad et al. (2020) and Puchwein et al. (in preparation). The main advantage of this approach is that since the post-processing step using the ATON radiative transfer simulations is computationally cheap compared to the hydrodynamical simulations, we may empirically calibrate the source model to yield a reionization history that is consistent with a wide range of observational constraints. This avoids many of the uncertainties associated with direct hydrodynamical modelling of the source population.

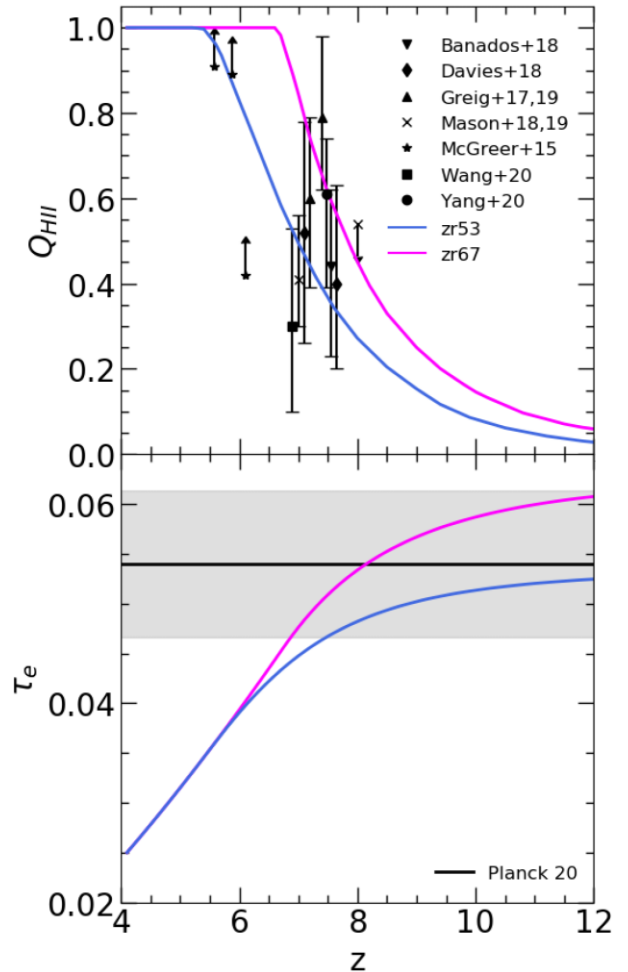
Note, however, the relatively small volume of our simulations means the patchy structure of reionization will not be fully captured on scales larger than the box size. This will lead to smaller neutral islands and an earlier percolation of ionized regions relative to simulations performed in a larger volume (Iliev et al. 2014; Kaur, Gillet & Mesinger 2020). We therefore adjust the ionizing emissivity in the models by hand to achieve a given reionization history; this scaling is equivalent to varying the escape fraction of ionizing photons. In addition, while our ATON simulations self-consistently follow the propagation of ionizing photons using a  $2048^3$  Cartesian grid, self-shielded regions below the size of the grid cells ( $\lesssim 20h^{-1}$  ckpc) will not be resolved. We attempt to partially correct for this by implementing a correction for the self-shielding of dense gas in all our simulations in post-processing, using the results of Chardin, Kulkarni & Haehnelt (2018). We find, however, that this correction makes almost no difference to our final results, as the majority of the strong 21-cm absorbers in our simulations arise from the diffuse IGM.

We consider two different reionization histories in this work, where the IGM becomes fully ionized at either  $z = 5.3$  (model zr53) or 6.7 (model zr67). The filling fraction of ionized gas and the Thomson scattering optical depth predicted by these models are displayed in Fig. 1. Both of these models are consistent with current observational constraints on the timing of reionization. As already discussed, the reionization model that ends at  $z = 5.3$  is furthermore consistent with the large fluctuations in the Ly $\alpha$  forest transmission observed at  $z \simeq 5.5$  (Becker et al. 2015; Kulkarni et al. 2019; Keating et al. 2020). Finally, we also use a simulation (zr53-homog) that has been constructed to give exactly the same globally averaged reionization history as the zr53 model, but using a spatially uniform ionizing background. A comparison between the zr53 and zr53-homog simulations therefore allows us to estimate the uncertain effect that pressure smoothing (from e.g. UV photoheating) may have on the gas in the pre-reionization IGM (see Section 4.2). All the simulations used in this work are listed in Table 1, where the final two models listed are used in Appendix A only.

## 2.2 Heating of neutral gas by the X-ray and Ly $\alpha$ backgrounds

Absorption features in the 21-cm forest arise from neutral hydrogen in the IGM. In addition to modelling the inhomogeneous reionization of the IGM by UV photons, we must therefore also consider the temperature and ionization state of gas that is optically thick to Lyman continuum photons. This heating is attributable to adiabatic compression and shocks – which are already included within our hydrodynamical simulations – and the X-ray and (to a lesser extent) Ly $\alpha$  radiation backgrounds at high redshift (Ciardi, Salvaterra & Di Matteo 2010), which are not. Hence, we now describe the procedure we use to include spatially uniform X-ray and Ly $\alpha$  heating in our simulations, by recalculating the density-dependent temperature and ionization state of the neutral gas in our hybrid simulations in post-processing. Further details on the model we use are also provided in Appendix B.

As we do not directly model the star formation rate in our simulations, rather than using a detailed model for the number and spectral energy distribution of X-ray sources at high redshift, for simplicity and ease of comparison to the existing literature we instead follow the approach introduced by Furlanetto (2006b) for parametrizing the comoving X-ray background emissivity. This uses the observed relationship between the star formation rate, SFR, and hard X-ray band luminosity (2–10 keV) for star-forming galaxies at  $z = 0$  (Gilfanov, Grimm & Sunyaev 2004; Lehmer et al. 2016).



**Figure 1.** Top panel: the filling factor of ionized gas in the zr53 (blue solid curve) and zr67 (fuchsia solid curve) simulations, compared to observational constraints from dark gaps in the Ly $\alpha$  forest (McGreer, Mesinger & D’Orolicio 2015), the damping wing in high-redshift quasar spectra (Greig et al. 2017; Davies et al. 2018; Bañados et al. 2018a; Greig, Mesinger & Bañados 2019; Wang et al. 2020; Yang et al. 2020a) and Ly $\alpha$  emitting galaxies (Mason et al. 2018, 2019). For clarity of presentation, a small offset has been added to the redshifts of some of the data points. Bottom panel: the Thomson scattering optical depth to CMB photons. The black line with the shaded region shows the Planck Collaboration VI (2020) measurement.

**Table 1.** Hydrodynamical simulations used in this work.

| Name         | Reionization              | $z_r$     | Star formation |
|--------------|---------------------------|-----------|----------------|
| zr53         | Hybrid RT/hydro           | 5.3       | VHS04          |
| zr53-homog   | Homogeneous, matches zr53 | 5.3       | VHS04          |
| zr67         | Hybrid RT/hydro           | 6.7       | VHS04          |
| QLy $\alpha$ | Rapid, optically thin     | $\sim 15$ | VHS04          |
| PS13         | Rapid, optically thin     | $\sim 15$ | PS13           |

*Notes.* From the left- to right-hand side, the columns give the name of the simulation, the nature of the reionization model, the redshift when the IGM is fully reionized, and the prescription for converting dense gas into collisionless star particles, which follows either Viel et al. (2004) or Puchwein & Springel (2013). The first three simulations are part of the Sherwood–Relics suite. The final two models are optically thin simulations with rapid reionization at  $z \simeq 15$ , and are described in Appendix A. All models have a volume of  $(40h^{-1}\text{cMpc})^3$  and include  $2 \times 2048^3$  dark matter and gas particles.

Furlanetto (2006b) adopt the normalization ,

$$L_X = 3.4 \times 10^{40} \text{ erg s}^{-1} f_X \left( \frac{\text{SFR}}{1 \text{ M}_\odot \text{ yr}^{-1}} \right), \quad (1)$$

for the total X-ray luminosity at photon energies  $> 0.2 \text{ keV}$ , assuming a power-law spectral index  $\alpha_X = 1.5$ . The X-ray efficiency,  $f_X$ , parametrizes the large uncertainty in the extrapolation of equation (1) toward higher redshift. Using the conversion  $\epsilon_{X,0.2\text{keV}} = L_X(\alpha_X - 1)/\nu_{0.2\text{keV}}$ , the corresponding comoving X-ray emissivity is

$$\epsilon_{X,\nu}(z) = 3.5 \times 10^{21} f_X \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ cMpc}^{-3} \times \left( \frac{\nu}{\nu_{0.2\text{keV}}} \right)^{-\alpha_X} \left( \frac{\rho_{\text{SFR}}(z)}{10^{-2} \text{ M}_\odot \text{ yr}^{-1} \text{ cMpc}^{-3}} \right). \quad (2)$$

We assume a power-law spectrum with  $\alpha_X = 1.5$ , and use the fit to the observed comoving star formation rate density from Puchwein et al. (2019) (their equation 21), where

$$\rho_{\text{SFR}}(z) = 0.01 \text{ M}_\odot \text{ yr}^{-1} \text{ cMpc}^{-3} \frac{(1+z)^{2.7}}{1 + [(1+z)/3.0]^{5.35}}. \quad (3)$$

We assume that  $\rho_{\text{SFR}} = 0$  at redshifts  $z > z_* = 14$ , and have verified that adopting  $z_* > 14$  does not change our predictions for 21-cm absorption at  $z \leq 12$ .

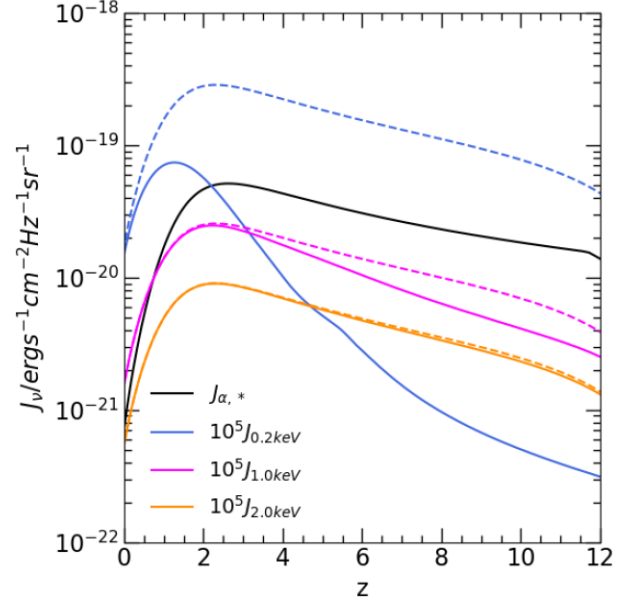
The UV background emissivity at the Ly  $\alpha$  wavelength from stars in our model is instead given by

$$\epsilon_\alpha(z) = 8.7 \times 10^{25} f_\alpha \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ cMpc}^{-3} \times \left( \frac{\rho_{\text{SFR}}(z)}{10^{-2} \text{ M}_\odot \text{ yr}^{-1} \text{ cMpc}^{-3}} \right), \quad (4)$$

where we have used the conversion between SFR and UV luminosity at  $1500 \text{ \AA}$  from Madau & Dickinson (2014) and assumed a flat spectrum ( $\epsilon_\nu \propto \nu^0$ ) in the UV, where the Ly  $\alpha$  efficiency  $f_\alpha$  parametrizes the uncertain amplitude. We adopt  $f_\alpha = 1$  as the fiducial value in this work, but note that this parameter is uncertain and the Ly  $\alpha$  emissivity should furthermore vary spatially (see e.g. fig. 4 in Semelin 2016). For illustrative purposes we therefore also show some results for the much smaller value of  $f_\alpha = 0.01$  (but note that, in practice,  $f_\alpha$  and the reionization history are not fully decoupled). The primary effect of increasing (decreasing) the Ly  $\alpha$  efficiency is to produce a tighter (weaker) coupling of the H I spin and kinetic temperatures. A smaller value of  $f_\alpha$  may be more appropriate for absorbers that are distant from the sources of Ly  $\alpha$  background photons. Instead of a flat UV spectrum we also considered the power-law population II and III spectra used by Pritchard & Furlanetto (2006), but the strength of the Ly  $\alpha$  coupling in our model is not very sensitive to this choice at the redshifts we consider.

With these emissivities in hand, we may evaluate the solution to the cosmological radiative transfer equation (see equation (B1) in Appendix B) to obtain the X-ray specific intensity at photon energies  $0.2\text{--}30 \text{ keV}$  (Pritchard & Loeb 2012). Similarly, we obtain the specific intensity of the Ly  $\alpha$  background by evaluating equation (B7), following Pritchard & Furlanetto (2006). Fig. 2 shows the redshift evolution of the specific intensity of the Ly  $\alpha$  background from stellar emission,  $J_{\alpha,*}(z)$ , and the specific intensity of the X-ray background at three different energies,  $0.2, 1,$  and  $2 \text{ keV}$ . The dashed curves show the X-ray specific intensities in the optically thin limit, i.e. when the optical depth of the intervening IGM to X-ray photons is set to zero in equation (B1). Note that  $J_{2.0\text{keV}}$  remains almost unchanged in the optically thin limit, implying the IGM is transparent to photons emitted with energies  $\geq 2 \text{ keV}$  at  $z \lesssim 10$  (cf. McQuinn 2012).

The unresolved soft X-ray background at  $z = 0$  places an upper limit on the contribution of high-redshift sources to the

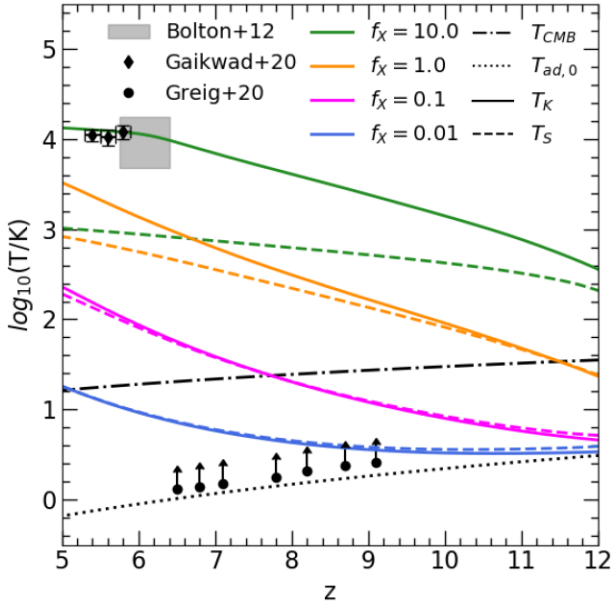


**Figure 2.** The redshift evolution of the specific intensity of the Ly  $\alpha$  background from stars for a Ly  $\alpha$  efficiency  $f_\alpha = 1$  (solid black curve) and the specific intensity of the X-ray background for photon energies  $0.2$  (blue curve),  $1$  (fuchsia curve), and  $2 \text{ keV}$  (orange curve), assuming an X-ray efficiency of  $f_X = 1$ . The X-ray specific intensities have been multiplied by a factor of  $10^5$  for presentation purposes. For comparison, the dashed curves show the X-ray specific intensities evaluated in the optically thin limit.

hard X-ray background, since these photons may redshift without significant absorption to  $z = 0$  (Dijkstra, Haiman & Loeb 2004; McQuinn 2012). When assuming  $f_X = 1.8$ , integrating our model specific intensity in the soft X-ray band ( $0.5\text{--}2 \text{ keV}$ ) at  $z = 0$  yields  $J_{0.5\text{--}2\text{keV}} = 2.9 \times 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ deg}^{-2}$ . This value is consistent with the unresolved soft X-ray background obtained from Chandra observations of the COSMOS legacy field,  $J_{0.5\text{--}2\text{keV}} = 2.9 \pm 0.16 \times 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ deg}^{-2}$  (Cappelluti et al. 2017). Note, however, the  $z = 0$  soft X-ray background does not provide a direct constraint on the very uncertain soft X-ray background at high redshift (see e.g. Dijkstra et al. 2012; Fialkov et al. 2017). Recently, Greig et al. (2021a) have presented the first weak, model-dependent lower limits on the soft X-ray background emissivity at  $6.5 \leq z \leq 8.7$  using the Murchison Widefield Array (MWA) upper limits on the 21-cm power spectrum (Trott et al. 2020), where  $\epsilon_{X,0.5\text{--}2\text{keV}} \gtrsim 10^{34.5} \text{ erg s}^{-1} \text{ cMpc}^{-3}$ . For comparison, for an X-ray efficiency of  $f_X = 0.01$ , our X-ray background model gives  $\epsilon_{X,0.5\text{--}2\text{keV}} = 10^{36.0} \text{ erg s}^{-1} \text{ cMpc}^{-3}$  at  $z = 8.1$ , which is well above the Greig et al. (2021a) lower limit.

Given the specific intensities of the X-ray and Ly  $\alpha$  radiation backgrounds, we next compute the thermal evolution of the IGM that remains optically thick to UV photons, but is heated by X-ray and Ly  $\alpha$  backgrounds that are assumed to be spatially uniform on the scale of our simulated volume.<sup>2</sup> We follow the procedure described in Appendix B for this purpose. Fig. 3 displays the temperature

<sup>2</sup>The mean-free path to X-ray photons is  $\lambda_X = 5 \text{ cMpc } x_{\text{H I}}^{-1} (1 + \delta)^{-1} (E/0.2 \text{ keV})^3 [(1+z)/10]^{-2}$ . Fluctuations in the temperature of soft X-ray heated gas on  $\sim 10 \text{ cMpc}$  scales are thus expected (Pritchard & Furlanetto 2007; Ross et al. 2017; Eide et al. 2018). These fluctuations would not, however, be adequately captured in our small simulation volume.



**Figure 3.** The redshift evolution of the gas kinetic temperature,  $T_K$ , (solid curves) and spin temperature,  $T_S$ , (dashed curves) at mean density following X-ray background heating by photons with  $E = 0.2\text{--}30\text{ keV}$ . The different coloured curves correspond to efficiency parameter  $f_X = 10$  (green curves), 1 (orange curves), 0.1 (fuchsia curves), and 0.01 (blue curves). For comparison, the CMB temperature,  $T_{\text{CMB}} = 2.73\text{ K}(1+z)$  corresponds to the dot-dashed curve, and the kinetic temperature for adiabatic heating and cooling only,  $T_{\text{ad},0} = 2.73\text{ K}(1+z)^2/(1+z_{\text{dec}})$ , is shown by the dotted curve. We assume the gas thermally decouples from the CMB at  $z_{\text{dec}} = 147.8$  (Furlanetto et al. 2006). The filled diamonds and grey shading correspond to the gas kinetic temperature measurements from Ly $\alpha$  transmission spikes in quasar spectra (Gaikwad et al. 2020) and Ly $\alpha$  absorption lines in quasar proximity zones (Bolton et al. 2012), respectively. The filled circles show the model-dependent lower limits on the H I spin temperature obtained from LOFAR (Greig et al. 2021b) and MWA (Greig et al. 2021a).

evolution of a gas parcel at mean density for four different values of the X-ray efficiency parameter  $f_X$ . An approximate lower limit on  $f_X$  is provided by the recent constraints on the spin temperature from upper limits on the 21-cm power spectrum at  $z \simeq 9.1$  obtained with LOFAR (Mertens et al. 2020), and at  $z = 6.5\text{--}8.7$  from MWA (Trott et al. 2020). These data disfavour very cold reionization models with no X-ray heating (Ghara et al. 2020; Mondal et al. 2020; Greig et al. 2021a,b). An approximate upper limit on  $f_X$  at  $z > 6$  is provided by Ly $\alpha$  absorption measurements of the kinetic temperature at  $z \simeq 5\text{--}6$ , after the IGM has been photoionized and heated by UV photons (Bolton et al. 2012; Gaikwad et al. 2020). These data are consistent with  $f_X \simeq 10$ . Adopting larger X-ray efficiencies in our model would overheat the low-density IGM by  $z = 6$ .

### 3 THE 21-CM FOREST OPTICAL DEPTH

We now turn to the calculation of the 21-cm optical depth. The 21-cm line arises from the hyperfine structure of the hydrogen atom, and is determined by the relative orientation of the proton and electron spin, where the ground-state energy level is split into a singlet and triplet state. A photon with rest-frame wavelength  $\lambda_{21} = 21.11\text{ cm}$ , or equivalently frequency  $\nu_{21} = 1420.41\text{ MHz}$ , can induce a transition between these two states.

In the absence of redshift space distortions, the optical depth to 21-cm photons at redshift  $z$  is

$$\begin{aligned} \tau_{21}(z) &= \frac{3h_p c^3 A_{10}}{32\pi v_{21}^2 k_B} \frac{n_{\text{HI}}(z)}{T_S(z)H(z)} \\ &= 0.27 x_{\text{HI}} \left(\frac{1+\delta}{10}\right) \left(\frac{T_S}{10\text{ K}}\right)^{-1} \left(\frac{1+z}{10}\right)^{3/2}, \end{aligned} \quad (5)$$

where  $n_{\text{HI}}$  is the H I number density,  $T_S$  is the spin temperature,  $A_{10} = 2.85 \times 10^{-15}\text{ s}^{-1}$  is the Einstein spontaneous emission coefficient for the hyperfine transition,  $\delta$  is the gas overdensity, and  $H(z)$  is the Hubble parameter (Madau, Meiksin & Rees 1997). Note the factor of 0.27 in the second equality is cosmology-dependent. Absorption will therefore be most readily observable for dense, cold, and significantly neutral hydrogen gas. The H I spin temperature, a measure of the relative occupation numbers of the singlet and triplet states, is (Field 1958)

$$T_S^{-1} = \frac{T_{\text{CMB}}^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}, \quad (6)$$

where  $T_{\text{CMB}} = 2.73(1+z)\text{ K}$  is the temperature of the cosmic microwave background (CMB; Fixsen 2009),  $T_\alpha$  is the Ly $\alpha$  colour temperature, and  $x_c$ ,  $x_\alpha$  are the coupling coefficients for collisions and Ly $\alpha$  photon scattering, respectively. If  $x_c + x_\alpha \gg 1$ , the H I spin temperature is coupled to the gas kinetic temperature, and if  $x_c + x_\alpha \ll 1$  it is coupled to the CMB temperature.

The collisional coupling coefficient is

$$x_c = \frac{T_\star}{A_{10} T_{\text{CMB}}} (\kappa_{10}^{\text{HH}} n_{\text{H}} + \kappa_{10}^{\text{eH}} n_{\text{e}} + \kappa_{10}^{\text{pH}} n_{\text{p}}), \quad (7)$$

where  $T_\star = h_p v_{21}/k_B$ , and  $\kappa_{10}^{\text{HH}}$ ,  $\kappa_{10}^{\text{eH}}$ ,  $\kappa_{10}^{\text{pH}}$  are the temperature-dependent de-excitation rates for collisions between hydrogen atoms, electrons and hydrogen atoms, and protons and hydrogen atoms, respectively. We use the convenient fitting functions to the de-excitation rates from Kuhlen, Madau & Montgomery (2006) and Liszt (2001), modified to better agree with tabulated values for  $\kappa_{10}^{\text{HH}}$  (Furlanetto et al. 2006),  $\kappa_{10}^{\text{eH}}$  (Furlanetto & Furlanetto 2007a), and  $\kappa_{10}^{\text{pH}}$  (Furlanetto & Furlanetto 2007b) over the range  $1 \leq T_K \leq 10^4\text{ K}$ .

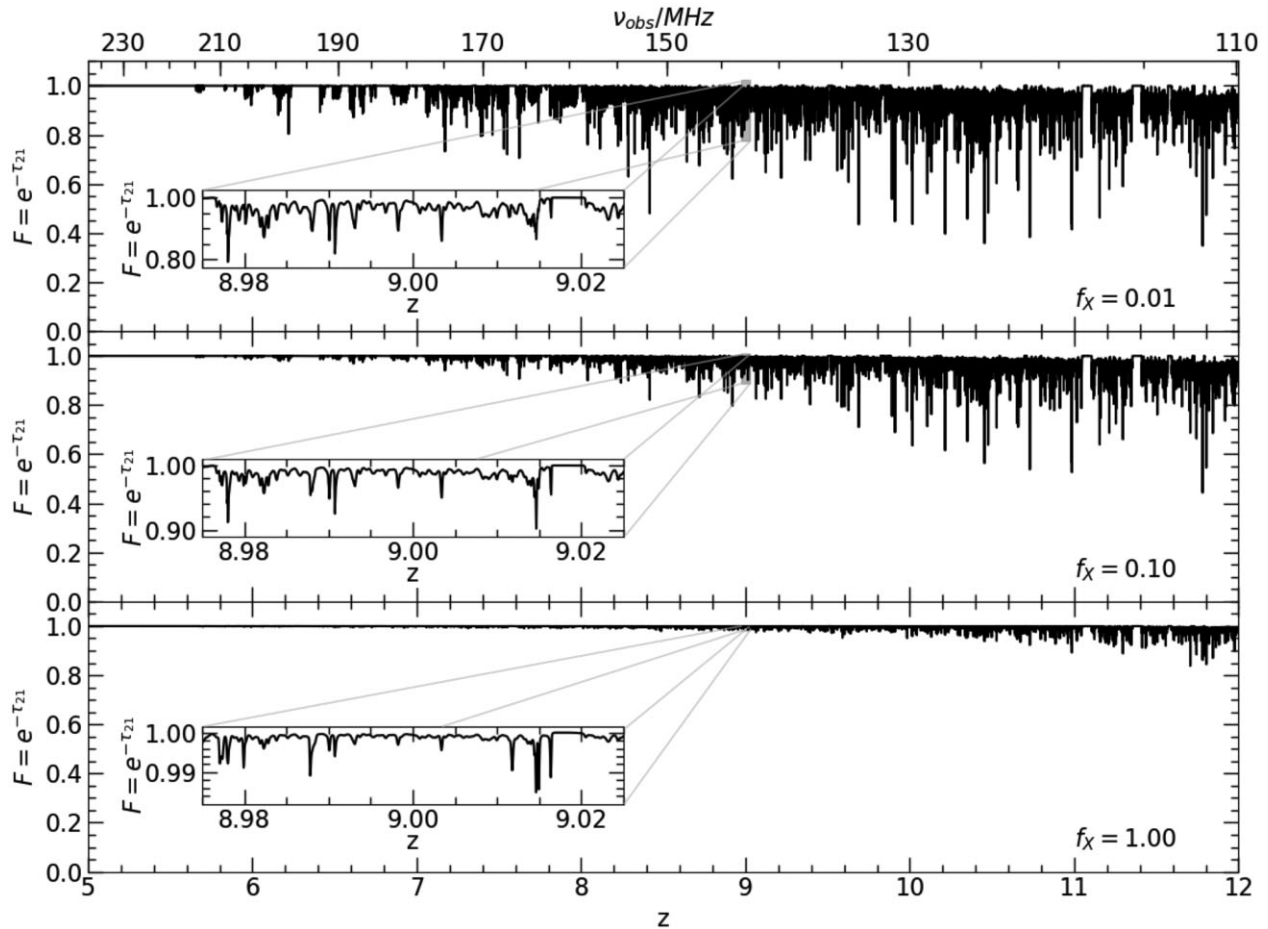
The coupling coefficient for Ly $\alpha$  scattering is (Wouthuysen 1952; Field 1958; Madau et al. 1997)

$$x_\alpha = \frac{2T_\star \lambda_\alpha^3 \Lambda_\alpha}{9T_{\text{CMB}} A_{10} h_p c} S_\alpha J_\alpha, \quad (8)$$

where  $\lambda_\alpha = 1215.67\text{ \AA}$ ,  $\Lambda_\alpha = 6.265 \times 10^8\text{ s}^{-1}$  is the Einstein spontaneous emission coefficient for the Ly $\alpha$  transition,  $S_\alpha$  is a factor of order unity that corrects for the spectral distortions in the Ly $\alpha$  spectrum, and  $J_\alpha$  is the proper Ly $\alpha$  specific intensity in units  $\text{erg s}^{-1}\text{ cm}^{-2}\text{ Hz}^{-1}\text{ sr}^{-1}$ . We use the fits provided by Hirata (2006) to calculate  $T_\alpha$  and  $S_\alpha$ , where  $T_S$ ,  $T_\alpha$ , and  $S_\alpha$  must be solved for iteratively.

In this work, we also include the effect of redshift space distortions on the 21-cm forest absorption features. In our calculation of the 21-cm optical depth, we therefore include a convolution with the Gaussian line profile and incorporate the gas peculiar velocities from our hybrid RT/hydrodynamical simulations. The optical depth in equation (5) may then be calculated in discrete form as (e.g. Furlanetto & Loeb 2002)

$$\begin{aligned} \tau_{21,i} &= \frac{3h_p c^3 A_{10}}{32\pi^{3/2} v_{21}^2 k_B} \frac{\delta v}{H(z)} \\ &\times \sum_{j=1}^N \frac{n_{\text{HI},j}}{b_j T_{S,j}} \exp\left(-\frac{(v_{\text{H},i} - u_j)^2}{b_j^2}\right), \end{aligned} \quad (9)$$



**Figure 4.** The redshift evolution of the 21-cm forest transmission,  $F = e^{-\tau_{21}}$ , in the zr53 simulation for a Ly  $\alpha$  efficiency  $f_{\alpha} = 1$  and an X-ray efficiency of  $f_X = 0.01$  (top panel), 0.1 (middle panel), and 1 (bottom panel). The inset displays a zoom-in on part of the 21-cm forest at redshift  $z \sim 9$  – note the different scales on the vertical axes of the inset. The incidence of gaps in the 21-cm forest, which are associated with large regions of ionized gas, increases toward lower redshift, and become particularly apparent in the  $f_X = 0.01$  model at redshift  $z < 7$ . No instrumental features have been added to the spectra.

for pixel  $i$  with Hubble velocity  $v_{H,i}$  and velocity width<sup>3</sup>  $\delta v$ . Here  $b = (2k_B T_K / m_H)^{1/2}$  is the Doppler parameter,  $T_K$  is the gas kinetic temperature,  $u_j = v_{H,j} + v_{pec,j}$ , and  $v_{pec}$  is the peculiar velocity of the gas. We evaluate equation (9) in our simulations by extracting a total of 5000 periodic lines of sight, drawn parallel to the simulation box axes at redshift intervals of  $\Delta z = 0.1$  over the range  $5 \leq z \leq 12$ . The total path-length we use to make our mock 21-cm forest spectra at each output redshift is therefore  $200h^{-1}$  cGpc.

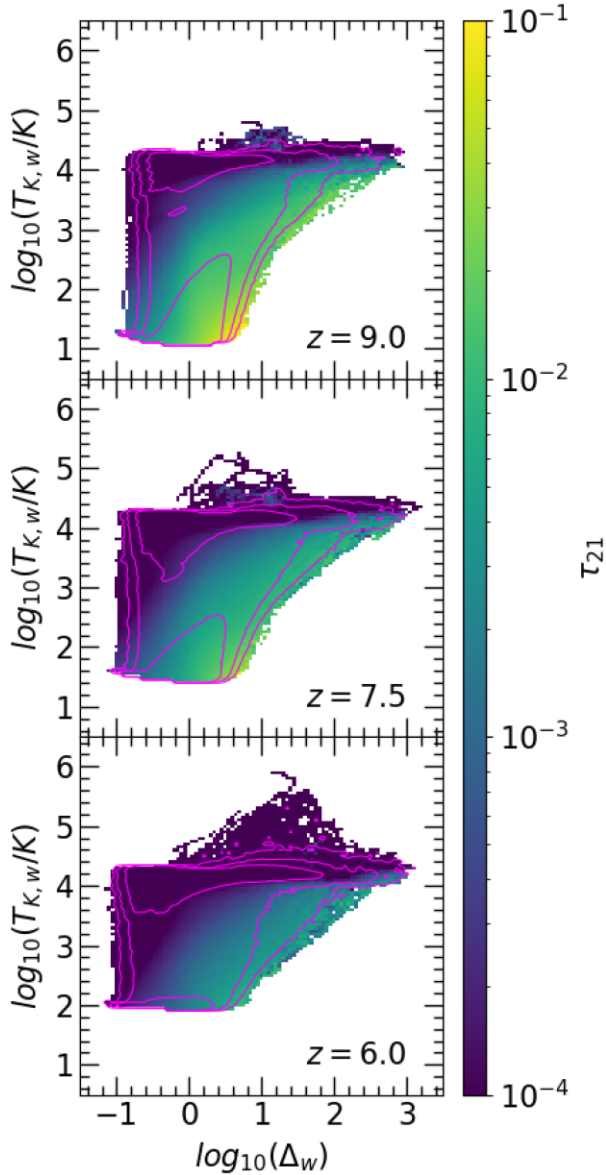
The redshift evolution of the transmission,  $F = e^{-\tau_{21}}$ , for a random selection of 21-cm forest spectra drawn from the zr53 simulation is shown in Fig. 4, for three different X-ray efficiencies. No instrumental features have been added to the simulated data. The detailed small-scale structure of the 21-cm absorption is displayed in the insets. One can see the strong effect that X-ray heating has on the the 21-cm absorption as the X-ray efficiency parameter is increased from  $f_X = 0.01$  in the top panel, to  $f_X = 1$  in the bottom panel (cf.

Xu et al. 2011; Mack & Wyithe 2012). The redshift evolution due to the increasing filling factor of warm ( $T_K \sim 10^4$  K), photoionized gas is also apparent. In particular, the occurrence of gaps in the 21-cm forest absorption due to extended regions of ionized gas increases toward lower redshift.

In order to better identify the gas associated with the absorption, we calculate the optical depth weighted density,  $\Delta_w = 1 + \delta_w$ , and optical depth weighted kinetic temperature,  $T_{K,w}$ , for each pixel in our zr53 mock spectra for  $f_X = 0.1$ . This is analogous to the approach used to study the properties of gas responsible for absorption in the Ly  $\alpha$  forest (Schaye et al. 1999); peculiar motions (and to a much lesser extent, line broadening) would otherwise distort the mapping between 21-cm optical depth, temperature, and gas density. The results are shown in Fig. 5, where the temperature–density plane is displayed for the zr53 simulation at three different redshifts:  $z = 9$  (top panel), 7.5 (middle panel), and 6 (bottom panel). The colour bar and contours show the average 21-cm optical depth and the relative number density of the pixels, respectively.

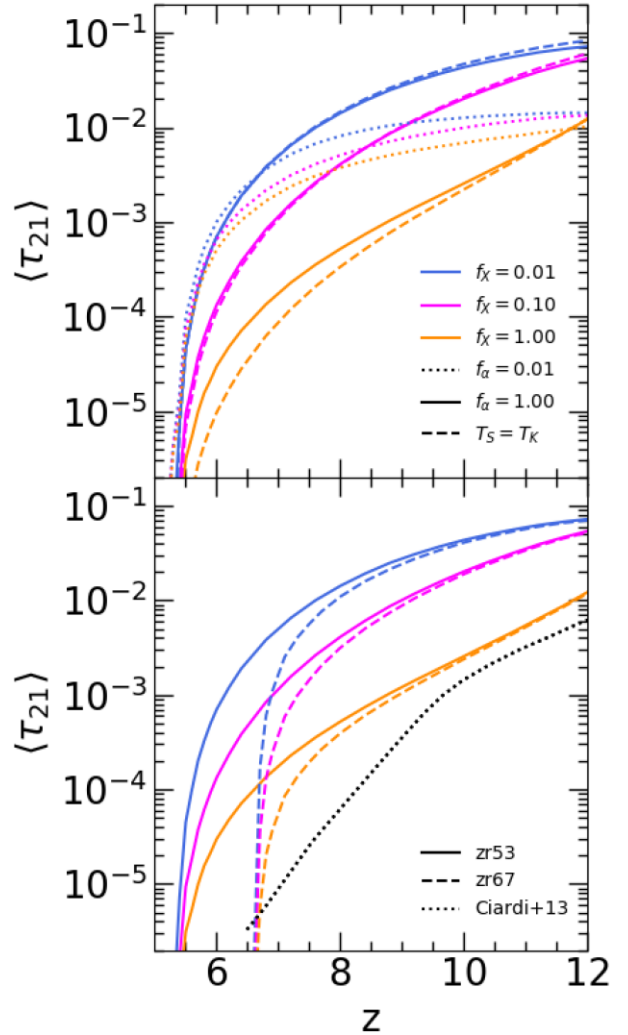
The gas distribution in Fig. 5 is bimodal, with the bulk of the pixels associated with either warm ( $T_K \sim 10^4$  K), photoionized gas or cold ( $T_K \leq 10^2$  K), significantly neutral regions (see also Ciardi et al. 2013; Semelin 2016). The plume of gas at intermediate temperatures has been heated by shocks from structure formation. Note, furthermore, that in this very late reionization model, the IGM

<sup>3</sup>Note the width of the pixel must be smaller than the typical thermal width of an absorber,  $\Delta v_{21} = 0.61(T_K/10^2 \text{ K})^{1/2}$  kHz, to ensure the optical depths obtained using equation (9) are converged. In this work, we resample the simulation outputs using linear interpolation to achieve the required pixel size. Alternatively, the line profile may be evaluated using error functions (Meiksin 2011; Hennawi et al. 2021).



**Figure 5.** The optical depth weighted temperature–density plane for gas in the zr53 simulation at redshift  $z = 9$  (top panel), 7.5 (middle panel), and 6 (bottom panel), for an X-ray efficiency  $f_X = 0.1$  and Ly  $\alpha$  efficiency  $f_\alpha = 1$ . The colour scale shows the average 21-cm optical depth at each point in the plane. The number density of points increases by 1 dex for each contour level.

is still not fully ionized by  $z = 6$ . The largest optical depths in the model arise not from the highest density gas, but the cold, diffuse IGM with  $3 < \Delta < 10$ . This is because gas at higher densities is typically reionized early due to proximity to the ionizing sources, and also because gas around haloes (with  $\Delta \gtrsim 100$ ) is shock-heated and partially collisionally ionized. Note again, however, there is no cold, star-forming gas in this simulation – for further discussion of this point; see Appendix A. Toward lower redshift, the increase in the minimum kinetic temperature of the neutral gas due to X-ray heating, the partial ionization of the H I by secondary electrons and collisions, and the decrease in the proper number density of gas at fixed overdensity, all conspire to lower the maximum optical depth. The contours furthermore show that the regions with the largest optical depths are at least 100 times rarer than the bulk of the cold,



**Figure 6.** Redshift evolution of the volume-averaged 21-cm optical depth in the zr53 model (solid curves) for a Ly  $\alpha$  efficiency  $f_\alpha = 1$  and an assumed X-ray efficiency of  $f_X = 0.01$  (blue curves), 0.1 (fuchsia curves), and 1 (orange curves). In the upper panel, this is compared to results from the same hybrid RT/hydrodynamical simulation, but with 21-cm optical depths calculated under the assumption of strong (i.e.  $T_S = T_K$ , shown by the dashed curves) and weak Ly  $\alpha$  coupling ( $f_\alpha = 0.01$ , shown by the dotted curves). In the lower panel, the dashed curves instead show  $\langle \tau_{21} \rangle$  for the hybrid model with an earlier end to reionization (zr67). The dotted black curve in the lower panel corresponds to the RT+Ly  $\alpha$ +x model from fig. 2 of Ciardi et al. (2013).

neutral gas. Nevertheless, in this very late reionization model, it remains possible that some detectable 21-cm absorption may persist as late as  $z \simeq 6$ . We now explore this possibility in more detail.

## 4 THE DETECTABILITY OF 21-CM FOREST ABSORPTION FOR VERY LATE REIONIZATION

### 4.1 The volume-averaged 21-cm optical depth

We first consider the redshift evolution of the volume-averaged 21-cm optical depth,  $\langle \tau_{21} \rangle$ , in the zr53 simulation, displayed as the solid curves in Fig. 6 for our fiducial model with  $f_\alpha = 1$ . In the upper panel, we test the common assumption that, as a result of the Wouthuysen–Field effect, the spin temperature becomes strongly coupled to the

gas kinetic temperature during the later stages of reionization, such that  $T_S = T_K$  (e.g. Xu et al. 2009; Mack & Wyithe 2012; Ciardi et al. 2013). This is shown by the dashed curves in the upper panel of Fig. 6. As also noted by Semelin (2016), a full calculation of  $T_S$  using equation (6) can either reduce or enhance 21-cm optical depths relative to the value obtained assuming strong coupling. This is caused by a partial coupling of the spin temperature to the CMB temperature; if  $T_K < T_{\text{CMB}}$ , the full calculation will result in a higher spin temperature and smaller 21-cm optical depth, and vice versa.

This can be observed in Fig. 6 for  $f_X = 0.1$  (fuchsia curves), where  $\langle \tau_{21} \rangle$  for the full calculation assuming  $f_\alpha = 1$  (solid curves) is smaller than the  $T_S = T_K$  case (dashed curves) at  $z \gtrsim 8$ , but is greater at lower redshifts. This coincides with the temperature evolution shown in Fig. 3, particularly the transition from  $T_K < T_{\text{CMB}}$  (and  $T_S > T_K$ ) at  $z > 8$  to  $T_K > T_{\text{CMB}}$  (and  $T_S < T_K$ ) at  $z < 8$ . Similarly, in the case of a weaker ( $f_X = 0.01$ , blue curves) or stronger ( $f_X = 1$ , orange curves) X-ray background, the full  $T_S$  calculation, respectively, decreases or increases  $\langle \tau_{21} \rangle$  relative to the strong coupling approximation. The dotted curves furthermore show the  $\langle \tau_{21} \rangle$  redshift evolution for significantly weaker Ly  $\alpha$  coupling, with  $f_\alpha = 0.01$ . In this case  $T_S$  is now decoupled from  $T_K$  and has a value similar to  $T_{\text{CMB}}$ . The weak coupling means  $\langle \tau_{21} \rangle$  is significantly increased in the models with efficient X-ray heating. Hence, while the assumption of strong coupling,  $T_S = T_K$ , remains a reasonable approximation if  $f_\alpha = 1$ , this will not be the case if the background Ly  $\alpha$  emissivity is significantly overestimated in our fiducial model (i.e.  $f_\alpha \ll 1$ ).

The lower panel of Fig. 6 instead shows  $\langle \tau_{21} \rangle$  for the two different reionization histories in Fig. 1. Both of these reionization models are broadly consistent with existing constraints on the timing of reionization, and the zr53 model furthermore successfully reproduces the large fluctuations in the Ly  $\alpha$  forest opacity at  $z = 5.5$  (Kulkarni et al. 2019). For comparison, we also show  $\langle \tau_{21} \rangle$  from Ciardi et al. (2013) as the dotted curve. This includes X-ray and Ly  $\alpha$  heating following Ciardi et al. (2010), and is most similar to our zr67 simulation with  $f_X \simeq 1$ . The differences between this work and Ciardi et al. (2013) are due to different assumptions for the X-ray emissivity and the reionization history. A later end to reionization means  $\langle \tau_{21} \rangle$  in Fig. 6 remains significantly larger than earlier reionization models at redshifts  $6 \lesssim z \lesssim 7$ . If reionization does indeed complete late, such that large neutral islands persist in the IGM at  $z \simeq 6$  (e.g. Lidz et al. 2007; Mesinger 2010), this suggests 21-cm forest absorption lines may be more readily observable than previously thought at these redshifts.

#### 4.2 The differential number density of 21-cm absorption lines

We now consider the number density of individual absorption lines in our high-resolution mock spectra. We present this as the total number of lines,  $N$ , within a given optical depth bin, per unit redshift (see also Furlanetto 2006a; Shimabukuro et al. 2014), where

$$f(\tau_{21}, z) = \frac{\partial^2 N}{\partial \tau_{21} \partial z}. \quad (10)$$

The absorption lines in our simulated 21-cm forest spectra are identified following a similar method to Garzilli, Theuns & Schaye (2015), who identify absorption lines in mock Ly  $\alpha$  forest spectra as local optical depth maxima located between two minima. In this work, we require that the local maxima must have a prominence (i.e. be higher by a certain value than the minima) that corresponds to a factor of 1.001 difference in the transmitted flux,  $F = e^{-\tau_{21}}$ , between the line base and peak. We then define the optical depth for each identified line as being equal to the local maximum. We find this

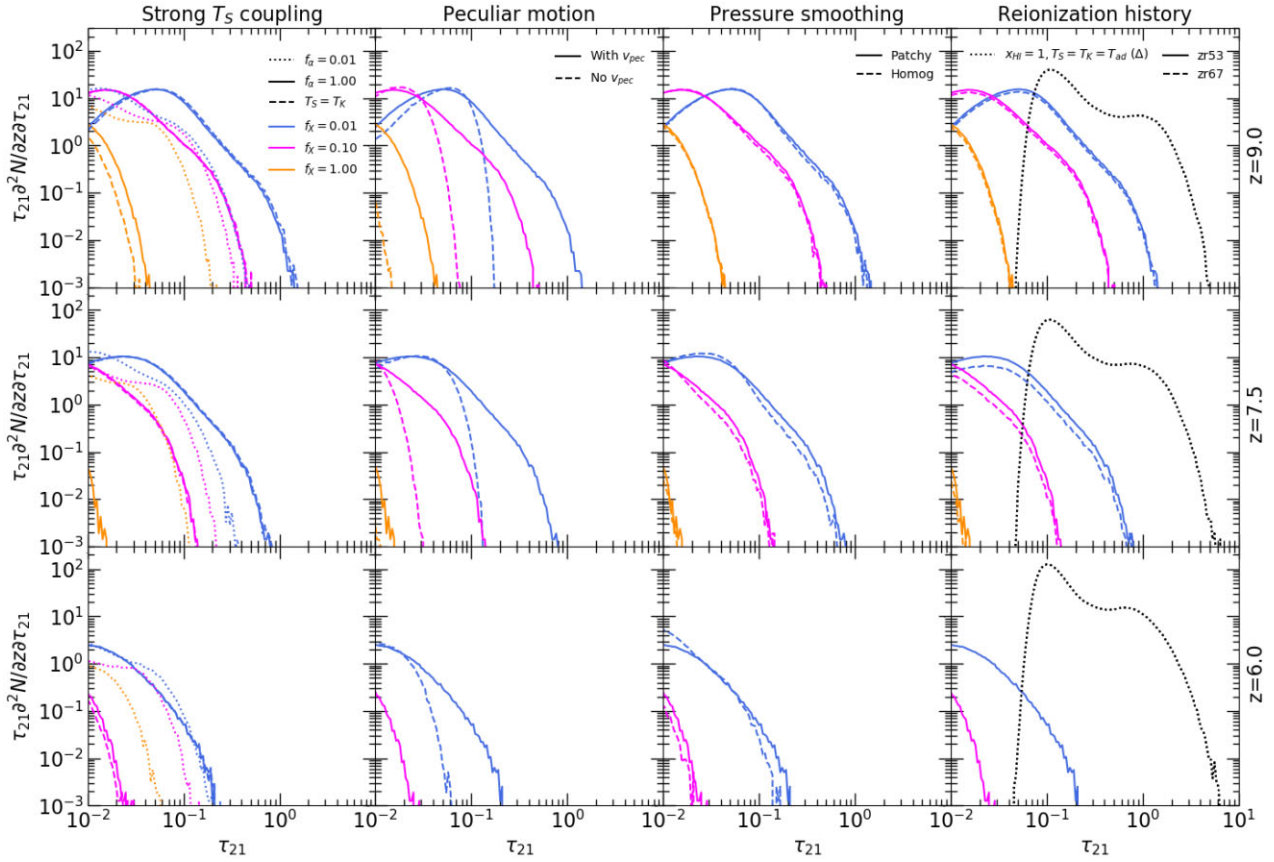
method is robust for lines with  $\tau_{21} \geq 10^{-2}$  (i.e.  $F = e^{-\tau_{21}} \leq 0.99$ ), but for optical depths below this threshold the number of lines is sensitive to the choice for the prominence, and is thus unreliable.

The number density distributions,  $\tau_{21} f(\tau_{21}, z)$ , for different model parameters at three different redshifts,  $z = 9, 7.5$ , and 6, are displayed for our fiducial model with  $f_\alpha = 1$  in Fig. 7 (for an illustration of the effect of these model parameter variations on individual absorbers, see Appendix C). Each column corresponds to a different model parameter choice, each row shows a different redshift, and in each panel we show the distribution for three X-ray efficiencies:  $f_X = 0.01$  (blue curves), 0.1 (fuchsia curves), and 1 (orange curves). The peak of the distribution is at  $\tau_{21} \leq 0.1$ , and it shifts to lower amplitudes and smaller optical depths as the IGM reionizes and the spin temperature of the X-ray heated gas increases. The distribution also has an extended tail toward higher optical depths. While strong 21-cm absorbers will be rare, this suggests that for  $f_X \sim 0.1$ , features with a transmission of  $F = e^{-\tau_{21}} \simeq 0.9$  should still be present at  $z = 7.5$  in the late reionization model (see also Fig. 4).

In the first column of Fig. 7, we re-examine the effect of strong Ly  $\alpha$  coupling on the distribution of 21-cm optical depths. As was the case for the volume-averaged optical depth in Fig. 6, the impact is relatively modest for low X-ray efficiencies: for  $f_X = 0.01$  at  $z = 6$ , the two cases are almost identical. For  $f_X = 1$ , however, the abundance of features with  $\tau_{21} \geq 0.01$  for  $T_S = T_K$  is more than 50 per cent smaller than the full calculation at  $z = 9$ . In either case, however, by  $z = 7.5$  most gas in the  $f_X = 1$  model has  $\tau_{21} < 10^{-2}$ , and will therefore be challenging to detect directly. However, the dotted curves also demonstrate that if  $f_\alpha = 0.01$ , the weak coupling of  $T_S$  to  $T_K$  allows strong 21-cm absorbers to still be observable at  $z = 6$ , even for  $f_X = 1$ .

We consider the effect of gas peculiar velocities on the 21-cm forest in the second column of Fig. 7. Redshift space distortions are well known to impact on the observability of the high-redshift 21-cm signal (Bharadwaj & Ali 2004; Mao et al. 2012; Majumdar et al. 2020). We do this by creating mock 21-cm spectra that ignore the effect of gas peculiar motions, such that  $v_{\text{pec}} = 0$  in equation (9). The results are shown by the dashed curves. While the position of the peak in the number density distribution is unchanged, the high optical depth tail is strongly affected, particularly for inefficient X-ray heating. Ignoring peculiar velocities within 21-cm forest models can therefore significantly reduce the incidence of the strongest absorbers, and this will have a negative impact on the predicted observability of the 21-cm forest. Qualitatively, this agrees with the assessment of Semelin (2016), who also included the effect of gas peculiar motions in their models.

As our hybrid simulations self-consistently model the hydrodynamical response of gas to photoheating by the inhomogeneous UV radiation field, we may also estimate the effect of (the lack of) pressure (Jeans) smoothing on the 21-cm forest. Inhomogeneous reionization introduces large-scale gas temperature fluctuations in the IGM (Keating, Puchwein & Haehnelt 2018), and these lead to differences in the local gas pressure that smooth the structure of the IGM on different scales (e.g. Gnedin & Hui 1998; Kulkarni et al. 2015; Nasir, Bolton & Becker 2016; D’Aloisio et al. 2020). In the absence of significant X-ray heating, the neutral gas responsible for the 21-cm forest should therefore experience minimal pressure smoothing compared to the photoionized IGM. We therefore compare the results of our zr53 model to the zr53-homog simulation in the third column of Fig. 7. The latter model has exactly the same initial conditions and volume-averaged reionization history as zr53, but all the gas in the simulation volume is instead heated simultaneously (i.e. we do not follow the radiative transfer for UV photons).



**Figure 7.** The differential number density of absorption lines in synthetic 21-cm forest spectra. Each row shows the distribution at redshift  $z = 9$  (top row), 7.5 (middle row), and 6 (bottom row) for our fiducial model with Ly  $\alpha$  efficiency  $f_\alpha = 1$ , and in each panel the distribution is shown for three X-ray efficiencies,  $f_X = 0.01$  (blue curves), 0.1 (fuchsia curves), and 1 (orange curves). Each column displays the zr53 simulation (solid curves) compared to models where one of the parameter choices is varied (dashed curves). These parameters are, from left to right, the assumption of strong Ly  $\alpha$  coupling (i.e.  $T_S = T_K$ ), neglecting the effect of peculiar velocities (i.e.  $v_{pec} = 0$ ), pressure smoothing due to a uniform rather than patchy UV photoheating rate (i.e. the zr53-homog model) and an earlier end to reionization (the zr67 model). In the first column, we also show the number density distribution for very weak Ly  $\alpha$  coupling (i.e.  $f_\alpha = 0.01$ , dotted curves). The black dotted curves in the last column show the case of no reionization or X-ray heating (i.e.  $T_S = T_K = T_{ad} = 2.73 \text{ K} (1 + \delta)^{2/3} (1 + z)^2 / (1 + z_{dec})$ , where  $z_{dec} = 147.8$  (Furlanetto et al. 2006), and  $x_{H1} = 1$ ).

The dashed curves in the third column of Fig. 7 show the line density distribution obtained from the density and peculiar velocity fields in the zr53-homog model (differences due to  $x_{H1}$ ,  $T_K$ , and  $T_S$  in the two models have been removed). We observe that there is a small, redshift-dependent difference between the two distributions, such that the simulation with the homogeneous UV background exhibits fewer strong absorption lines. This is because the gas responsible for the highest optical depths in the 21-cm forest (see Fig. 5) is still cold within the hybrid model, and hence has slightly higher density due to the smaller pressure smoothing scale.

We caution, however, that this comparison will still not fully capture the effect of pressure smoothing on 21-cm forest absorbers. For reference, the comoving pressure smoothing scale in the IGM is (Gnedin & Hui 1998; Garzilli et al. 2015)

$$\begin{aligned} \lambda_p &= f_J \frac{\lambda_J}{2\pi} = f_J \left( \frac{10k_B T_K}{9\mu m_H (1 + \delta) H_0^2 \Omega_m (1 + z)} \right)^{1/2} \\ &= 1.5 h^{-1} \text{ cpc} f_J \left[ \left( \frac{10}{1 + \delta} \right) \left( \frac{1.22}{\mu} \right) \left( \frac{T_K}{10^2 \text{ K}} \right) \left( \frac{1 + z}{10} \right) \right]^{1/2}, \end{aligned} \quad (11)$$

where  $\lambda_J$  is the Jeans scale,  $\mu$  is the mean molecular weight of hydrogen and helium assuming primordial composition ( $\mu = 1.22$  for

fully neutral gas,  $\mu = 0.59$  for fully ionized), and  $f_J = \lambda_p / \lambda_J$  is a factor of order unity that accounts for the finite time required for gas to dynamically respond to a change in pressure. For comparison, the mean interparticle separation and gravitational softening length in our simulations are  $19.5$  and  $0.78 h^{-1} \text{ cpc}$ , respectively. Equation (11) thus implies that the pressure smoothing scale for typical 21-cm forest absorbers is not fully resolved in our simulations (see also Emberson, Thomas & Alvarez 2013). We furthermore do not capture the 21-cm absorption from minihaloes with  $M < 2.5 \times 10^7 M_\odot$  (Furlanetto 2006a). Larger differences could then be observed in Fig. 7 for fully resolved gas. On the other hand, although we follow the dynamical response of gas to heating by UV photons, the X-ray heating of the neutral gas in our hybrid simulation is applied in post-processing. It is therefore decoupled from the hydrodynamics, and this may then underestimate the impact of pressure smoothing on cold gas for high X-ray efficiencies. Regardless of these modelling uncertainties, however, this suggests that the effect of the pressure smoothing scale on the 21-cm forest in the diffuse IGM remains small compared to the substantial impact of X-ray heating on the spin temperature at  $z \leq 10$ .

Finally, in the fourth column of Fig. 7 the effect of the reionization history is displayed for the zr53 (solid curves) and zr67 (dashed curves) simulations. For comparison, the dotted curves also show

the line number density distribution under the assumption of no reionization or X-ray heating (i.e.  $T_S = T_K = T_{\text{ad}}$  and  $x_{\text{H I}} = 1$ ). As expected, the two reionization models are significantly different at  $z = 6$ ; there are no strong absorption features with  $\tau_{21} > 10^{-2}$  in zr67 model, as reionization has already completed by this time. At  $z = 7.5$ , one can see that there are also fewer absorption features in the zr67 model due to the larger volume of ionized gas. However, the differences between the two models become smaller with increasing redshift. This again demonstrates that for reionization models that complete at  $z < 6$ , the 21-cm forest may remain observable if sufficiently bright radio sources exist at  $6 < z < 7$ . Alternatively, a null-detection could place an interesting limit on the very uncertain X-ray background (e.g. Mack & Wyithe 2012). We now investigate this possibility further.

### 4.3 Detectability of strong 21-cm forest absorbers at redshift $z = 6$ for late reionization and X-ray heating

A detection of the 21-cm forest relies on the identification of objects at high redshift that are sufficiently radio bright to act as background sources. Based on a model for the radio galaxy luminosity function at  $z > 6$ , Saxena et al. (2017) predict around one radio source per 400 square degrees at a flux density limit of  $S_{150\text{MHz}} = 3.5$  mJy, and at least  $\sim 30$  bright sources with  $S_{150\text{MHz}} > 15$  mJy (see also Bolgar et al. 2018). Ongoing observational programmes such as the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al. 2017; Kondapally et al. 2021), the Giant Metrewave Radio Telescope (GMRT) all-sky radio survey at 150 MHz (Intema et al. 2017), and the Galactic and Extragalactic All-sky Murchison Widefield Array survey (GLEAM; Wayth et al. 2015) should furthermore detect hundreds of bright  $z > 6$  radio sources. Encouragingly, a small number of radio-loud sources have already been identified at  $z > 5.5$  (e.g. Bañados et al. 2018b), including the  $z = 6.1$  blazar PSO J0309 + 27 with a flux density  $S_{147\text{MHz}} = 64.2 \pm 6.2$  mJy (Belladitta et al. 2020).

We now use our hydrodynamical simulations to assess the feasibility of detecting the 21-cm forest in late reionization models, assuming  $f_\alpha = 1$ . We shall calculate the minimum redshift path-length,  $\Delta z_{\text{min}}$ , necessary for detecting a single, strong (i.e.  $\tau_{21} > 0.01$ ) absorption line with a minimum transmission at some arbitrary threshold  $F_{\text{th}} = e^{-\tau_{21, \text{th}}}$ . For a signal-to-noise ratio (S/N), the minimum flux density contrast,  $\Delta S_{\text{min}}$ , detectable by an interferometric radio array is then (e.g. Ciardi et al. 2015a),

$$\Delta S_{\text{min}} = S_{\text{min}} - S_{\text{abs}} = \frac{2k_{\text{B}} T_{\text{sys}}}{A_{\text{eff}} \sqrt{\Delta \nu t_{\text{int}}}} S/N, \quad (12)$$

where  $T_{\text{sys}}$  is the system temperature,  $\Delta \nu$  is the bandwidth,  $A_{\text{eff}}$  is the effective area of the telescope,  $t_{\text{int}}$  is the integration time, and  $S_{\text{min}}$  is the minimum intrinsic flux density a radio source must have to allow detection of a 21-cm absorption feature with a minimum at a flux density of  $S_{\text{abs}} = S_{\text{min}} e^{-\tau_{21, \text{th}}}$ . Adopting some representative values in equation (12), the minimum flux density required to detect a 21-cm absorption feature with a minimum transmission  $F_{\text{th}}$  is therefore

$$S_{\text{min}} = 10.3 \text{ mJy} \left( \frac{0.01}{1 - F_{\text{th}}} \right) \left( \frac{S/N}{5} \right) \left( \frac{5 \text{ kHz}}{\Delta \nu} \right)^{1/2} \left( \frac{1000 \text{ hr}}{t_{\text{int}}} \right)^{1/2} \times \left( \frac{1000 \text{ m}^2 \text{ K}^{-1}}{A_{\text{eff}}/T_{\text{sys}}} \right). \quad (13)$$

In what follows, we shall adopt values for the sensitivity,  $A_{\text{eff}}/T_{\text{sys}}$ , in equation (12) appropriate for LOFAR, SKA1-low, and SKA2, where

$A_{\text{eff}}/T_{\text{sys}} \simeq 80, 600, \text{ and } 5500 \text{ m}^2 \text{ K}^{-1}$ , respectively<sup>4</sup> (Braun et al. 2019). Additionally, to approximately model the effect of spectral resolution on the data we convolve our mock spectra with a boxcar function. Following the bandwidths adopted in Ciardi et al. (2015b), we assume boxcar widths of 10 and 5 kHz for LOFAR and SKA1-low, respectively. For a more futuristic measurement with SKA2, we assume a smaller bandwidth and adopt a boxcar width of 1 kHz.

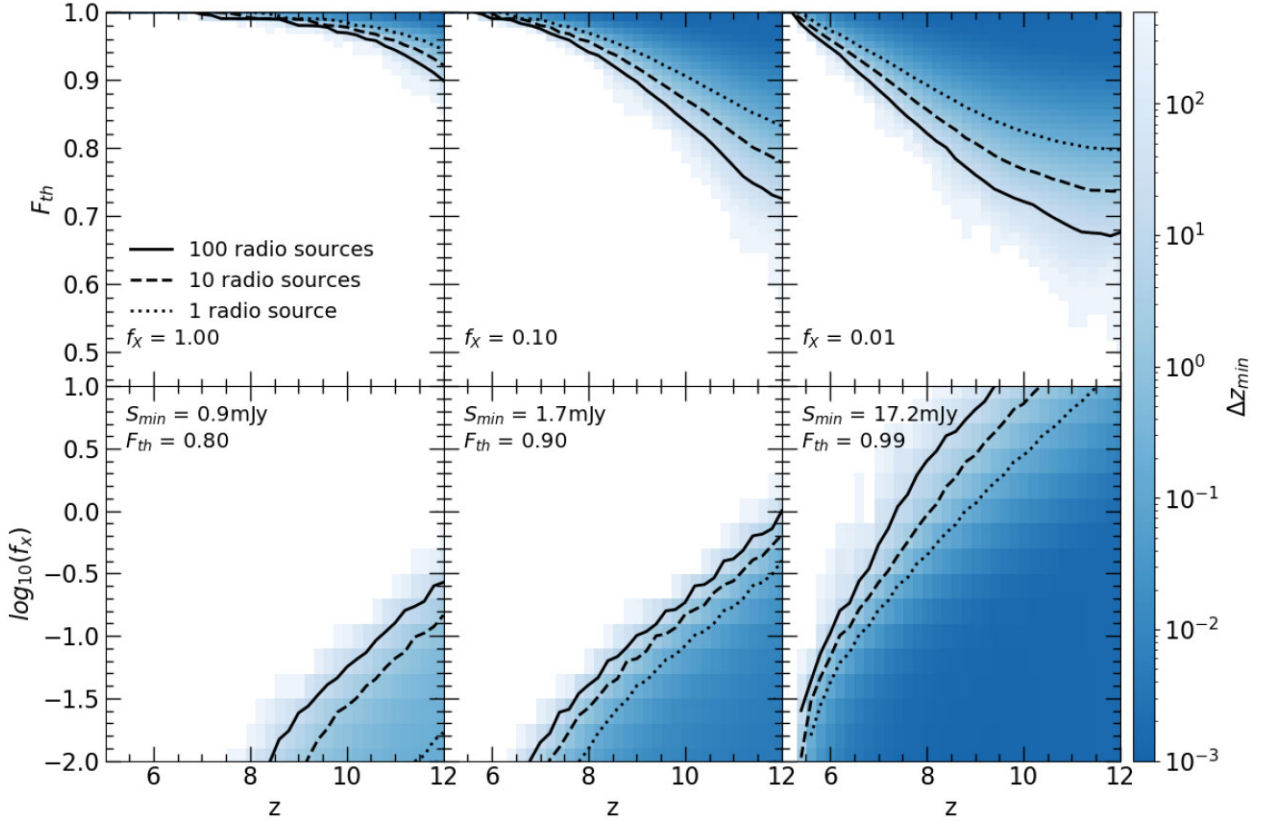
First, in Fig. 8, we show the minimum redshift path-length  $\Delta z_{\text{min}}$  required to detect a single 21-cm absorption line in the minimum transmission threshold  $F_{\text{th}}$ -redshift plane for three different X-ray efficiencies  $f_X$  (upper panels), or in the  $f_X$ -redshift plane for three different transmission thresholds  $F_{\text{th}}$  (lower panels). Note that for now we assume a sufficient number of background radio sources exists for such a measurement; we consider the issue of detectability at  $z = 6$  further in Fig. 9. The mock spectra used in Fig. 8 are drawn from the zr53 simulation and have been convolved with a boxcar of width 5 kHz (i.e. our assumed SKA1-low bandwidth). Unshaded white regions indicate where no absorbers are present over our total simulated path-length of  $200h^{-1}$  cGpc. Fig. 8 shows that no absorption features with  $F_{\text{th}} \leq 0.77$  should be present at  $z \lesssim 8$  for even a very low X-ray efficiency of  $f_X = 0.01$  in the late reionization model. Similarly, almost no strong 21-cm absorption with  $F_{\text{th}} \lesssim 0.99$  will exist at  $z < 7$  for  $f_X \geq 1$ . This highlights the challenging nature of 21-cm forest measurements from the diffuse IGM, even if reionization ends very late, and also how sensitive the 21-cm forest absorption is to X-ray heating. Proposals to use the 21-cm forest as a sensitive probe for distinguishing between different cosmological or dark matter models using the diffuse IGM are therefore likely to be restricted to very high redshifts, prior to any substantial X-ray heating of the IGM.

As a reference, the black curves in Fig. 8 correspond to the redshift path-length obtainable by a hypothetical observation of 1, 10, or 100 radio sources of sufficient brightness in redshift bins of width  $\Delta z = 0.2$  (i.e. an observation of  $N$  radio sources provides a total redshift path-length of  $0.2N$ ).<sup>5</sup> A null-detection over this path-length would provide a model-dependent lower limit on the X-ray background emissivity, such that  $f_X \geq f_{X, \text{max}}$ , where  $f_{X, \text{max}}$  is the maximum X-ray efficiency that retains at least one strong absorption feature with  $F \leq F_{\text{th}}$ . From the lower middle panel in Fig. 8, the null-detection of a feature with  $F_{\text{th}} < 0.9$  at  $z = 9$  in 1 (10) radio source(s) implies  $f_{X, \text{max}} \simeq 0.04$  ( $f_{X, \text{max}} \simeq 0.07$ ). The parameter space that lies below the black curves would then be disfavoured.

In practice, however, radio telescope sensitivity, spectral resolution and the availability of sufficiently bright background radio sources will impact upon the detectability of strong lines. We quantify this in Fig. 9, where similarly to Fig. 8 we show  $\Delta z_{\text{min}}$ , but now in the  $f_X$ - $F_{\text{th}}$  plane at redshift  $z = 6$ . This is shown for our LOFAR (left-hand panel), SKA1-low (middle panel), and SKA2 (right-hand panel) model assumptions, where we have convolved the synthetic spectra with a boxcar of width 10, 5, and 1 kHz, respectively. The minimum intrinsic source flux density,  $S_{\text{min}}$ , required to detected

<sup>4</sup>Note that in reality, the sensitivity  $A_{\text{eff}}/T_{\text{sys}}$  is frequency-dependent. However, over the frequency range we consider,  $142 \text{ MHz} \leq \nu_{21}/(1+z) \leq 203 \text{ MHz}$ , this dependence is reasonably weak. See fig. 8 in Braun et al. (2019) for further details.

<sup>5</sup>The choice of  $\Delta z = 0.2$  is somewhat arbitrary - we require a bin that is small enough that redshift evolution is not significant, but large enough to probe a reasonable path-length. For reference, increasing the bin size to  $\Delta z = 0.4$  would approximately halve the number of background sources required to detect a single absorber with  $F_{\text{th}}$ , assuming minimal redshift evolution across the bin.

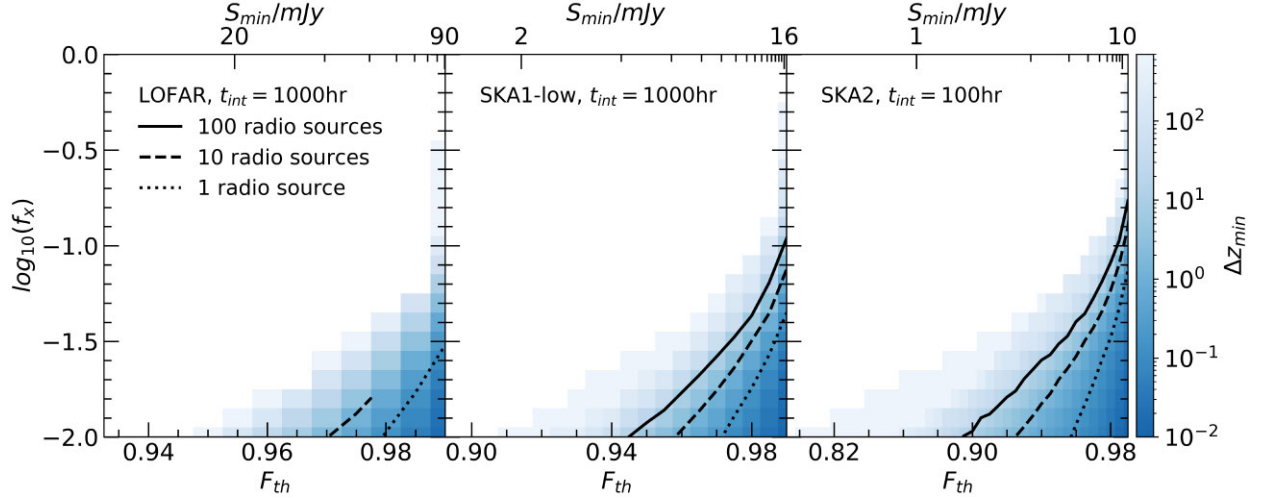


**Figure 8.** The minimum redshift path-length,  $\Delta z_{\min}$ , required to observe a single 21-cm absorption feature in the zr53 simulation *assuming that a sufficient number of background radio sources exist*. The mock spectra have been convolved with a boxcar of width 5 kHz to approximately model the effect of spectral resolution on the lines. In the upper panels, we show  $\Delta z_{\min}$  in the  $F_{\text{th}}-z$  plane for an X-ray efficiency factor of  $f_X = 1$  (left-hand panel), 0.1 (middle panel), and 0.01 (right-hand panel). In the lower panels, we instead show  $\Delta z_{\min}$  in the  $f_X-z$  plane for a 21-cm absorption feature with minimum transmission  $F \leq F_{\text{th}} = 0.8$  (left-hand panel), 0.9 (middle panel), and 0.99 (right-hand panel). Here we also note the minimum intrinsic flux density,  $S_{\min}$ , that a background radio source must have such that an absorption line with a minimum at  $F \leq F_{\text{th}}$  is detectable with SKA1-low at an S/N = 5 and integration time of  $t_{\text{int}} = 1000$  hr (see equation 12). The unshaded white regions are where no absorbers are present over our total simulated path-length of  $200h^{-1}$  cGpc. The thick black curves in each panel track the redshift path-length that would be covered by the observation of 1 (dotted), 10 (dashed), and 100 (solid) radio sources assuming redshift bins of width  $\Delta z = 0.2$ .

a line with  $F_{\text{th}}$  has also been calculated using equation (12) and is displayed on the horizontal top axis. Here we assume a strong absorption line with minimum transmission  $F_{\text{th}}$  is detected with S/N = 5 for an integration time of  $t_{\text{int}} = 1000$  hr with LOFAR and SKA1-low, and  $t_{\text{int}} = 100$  hr with SKA2. First, one can see that if using a more sensitive telescope with higher spectral resolution it is possible to detect deeper, narrower absorption features. Moreover, tighter constraints on the X-ray efficiency  $f_X$  may also be obtained. For example, at  $z = 6$ , there are no absorption features with  $F \leq 0.95$  for  $f_X > 0.01$  if observed by LOFAR. However, this increases to  $f_X > 0.025$  for SKA1-low and  $f_X > 0.05$  for SKA2. The minimum source flux density required to detect an absorption feature at fixed S/N = 5 also decreases significantly, thus increasing the number of potentially suitable background radio sources.

We quantify this in more detail in Tables 2–4, where we list the maximum X-ray efficiency,  $f_{X,\text{max}}$ , that retains at least one 21-cm absorption feature at  $z = 6$  with a transmission minimum  $F \leq F_{\text{th}}$  over a path-length of  $\Delta z = 0.2, 2,$  or  $20$  in the zr53 simulation. This corresponds to  $N = 1, 10,$  and  $100$  sources, respectively, for redshift bins of width  $\Delta z = 0.2$ . We also give the minimum flux density,  $S_{\min}$ , required to detect an absorption line with  $F_{\text{th}}$  at S/N = 5. Additionally, we give the expected number of background sources in the sky

at  $z \simeq 6$  with  $S_{\min}$  reported by Saxena et al. (2017) for an observing time of 100 hr with the standard LOFAR configuration (see their fig. 11). As a quantitative example, using Table 2 (SKA1-low), for 10 background sources with  $S_{203\text{MHz}} = 3.4$  mJy, on average we would expect to detect at least one 21-cm absorption line with  $F < 0.95$  at  $z = 6.0 \pm 0.1$  if  $f_X \leq 0.007$ . A null-detection would instead imply a lower limit of  $f_X > 0.007$ . Within our model, this X-ray efficiency may be converted to an estimate of the soft X-ray band emissivity at 0.5–2 keV, where  $\epsilon_{X,0.5-2\text{keV}} = 10^{38.3} f_X \text{ erg s}^{-1} \text{ cMpc}^{-3}$  at  $z = 6$ . Hence  $f_X > 0.007$  corresponds to  $\epsilon_{X,0.5-2\text{keV}} > 10^{36.1} \text{ erg s}^{-1} \text{ cMpc}^{-3}$ . Alternatively, from Table 3 (LOFAR), the null-detection of an absorption line with  $F < 0.95$  at  $z = 6.0 \pm 0.1$  in the spectra of 10 radio bright sources with  $S_{203\text{MHz}} = 18.2$  mJy would imply a slightly weaker constraint of  $f_X > 0.001$  and  $\epsilon_{X,0.5-2\text{keV}} > 10^{35.3} \text{ erg s}^{-1} \text{ cMpc}^{-3}$ . This suggests that lower limits on the soft X-ray background emissivity at high redshift from a null-detection of the 21-cm forest may complement existing constraints from upper limits on the 21-cm power spectrum (Greig et al. 2021a). We note, however, these results are highly model-dependent. If the Ly $\alpha$  coupling is very weak (i.e. if  $f_\alpha \ll 1$ ), or there is a significant contribution to the 21-cm forest absorption from unresolved small-scale structure, the  $f_{X,\text{max}}$  values in Table 2–4 will translate to lower limits on  $f_X$  that are conservative.



**Figure 9.** As for Fig. 8, but now the minimum redshift path-length  $\Delta z_{\min}$  is shown in the  $f_X$ - $F_{\text{th}}$  plane at redshift  $z = 6$ . The mock spectra have been convolved with a boxcar of width 10 (left-hand panel), 5 (middle panel), and 1 kHz (right-hand panel) to approximately model the effect of our assumed bandwidths for LOFAR, SKA1-low, and SKA2, respectively. Note the scale on the horizontal axis is different in each panel. The upper horizontal axis now also shows the minimum intrinsic flux density,  $S_{\min}$ , required for a background radio source, such that a line with minimum transmission  $F \leq F_{\text{th}}$  is detectable at an  $S/N = 5$  by LOFAR with an integration time of  $t_{\text{int}} = 1000$  hr (left-hand panel), by SKA1-low with  $t_{\text{int}} = 1000$  hr (middle panel), and by SKA2 with  $t_{\text{int}} = 100$  hr (right-hand panel). In the left-hand panel (LOFAR), the thick dashed curve is truncated where the number of available background radio sources predicted by Saxena, Röttgering & Rigby (2017) with  $S_{\min}$  at  $z \simeq 6$  falls below 10. The curve for 100 sources (solid) is not shown, as this exceeds the expected radio source number from Saxena et al. (2017) at the required  $S_{\min}$ .

**Table 2.** The maximum X-ray background efficiency,  $f_{X, \text{max}}$ , that retains at least one strong 21-cm absorption feature with transmission  $F \leq F_{\text{th}}$  in our synthetic 21-cm forest spectra, for a redshift path-length corresponding to  $N$  bright radio sources covering a redshift bin of width  $\Delta z = 0.2$ , centred at redshift  $z = 6$ .

| $F_{\text{th}} = e^{-\tau_{21, \text{th}}}$ | $S_{\min}(\text{mJy})$ | $z = 6$     |             | $f_{X, \text{max}}, z = 6$ |             |
|---|------------------------|-------------|-------------|----------------------------|-------------|
|   |                        | $N_{S17}$   | $N = 1$     | $N = 10$                   | $N = 100$   |
| 0.99  | 17.2                   | $\sim 100$  | 0.045       | 0.075                      | 0.109       |
| 0.95  | 3.4                    | $\sim 2400$ | $< 10^{-3}$ | 0.007                      | 0.012       |
| 0.9   | 1.7                    | $\sim 6100$ | $< 10^{-3}$ | $< 10^{-3}$                | $< 10^{-3}$ |

*Notes.* The mock spectra have been convolved with a boxcar of width 5 kHz to approximately model the effect of observed bandwidth on the lines. The minimum intrinsic flux density of the background source,  $S_{\min}$ , required to detect a line with  $F_{\text{th}}$  at an  $S/N = 5$  with SKA1-low is calculated using equation (12), assuming a bandwidth  $\Delta\nu = 5$  kHz, sensitivity  $A_{\text{eff}}/T_{\text{sys}} = 600 \text{ m}^2 \text{ K}^{-1}$ , and integration time of  $t_{\text{int}} = 1000$  hr. The expected number of radio sources in the sky with  $S_{\min}$  at  $z = 6$ ,  $N_{S17}$ , are estimated from Saxena et al. (2017) (their fig. 11). In the event of a null-detection of an absorption feature with  $F_{\text{th}}$ , the  $f_{X, \text{max}}$  values give a (model-dependent) lower limit on the X-ray efficiency.

**Table 3.** As for Table 2, except the mock 21-cm forest spectra are now smoothed with a boxcar of width 10 kHz and the minimum source flux densities,  $S_{\min}$ , have been computed for LOFAR using a bandwidth  $\Delta\nu = 10$  kHz, sensitivity  $A_{\text{eff}}/T_{\text{sys}} = 80 \text{ m}^2 \text{ K}^{-1}$ , and integration time of  $t_{\text{int}} = 1000$  hr.

| $F_{\text{th}} = e^{-\tau_{21, \text{th}}}$ | $S_{\min}(\text{mJy})$ | $z = 6$    |             | $f_{X, \text{max}}, z = 6$ |             |
|---|------------------------|------------|-------------|----------------------------|-------------|
|   |                        | $N_{S17}$  | $N = 1$     | $N = 10$                   | $N = 100$   |
| 0.99  | 91.0                   | $\sim 1$   | 0.030       | –                          | –           |
| 0.95  | 18.2                   | $\sim 90$  | $< 10^{-3}$ | 0.001                      | –           |
| 0.9   | 9.1                    | $\sim 420$ | $< 10^{-3}$ | $< 10^{-3}$                | $< 10^{-3}$ |

*Note.* Dashes mean that  $f_{X, \text{max}}$  is not measurable due to the lack of expected sources.

**Table 4.** As for Table 2, except the mock 21-cm forest spectra are now smoothed with a boxcar of width 1 kHz and the minimum source flux densities,  $S_{\min}$ , have been computed for SKA2 using a bandwidth  $\Delta\nu = 1$  kHz, sensitivity  $A_{\text{eff}}/T_{\text{sys}} = 5500 \text{ m}^2 \text{ K}^{-1}$ , and integration time of  $t_{\text{int}} = 100$  hr.

| $F_{\text{th}} = e^{-\tau_{21, \text{th}}}$ | $S_{\min}(\text{mJy})$ | $z = 6$     |             | $f_{X, \text{max}}, z = 6$ |             |
|---|------------------------|-------------|-------------|----------------------------|-------------|
|   |                        | $N_{S17}$   | $N = 1$     | $N = 10$                   | $N = 100$   |
| 0.99  | 13.2                   | $\sim 190$  | 0.074       | 0.125                      | 0.172       |
| 0.95  | 2.6                    | $\sim 3600$ | 0.007       | 0.020                      | 0.031       |
| 0.9   | 1.3                    | $\sim 8000$ | $< 10^{-3}$ | 0.004                      | 0.011       |
| 0.8   | 0.7                    | $\sim 5300$ | $< 10^{-3}$ | $< 10^{-3}$                | $< 10^{-3}$ |

## 5 CONCLUSIONS

We have used very high resolution hydrodynamical simulations combined with a novel approach for modelling patchy reionization to model the 21-cm forest during the epoch of reionization. Our simulations have been performed as part of the Sherwood–Relics simulation programme (Puchwein et al., in preparation). In particular, we have considered the observability of strong ( $\tau_{21} > 10^{-2}$ ) 21-cm absorbers in a late reionization model consistent with the large Ly  $\alpha$  forest transmission fluctuations observed at  $z = 5.5$  (Becker et al. 2015), where large neutral islands of intergalactic gas persist until  $z \simeq 6$  (Kulkarni et al. 2019; Keating et al. 2020). We also explore a wide range of assumptions for X-ray heating in the pre-reionization intergalactic medium (IGM), and have assessed the importance of several common modelling assumptions for the predicted incidence of strong 21-cm absorbers. Our key results are summarized as follows:

- (i) In a model of late reionization ending at  $z = 5.3$ , for an X-ray efficiency parameter  $f_X \lesssim 0.1$  (i.e. for relatively modest X-ray pre-heating of neutral hydrogen gas, such that the gas kinetic temperature  $T_K \lesssim 10^2 \text{ K}$ ) strong 21-cm absorption lines with optical depths  $\tau_{21} \geq 0.01$  situated in neutral islands of intergalactic gas should persist until  $z = 6$ . In this case, the 21-cm absorbers with the largest optical

depths should arise from cold, diffuse gas with overdensities  $3 < \Delta < 10$  and kinetic temperatures  $T_k < 10^2$  K. A null-detection of 21-cm forest absorbers at  $z = 6$  may therefore place a valuable lower limit on the high-redshift soft X-ray background and/or the kinetic temperature of the diffuse pre-reionization IGM in the neutral islands. With  $\sim 10$  radio-loud active galactic nuclei now known at  $5.5 < z < 6.5$  (e.g. Bañados et al. 2018b; Liu et al. 2021) and the prospect of more radio-loud sources being identified in the next few years, this possibility merits further investigation.

(ii) By far the largest uncertainty in models of the 21-cm forest is the heating of the pre-reionization IGM by the soft X-ray background (see also Mack & Wyithe 2012). In the absence of strong constraints on the soft X-ray background at  $z \geq 6$ , proposals to use the 21-cm forest to distinguish between cosmological models (where differences between competing models are small compared to the effect of X-ray heating) will likely be restricted to redshifts prior to the build-up of the soft X-ray background. Uncertainties in the strength of the Wouthuysen–Field coupling will also be important to consider if the Ly $\alpha$  background is significantly weaker than expected from extrapolating the observed star formation rate density to  $z > 6$ . In contrast, we find the effect of uncertain pressure/Jeans smoothing on the 21-cm absorption from the diffuse IGM should remain comparatively small.

(iii) Models of the 21-cm forest must include the effect of gas peculiar motions on absorption line formation to accurately predict the incidence of strong absorption features (see also Semelin 2016). Ignoring redshift space distortions reduces the incidence of the strongest 21-cm forest absorbers, and results in a maximum optical depth in the 21-cm forest that is up to a factor of  $\sim 10$  smaller compared to a model that correctly incorporates gas peculiar velocities.

(iv) We present model-dependent estimates for the minimum redshift path-length required to detect a single, strong 21-cm forest absorption feature as a function of redshift and X-ray efficiency parameter,  $f_X$  within a late reionization model that ends at redshift  $z = 5.3$ . At  $z = 6.0 \pm 0.1$  for an integration time of  $t_{\text{int}} = 1000$  hr per background radio source, a null-detection of 21-cm forest absorbers with  $F < 0.95$  at an  $S/N = 5$  in the spectra of 10 radio sources with  $S_{203\text{ MHz}} > 3.4$  mJy ( $> 18.2$  mJy) using SKA1-low (LOFAR) implies a soft X-ray background emissivity  $\epsilon_{X,0.5-2\text{ keV}} > 10^{36.1(35.3)} \text{ erg s}^{-1} \text{ cMpc}^{-3}$ . As the soft X-ray background at high redshift is still largely unconstrained, this suggests lower limits on the X-ray emissivity from a null-detection of the 21-cm forest could provide a valuable alternative constraint that complements existing and forthcoming constraints from upper limits on the 21-cm power spectrum.

While the calculation we present in this work is illustrative, a more careful forward modelling of the 21-cm absorption data is still required. We have not considered how to recover absorption features from noisy data beyond the simple  $S/N$  calculation adopted here, or how an imperfect knowledge of the radio source continuum and/or radio background might impact upon the detectability of 21-cm absorbers. Uncertainties in other parameters such as the reionization history and the Ly $\alpha$  background emissivity should furthermore be marginalized over to obtain a robust lower limit on the soft X-ray background. Our simulations do not account for the absorption from unresolved minihaloes with masses  $< 2.5 \times 10^7 M_\odot$ , and will lack coherent regions of neutral gas on scales greater than our box size of  $40h^{-1}$  cMpc. On the other hand, even a modest amount of feedback, either in the form of photoevaporation (Park et al. 2016; Nakatani, Fialkov & Yoshida 2020) or feedback from star formation (Meiksin

2011) will substantially reduce the absorption signature from minihaloes. These feedback effects may be particularly important during the final stages of reionization at  $z \simeq 6$ , where any remaining 21-cm absorption should arise from neutral islands in the diffuse IGM.

More detailed models of the 21-cm forest will require either radiation-hydrodynamical simulations that encompass a formidable dynamic range, and/or multiscale, hybrid approaches that adopt sub-grid models for unresolved absorbers and their response to feedback. Both must furthermore cover a very large and uncertain parameter space. Nevertheless, we conclude that if reionization completes at  $z < 6$ , the prospects for using SKA1-low or possibly LOFAR to place an independent constraint on the soft X-ray background using strong absorbers in the 21-cm forest are encouraging.

## ACKNOWLEDGEMENTS

We thank Benedetta Ciardi, Margherita Molaro, and Shikhar Mittal for helpful discussions. We also thank an anonymous referee for their insightful comments. The simulations used in this work were performed using the Joliot Curie supercomputer at the Tré 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](http://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. This work has made use of MATPLOTLIB (Hunter 2007), ASTROPY (Astropy Collaboration et al. 2013), NUMPY (Harris et al. 2020), and SCIPY (Virtanen et al. 2020). TŠ is supported by the University of Nottingham Vice Chancellor’s Scholarship for Research Excellence (EU). JSB acknowledges the support of a Royal Society University Research Fellowship. JSB and NH are also supported by STFC consolidated grant ST/T000171/1. MGH acknowledges support from the UKRI STFC (grant numbers ST/N000927/1 and ST/S000623/1). We thank Volker Springel for making P-GADGET-3 available.

## DATA AVAILABILITY

All data and analysis code used in this work are available from the first author on reasonable request. An open access preprint of this paper will be made available at arXiv.org.

## REFERENCES

- Aldrovandi S. M. V., Pequignot D., 1973, *A&A*, 25, 137  
 Astropy Collaboration et al., 2013, *A&A*, 558, A33  
 Aubert D., Teyssier R., 2008, *MNRAS*, 387, 295  
 Bañados E. et al., 2018a, *Nature*, 553, 473  
 Bañados E., Carilli C., Walter F., Momjian E., Decarli R., Farina E. P., Mazzucchelli C., Venemans B. P., 2018b, *ApJ*, 861, L14  
 Bañados E. et al., 2021, *ApJ*, 909, 80  
 Barkana R., Loeb A., 2005, *ApJ*, 626, 1  
 Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, *MNRAS*, 447, 3402

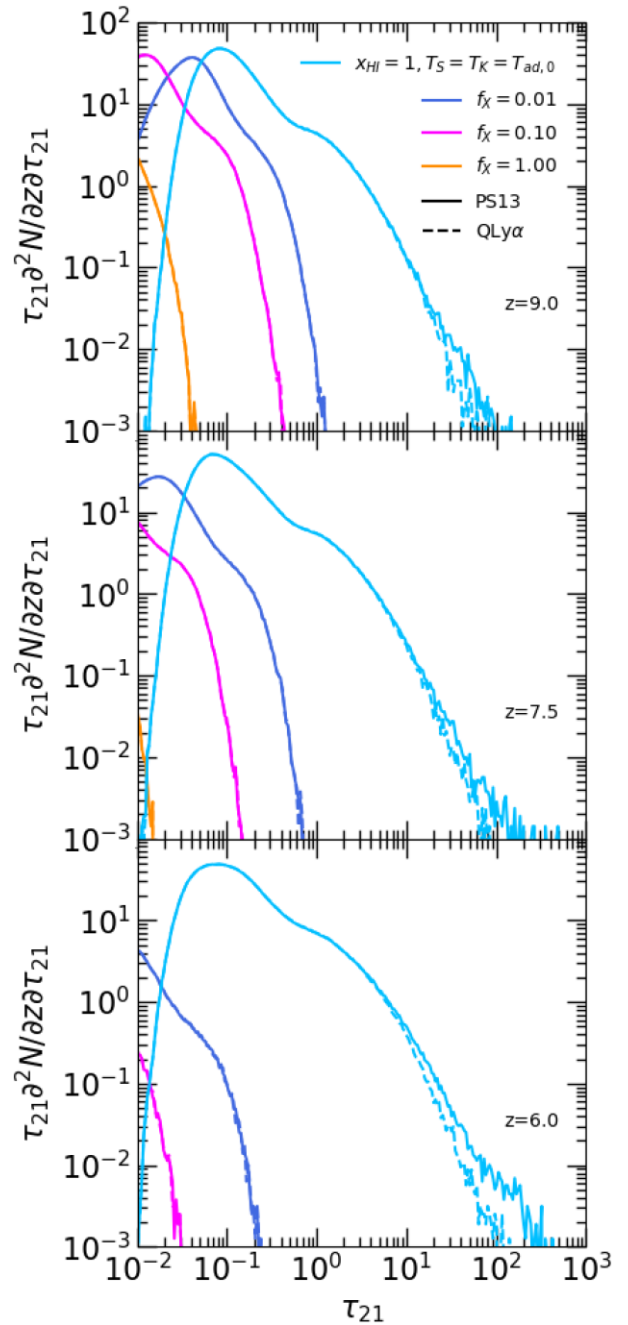
- Becker G. D., Davies F. B., Furlanetto S. R., Malkan M. A., Boera E., Douglass C., 2018, *ApJ*, 863, 92
- Becker G. D., D'Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, *MNRAS*, preprint (arXiv:2103.16610)
- Belladitta S. et al., 2020, *A&A*, 635, L7
- Bharadwaj S., Ali S. S., 2004, *MNRAS*, 352, 142
- Bolgar F., Eames E., Hottier C., Semelin B., 2018, *MNRAS*, 478, 5564
- Bolton J. S., Haehnelt M. G., 2007, *MNRAS*, 374, 493
- Bolton J. S., Becker G. D., Raskutti S., Wyithe J. S. B., Haehnelt M. G., Sargent W. L. W., 2012, *MNRAS*, 419, 2880
- Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T.-S., Meiksin A., Regan J. A., Viel M., 2017, *MNRAS*, 464, 897
- Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, *MNRAS*, 479, 1055
- Braun R., Bonaldi A., Bourke T., Keane E., Wagg J., 2019, preprint (arXiv:1912.12699)
- Cappelluti N. et al., 2017, *ApJ*, 837, 19
- Carilli C. L., Gnedin N. Y., Owen F., 2002, *ApJ*, 577, 22
- Cen R., 1992, *ApJS*, 78, 341
- Chardin J., Puchwein E., Haehnelt M. G., 2017, *MNRAS*, 465, 3429
- Chardin J., Kulkarni G., Haehnelt M. G., 2018, *MNRAS*, 478, 1065
- Chen X., Miralda-Escudé J., 2004, *ApJ*, 602, 1
- Chuzhoy L., Shapiro P. R., 2007, *ApJ*, 655, 843
- Ciardi B. et al., 2013, *MNRAS*, 428, 1755
- Ciardi B. et al., 2015b, *MNRAS*, 453, 101
- Ciardi B., Salvaterra R., 2007, *MNRAS*, 381, 1137
- Ciardi B., Salvaterra R., Di Matteo T., 2010, *MNRAS*, 401, 2635
- Ciardi B., Inoue S., Mack K., Xu Y., Bernardi G., 2015a, *Proc. Sci.*, 21-cm forest with the SKA. SISSA, Trieste, PoS(AASKA14)006
- D'Aloisio A., McQuinn M., Trac H., 2015, *ApJ*, 813, L38
- D'Aloisio A., McQuinn M., Trac H., Cain C., Mesinger A., 2020, *ApJ*, 898, 149
- Davies F. B. et al., 2018, *ApJ*, 864, 142
- Davies F. B., Furlanetto S. R., 2016, *MNRAS*, 460, 1328
- Dijkstra M., Haiman Z., Loeb A., 2004, *ApJ*, 613, 646
- Dijkstra M., Gilfanov M., Loeb A., Sunyaev R., 2012, *MNRAS*, 421, 213
- Eide M. B., Graziani L., Ciardi B., Feng Y., Kakiichi K., Di Matteo T., 2018, *MNRAS*, 476, 1174
- Eilers A.-C., Davies F. B., Hennawi J. F., Prochaska J. X., Lukić Z., Mazzucchelli C., 2017, *ApJ*, 840, 24
- Eilers A.-C., Davies F. B., Hennawi J. F., 2018, *ApJ*, 864, 53
- Emberson J. D., Thomas R. M., Alvarez M. A., 2013, *ApJ*, 763, 146
- Ewall-Wice A., Dillon J. S., Mesinger A., Hewitt J., 2014, *MNRAS*, 441, 2476
- Fialkov A., Cohen A., Barkana R., Silk J., 2017, *MNRAS*, 464, 3498
- Field G. B., 1958, *Proc. IRE*, 46, 240
- Fixsen D. J., 2009, *ApJ*, 707, 916
- Furlanetto S. R., 2006a, *MNRAS*, 370, 1867
- Furlanetto S. R., 2006b, *MNRAS*, 371, 867
- Furlanetto S. R., Furlanetto M. R., 2007a, *MNRAS*, 374, 547
- Furlanetto S. R., Furlanetto M. R., 2007b, *MNRAS*, 379, 130
- Furlanetto S. R., Loeb A., 2002, *ApJ*, 579, 1
- Furlanetto S. R., Pritchard J. R., 2006, *MNRAS*, 372, 1093
- Furlanetto S. R., Stoever S. J., 2010, *MNRAS*, 404, 1869
- Furlanetto S. R., Peng Oh S., Briggs F. H., 2006, *Phys. Rep.*, 433, 181
- Gaikwad P. et al., 2020, *MNRAS*, 494, 5091
- Garzilli A., Theuns T., Schaye J., 2015, *MNRAS*, 450, 1465
- Ghara R. et al., 2020, *MNRAS*, 493, 4728
- Gilfanov M., Grimm H. J., Sunyaev R., 2004, *MNRAS*, 347, L57
- Gnedin N. Y., Hui L., 1998, *MNRAS*, 296, 44
- Greig B. et al., 2021b, *MNRAS*, 501, 1
- Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, *MNRAS*, 466, 4239
- Greig B., Mesinger A., Bañados E., 2019, *MNRAS*, 484, 5094
- Greig B., Trott C. M., Barry N., Mutch S. J., Pindor B., Webster R. L., Wyithe J. S. B., 2021a, *MNRAS*, 500, 5322
- Haardt F., Madau P., 1996, *ApJ*, 461, 20
- Harris C. R. et al., 2020, *Nature*, 585, 357
- Hennawi J. F., Davies F. B., Wang F., Oñorbe J., 2021, *MNRAS*, preprint (arXiv:2007.15747)
- Hirata C. M., 2006, *MNRAS*, 367, 259
- Hsyu T., Cooke R. J., Prochaska J. X., Bolte M., 2020, *ApJ*, 896, 77
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Ighina L., Belladitta S., Caccianiga A., Broderick J. W., Drouart G., Moretti A., Seymour N., 2021, *A&A*, 647, L11
- Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, *MNRAS*, 439, 725
- Intema H. T., Jagannathan P., Mooley K. P., Frail D. A., 2017, *A&A*, 598, A78
- Ioka K., Mészáros P., 2005, *ApJ*, 619, 684
- Kashino D., Lilly S. J., Shibuya T., Ouchi M., Kashikawa N., 2020, *ApJ*, 888, 6
- Katz N., Weinberg D. H., Hernquist L., 1996, *ApJS*, 105, 19
- Kaur H. D., Gillet N., Mesinger A., 2020, *MNRAS*, 495, 2354
- Keating L. C., Haehnelt M. G., Becker G. D., Bolton J. S., 2014, *MNRAS*, 438, 1820
- Keating L. C., Puchwein E., Haehnelt M. G., 2018, *MNRAS*, 477, 5501
- Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, *MNRAS*, 491, 1736
- Kondapally R., Best P. N., Hardcastle M. J., 2021, *A&A*, 648, A3
- Kuhlen M., Madau P., Montgomery R., 2006, *ApJ*, 637, L1
- Kulkarni G., Hennawi J. F., Oñorbe J., Rorai A., Springel V., 2015, *ApJ*, 812, 30
- Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, *MNRAS*, 485, L24
- Lehmer B. D. et al., 2016, *ApJ*, 825, 7
- Lidz A., McQuinn M., Zaldarriaga M., Hernquist L., Dutta S., 2007, *ApJ*, 670, 39
- Liszt H., 2001, *A&A*, 371, 698
- Liu Y. et al., 2021, *ApJ*, 908, 124
- McGreer I. D., Mesinger A., D'Odorico V., 2015, *MNRAS*, 447, 499
- McQuinn M., 2012, *MNRAS*, 426, 1349
- Mack K. J., Wyithe J. S. B., 2012, *MNRAS*, 425, 2988
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Madau P., Efstathiou G., 1999, *ApJ*, 517, L9
- Madau P., Meiksin A., Rees M. J., 1997, *ApJ*, 475, 429
- Majumdar S., Kamran M., Pritchard J. R., Mondal R., Mazumdar A., Bharadwaj S., Mellema G., 2020, *MNRAS*, 499, 5090
- Mao Y., Shapiro P. R., Mellema G., Iliev I. T., Koda J., Ahn K., 2012, *MNRAS*, 422, 926
- Mason C. A. et al., 2019, *MNRAS*, 485, 3947
- Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, *ApJ*, 856, 2
- Meiksin A., 2011, *MNRAS*, 417, 1480
- Meiksin A., 2020, *MNRAS*, 491, 4884
- Mertens F. G. et al., 2020, *MNRAS*, 493, 1662
- Mesinger A., 2010, *MNRAS*, 407, 1328
- Mirocha J., 2014, *MNRAS*, 443, 1211
- Mittal S., Kulkarni G., 2020, *MNRAS*, 503, 4264
- Mondal R. et al., 2020, *MNRAS*, 498, 4178
- Nakatani R., Fialkov A., Yoshida N., 2020, *ApJ*, 905, 151
- Nasir F., D'Aloisio A., 2020, *MNRAS*, 494, 3080
- Nasir F., Bolton J. S., Becker G. D., 2016, *MNRAS*, 463, 2335
- Oh S. P., 2002, *MNRAS*, 336, 1021
- Oñorbe J., Davies F. B., Lukić Z., Hennawi J. F., Sorini D., 2019, *MNRAS*, 486, 4075
- Pagano L., Delouis J. M., Mottet S., Puget J. L., Vibert L., 2020, *A&A*, 635, A99
- Park H., Shapiro P. R., Choi J.-h., Yoshida N., Hirano S., Ahn K., 2016, *ApJ*, 831, 86
- Planck CollaborationXVI, 2014, *A&A*, 571, A16
- Planck CollaborationVI, 2020, *A&A*, 641, A6
- Pritchard J. R., Furlanetto S. R., 2006, *MNRAS*, 367, 1057
- Pritchard J. R., Furlanetto S. R., 2007, *MNRAS*, 376, 1680
- Pritchard J. R., Loeb A., 2012, *Rep. Prog. Phys.*, 75, 086901
- Puchwein E., Springel V., 2013, *MNRAS*, 428, 2966 (PS13)

- Puchwein E., Bolton J. S., Haehnelt M. G., Madau P., Becker G. D., Haardt F., 2015, *MNRAS*, 450, 4081
- Puchwein E., Haardt F., Haehnelt M. G., Madau P., 2019, *MNRAS*, 485, 47
- Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021, *MNRAS*, preprint (arXiv:2101.09033)
- Ross H. E., Dixon K. L., Iliev I. T., Mellema G., 2017, *MNRAS*, 468, 3785
- Saxena A., Röttgering H. J. A., Rigby E. E., 2017, *MNRAS*, 469, 4083
- Schaye J., Theuns T., Leonard A., Efstathiou G., 1999, *MNRAS*, 310, 57
- Semelin B., 2016, *MNRAS*, 455, 962
- Shimabukuro H., Ichiki K., Inoue S., Yokoyama S., 2014, *Phys. Rev. D*, 90, 83003
- Shimwell T. W. et al., 2017, *A&A*, 598, A104
- Shull J. M., van Steenberg M. E., 1985, *ApJ*, 298, 268
- Smith D. J. B. et al., 2016, in Reylé C., Richard J., Cambrésy L., Deleuil M., Pécontal E., Tresse L., Vauglin I., eds, SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, p. 271
- Springel V., 2005, *MNRAS*, 364, 1105
- Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998, *MNRAS*, 301, 478
- Thyagarajan N., 2020, *ApJ*, 899, 16
- Trott C. M. et al., 2020, *MNRAS*, 493, 4711
- Verner D. A., Ferland G. J., 1996, *ApJS*, 103, 467
- Verner D. A., Ferland G. J., Korista K. T., Yakovlev D. G., 1996, *ApJS*, 465, 487
- Viel M., Haehnelt M. G., Springel V., 2004, *MNRAS*, 354, 684
- Villanueva-Domingo P., Ichiki K., 2021, preprint (arXiv:2104.10695)
- Villanueva-Domingo P., Mena O., Miralda-Escudé J., 2020, *Phys. Rev. D*, 101, 083502
- Virtanen P. et al., 2020, *Nat. Methods*, 17, 261
- Voronov G., 1997, *At. Data Nucl. Data Tables*, 65, 1
- Wang F. et al., 2020, *ApJ*, 896, 23
- Wayth R. B. et al., 2015, *Publ. Astron. Soc. Aust.*, 32, e025
- Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, *MNRAS*, 485, 1350
- Weymann R., 1965, *Phys. Fluids*, 8, 2112
- Wouthuysen S. A., 1952, *AJ*, 57, 31
- Xu Y., Chen X., Fan Z., Trac H., Cen R., 2009, *ApJ*, 704, 1396
- Xu Y., Ferrara A., Chen X., 2011, *MNRAS*, 410, 2025
- Yang J. et al., 2020a, *ApJ*, 897, L14
- Yang J. et al., 2020b, *ApJ*, 904, 26

## APPENDIX A: TEST OF THE PRESCRIPTION FOR CONVERTING DENSE GAS INTO COLLISIONLESS PARTICLES

As discussed in Section 2.1, we adopt a simplified scheme for the treatment of dense, star-forming gas in the Sherwood–Relics simulations, where all gas particles with density  $\Delta > 1000$  and temperature  $T_K < 10^5$  K are converted to collisionless star particles (Viel et al. 2004). As a consequence, very dense, cold halo gas is not included in the Sherwood–Relics models. We test whether this approximation affects our results for the 21-cm forest in Fig. A1. Here we compare two models drawn from the Sherwood simulation suite (Bolton et al. 2017) that use the same box size, mass resolution, and initial conditions as the other simulations used in this work. These two additional simulations use either the simplified scheme used in this study (QLy $\alpha$ ) or the star formation and energy-driven winds prescription of Puchwein & Springel (2013, hereafter PS13). The only difference between these two models is the incorporation of dense, star-forming gas within the PS13 simulation.

In Fig. A1, we show the differential line number density distribution obtained after applying the neutral fraction, gas kinetic and spin temperature from the patchy zr53 simulation to the native density and peculiar velocity fields from the QLy $\alpha$  and PS13 models. As before, we consider three different X-ray efficiencies. We observe



**Figure A1.** The differential line number density distribution for 21-cm forest absorption features in simulations that use two different implementations for the treatment of dense gas. The dashed curves displays the simplified approach used in this work (the QLy $\alpha$  simulation), whereas the solid curves use the subgrid star formation and feedback model from Puchwein & Springel (2013) (the PS13 simulation, solid line). The results are shown at three different redshifts:  $z = 9$  (top panel), 7.5 (middle panel), and 6 (bottom panel). The X-ray efficiencies are  $f_X = 0.01$  (blue curves), 0.1 (fuchsia curves), and 1 (orange curves). For comparison, the cyan curves show the distribution for fully neutral, unheated gas with temperature equal to the adiabatic temperature at mean density (i.e.  $x_{\text{HI}} = 1$  and  $T_S = T_K = T_{\text{ad},0}$ ).

little to no difference in the statistics of the 21-cm forest computed from these two simulations. This is because the highest density gas is usually located close to ionizing sources, and so is often too hot, ionized or rare to exhibit significant amounts of strong absorption in

the hyperfine line. This is further illustrated by the cyan curves in Fig. A1, where the mock 21-cm forest spectra are instead computed assuming a fully neutral, isothermal gas with  $T_S = T_K = T_{\text{ad},0}$ , where  $T_{\text{ad},0} = 2.73 \text{ K}(1+z)^2/148.8$  is the gas temperature assuming adiabatic cooling at the mean density. Small differences due to the presence of the high-density gas in the PS13 simulation are now apparent in the tail of the distribution at  $\tau_{21} \gtrsim 10$ . However, if we also include the adiabatic heating of the gas by compression, such that  $T_{\text{ad}} = T_{\text{ad},0}(1+\delta)^{2/3}$ , these models become almost identical. We conclude that the approximate treatment of dense, star-forming gas we adopt in this work should not significantly change our key results. The relative rarity of 21-cm absorption from cold gas within massive haloes suggests this population will in any case be completely dominated by 21-cm absorbers from the diffuse IGM and/or minihaloes during reionization.

## APPENDIX B: CALCULATION OF THE X-RAY AND Ly $\alpha$ SPECIFIC INTENSITIES, GAS KINETIC TEMPERATURE, AND IONIZATION STATE

In this section we describe the model for the X-ray heated IGM introduced in Section 2.2. The X-ray background is primarily responsible for ionizing and heating the intergalactic medium prior to reionization. The proper specific intensity of the X-ray background,  $J_{X,\nu}$  [erg s<sup>-1</sup> cm<sup>-2</sup> Hz<sup>-1</sup> sr<sup>-1</sup>], is given by the solution to the cosmological radiative transfer equation (Haardt & Madau 1996; Mirocha 2014):

$$J_{X,\nu}(z) = \frac{c(1+z)^3}{4\pi} \int_z^{z_*} \frac{\epsilon_{X,\nu'}(z')}{H(z')(1+z')} e^{-\bar{\tau}_\nu(z,z')} dz', \quad (\text{B1})$$

where  $\epsilon_{X,\nu}$  is the comoving X-ray emissivity,  $z_*$  is the redshift when X-ray emitting sources first form, and the emission frequency,  $\nu'$ , of a photon emitted at redshift  $z'$  and observed at frequency  $\nu$  and redshift  $z$  is

$$\nu' = \nu \frac{(1+z')}{(1+z)}. \quad (\text{B2})$$

The optical depth encountered by a photon observed at frequency  $\nu$  is

$$\bar{\tau}_\nu(z, z') = c \sum_i \int_z^{z'} \frac{\bar{n}_i(z'') \sigma_{\nu,i}}{H(z'')(1+z'')} dz'', \quad (\text{B3})$$

where the sum is over the species  $i = \text{H I}, \text{He I}, \text{He II}$ , and  $\sigma_{\nu,i}$  are the photoionization cross-sections (Verner et al. 1996).

The photoionization rates for species  $i = \text{H I}, \text{He I}, \text{He II}$  are

$$\Gamma_i = 4\pi \int_{\nu_i}^{\infty} \frac{J_{X,\nu}}{h_p \nu} \sigma_{\nu,i} d\nu + 4\pi \sum_j \Phi_i(h_p(\nu - \nu_j), x_e) \int_{\nu_j}^{\infty} \frac{J_{X,\nu}}{h_p \nu} \sigma_{\nu,j} d\nu, \quad (\text{B4})$$

where  $\nu_i$  is the frequency of the ionization threshold for species  $i$ . The second term in equation (B4) arises from secondary ionizations due to collisions with energetic photoelectrons, where  $\Phi_i$  is the number of secondary ionizations per primary photoelectron of energy  $h_p(\nu - \nu_i)$  for a free electron fraction of  $x_e$  (Shull & van Steenberg 1985). The corresponding photoheating rates are

$$\mathcal{H}_i = 4\pi n_i \phi_{\text{heat}}(h_p(\nu - \nu_i), x_e) \int_{\nu_i}^{\infty} \frac{J_{X,\nu}(\nu - \nu_i)}{\nu} \sigma_{\nu,i} d\nu, \quad (\text{B5})$$

where  $\phi_{\text{heat}}$  is the fraction of the primary photoelectron energy that contributes to the heating of the gas. We use the tabulated results from Furlanetto & Stoever (2010) for  $\Phi_i$  and  $\phi_{\text{heat}}$ .

The Compton scattering of free electrons off X-ray background photons will also heat the IGM (Madau & Efstathiou 1999). The Compton heating rate is

$$\mathcal{H}_C = \frac{4\pi n_e \sigma_T}{m_e c^2} \int_0^{\infty} J_{X,\nu} (h_p \nu - 4k_B T_K) d\nu, \quad (\text{B6})$$

where  $\sigma_T = 6.65 \times 10^{-25} \text{ cm}^2$  is the Thomson cross-section, appropriate for X-ray photons with energy  $\lesssim 100 \text{ keV}$  (i.e. relativistic effects may be ignored).

The Ly  $\alpha$  background has two contributions: emission from stars, and Ly  $\alpha$  photons produced by the excitation of H I atoms by X-ray photons. The proper Ly  $\alpha$  specific intensity from stars,  $J_{\alpha,*}$ , requires consideration of both Ly  $\alpha$  and higher order Lyman series photons. This is because the Ly  $\alpha$  photons redshift into resonance at redshift  $z$  and generate Ly  $\alpha$  photons via a series of radiative cascades to lower energies (Pritchard & Furlanetto 2006), such that

$$J_{\alpha,*}(z) = \frac{c(1+z)^3}{4\pi} \sum_{n=2}^{n_{\text{max}}} f_n \int_z^{z_{\text{max}}(n)} \frac{\epsilon_{\alpha,\nu'_n}(z')}{H(z')(1+z')} dz', \quad (\text{B7})$$

where  $f_n$  is the probability of producing a Ly  $\alpha$  photon from a cascade from level  $n$ ,  $\nu'_n$  is the emission frequency at redshift  $z'$  that corresponds to absorption by level  $n$  at redshift  $z$ ,

$$\nu'_n = \nu_{\text{LL}} \left(1 - \frac{1}{n^2}\right) \left(\frac{1+z'}{1+z}\right), \quad (\text{B8})$$

and  $z_{\text{max}}(n)$  is the maximum redshift from which an emitted photon will redshift into the Ly  $\alpha$  resonance,

$$z_{\text{max}}(n) = (1+z) \frac{1 - (n+1)^{-2}}{1 - n^{-2}} - 1. \quad (\text{B9})$$

We use the tabulated values for  $f_n$  from Pritchard & Furlanetto (2006) and assume  $n_{\text{max}} = 23$  (Barkana & Loeb 2005). The contribution from H I excitation by X-ray photons is (Pritchard & Furlanetto 2007)

$$J_{\alpha,X}(z) = \frac{\lambda_\alpha}{4\pi H(z)} \sum_i \frac{\phi_\alpha(h_p(\nu - \nu_i), x_e)}{\phi_{\text{heat}}(h_p(\nu - \nu_i), x_e)} \mathcal{H}_i, \quad (\text{B10})$$

where  $\phi_\alpha$  is the fraction of the primary photoelectron energy that is deposited in Ly  $\alpha$  photons (Furlanetto & Stoever 2010).

The Ly  $\alpha$  background photons will also heat the IGM by scattering off H I atoms (Chen & Miralda-Escudé 2004; Chuzhoy & Shapiro 2007; Ciardi & Salvaterra 2007; Mittal & Kulkarni 2020), although this effect is usually small compared to heating by X-ray photons (Ciardi et al. 2010). The Ly  $\alpha$  heating rate is

$$\mathcal{H}_\alpha = \frac{4\pi b H(z)}{c \lambda_\alpha} (J_{\alpha,*c}(z) I_c + [J_{\alpha,*i}(z) + J_{\alpha,X}(z)] I_i), \quad (\text{B11})$$

where  $J_{\alpha,*c}$  is the specific intensity of continuum ( $n = 2$ ) Ly  $\alpha$  photons,  $J_{\alpha,*i}$  is the specific intensity of recombination photons injected at the line centre ( $n > 2$ ), and  $I_c, I_i$  are the integrals over the Ly  $\alpha$  line profile. We use the approximations provided by Furlanetto & Pritchard (2006) for  $I_c$  and  $I_i$ .

Given the photoionization and heating rates, the evolution of the ionization and thermal state of the IGM at fixed gas density may be obtained by solving four coupled differential equations (Bolton & Haehnelt 2007). We assume all gas is initially neutral and has a kinetic temperature set by adiabatic heating and cooling only:

$$T_{\text{ad}} = T_{\text{ad},0} (1+\delta)^{2/3} = 2.73 \text{ K} \frac{(1+z)^2 (1+\delta)^{2/3}}{(1+z_{\text{dec}})}, \quad (\text{B12})$$

where we assume the gas with overdensity  $\delta$  thermally decouples from the CMB at  $z_{\text{dec}} = 147.8$  (Furlanetto et al. 2006). The first three differential equations then describe the number density of ionized hydrogen, singly ionized, and double ionized helium:

$$\frac{dn_{\text{H II}}}{dt} = n_{\text{H I}}(\Gamma_{\text{H I}} + n_e \Gamma_{\text{c,H I}}) - n_{\text{H II}} n_e \alpha_{\text{H II}}, \quad (\text{B13})$$

$$\frac{dn_{\text{He II}}}{dt} = n_{\text{He I}}(\Gamma_{\text{He I}} + n_e \Gamma_{\text{c,He I}}) - n_{\text{He II}} n_e (\alpha_{\text{He II}} + \alpha_d) - \frac{dn_{\text{He III}}}{dt}, \quad (\text{B14})$$

$$\frac{dn_{\text{He III}}}{dt} = n_{\text{He II}}(\Gamma_{\text{He II}} + n_e \Gamma_{\text{c,He II}}) - n_{\text{He III}} n_e \alpha_{\text{He III}}, \quad (\text{B15})$$

where  $\alpha_i$  and  $\Gamma_{\text{c},i}$  are, respectively, the recombination rates (Verner & Ferland 1996) and collisional ionization rates (Voronov 1997) for species  $i = \text{H I}, \text{He I}, \text{He II}$ , and  $\alpha_d$  is the He II dielectronic recombination coefficient (Aldrovandi & Pequignot 1973). The number density of neutral hydrogen, neutral helium and free electrons is

$$n_{\text{H I}} = n_{\text{H}} - n_{\text{H II}}, \quad (\text{B16})$$

$$n_{\text{He I}} = \frac{Y_p}{4(1 - Y_p)} n_{\text{H}} - n_{\text{He II}} - n_{\text{He III}}, \quad (\text{B17})$$

$$n_e = n_{\text{H II}} + n_{\text{He II}} + 2n_{\text{He III}}, \quad (\text{B18})$$

where  $Y_p = 0.24$  is the primordial helium fraction by mass (Hsyu et al. 2020). The fourth differential equation describes the kinetic temperature for gas at fixed overdensity:

$$\frac{dT_K}{dt} = \frac{2\mu m_{\text{H}}}{3k_{\text{B}}\rho} [\mathcal{H}_{\text{tot}} - \Lambda_{\text{tot}}] + \frac{T_K}{\mu} \frac{d\mu}{dt} - 2H(t)T_K, \quad (\text{B19})$$

where  $\mu$  is the mean molecular weight, and  $\mathcal{H}_{\text{tot}}$  and  $\Lambda_{\text{tot}}$  are the total heating and cooling rates per unit volume, respectively.<sup>6</sup>

<sup>6</sup>Note that for X-ray heated gas, we can safely neglect the missing  $[2T_K/3(1 + \delta)](d\delta/dt)$  term in equation (B19), as this is small compared to the photoheating term for gas in the diffuse IGM. Instead, prior to any X-ray or Ly  $\alpha$  heating, we just assume the gas kinetic temperature follows the solution of equation (B19) for adiabatic heating and cooling (i.e. equation B12). This simplification is advantageous, as a non-local calculation of the heating from adiabatic compression is significantly more complex (see also the discussion of this point in Villanueva-Domingo, Mena & Miralda-Escudé 2020). In practice, however, we find that even if we ignore the heating from adiabatic compression and assume an initially isothermal IGM, the change to our results is negligible. This is because all our heating models have experienced appreciable X-ray and Ly  $\alpha$  heating by  $z = 6$ . Finally, recall also that heating from adiabatic compression and shocks for gas with temperatures greater than the  $T_K$  predicted by equation (B19) is already included self-consistently within our hydrodynamical simulations.

The total heating rate is

$$\mathcal{H}_{\text{tot}} = \mathcal{H}_\alpha + \mathcal{H}_C + \sum_i \mathcal{H}_i, \quad (\text{B20})$$

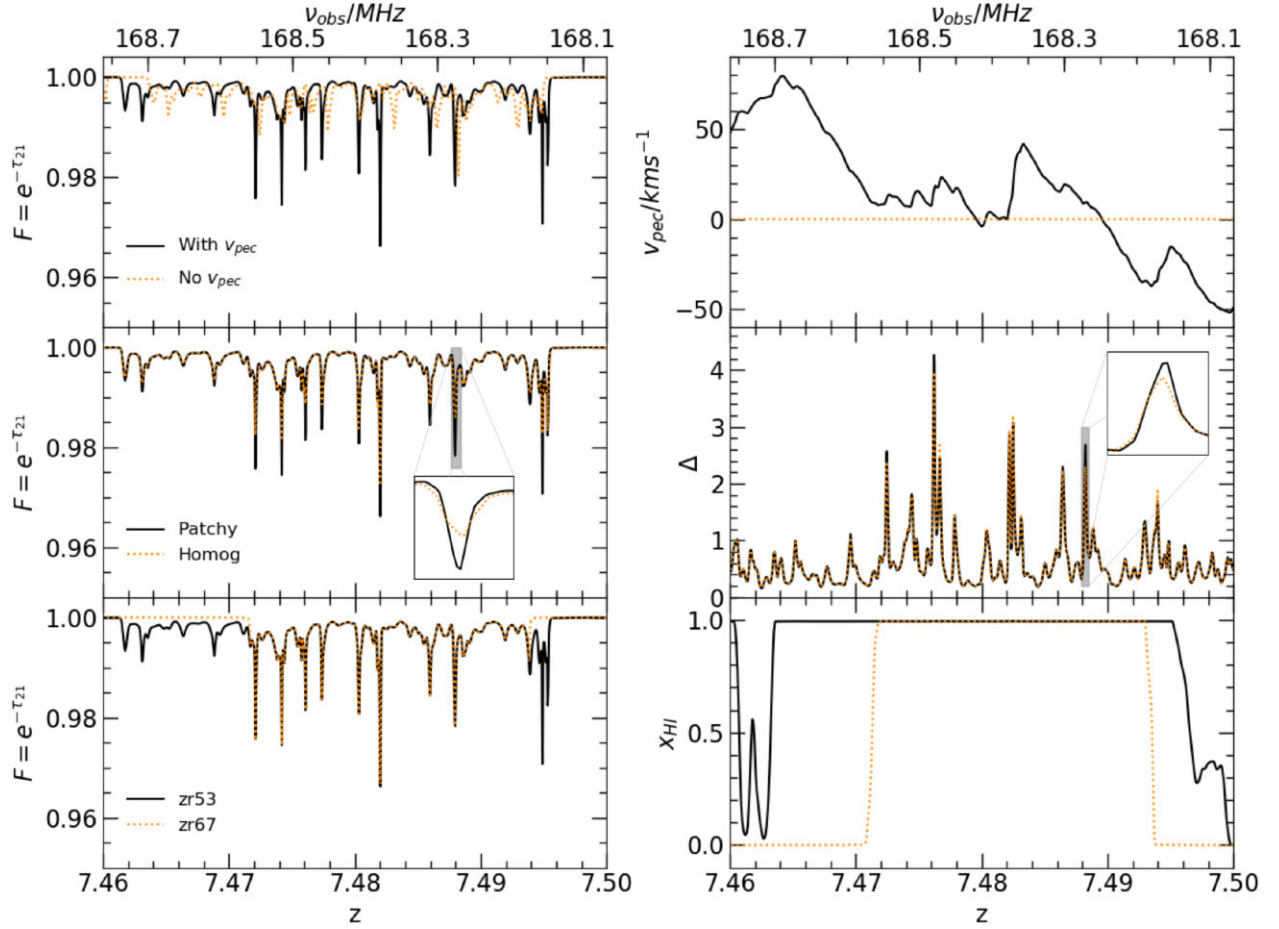
and the total cooling rate is

$$\Lambda_{\text{tot}} = \sum_i \Lambda_{\text{c},i} + \sum_i \Lambda_{\text{ex},i} + \sum_j \Lambda_{\text{rec},j} + \Lambda_{\text{brem}} + \Lambda_{\text{C}}, \quad (\text{B21})$$

where the sums are over species  $i = \text{H I}, \text{He I}, \text{He II}$  and  $j = \text{H II}, \text{He II}, \text{He III}$ . We consider contributions to the total cooling rate from collisional ionization, collisional excitation, recombination, bremsstrahlung, and inverse-Compton scattering of electrons off CMB photons, respectively (cf. Katz, Weinberg & Hernquist 1996). We use the collisional excitation cooling rates from Cen (1992), the inverse-Compton cooling rate from Weymann (1965) and the bremsstrahlung cooling rate from Theuns et al. (1998). The recombination and collisional ionization cooling rates are derived from the Verner & Ferland (1996) and Voronov (1997) fits, respectively.

### APPENDIX C: EFFECT OF PECULIAR VELOCITIES, PRESSURE SMOOTHING, AND REIONIZATION REDSHIFT ON INDIVIDUAL 21-CM ABSORBERS

In Section 4.2 and Fig. 7, the expected differences in the incidence of strong 21-cm absorbers for different modelling assumptions was discussed. This included the effect of gas peculiar velocities, pressure (Jeans) smoothing, and the timing of reionization on the 21-cm forest. In Fig. C1, the impact of these modelling assumptions on individual 21-cm absorbers is illustrated.



**Figure C1.** Left-hand panels: an example line of sight drawn from our mock 21-cm forest spectra at  $z \sim 7.5$  for an X-ray efficiency of  $f_X = 0.1$  and Ly  $\alpha$  efficiency  $f_\alpha = 1$ . From the top to bottom, we show the zr53 model with the solid black curve, and compare this to several model parameter variations (orange dotted curves): gas peculiar velocities set to zero (top panels), pressure smoothing under the assumption of homogeneous heating in the zr53-homog simulation (middle panels), and an earlier end to reionization in the zr67 model (bottom panels). Right-hand panels: the quantities responsible for the observed differences between the spectra displayed in the left-hand column. From the top to bottom, these are the gas peculiar velocity,  $v_{\text{pec}}$ , normalized gas density  $\Delta = (1 + \delta) = \rho/(\rho_0)$ , and hydrogen neutral fraction,  $x_{\text{HI}}$ . In the middle panels, we also present a zoomed-in view of an absorption feature (left-hand panels) and the associated density peak (right-hand panels), which has been broadened by pressure smoothing in the simulation with homogeneous heating.

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