Lyman-α transmission properties of the intergalactic medium in the CoDaII simulation

Max Gronke, Pierre Ocvirk, Charlotte Mason, Jorryt Matthee, Sarah Bosman, Jenny Sorce, Joseph Lewis, Kyungjin Ahn, Taha Dawoodbhoy, Ilian Iliev, et al.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **508,** 3697–3709 (2021) https://doi.org/10.1093/mnras/stab2762
Advance Access publication 2021 September 25

Lyman-α transmission properties of the intergalactic medium in the CoDaII simulation

Max Gronke [1,2★†] Pierre Ocvirk [3] Charlotte Mason [4†] Jorryt Matthee [5] Sarah E. I. Bosman [6] Jenny G. Sorce,[7,8,9] Joseph Lewis [3] Kyungjin Ahn,[10] Dominique Aubert,[3] Taha Dawoodbhoy [12] Ilian T. Iliev [11] Paul R. Shapiro[12] and Gustavo Yepes [13,14]

[1]*Department of Physics, University of California, CA 93106 Santa Barbara, USA*
[2]*Department of Physics & Astronomy, Johns Hopkins University, MD 21218 Baltimore, USA*
[3]*Observatoire Astronomique de Strasbourg, Université de Strasbourg, 11 rue de l'Université, F-67000 Strasbourg, France*
[4]*Center for Astrophysics,| Harvard & Smithsonian, 60 Garden St, MA 02138 Cambridge, USA*
[5]*Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland*
[6]*Department of Physics and Astronomy, University College London, Gower Street, WC1E 6BT London, UK*
[7]*Centre de Recherche Astrophysique de Lyon, Univ Lyon, ENS de Lyon, Univ Lyon1, CNRS, UMR5574, F-69007 Lyon, France*
[8]*Univ Lyon, Univ Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230 Saint-Genis-Laval, France*
[9]*Leibniz-Institut für Astrophysik (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany*
[10]*Department of Earth Sciences, Chosun University, Gwangju 61452, Korea*
[11]*Astronomy Centre, Department of Physics & Astronomy, University of Sussex, Falmer, BN1 9QH Brighton, Uk*
[12]*Department of Astronomy, The University of Texas at Austin, TX 78712-1205, USA*
[13]*Departamento de Física Teórica M-8, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*
[14]*Centro de Investigación Avanzada en Física Fundamental (CIAFF), Universidad Autónoma de Madrid, E-28049 Madrid, Spain*

Accepted 2021 September 14. Received 2021 September 14; in original form 2020 April 29

ABSTRACT

The decline in abundance of Lyman-α (Lyα) emitting galaxies at $z \gtrsim 6$ is a powerful and commonly used probe to constrain the progress of cosmic reionization. We use the CoDaII simulation, which is a radiation hydrodynamic simulation featuring a box of ∼94 comoving Mpc side length, to compute the Lyα transmission properties of the intergalactic medium (IGM) at $z \sim 5.8$ to 7. Our results mainly confirm previous studies, i.e. we find a declining Lyα transmission with redshift and a large sightline-to-sightline variation. However, motivated by the recent discovery of *blue* Lyα peaks at high redshift, we also analyse the IGM transmission on the blue side, which shows a rapid decline at $z \gtrsim 6$ of the blue transmission. This low transmission can be attributed not only to the presence of neutral regions but also to the residual neutral hydrogen within ionized regions, for which a density even as low as $n_{\rm HI} \sim 10^{-9}\,{\rm cm}^{-3}$ (sometimes combined with kinematic effects) leads to a significantly reduced visibility. Still, we find that ∼ 1 per cent of sightlines towards $M_{1600{\rm AB}} \sim -21$ galaxies at $z \sim 7$ are transparent enough to allow a transmission of a blue Lyα peak. We discuss our results in the context of the interpretation of observations.

Key words: dark ages, reionization, first stars – intergalactic medium – galaxies: formation.

1 INTRODUCTION

While baryonic astrophysics in today's Universe is mainly governed by effects which are small-scale – for cosmological standards – this was not always the case. A few billion years ago, at $z \sim 6$ the 'Epoch of Reionization' (EoR) was nearly complete. During this period, the vast majority of atoms in the Universe underwent the same transition: from neutral to ionized (for reviews, see e.g. Robertson et al. 2010; Mesinger 2016; Dayal & Ferrara 2018). While this approximate picture stands on relatively firm grounds – largely owing to the measurement of the Thompson optical depth by cosmic microwave background (CMB) experiments (Hinshaw et al. 2013; Planck Collaboration 2020) – the details are yet unclear. This is because the study of the EoR faces a number of challenges. On the observational side, because this period is so far away, thus, major obstacles do exist, for instance, for current and future 21cm experiments.

But also on the theoretical side, the study of the EoR is fairly challenging. Not only is it, as mentioned above, a baryonic process and consequentially rather messy – but it also spans a wide range of scales: the ionizing photons have to be produced (which can involve elaborate stellar modelling; e.g. Eldridge & Stanway 2016), then escape their birth cloud (e.g. Kakiichi & Gronke 2021; Kimm et al. 2019), and eventually the galaxy as well as its immediate surroundings (e.g. Dove & Shull 1994; Paardekooper, Khochfar & Dalla Vecchia 2015; Lewis et al. 2020); each of these steps requires in principle knowledge of sub-parsec baryonic physics. Afterwards, the ionizing radiation might traverse ≳Mpc sized already ionized

★ E-mail: maxbg@jhu.edu
† Hubble fellow

regions in order to affect neutral atoms far away from the source. Thus, requiring the consideration of truly cosmological scales.

In recent years, however, progress has been made on both frontiers. The observational constraints of the EoR are improving continuously. Apart from the previously mentioned integrated CMB constraints, quasar (Fan et al. 2006; Becker, Bolton & Lidz 2015; Bosman et al. 2018; Davies et al. 2018), and gamma-ray bursts (Chornock et al. 2013) spectra yield tighter, and tighter constraints. Measurements of galactic emission help to constrain the evolution of the ionized fraction better – mainly thanks to the Lyman-α (Lyα) line (for a review, see Dijkstra 2014). The change in the Lyα equivalent width distribution, and the change in the clustering statistics of Lyα emitters are nowadays commonly used as astrophysical EoR constraints (Furlanetto, Zaldarriaga & Hernquist 2006; McQuinn et al. 2007; Kakiichi et al. 2016; Mason et al. 2018a; Ouchi et al. 2018).

This progress is built upon arduous observational work. While for decades, the detection of extra-galactic Lyα radiation was only conjectured (Partridge & Peebles 1967; Koo & Kron 1980) later hundreds of Lyα emitting galaxies (LAEs) were detected (e.g. Møller & Warren 1993; Hu, Cowie & McMahon 1998; Rhoads et al. 2000). Now this field is pushed towards higher redshift, and a continuously increasing amount and quality of Lyα data is available to the community (e.g. Fontana et al. 2010; Jung et al. 2018; Sobral et al. 2018; Hoag et al. 2019).

One particularly interesting development driven by new instruments and telescopes is the availability of high-resolution high-signal-to-noise Lyα spectra at $z \gtrsim 5$ – unthinkable even just a decade ago. Thanks to this progress, it is now possible to study Lyα spectral properties even at these highest redshifts which lead to the somewhat surprising detection of *blue* Lyα peaks at $z \gtrsim 6$ (Hu et al. 2016; Matthee et al. 2018; Songaila et al. 2018; Bosman et al. 2020). Surprising, because prior to that it was commonly assumed (e.g. in models constraining the progress of the EoR) that all flux blueward of the Lyα line is absorbed at this high-z due to the high neutral fraction of the IGM. Studying the detectability of blue peaks is one of the main focus of this work.

Not just on the observational front but also computationally the EoR community is making steady progress due to the ever increasing power of supercomputers. Several groups manage now to run hydrodynamical cosmological volume boxes with radiative transfer to follow the evolution of the ionizing regions closely (e.g. on the fly in Gnedin 2014; Rosdahl et al. 2018, or in post-processing in Kulkarni et al. 2019). This allows the community now to predict the evolution of observables as well as to estimate the scatter around these trends. One recent example of such simulations is the CoDaII (Ocvirk et al. 2020) simulation which uses the RAMSES-CUDATON code (Ocvirk et al. 2016).

In this work, we use the CoDaII simulation in order to study the evolution of Lyα observables at $6 \lesssim z \lesssim 7$. This simulation features full coupled ionizing radiative transfer as well as a representative volume to study the Lyα observables at hand. In particular, we focus on the disappearance of Lyman-α emitting galaxies (Lyman-α emitters; LAEs) at $z \gtrsim 6$ due to the increasing neutral IGM, and how this intergalactic absorption affects the Lyα line shape. In contrast to earlier studies, we focus in particular on the detectability of blue peaks at these redshifts – motivated by the recent observations mentioned above.

This paper is structured as follows: in Section 2 we lay out our method, that is, we provide technical background of the CoDaII simulation (Section 2.1), and explain how we generate the transmission curves (Section 2.2). In Section 3, we present our results, which we discuss in some context in Section 4, before we conclude in Section 5.

2 METHOD

2.1 The CoDaII simulation

The CoDaII simulation is a fully coupled radiation-hydrodynamical simulation of galaxy formation during the EoR. It was performed on Titan, at Oak Ridge National Laboratory, with RAMSES-CUDATON (Ocvirk et al. 2016). It is fully presented in Ocvirk et al. (2020) and therefore we recall here only the main points relevant to this study. The simulation describes the evolution of a box of $\sim$94.5 comoving Mpc (cMpc) on a side, i.e. large enough to model global reionization, from $z = 150$ down to $z = 5.8$ (see, however, Iliev et al. 2014, for a discussion of the required boxsize in order to capture EoR fluctuations). The simulation grid is 4096^3, allowing us to resolve haloes down to $10^8 M_\odot$ (with a dark matter particle resolution of $\sim 4 \times 10^5 M_\odot$) and therefore providing a good sampling of the various halo masses contributing to cosmic reionization, as shown in Lewis et al. (2020). Another advantage of the CoDaII simulation is that it contains a statistically representative population of massive haloes, up to $10^{12} M_\odot$ at $z = 6$, owing to its very large volume. Since the mass of dark matter haloes in which objects with blue Lyα peak mentioned above reside is unknown, it is crucial that our methodology allows us to investigate such a broad range of masses.

Also, CoDaII reproduces a number of observables of the high redshift Universe, in particular its reionization history (i.e. the evolution of $x_{\rm HII}$ with redshift), and the UV luminosity functions at $z = 6$ and above, which are particularly important for the investigations presented in this paper. Finally, the spatial resolution (better than 3.3 kpc physical) is well suited to computing Lyα transmission spectra. However, such a cell size may seem to allow resolving self-shielded systems only marginally. Indeed, Rahmati & Schaye (2018) showed that such systems are typically 1–10 pkpc at $z = 6-7$. It is therefore possible that we are missing a contribution to opacity from those self-shielded systems below our resolution limit. While we cannot offer a clear way of quantifying such potential missing opacity from the simulation, we can try to gain some insight into this aspect from the resolution study of Ocvirk et al. (2019), fig. 7 and table 1. The latter quantifies the increase of residual neutral fraction after overlap in simulations with twice and four times the spatial resolution of CoDaII, i.e. with 8 and 64 times higher mass resolution. The most resolved simulation in that study has 0.92 pkpc cell size, i.e. marginally resolving the minimum size quoted by Rahmati & Schaye (2018). If self-shielded systems are very sensitive to resolution and hold a large amount of the global HI, we should see a significant increase in the residual $x_{\rm HI}$ when increasing resolution. Instead, the increase found is rather limited (about a factor of 2), despite increasing the mass resolution by a factor of 64. Following this argument, the amount of HI to account for unresolved self-shielded systems in CoDaII could also be an order unity effect. Generally, combining large volumes with spatial resolution high enough to capture self-shielded systems remains a challenge, and thus also a limitation of our methodology. Cosmic Dawn III will provide interesting insight into this aspect, with a factor of two increase in resolution and an improved calibration (Lewis et al., in preparation).

CoDaII performs explicit radiative transfer of ionizing radiation from one cell to another, across the simulation volume. However, the amount of ionizing radiation released by each star particle into the cell in which it forms was assumed to be reduced from the intrinsic

photon luminosity of its stars by the bound–free Lyman continuum opacity of the unresolved, subgrid-scale interstellar birth-cloud of the stars. To account for this extra opacity, we adopted a fixed birth-cloud escape fraction of $f_{\rm LyC} = 0.42$ which the intrinsic stellar luminosity of each star was reduced when assigning an ionizing photon luminosity to the stars in each star particle. This value was tuned to reproduce a range of observables of the global ionization history, although it somewhat underpredicts the neutral fraction at the tail end of the EoR (see Ocvirk et al. 2020, for a full discussion). Notably, CoDaII does use the full speed of light in its ionizing radiative transfer routines, and, thus, does not suffer from problems stemming from the usage of the 'reduced speed of light approximation' (in combination with the M1 closure relation; Deparis et al. 2019; Ocvirk et al. 2019).

For the generation of the UV magnitudes cited ($M_{1600\rm AB}$) the BPASS $Z = 10^{-3}$ binary population and spectral synthesis model was used (Eldridge et al. 2017), assuming no dust extinction.

2.2 Generation of transmission curves

We analysed the simulation snapshots at $z \sim \{5.8,\ 6,\ 6.26,\ 6.55,\ 7\}$ (with average neutral fractions of $\log_{10}(\langle x_{\rm HI}\rangle) \sim \{-5.15,\ -4.92,\ -1.41,\ -0.64,\ -0.30\}$). At each of these snapshots, we randomly selected 50 haloes for each 0.5 magnitude bin with $M_{\rm UV} < -18$. For each halo, we drew random lines of sight for which we generated transmission curves around the Lyα wavelength. In particular,

(i) we cut out a spherical region $<1.5R_{200}$ (i.e. set all the cells within that radius to be fully ionized) around the halo position (as given by the halo finder), where R_{200} is the radius in which the average density is 200 times the mean matter density at that redshift (as defined in Ocvirk et al. 2020, equation 2). We chose the cutout of $1.5R_{200}$ for a number of (connected) reasons. The main reason is that the circumgalactic medium (CGM) is believed to span 1−2 virial radii around each galaxy (e.g. Tumlinson, Peeples & Werk 2017) and current cosmological simulations are not able to resolve the cold neutral gas within it (Fielding et al. 2016; Liang, Kravtsov & Agertz 2016); in fact, it has been shown that the HI content of such simulations in the CGM is non-converged (Suresh et al. 2019; Hummels et al. 2019; Peeples et al. 2019; van de Voort et al. 2019). Since Lyα radiative transfer is dependent on small-scale structure within the HI (e.g. Neufeld 1991; Gronke et al. 2017) this non-convergence as well as underresolving these structures is highly problematic when performing a full Lyα radiative transfer. We, thus, treat the radiative transfer processes within the interstellar and circumgalactic medium as a 'black box', and concentrate on the impact of the IGM. The second reason for the cutout is a more technical reason. Even if we could resolve all the HI structure within in the ISM and CGM perfectly, photons scattered within $\lesssim 1.5R_{200}$ have a non-negligible probability to scatter back into the line of sight (Laursen, Sommer-Larsen & Razoumov 2011; Jensen et al. 2014), i.e. we could not treat the scattering process as absorption.

(ii) we compute the gas mass weighted mean velocity within $<R_{200}$ which we use as systemic redshift for each halo, that is, effectively shifting all velocities with respect to this one.

(iii) we drew 100 random sightlines around each halo not intercepting any domain boundary. Along each, we calculate the transmission curve using `trident` (Hummels, Smith & Silvia 2017) taking the neutral hydrogen number density $n_{\rm HI}$, the temperature T, and the gas velocity $\mathbf{v}$ for each cell into account. We integrate from the halo position to a distance of $5000\,{\rm km\,s^{-1}}/H(z)$ away from the source which is sufficient to have converged transmission curve (see e.g. figure 1 of Mason & Gronke 2020). We tested this assumption by finding that a subset of the transmission curves were unchanged if we lowered this threshold to $\sim 2000\,{\rm km\,s^{-1}}$.

In summary, this procedure allows us to be agnostic about the intrinsic line shape emergent from the Lyα emitting galaxy.

3 RESULT

3.1 Illustrative example sightlines

Fig. 1 shows some arbitrary transmission curves originating from a randomly chosen halo (left column) and in corresponding colour in the right column the neutral hydrogen number density as a function of distance from the source. This illustrates how Lyα photons are absorbed in the IGM. At $z \sim 7$ (lower row in Fig. 1), for instance, one notices the characteristic wing of the absorbing Voigt profile at $v \gtrsim 200\,{\rm km\,s^{-1}}$. Furthermore, though, the transmission curves clearly exhibit some resonant absorption closer to $v \sim 0$. This is in spite of the fact that the edge of the ionized bubble is in this example for most sightlines at $d \sim 6{\rm cMpc} \sim 640\,{\rm km\,s^{-1}}/H(z)$, i.e. too far shifted to be responsible for the resonant absorption. However, as can be seen in the right panel, inside this ionized region, patches of $n_{\rm HI} \gtrsim 10^{-8}\,{\rm cm^{-3}}$ exist which are responsible for the resonant absorption. These patches can be fairly close to the emitting galaxy, and thus, likely infalling causing absorption even on the red side of the spectrum (Dijkstra, Lidz & Wyithe 2007; Iliev et al. 2008).

These patches with a 'large' HI number density of $n_{\rm HI} \gtrsim 10^{-8}\,{\rm cm^{-3}}$ can cause resonant absorption also at lower redshifts which can be seen in the upper panels of Fig. 1. In these examples, the patches lead to a rather noisy Lyα transmission curve on the blue side.

Apart for the neutral region, and neutral patches inside the ionized region, also the residual neutral fraction inside the ionized bubbles can cause significant absorption. We illustrate this in the lower row of Fig. 1 where we modified one sightline by setting $n_{\rm HI}(<5\,{\rm cMpc}) = 0$. The solid and dashed purple curve in Fig. 1 shows the unmodified and altered sightline, respectively. Clearly the region of wing absorption ($\gtrsim 200\,{\rm km\,s^{-1}}$) is unchanged. However, due to the residual neutral part inside this region the spectrum blueward of $\sim 200\,{\rm km\,s^{-1}}$ shows zero transmission.

That even a small neutral hydrogen number density is sufficient in order to cause significant resonant absorption can be seen by this simple estimate:

$$n_{\rm HI} \sim \frac{\tau}{d\sigma_0} \sim \tau \frac{H(z)}{\sigma_0 v_{\rm core}} \tag{1}$$

where σ_0 and $v_{\rm core}$ are the Lyα cross-section at line center and the width of the core region which take values of $\sim 5 \times 10^{-14}\,{\rm cm^2}$ and $\sim 80\,{\rm km\,s^{-1}}$ at $T \sim 10^4\,{\rm K}$, respectively (e.g. Dijkstra 2017). For $z \sim 6$ ($z \sim 7$) this yields $n_{\rm HI} \sim 2 \times 10^{-10}\tau\,{\rm cm^{-3}}$ ($\sim 4 \times 10^{-10}\tau\,{\rm cm^{-3}}$). Equation (3) in Weinberg et al. (1997) is a more precise estimate which yields similar values. In Appendix A we investigate this numerically.

Fig. 2 illustrates the three different absorption causes. It shows a slice around a halo at $z \sim 7$. Clearly, the large remaining neutral regions are visible. Furthermore, also the remaining neutral absorbers inside the ionized regions can be seen. Most importantly, though, the neutral hydrogen number density even inside the ionized regions shows a lot of structure with only small patches exhibiting $n_{\rm HI} \lesssim 10^{-9}\,{\rm cm^{-3}}$ required for a non-absorption at this redshift.

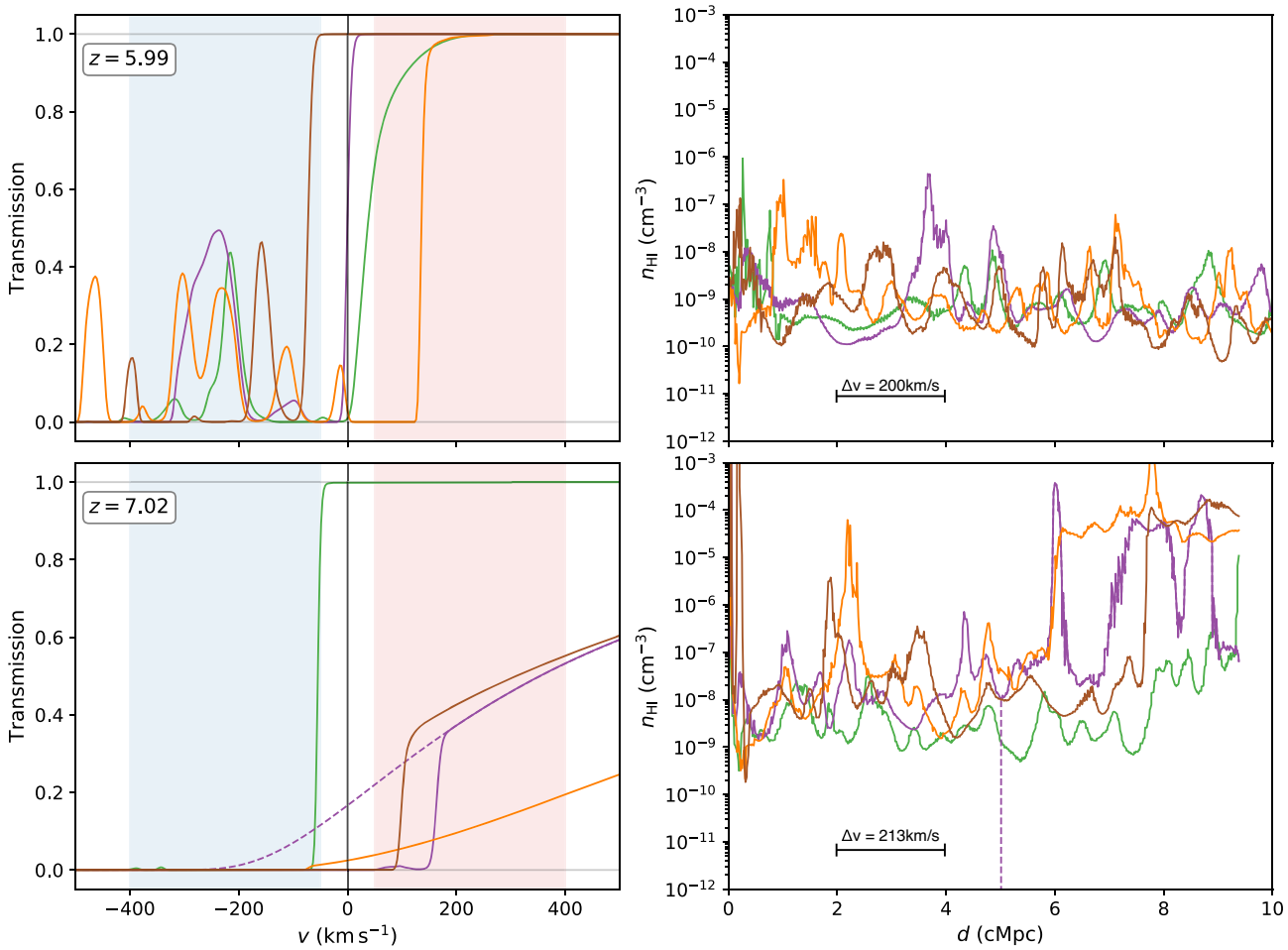

Figure 1. Examples of transmission curves (left-hand panels) and the corresponding neutral hydrogen number densities (right-hand panels) of an arbitrary halo at $z \sim 6$ (top row) and $z \sim 7$ (bottom row) with $R_{200} \sim 0.3$ cMpc ($M \sim 2 \times 10^{11}\,\mathrm{M}_\odot\,h^{-1}$, $M_{1600\mathrm{AB}} \sim -22$ at $z \sim 6$). The purple dashed lines in the lower row show an example where we set $n_{\mathrm{HI}}(< 5\,\mathrm{cMpc}) = 0$ for illustration purposes. The blue and red shaded region in the left column mark the $50 < |v|/(\mathrm{km\,s^{-1}}) < 400$ region which we use as 'blue' and 'red' side in the further analysis. In the right-hand panels, we mark the velocity corresponding to the Hubble flow of ~ 2 cMpc at that redshift.

3.2 Lyα transmission statistics

Fig. 3 shows the median and difference to the 16th and 84th percentile of the maximum transmission $T_{\mathrm{max,blue}}$ within $v \in [-400, -50]\,\mathrm{km\,s^{-1}}$ which we define as the 'blue side'. The lower panel of Fig. 3 shows the integrated transmission in this wavelength range $T_{\mathrm{int.blue}} \equiv \int T(v)\mathrm{d}v/\Delta v_{\mathrm{blue}}$. Several points are clear from Fig. 3: *(i)* at $z \gtrsim 7$ essentially all Lyα transmission emergent on the blue side of the spectrum is absorbed by the IGM in the CoDAII simulation with some rare peaks being transmitted, *(ii)* at $z \sim 6$ some transmission on the blue side is allowed, *(iii)* but even at this later times there is a strong sightline-to-sightline variation for all the haloes, and *(iv)* overall, there seems to be no clear dependence on the UV magnitude.

Fig. 4 shows the same statistics but for the red side ($v \in [50, 400]\,\mathrm{km\,s^{-1}}$) of the Ly$\alpha$ transmission curve. As expected, the overall transmission is much larger with a $T_{\mathrm{max,\,red}} \sim 1$ for $z \lesssim 6$, and even the integrated transmission reaching mostly $\gtrsim 80$ per cent of its maximum value at $z \sim 6$. Also at higher redshifts a large fraction of the redward flux is transmitted reaching $\sim 0 - 80$ per cent at $z \sim 7$. As for the blue side, we found a rather large sightline-to-sightline variation.

Note that in both Figs 3 and 4 we show only 100 randomly selected haloes for visualization purposes. Fig. 5 shows instead the full distribution for all the analysed haloes of these transmission statistics. Apart from the confirmation of the main findings stated above, the evolution in the transmission properties is clearly visible. In addition, there seems to be a slight brightness dependence of $T_{\mathrm{int.,\,red}}$ with the more massive haloes having a larger probability that their Lyα is not transmitted on the red-side. We attribute this effect to the larger cosmological infall velocity (with respect to the (Lyα) emitting regions) for these haloes which leads to an increased absorption on the red side (see Section 3.1). Though, note that Lya from these massive haloes may be emitted anyway with a larger velocity offset due to more scattering in the ISM (Shibuya et al. 2014; Stark et al. 2017; Mason et al. 2018a).

Another representation of the integrated transmission is shown in Fig. 6 – which, assuming an intrinsic Lyα equivalent width (EW) distribution and respective Lyα escape fraction – could be translated to an observed EW distribution of Lyman break galaxies. On the red side, the evolution from a bimodal distribution at $z \sim 7$ to a unimodal distribution with most sightlines yielding $T_{\mathrm{int.,\,red}} \sim 1$ at $z \sim 6$ can be observed. The integrated blue transmission shows a non-negligible

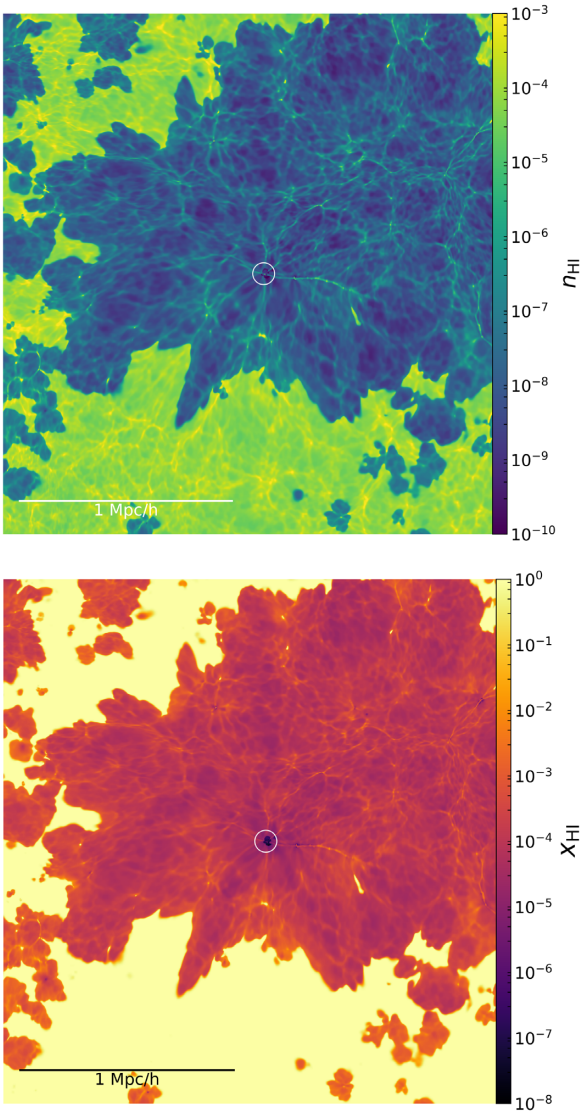

Figure 2. Slice through the simulation around one $M \sim 7 \times 10^{11}\,M_\odot$ ($M_{1600AB} \sim -21.1$) halo at $z \sim 7$ (marked with a white circled indicating $2R_{200}$). The upper and lower panel shows the neutral hydrogen number density and the neutral fraction, respectively. The plots illustrate the different ways of absorption discussed in Section 3.1: *(i)* wing absorption by the large, remaining neutral regions; *(ii)* absorption through the neutral patches; or *(iii)* resonant absorption by the residual neutral hydrogen in the otherwise ionized regions.

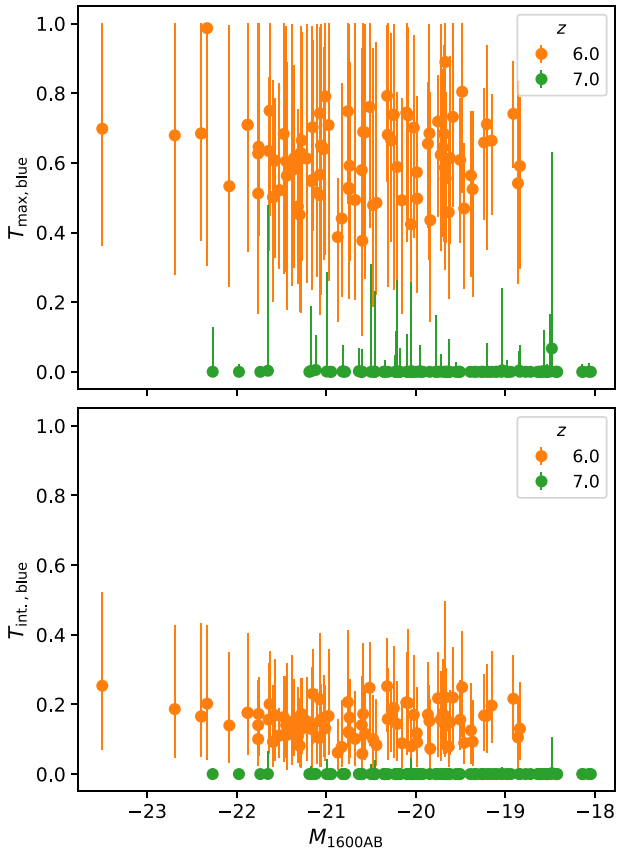

Figure 3. IGM transmission on the blue side of 100 randomly selected haloes. Each point and error bars represent the median, 16th and 84th percentile drawn from 100 sightlines for each halo.

tail with $T_{\rm int.,\,blue} > 0$ which becomes more prominent at later times. As we discuss in Section 3.3 this tail is important for observed blue peaks at high-z.

Focusing on the red side (right column) of Fig. 6, it is interesting that at earlier times ($z \sim 6.5$), the transmission statistics are more bimodal (see also Mesinger et al. 2015). I.e., the transmission is either very large or close to zero. This makes sense since at such high-z the $T(v)$ is essentially a step function due to the large HI cross-section, and the fact that $T \propto e^{-\tau}$. Specifically, one can write $T(v) \sim 0$ for $v \lesssim v_{\rm cutoff}$ where $v_{\rm cutoff}$ is commonly set by the infall velocity; as discussed in Section 3.1. This distribution might be relevant for searches of Lyα emitting galaxies which optimally should target as many sources as possible assuming a unity IGM transmission. At $z \gtrsim 7$, on the other hand, we find the integrated transmission peaks at 0.

Fig. 7 summarizes the redshift evolution of these statistics. Here, we show the percentiles of the medians for every halo as a function of redshift. We show both the maximum and normalized integrated transmission. Fig. 7 shows that the decline on the blue side is rather rapid with no flux transmitted at $\gtrsim 6.5$. The decline on the red side, on the other hand, is offset by $\Delta z \sim 0.6$ towards higher redshifts.

In Fig. 8, we show the evolution of the fraction of sightlines with a *total* integrated (total) transmission >0.2 (within $50 < |v|/({\rm km\,s^{-1}}) < 400$). Since this observed equivalent width is a product of the intrinsic equivalent width, the galactic Lyα escape fraction, and the integrated intergalactic transmission, this can be understood as the impact of the IGM on the 'Lyα emitter fraction'.[1] Between $z \sim 6.5$ and $z \sim 7$ we see a rapid decline for all halo sizes. The most UV luminous haloes show a slightly larger integrated transmission which is likely due to their larger virial radius, i.e. whether or not this signature holds will heavily depend on the CGM evolution (see the discussion Section 4.2 on that matter) – but could also be due to them residing in more ionized regions (Mason et al. 2018b).

[1]The 'Lyα emitter fraction' is defined as the fraction of Lyman break galaxies with an Lyα EW above some threshold – usually 20 Å. Our choice of 0.2 as a cutoff is motivated by this, and a 'common' intrinsic Lyα EW of $\sim$ 100 Å(Schaerer 2003).

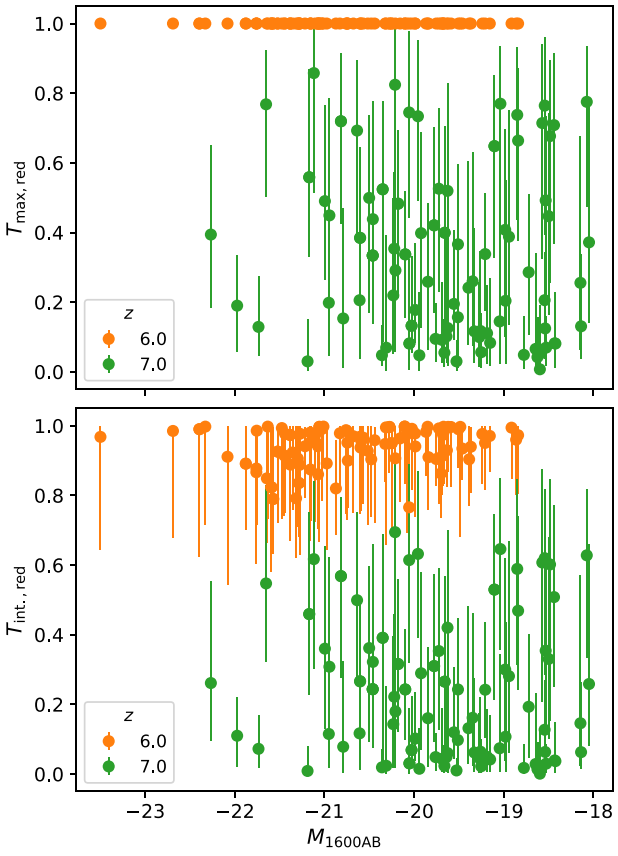

Figure 4. IGM transmission on the red side of the 100 randomly selected haloes. Each point and error bar s represents the median, 16th and 84th percentile drawn from 100 sightlines for each halo.

3.3 Comparison to observed blue peaks at high-z

Recently, several Lyα spectra at $z \gtrsim 6$ with a prominent blue peak have been observed (Songaila et al. 2018). Specifically, 'Aerith B' (Bosman et al. 2020), 'COLA1' (Hu et al. 2016; Matthee et al. 2018), and 'NEPLA4' (Songaila et al. 2018) are well-studied examples which we want to compare to our findings.

'COLA1' is a luminous Lyα emitter at $z = 6.59$ located in the well-studied COSMOS field. Besides being exceptionally bright in Lyα ($L_{\rm Ly\alpha} = 4 \times 10^{43}$ erg s^{-1}) it appears reasonably bright in the UV continuum with $M_{\rm UV} = -21.6 \pm 0.3$. Contrarily to other galaxies with similar UV luminosity, COLA1 appears particularly compact (Matthee et al. 2018).

'NEPLA4' is a Lyα emitter at $z \sim 6.54$ with a line shape resembling that of 'COLA1'. Since its location is in the currently less-studied NEP field the rest-frame UV magnitude is, unfortunately, unknown.

Lastly, 'Aerith B' is a bright ($M_{\rm AB} = -21.0^{+0.3}_{-0.2}$) Lyman-break galaxy at $z \approx 5.79$ displaying strong Lyα emission ($EW_{\rm rest} = 64^{+39}_{-24}$Å). Aerith B is situated within the ionization cone of a neighbourhing bright quasar at a distance $d \sim 750$ proper kpc, with an estimated resulting UV intensity at its location of $J_{21} = 406 \pm 40$: a factor $\sim$100 higher than the cosmic peak of the UVB. Unlike COLA1 and NEPLA4, the velocity separation between the red and blue peak of the Lyα line is large in Aerith B ($\Delta v = (580 \pm 80)$ km s^{-1}), indicating the galaxy is unlikely a significant Lyman continuum leaker (Dijkstra, Gronke & Venkatesan 2016; Izotov et al. 2016).

Fig. 9 shows the spectra of these three objects (specifically from Songaila et al. 2018; Matthee et al. 2018; Bosman et al. 2020, from left to right) alongside with transmission curves of 'similar' haloes, i.e. within ± 0.25 UV magnitude of the $M_{\rm UV} \approx -21.6$ and $M_{\rm UV} \approx -21.0$ 'COLA1' and 'Aerith B', respectively, have been associated with. In Fig. 9 we show the median transmission curve as well as the 16th and 84th percentile of the distribution stemming of 100 sightlines for the 25 [184] haloes falling in the right $M_{\rm UV}$ range for 'COLA1' ['Aerith B']. It is clear that finding an object such as 'COLA1' or 'Nepla4' in the CoDaII simulation is extremely unlikely. On the other hand, while a Lyα spectrum such as shown by 'Aerith B' is still far from common, it seems at least not entirely unlikely from Fig. 9. However, the median transmission curve can be misleading, and one should directly compare the impact of the IGM on the Lyα spectral properties (for a detailed discussion of this effect, see Byrohl & Gronke 2020).

In Figs 10 and 11 we show the distribution of all the integrated flux on the blue side $F_{\rm int.,\,blue}$ and the maximum width of a transmitted blue peak $W_{\rm blue,\,max.}(T > 0.5)$ (defined as an uninterrupted transmission $T > 0.5$), respectively. Specifically, we show the inverted cumulative distribution functions of the sightlines originating from similar haloes as 'COLA1', 'Aerith B', and 'NEPLA4'. Since, as described above, for the latter the UV magnitude is unknown, we use our full distribution at that redshift.

Figs 10 and 11 quantify the statement of 'extremely unlikely' made above better. Given the blue peak height observed and making the – quite optimistic – assumption of an intrinsically symmetric spectrum, one requires $T \gtrsim 0.5$ on the blue so in order to explain the observations. Fig. 10 shows that for such haloes this occurs in $\sim 1 - 10$ per cent of all the sightlines. Maybe more revealing is Fig. 11 where we show the distribution of the maximum blue peak width transmitted at least 50 per cent by the IGM. In this figure, we also show as vertical lines the approximate blue peak width of the three observed objects studied. Due to the wide peak of 'Aerith B' the probability of finding an 'Aerith B' like object in CoDaII is ~ 1 per cent. Inversely, since 'NEPLA4' shows a blue peak width of only ~ 80 km s^{-1}, the probability of transmitting such a blue peak at this redshift is ~ 10 per cent. Note, again, however, that due to the lack of UV information we show the full $W_{\rm blue,\,max.}(T > 0.5)$ distribution at $z \sim 6$. We also caution that firstly the assumption made of the intrinsic spectrum might be unlikely itself since most Lyα emitters at lower redshift possess a mostly Lyα spectrum asymmetric towards the red side (e.g. Yang et al. 2016; Erb et al. 2014), and secondly the systemic redshift of the sources is unknown yielding the possibility of two red peaks being detected. We discuss these caveats and the likelihood of observing more blue peaks (or two red peaks) at high-z in Section 4.3.

Unlike COLA1 and NEPLA4, Aerith B is known a priori to be located in a very biased environment – an ionized quasar proximity zone – which is not captured by CoDaII due to the non-inclusion of quasars as ionizing sources. It is then interesting that the simulation predicts LAEs similar to Aerith B have a non-zero probability of being detected 'in the field' at $z = 5.8$. This may be related to the fact that CoDa II could overpredict J21 after overlap by a factor $\sim$10, and similarly underpredicts the neutral fraction by the same factor (see Ocvirk et al. 2020, figure 3). Despite observations of LAEs being more extensive at $z \sim 6$ than at $z \sim 6.5$, no other such objects are currently known.

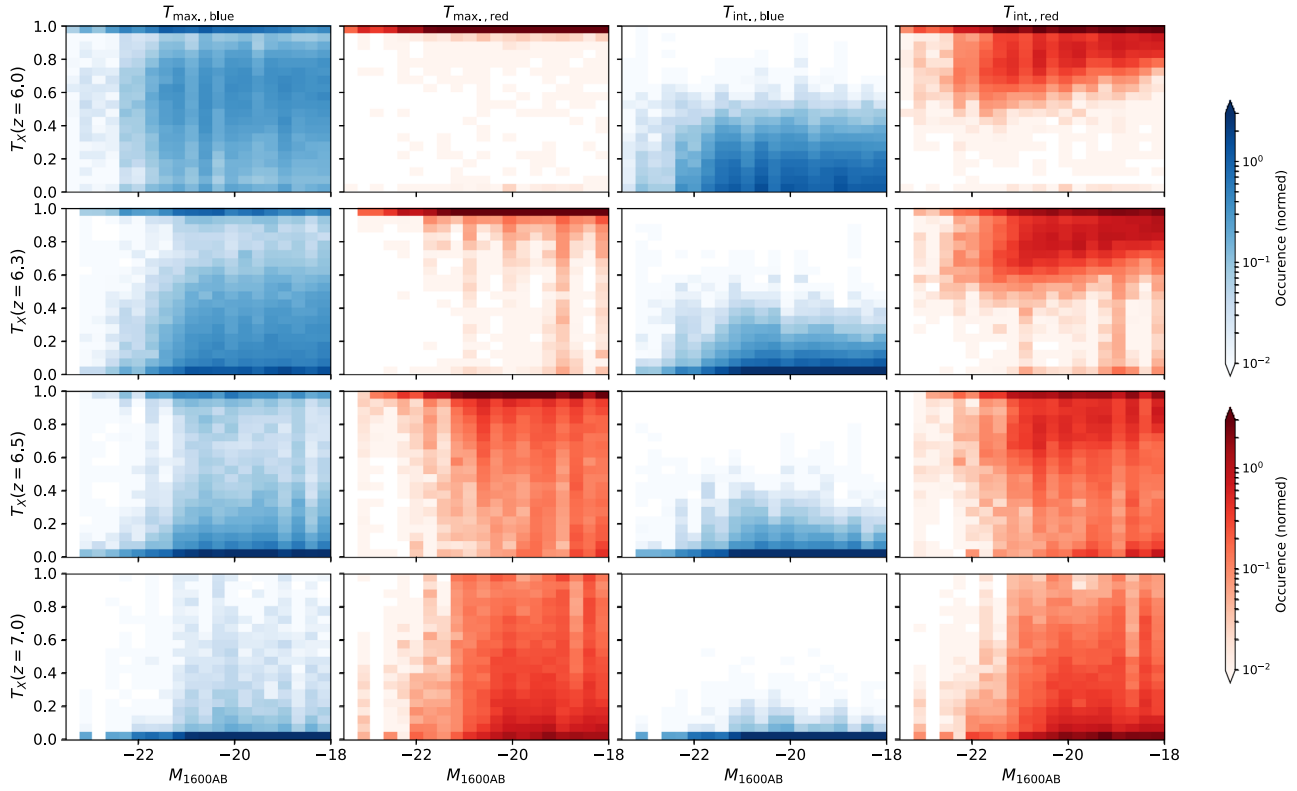

Figure 5. Distributions of the transmission properties. Shown are the maximum transmission on the blue and red sides (first and second column, respectively), and the normalized integrated transmission on both sides (third and forth column). Each row represents a redshift as denoted on the left side. For representation purposes, each 2D histogram is normalized and the colour coding is logarithmically scaled.

4 DISCUSSION

4.1 Implications of core-scatterings in an ionized IGM

As we showed in Section 3, absorption in the core of the line due to neutral hydrogen atoms in the ionized regions of the IGM is responsible for a large fraction of the absorbed flux. They lead to essentially zero transmission on the blue side, and – due to cosmological infall – can also heavily affect the red side. This means that there is no simple relation anymore between the size of the ionized region, and the absorption pattern, and has, thus, several implications for the study of the EoR, which we want to discuss here.

The simple picture of the size of the ionized bubbles, and the cut-off of the Lyα line linked to the edge of the bubble often put forward in the literature is not valid anymore if there is a significant ($n_{\rm HI} \gtrsim 10^{-10}\,{\rm cm}^{-3}$) neutral hydrogen component present. This is the case in the CoDaII simulation at $z \gtrsim 6.5$.[2] Instead, the visibility is mainly set by the (size of the) highly ionized regions around Lyα emitting sources with $n_{\rm HI} \lesssim 10^{-10}\,{\rm cm}^{-3}$, and the kinematics of the gas surrounding these galaxies. Due to the lack of quasars or other high-energy sources in CoDaII, the presence of these regions is mainly set by the redshift. This can be seen by the rapid development of the transmission on the blue side (cf. Section 3.2).

[2]The abundance of sinks of ionizing photons such as Lyman limit systems and the boxsize do affect these low neutral fractions (Iliev et al. 2014).

This affects all the Lyα observables commonly used to constrain the EoR, that is, the Lyα equivalent width distribution, LAE fraction, Lyα luminosity function, and the clustering statistics of LAEs.

4.2 Caveat: the omission of galactic and circumgalactic radiative transfer

The main caveat of this work is the focus on the transmission of the intergalactic medium, i.e. leaving out the Lyα radiative transfer processes in the ISM and CGM. Evolution of galactic and circumgalactic properties can change the Lyα escape fraction, and thus lower the observed Lyα luminosity. In particular, a lower Lyα escape fraction might mimic a lower IGM transmission, that is, a higher IGM neutral fraction / redshift.

Since Lyα photons are primarily destroyed by dust, two components can theoretically lower the Lyα escape fraction: *(a)* a larger dust content, and *(b)* a longer path-length of Lyα photons (through this dusty medium). While on an average larger dust content towards higher redshift seems unlikely, the latter option might be feasible. For instance, a lower ionization background leads to a larger HI column density in the CGM, thus increasing the path-length of Lyα photons. This scenario was discussed in Sadoun, Zheng & Miralda-Escudé (2017), who manage to reproduce the observed drop in the Lyα emitter fraction (e.g. Stark, Ellis & Ouchi 2011; Treu et al. 2013; Schenker et al. 2014).

However, while this scenario lowers the Lyα escape fraction, and consequently, the observed equivalent widths in agreement with observations, it also changes the other Lyα observables. In particular, a larger optical depth / path-length leads to an increased frequency

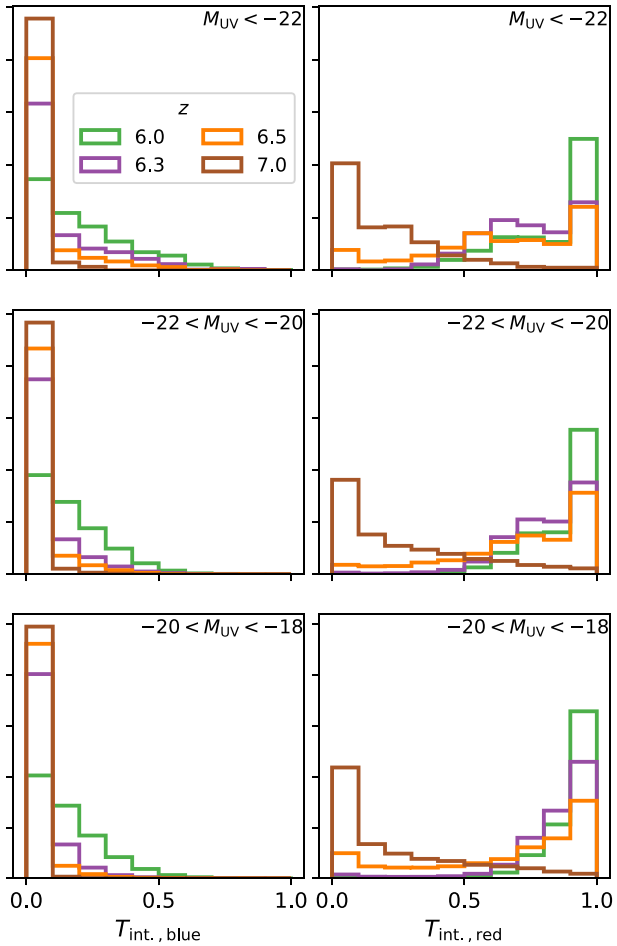

Figure 6. Normalized sightline distribution of integrated transmission split by UV magnitude and red / blue side. The colour coding corresponds to different redshifts.

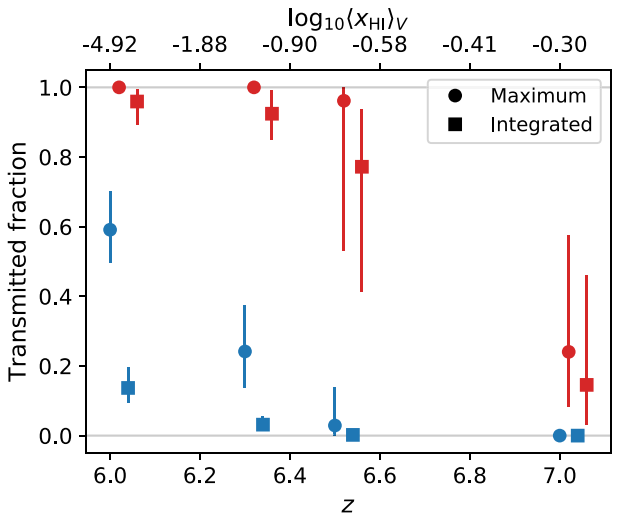

Figure 7. Evolution of transmitted maximum and integrated flux (circles and squares, respectively) on the red and blue side (in the corresponding colour) of all the selected haloes. Each point and error bars show the 16th, 50th and 84th percentile of the medians, i.e. represent the scatter between the sightlines. The secondary x-axis on top of the plot shows the volume averaged neutral fraction of the CoDaII simulation at the respective redshifts.

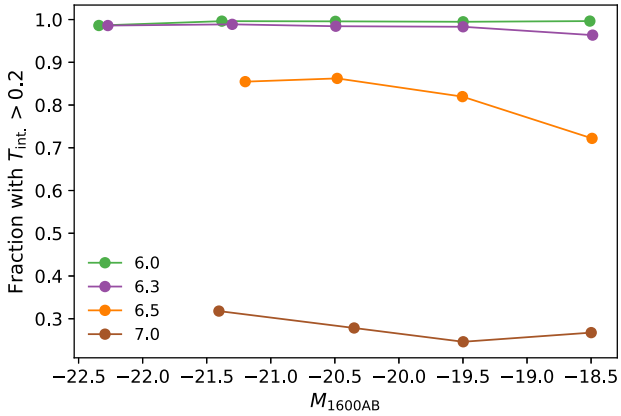

Figure 8. Evolution of fraction of haloes with total $T_{int.} > 0.2$ (i.e. $|v| < 400\,\mathrm{km\,s^{-1}}$). Shown are only the M_{UV} bins with at least 10 haloes.

diffusion, and hence, wider Lyα lines (e.g. Neufeld 1990). This seems in tension with observed Lyα spectra at high-z (Hu et al. 2010; Ouchi et al. 2010; Pentericci et al. 2018; Matthee et al. 2020), but further study is required to come to a firm conclusion.

While one solution to the caveat might seem the inclusion of the ISM & CGM into the radiative transfer calculations, we deliberately chose not to do so in this study. As preluded in Section 2.2, this is mainly due to three reasons: *(i)* Lyα radiative transfer is sensitive to sub-parsec structure inside the neutral hydrogen, such as its clumpiness (e.g. Neufeld 1991; Gronke et al. 2017), *(ii)* this structure exists (e.g. in the CGM Rauch, Sargent & Barlow 1999; Lan & Fukugita 2017) and *(iii)* CoDaII (as similar simulations) cannot resolve these scales in the ISM or CGM. In summary, while the ISM or CGM do have an effect on the Lyα line we do not resolve the relevant scales here, so we only consider the IGM effects. In fact, the inability to reproduce Lyα spectra of radiative transfer simulations using high-resolution galactic simulations might be due to this issue (see discussion in Gronke et al. 2017). Instead, we chose to follow a 'Russian doll' approach to tackle the multiscale problem of modelling Lyα observables, and focus on the intergalactic part in this study.

4.3 *Allez les bleus:* the curious cases of blue Lyα peaks at high-z

In Section 3.3 we compared our findings to observations of blue Lyα peaks at $z \gtrsim 6$. Specifically, we analysed the observed spectra of 'Aerith B' (Bosman et al. 2020), 'COLA1' (Hu et al. 2016; Matthee et al. 2018), and 'NEPLA4' (Songaila et al. 2018; cf. Fig. 9). Other authors did claim a detection of a blue Lyα peak at high redshift, such as the $\sim 450\,\mathrm{km\,s^{-1}}$ blueshifted peak at $z \sim 9.1$ described in Hashimoto et al. (2018). However, due to a lack of a (dominant) red peak, this is less certainly Lyα compared to the three objects we focused on here. Similarly, Songaila et al. (2018) present more 'complex' Lyα profiles with a blue wing – but not necessarily a clear blue peak. Better spectral resolution and sensitivity will help to clearly show which of these are blended peaks.

In the analysis presented in Section 3.3, we concluded that an IGM transmission required in order to observe these cases occurs in only a few per cent of the sightlines in CoDaII (mostly where motions allow for a blueward transmission). The crucial assumption leading to this number is the intrinsic symmetry of the Lyα spectrum. For lower redshift Lyα emitting galaxies, we know that in fact most Lyα lines are asymmetric towards the red side. At $z \sim 0.3$ high-resolution

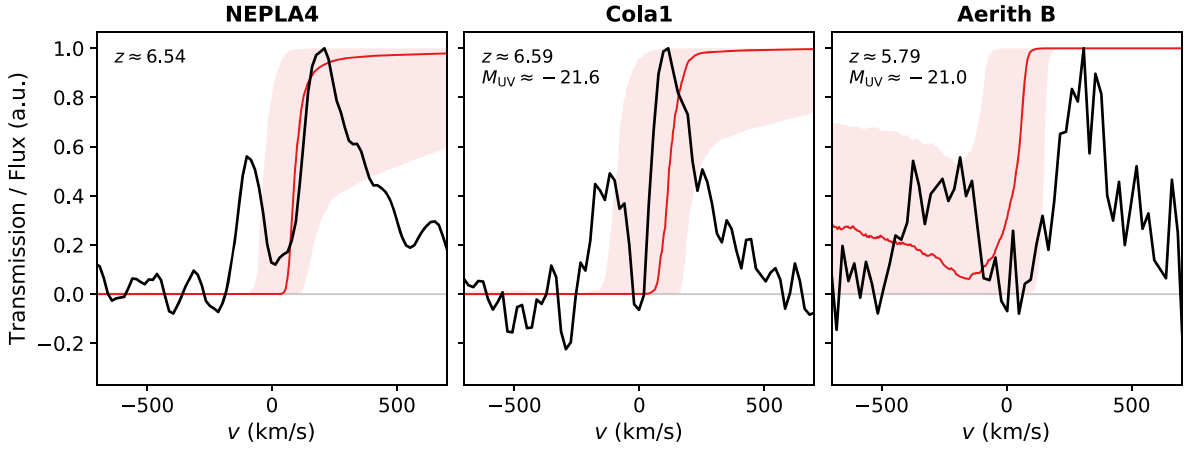

Figure 9. Lyα spectra of 'NEPLA4', 'COLA1', and 'Aerith B' (from left to right panel) with the median (and 16th to 84th percentiles as shaded region) transmission curves of similar haloes in CoDaII. Specifically, the central [right] panel shows sightlines (100 per halo, maximum of 200 haloes) originating from haloes within ±0.25 magnitude of the observed value at $z = 6.55$ [$z = 5.8$].

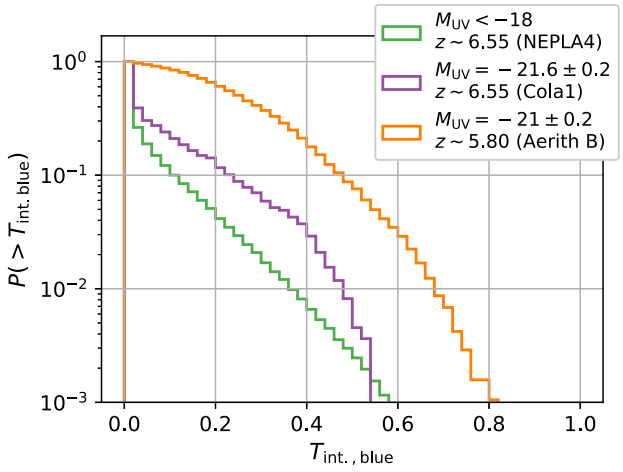

Figure 10. Inverse cumulative distribution function of the integrated flux on the blue side for haloes similar to 'COLA1', 'Aerith B', and 'EPLA4'.

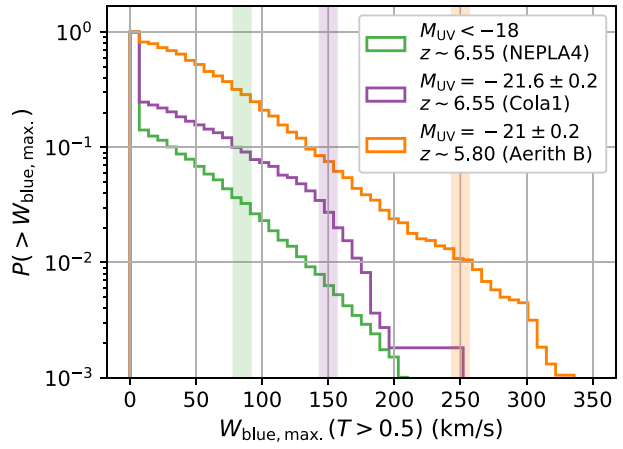

Figure 11. Inverse cumulative distribution function of the maximum width of a transmitted peak on the blue side.

spectroscopy of high-redshift analogues suggests that less than half of the flux is emerging on the blue side.[3] At higher redshifts, lower spectral resolution makes clear distinction in a statistically significant sample harder. Erb et al. (2014) found in their sample of 36 LAEs and 122 LBGs at $z \sim 2-3$ that only 6 possess a dominant blue side. Trainor et al. (2015) later quantified this number to be ∼ 10 per cent in their extended sample of 350 LAEs and 65 LBGs at the same redshift. They also measured in ∼ 45 per cent of the galaxies any blue peak with both fractions being same for their LAE and LBG sample. Similarly, Herenz et al. (2017) report that ∼ 35 per cent of the 237 from the MUSE Wide survey show a blue peak, and Yamada et al. (2012) find that approximately half their sample of 91 Lyα emitting galaxies at $z \sim 3.1$ possess a double peak.

Following this evolution, at higher redshift this ratio might become tilted even more towards the blue. In fact due to a larger gas infall at these early epochs some Lyα radiative transfer studies using cosmological simulations find spectra with all or most of the flux emerging on the blue side (e.g. Laursen, Razoumov & Sommer-Larsen 2009; Zheng et al. 2010). However, since simulations like these fail thus far to reproduce the Lyα spectral properties at lower redshift mentioned above (see comparison by Gronke et al. 2018), it is yet unclear whether this holds. Also, other observational quantities such as the large Lyα equivalent widths of LAEs at $z \gtrsim 6$ are hard to reconcile with the picture that a majority of the flux (on the blue) is absorbed by the IGM (Matthee et al. 2017). Nevertheless, the degeneracy between radiative transfer processes on galactic scales, and the IGM do exist (also see discussion in Section 4.2). Recently, Byrohl & Gronke (2020) have suggested a probe to break this degeneracy by the detection of multiple blue peaks, which might be feasible in the near future.

In conclusion, it is most likely that our assumption of a symmetric intrinsic spectrum is an oversimplification, and in reality the red peak in $z \gtrsim 6$ galaxies is dominant. This would decrease the number of observable blue peaks in CoDaII even further. For instance, if the

[3]For instance, the 12 'Green peas' (Henry et al. 2015; Yang et al. 2016) with Lyα spectrum taken show in ∼ half the galaxies a clear blue peak with always a dominant red size, and the 13 'LARS' galaxies (Östlin et al. 2014; Hayes et al. 2014; Rivera-Thorsen et al. 2015) with Lyα emission suggest have only ∼ one clearly detectable blue peak – which has still more flux on the red side.

average asymmetries of Lyα emitters with a blue peak of $\gtrsim 2/3$ of the flux being emitted on the red side (approximately consistent with the lower z studies mentioned above) holds, we would require $T_{\rm int.,\,blue}$ close to unity to explain the observations.

In this study, we compared out findings to three individual galaxies with significant blue flux. Due to a lack of high-resolution Lyα spectra it is too early too say how common such objects are; however, Songaila et al. (2018) state that 'roughly a quarter (two out of eight) of the LAEs have complex profiles with apparent blue wings' (note, however, that observations carried out using the 'Hyper-suprime cam' on the Subaru telescope suggest a lower blue peak fraction at high-z, although, no thorough analysis has been performed; Shibuya et al. 2014; Harikane et al. 2019). While this number is almost an order of magnitude larger than what we find here, there might be other reasons behind some of the detections.

As mentioned in Matthee et al. (2018), it is also thinkable that the observations of a double peak at high-z is not a red and a blue peak but instead two red peaks. This could be due to either two intrinsically emitted red peaks, or a wide red peak plus an absorption feature. While double backscattering causes nominally two red peaks,[4] the separation of these peaks is too small to yield two distinctive peaks; instead, they are blended causing merely an additional 'bump' towards the red (as e.g. in Matthee et al. 2020). In addition, in such a scenario the peak towards the blue is expected to be stronger.

However, the second case, i.e. the wide emission plus an absorption, might be more feasible and is in fact the case discussed extensively in Matthee et al. (2018). While at lower redshift a similar spectrum has not been observed[5] the increasing number density of absorbing systems towards higher redshift (Songaila & Cowie 2010; Noterdaeme et al. 2012; Crighton et al. 2015) might account for this. In Fig. 12 (as squares and dashed lines), we show the likelihood of such an event occurring leading to the observation of a 'blue' peak. Specifically, we demand that the red (blue) side is at least 150 km s^{-1} (100 km s^{-1}) wide with a transmission of $T(v) > 0.7$ (>0.3) and separated by at least 30 km s^{-1} and a maximum of 300 km s^{-1}.[6] Clearly, given the number of parameters involved in such a scenario, this is merely an example, and we chose the parameters to be conservative, i.e. demanding more realistic (wider) peaks or a larger peak separation will lower the number of 'fake' blue peaks. However, Fig. 12 shows that even with this choice of parameters such an event occurring is quite unlikely at all redshifts considered, especially towards higher z. Nevertheless, for $z \lesssim 6$ the probability is non-negligible highlighting the importance of systemic redshift measurements which can distinguish between these scenarios (see discussion in Matthee et al. 2018).

Fig. 12 also shows the probability of observing a 'real' blue peak, i.e. at $v < 0$ with the same parameters as above. As discussed before,

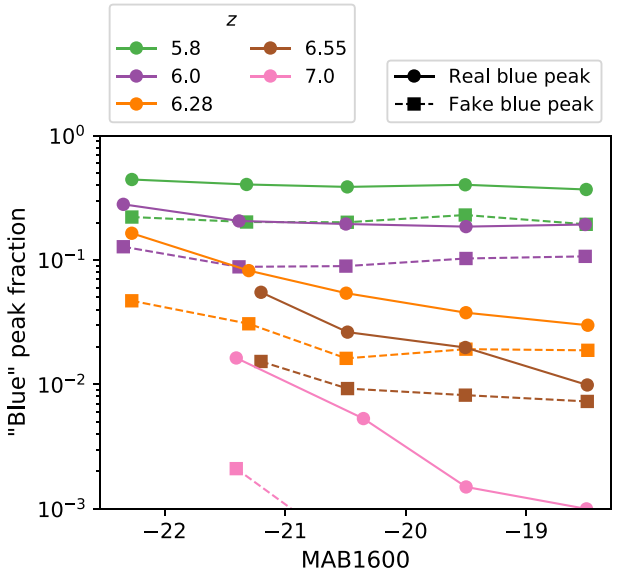

Figure 12. Fraction of blue peaks detectable. The *circles* and *solid* lines show the fraction of 'real' blue peaks, i.e. in the range $v \in [-50,\ -400]$ km s^{-1}, the *squares* connected by *dashed* lines show 'fake' blue peaks, i.e. when an intervening absorber could have caused a double peaked detection – even with two red peaks. See Section 4.3 for details.

at $z \gtrsim 6.5$ the likelihood is in the per cent-level. Noteworthy is the dependence on UV magnitude with a higher probability to see a blue peak in larger haloes. After the study of several individual skewers, we attribute this to the fact that a larger surrounding is cut-out around these objects due to a larger virial radius; implying whether or not this effect is real is linked to the CGM problematic discussed in Section 4.2. Furthermore, we note that in all the cases where a blue peak would be visible, the transmission curve is still a step function as discussed previously but with $v_{\rm cutoff} < 0$ due to relative motions between the emitting regions and the gas at larger radii. Whether this is realistic, e.g. due to large-scale outflows, or infalling galaxies into a potential well, depends highly on *(i)* what the emitting regions are, and *(ii)* how much radiative transfer processes act altering the surface brightness. While we do not address this in detail in this study (see Section 4.2), this is an interesting direction for future studies.

4.4 Comparison to previous work

Previous work on the impact of the IGM on the Lyα line shape can be grouped into three categories:

(i) (Semi-)Analytical work which uses a simplified model of the IGM to compute the transmission spectrum. Notable examples of this category are, e.g. Santos (2004) or Dijkstra et al. (2007), which found the importance of cosmological infall leading to absorption on the red part of the spectrum. They also highlight the impact of the size of the ionized region around Lyα emitters – which they usually assume to be fully ionized.

(ii) Work which extends on the first category by not assuming certain 'bubble sizes' but using the input of seminumerical simulations based on an excursion set method such as 21cmFAST (Mesinger, Furlanetto & Cen 2011) in order to map halo masses and redshift to bubble sizes. This approach is very powerful as it is relatively fast, and can, thus, be used to map observations of Lyα observables such as the EW distribution to different reionization histories and eventually constrain the global ionization fraction as, e.g. done in

[4]This is due to the fact that in a 'backscattering' event a Lyα photon's frequency is boosted by $\sim 2v$ where v is the bulk velocity of the scattering medium.

[5]Detections of indisputably triple peaked spectra are very rare. The 'Sunburst Arc' (Rivera-Thorsen et al. 2019) and 'Ion2' (Vanzella et al. 2019) are two recent examples from $z \sim 2-3$ (also see Vanzella et al. 2018; Izotov et al. 2018, for two more examples at different z). Both exhibit a peak at line-center, though, and thus are of different nature than the double peaked detections at $z \gtrsim 6$ but the at least spectrum of 'Ion 2' could likely be altered by intervening absorbers.

[6]Note that the separation here is not exactly between the peaks but the distance in velocity space between the points where the transmission falls below and raises above the 0.7 and 0.3 thresholds, respectively.

Mason et al. (2018a); Mason et al. (2019b, 2019a). One can combine dark-matter only or hydrodynamic cosmological simulations with ionizing radiative transfer in post-processing or using the semi-analytical ionization techniques (as recently done in, e.g. Weinberger et al. 2018). Importantly, in these later, seminumerical models the ionized regions are set to a constant ionized fraction (usually set by photoionization equilibrium Meiksin 2009).

(iii) There has also been an approach which simulated the EoR by post-processing large cosmological *N*-body simulations (e.g. McQuinn et al. 2007; Iliev et al. 2008; Jensen et al. 2013) or hydrodynamical simulations (e.g. Laursen et al. 2011; Laursen et al. 2019). Commonly, ionizing luminosities are assigned to each *N*-body halo and ray-traced across the density field of the intergalactic gas outside haloes. Iliev et al. (2008), e.g. used this approach to analyse the effect of intergalactic Lyα transmission on LAE observations in a simulation box more than $\sim$100 Mpc on a side, confirming the effect of intergalactic infall surrounding massive galaxies mentioned above. They also found that HII regions surrounding bright LAEs were filled with lower mass haloes clustered around the central galaxy which were also important ionization sources. They note that these 'proximity zones' can lead to transmission on the blue side by $z \lesssim 7$.

(iv) Lastly, there is the approach using a full radiation-hydrodynamics simulation as input, as done in this study (also see e.g. recent work by Garel et al. 2021; Park et al. 2021). Naturally, other studies employed this approach before us (e.g. Gnedin & Prada 2004) using different simulations as input. An advantage of these studies is a more realistic ionization morphology on small-scales affected by hydrodynamical backreaction, and better resolved fluctuations inside the ionized regions. A disadvantage of these earlier studies is that they come from much smaller simulation domains, too small to model the large-scale patchiness of reionization realistically, and with resolution too limited to resolve the smaller mass haloes that can contribute significantly to reionization, even in the neighbourhood of bright LAEs.

Overall, the findings of these previous studies are fairly consistent (e.g. in pointing out the large sightline-to-sightline variation, or the effect of infall) with differences owing to the specific choice of parameters. Previous work did not, to our knowledge, discuss the effect the IGM transmission on the blue side has on the observable Lyα line properties and, thus the associated importance of the residual neutral fraction – and its fluctuations – of the ionized regions in observing objects like COLA1 – but instead focused more on the global observables such as the LAE clustering, Lyα luminosity function, and EW distributions, and the effect the intrinsic line shape has on these statistical measures (see, in particular, Jensen et al. 2013, who used Gaussian as well as double peaked profiles as intrinsic spectrum).

Clearly, all of the approaches have their own advantages and disadvantages and bring progress in different ways. For instance, recently Mason & Gronke (2020) analysed the importance of the residual neutral fraction, and its impact on the blue side of the Lyα spectrum, in a simplified model [category *(i)* or *(ii)* above] and demonstrate under which conditions a blue peak is observable.

As we argue in Section 4.3, the occurrence of blue peaks at high-z is an interesting new way of testing simulations of the EoR against observations. Current full radiation-hydrodynamics simulations with focus on reionization, include large boxes – to the sacrifice of spatial resolution – such as CoDaII, and smaller ($\lesssim$ 10cMpc) boxes focusing on a selected number of haloes with higher spatial resolution (e.g. Rosdahl et al. 2018; Wu et al. 2019). In principle, all these simulations can be tested against Lyα observables, and in particular the occurrence of blue peaks. We see, however, three main obstacles which we want to caution against: *(i)* although some of the simulations have better resolution, as already discussed above, this is thus far still not enough to achieve convergence in HI (circum-)galactic properties, and hence to resolve structures likely relevant for full Lyα radiative transfer, *(ii)* naturally, in order to compare with observations of galaxies residing in more massive haloes such as 'Cola1' (i.e. $M_{\rm UV} \lesssim -21$ at $z \gtrsim 6$), a statistical relevant sample of such haloes is required setting a minimum boxsize, and *(iii)* as we found the Lyα transmission (in particular on the blue side) is sensitive to the fiducial neutral fraction which, hence, needs to be captured correctly by simulations in order to use this observable as a probe of the EoR. This raises a potential issue with the commonly used 'reduced speed of light approximation,' which is a numerical ingenuity to decrease the computational cost of radiative transfer in simulations but affecting the residual neutral fraction after overlap, as shown in Ocvirk et al. (2019).

In this wide landscape of theoretical realizations, CoDaII 's main advantages are its size and the use of the full speed of light. What is striking, though, is that CoDaII is too transparent as compared to the Fan et al. (2006) measurements (as shown and discussed in Ocvirk et al. 2020), but still not transparent enough to yield a large abundance of blue-peaked LAEs. Reproducing both of these aspects has never been done, and seems difficult, as we can judge from this study. It will certainly be an interesting challenge for future numerical simulations of the EoR.

5 CONCLUSION

We analysed the Lyα transmission properties of the CoDaII simulation which is a modern cosmological radiation-hydrodynamics simulation. Our findings can be summarized as follows:

(i) The transmission of blue Lyα flux rapidly declines with increasing redshift, due to residual neutral gas inside ionized bubbles which can lead to complete absorption for $n_{\rm HI} \gtrsim 10^{-9}$ cm^{-3} at $z \gtrsim 7$,

(ii) There is large sightline variation in blue flux transmission, but no clear $M_{\rm UV}$ dependence. This is mainly due to kinematic effects, i.e. outflows and/or a relative motion of the emitting galaxy to the surrounding IGM gas. Whether this holds in reality depends strongly on the circumgalactic gas, which we ignore in our analysis, as discussed in Section 4.2.

(iii) The transmission on the red side is greater than the blue side, but also has high sightline variation, in particular at $z \gtrsim 6.5$ when the transmission can vary from zero to unity for a given galaxy.

(iv) Te observed prevalence of blue peaks can provide an additional test for reionization simulations, but better observational statistic are required in order to do so. In CoDaII, we find for $M_{1600\rm AB} \sim -21$ galaxies, the opacity of the IGM allows the transmission of blue peaks through $\sim$ 20 per cent ($\sim$ 1 per cent) of lines of sight at $z \sim 6$ ($z \sim 7$).

ACKNOWLEDGEMENTS

The authors thank the referee for constructive feedback that improved the outcome of this study. We are grateful to Antoinette Songaila Cowie for sharing the 'NEPLA4' spectrum with us. This research has made use of NASA's Astrophysics Data System, and many open source projects such as trident (Hummels et al. 2017),

`IPython` (Pérez & Granger 2007), `SciPy` (Virtanen et al. 2020), `NumPy` (Walt, Colbert & Varoquaux 2011), `matplotlib` (Hunter 2007), `pandas` (McKinney 2010), and the `yt-project` (Turk et al. 2011). MG was supported by NASA through the NASA Hubble Fellowship grant *HST*-HF2-51409 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. MG acknowledges support from NASA grants *HST*-GO-15643.017, and *HST*-AR-15797.001 as well as XSEDE grant TG-AST180036. CAM acknowledges support by NASA Headquarters through the NASA Hubble Fellowship grant *HST*-HF2-51413.001-A. PRS was supported in part by U.S. NSF grant AST-1009799, NASA grant NNX11AE09G, and supercomputer resources from NSF XSEDE grant TG-AST090005 and the Texas Advanced Computing Center (TACC) at The University of Texas at Austin. JM acknowledges a Zwicky Prize Fellowship from ETH Zurich. GY acknowledges financial support by MICIU/FEDER under project grant PGC2018-094975-C21. SEIB acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 669253). ITI was supported by the Science and Technology Facilities Council [grants ST/I000976/1, ST/F002858/1, ST/P000525/1, and ST/T000473/1]; and The Southeast Physics Network (SEPNet). KA was supported by NRF-2016R1D1A1B04935414 and NRF-2016R1A5A1013277. KA also appreciates APCTP for its hospitality during completion of this work. PO acknowledges support from the French ANR funded project ORAGE (ANR-14-CE33-0016). ND and DA acknowledge funding from the French ANR for project ANR-12-JS05- 0001 (EMMA). The CoDa II simulation was performed at Oak Ridge National Laboratory/Oak Ridge Leadership Computing Facility on the Titan supercomputer (INCITE 2016 award AST031). Processing was performed on the Eos and Rhea clusters. Resolution study simulations were performed on Piz Daint at the Swiss National Supercomputing Center (PRACE Tier 0 award, project id pr37). The authors would like to acknowledge the High Performance Computing center of the University of Strasbourg for supporting this work by providing scientific support and access to computing resources. Part of the computing resources were funded by the Equipex EquipMeso project (Programme Investissements d'Avenir) and the CPER Alsacalcul/Big Data.

DATA AVAILABILITY

Data related to this work will be shared on reasonable request to the corresponding author.

REFERENCES

Becker G. D., Bolton J. S., Lidz A., 2015, PASA, 32, e045
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055
Bosman S. E. I., Kakiichi K., Meyer R. A., Gronke M., Laporte N., Ellis R. S., 2020, ApJ, 896, 49
Byrohl C., Gronke M., 2020, A&A, 642, L16
Chornock R., Berger E., Fox D. B., Lunnan R., Drout M. R., Fong W.-f., Laskar T., Roth K. C., 2013, ApJ, 774, 26
Crighton N. H. M. et al., 2015, MNRAS, 452, 217
Davies F. B. et al., 2018, ApJ, 864, 142
Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1
Deparis N., Aubert D., Ocvirk P., Chardin J., Lewis J., 2019, A&A, 622, A142
Dijkstra M., 2014, PASA, 31, e040
Dijkstra M., 2017, preprint (arXiv:1704.03416)
Dijkstra M., Lidz A., Wyithe J. S. B., 2007, MNRAS, 377, 1175
Dijkstra M., Gronke M., Venkatesan A., 2016, ApJ, 828, 71
Dove J. B., Shull J. M., 1994, ApJ, 423, 196
Eldridge J. J., Stanway E. R., 2016, MNRAS, 462, 3302
Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, PASA, 34, e058
Erb D. K. et al., 2014, ApJ, 795, 33
Fan X. et al., 2006, AJ, 132, 117
Fielding D., Quataert E., Mccourt M., Thompson T. A., 2016, MNRAS, 466, 3810
Fontana A. et al., 2010, ApJ, 725, L205
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2006, MNRAS, 365, 1012
Garel T., Blaizot J., Rosdahl J., Michel-Dansac L., Haehnelt M. G., Katz H., Kimm T., Verhamme A., 2021, MNRAS, 504, 1902
Gnedin N. Y., 2014, ApJ, 793, 29
Gnedin N. Y., Prada F., 2004, ApJ, 608, L77
Gronke M., Dijkstra M., McCourt M., Peng Oh S., 2017, A&A, 607, A71
Gronke M., Girichidis P., Naab T., Walch S., 2018, ApJ, 862, L7
Harikane Y. et al., 2019, ApJ, 883, 142
Hashimoto T. et al., 2018, Nature, 557, 392
Hayes M. et al., 2014, ApJ, 782, 6
Henry A., Scarlata C., Martin C. L., Erb D., 2015, ApJ, 809, 19
Herenz E. C. et al., 2017, A&A, 606, A12
Hinshaw G. et al., 2013, ApJS, 208, 19
Hoag A. et al., 2019, ApJ, 878, 12
Hummels C. B. et al., 2019, ApJ, 882, 156
Hummels C. B., Smith B. D., Silvia D. W., 2017, ApJ, 847, 59
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Hu E. M., Cowie L. L., McMahon R. G., 1998, ApJ, 502, L99
Hu E. M., Cowie L. L., Barger A. J., Capak P., Kakazu Y., Trouille L., 2010, ApJ, 725, 394
Hu E. M., Cowie L. L., Songaila A., Barger A. J., Rosenwasser B., Wold I. G. B., 2016, ApJ, 825, L7
Iliev I. T., Shapiro P. R., McDonald P., Mellema G., Pen U.-L., 2008, MNRAS, 391, 63
Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725
Izotov Y. I., Schaerer D., Thuan T. X., Worseck G., Guseva N. G., Orlitová I., Verhamme A., 2016, MNRAS, 461, 3683
Izotov Y. I., Worseck G., Schaerer D., Guseva N. G., Thuan T. X., Fricke Verhamme A., Orlitová I., 2018, MNRAS, 478, 4851
Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2013, MNRAS, 428, 1366
Jensen H., Hayes M., Iliev I., Laursen P., Mellema G., Zackrisson E., 2014, MNRAS, 15, 15
Jung I. et al., 2018, ApJ, 864, 103
Kakiichi K., Gronke M., 2021, ApJ, 908, 30
Kakiichi K., Dijkstra M., Ciardi B., Graziani L., 2016, MNRAS, 463, 4019
Kimm T., Blaizot J., Garel T., Michel-Dansac L., Katz H., Rosdahl J., Verhamme A., Haehnelt M., 2019, MNRAS, 486, 2215
Koo D. C., Kron R. T., 1980, PASP, 92, 537
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Lan T.-W., Fukugita M., 2017, ApJ, 850, 156
Laursen P., Razoumov A. O., Sommer-Larsen J., 2009, ApJ, 696, 853
Laursen P., Sommer-Larsen J., Razoumov A. O., 2011, ApJ, 728, 52
Laursen P., Sommer-Larsen J., Milvang-Jensen B., Fynbo J. P. U., Razoumov A. O., 2019, A&A, 627, A84
Lewis J. S. W. et al., 2020, MNRAS, 496, 4342
Liang C. J., Kravtsov A. V., Agertz O., 2016, MNRAS, 458, 1164
Mason C. A. et al., 2018b, ApJ, 857, L11
Mason C. A. et al., 2019a, MNRAS, 485, 3947
Mason C. A., Gronke M., 2020, MNRAS, 499, 1395
Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018a, ApJ, 856, 2
Mason C. A., Naidu R. P., Tacchella S., Leja J., 2019b, MNRAS, 489, 2669
Matthee J., Sobral D., Darvish B., Santos S., Mobasher B., Paulino-Afonso A., Röttgering H., Alegre L., 2017, MNRAS, 472, 772

Matthee J., Sobral D., Gronke M., Paulino-Afonso A., Stefanon M., Röttgering H., 2018, A&A, 619, A136
Matthee J., Sobral D., Gronke M., Pezzulli G., Cantalupo S., Röttgering H., Darvish B., Santos S., 2020, MNRAS, 492, 1778
McKinney W., 2010, in van der Walt S., Millman J., eds, Proceedings of the 9th Python in Science Conference. p. 51
McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, MNRAS, 381, 75
Meiksin A. A., 2009, Rev. Mod. Phys., 81, 1405
Mesinger A., 2016, Understanding the Epoch of Cosmic Reionization: Challenges and Progress,. Vol. 423. Springer
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, MNRAS, 446, 566
Møller P., Warren S. J., 1993, A&A, 270, 43
Neufeld D. A., 1990, ApJ, 350, 216
Neufeld D. A., 1991, ApJ, 370, L85
Noterdaeme P. et al., 2012, A&A, 547, L1
Ocvirk P. et al., 2016, MNRAS, 463, 1462
Ocvirk P. et al., 2020, MNRAS, 496, 4
Ocvirk P., Aubert D., Chardin J., Deparis N., Lewis J., 2019, A&A, 626, A77
Ouchi M. et al., 2010, ApJ, 723, 869
Ouchi M. et al., 2018, PASJ, 70, S13
Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544
Park H. et al., 2021, preprint (arXiv:2105.10770)
Partridge R. B., Peebles P. J. E., 1967, ApJ, 148, 377
Peeples M. S. et al., 2019, ApJ, 873, 129
Pentericci L. et al., 2018, A&A, 619, A147
Pérez F., Granger B. E., 2007, Computing in Science and Engineering, 9, 21
Planck Collaboration et al., 2020, A&A, 641, A6
Rahmati A., Schaye J., 2018, MNRAS, 478, 5123
Östlin G. et al., 2014, ApJ, 797, 11
Rauch M., Sargent W. L. W., Barlow T. A., 1999, ApJ, 515, 500
Rhoads J. E., Malhotra S., Dey A., Stern D., Spinrad H., Jannuzi B. T., 2000, ApJ, 545, L85
Rivera-Thorsen T. E. et al., 2015, ApJ, 805, 14
Rivera-Thorsen T. E. et al., 2019, Science, 366, 738
Robertson B. E., Ellis R. S., Dunlop J. S., McLure R. J., Stark D. P., 2010, Nature, 468, 49
Rosdahl J. et al., 2018, MNRAS, 479, 994
Sadoun R., Zheng Z., Miralda-Escudé J., 2017, ApJ, 839, 44
Santos M. R., 2004, MNRAS, 349, 1137
Schaerer D., 2003, A&A, 397, 527
Schenker M. A., Ellis R. S., Konidaris N. P., Stark D. P., 2014, ApJ, 795, 20
Shibuya T., Ouchi M., Nakajima K., Yuma S., Hashimoto T., Shimasaku K., Mori M., Umemura M., 2014, ApJ, 785, 64
Sobral D., Santos S., Matthee J., Paulino-Afonso A., Ribeiro B., Calhau J., Khostovan A. A., 2018, MNRAS, 476, 4725
Songaila A., Cowie L. L., 2010, ApJ, 721, 1448
Songaila A., Hu E. M., Barger A. J., Cowie L. L., Hasinger G., Rosenwasser B., Waters C., 2018, ApJ, 859, 91
Stark D. P. et al., 2017, MNRAS, 464, 469
Stark D. P., Ellis R. S., Ouchi M., 2011, ApJ, 728, L2
Suresh J., Nelson D., Genel S., Rubin K., Hernquist L., 2018, MNRAS, 483, 3
Trainor R. F., Steidel C. C., Strom A. L., Rudie G. C., 2015, ApJ, 809, 89
Treu T., Schmidt K. B., Trenti M., Bradley L. D., Stiavelli M., 2013, ApJ, 775, L29
Tumlinson J., Peeples M. S., Werk J. K., 2017, ARA&A, 55, 389
Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, ApJS, 192, 9
Vanzella E. et al., 2018, MNRAS, 476, L15
Vanzella E. et al., 2019, MNRAS, 1103, 1093
van de Voort F., Springel V., Mandelker N., van den Bosch F. C., Pakmor R., 2019, MNRAS, 482, L85
Virtanen P. et al., 2020, Nature Methods, 17, 261
Walt S. v. d., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Weinberger L. H., Kulkarni G., Haehnelt M. G., Choudhury T. R., Puchwein E., 2018, MNRAS, 479, 2564
Weinberg D. H., Miralda-Escudé J., Hernquist L., Katz N., 1997, ApJ, 490, 564
Wu X., Kannan R., Marinacci F., Vogelsberger M., Hernquist L., 2019, MNRAS, 488, 419
Yamada T., Matsuda Y., Kousai K., Hayashino T., Morimoto N., Umemura M., 2012, ApJ, 751, 29
Yang H., Malhotra S., Gronke M., Rhoads J. E., Dijkstra M., Jaskot A., Zheng Z., Wang J., 2016, ApJ, 820, 130
Zheng Z., Cen R., Trac H., Miralda-Escudé J., 2010, ApJ, 716, 574

APPENDIX A: CHANGING THE FIDUCIAL NEUTRAL FRACTION

Fig. A1 shows transmission curves for homogeneous neutral hydrogen number densities, and no peculiar motion at $z \sim 7$ (but an unchanged temperature). The estimate of $n_{\rm HI} \sim 2 \times 10^{-10} \tau\ {\rm cm}^{-3}$ (cf. Section 3) fits this result fairly well.

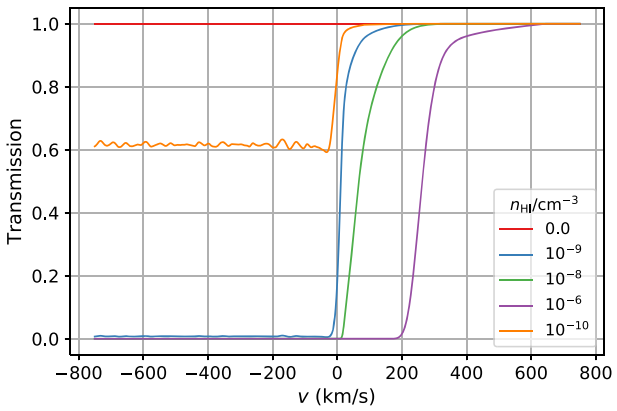

Figure A1. Transmission curves for an homogeneous IGM at $z \sim 7$.

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