Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **525,** 1664–1676 (2023) https://doi.org/10.1093/mnras/stad2376
Advance Access publication 2023 August 3

On the general nature of 21-cm-Lyman α emitter cross-correlations during reionization

Anne Hutter [1,2★] Caroline Heneka [3★] Pratika Dayal [4] Stefan Gottlöber,[5] Andrei Mesinger [6] Maxime Trebitsch [4] and Gustavo Yepes [7,8]

[1]*Cosmic Dawn Center (DAWN)*
[2]*Niels Bohr Institute, University of Copenhagen, Jagtvej 128, DK-2200 Copenhagen N, Denmark*
[3]*Institute of Theoretical Physics (ITP), Heidelberg University, Philosophenweg 16, D-69120 Heidelberg, Germany*
[4]*Kapteyn Astronomical Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands*
[5]*Leibniz-Institut für Astrophysik, An der Sternwarte 16, D-14482 Potsdam, Germany*
[6]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[7]*Departamento de Fisica Teorica, Modulo 8, Facultad de Ciencias, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[8]*CIAFF, Facultad de Ciencias, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*

Accepted 2023 July 31. Received 2023 July 21; in original form 2023 June 5

ABSTRACT

We explore how the characteristics of the cross-correlation functions between the 21-cm emission from the spin-flip transition of neutral hydrogen (H I) and early Lyman α (Ly α) radiation emitting galaxies (Ly α emitters, LAEs) depend on the reionization history and topology and the simulated volume. For this purpose, we develop an analytic expression for the 21-cm-LAE cross-correlation function and compare it to results derived from different ASTRAEUS and 21CMFAST reionization simulations covering a physically plausible range of scenarios where either low-mass ($\lesssim 10^{9.5}$ M$_\odot$) or massive ($\gtrsim 10^{9.5}$ M$_\odot$) galaxies drive reionization. Our key findings are: (i) the negative small-scale ($\lesssim$ 2 cMpc) cross-correlation amplitude scales with the intergalactic medium's (IGM) average H I fraction ($\langle\chi_{\rm H\,I}\rangle$) and spin-temperature weighted overdensity in neutral regions ($\langle(1+\delta)\rangle_{\rm H\,I}$); (ii) the inversion point of the cross-correlation function traces the peak of the size distribution of ionized regions around LAEs; (iii) the cross-correlation amplitude at small scales is sensitive to the reionization topology, with its anticorrelation or correlation decreasing the stronger the ionizing emissivity of the underlying galaxy population is correlated to the cosmic web gas distribution (i.e. the more low-mass galaxies drive reionization); (iv) the required simulation volume to not underpredict the 21-cm-LAE anticorrelation amplitude when the cross-correlation is derived via the cross-power spectrum rises as the size of ionized regions and their variance increases. Our analytic expression can serve two purposes: to test whether simulation volumes are sufficiently large, and to act as a fitting function when cross-correlating future 21-cm signal Square Kilometre Array and LAE galaxy observations.

Key words: methods: analytical – methods: numerical – galaxies: high-redshift – intergalactic medium – dark ages, reionization, first stars.

1 INTRODUCTION

Our Universe underwent the last major phase transition during its first billion years when the ultraviolet (UV) photons from the first stars and galaxies ionized the neutral hydrogen (H I) in the intergalactic medium (IGM). During this Epoch of Reionization (EoR), ionized regions grew and merged around galaxies until the IGM was ionized by $z \simeq 5.3$ (Dayal & Ferrara 2018; Keating et al. 2020; Bosman et al. 2021; Qin et al. 2021; Zhu et al. 2021). However, the exact timing of the reionization process and the topology of the ionized IGM, i.e. the evolution of the spatial distribution of ionized regions within the cosmic web structure, remain uncertain. Both go back to our limited knowledge about the properties of the first galaxies and whether the majority of H I ionizing photons emerged from the few massive galaxies in the densest regions or from the numerous low-mass galaxies that are more homogeneously distributed in the cosmic web structure.

In the past years, the rising number of observed high-redshift galaxies and precision measurements of the cosmic microwave background (CMB) have started to paint a picture wherein reionization has a midpoint around $z \simeq 7$–8 (Planck Collaboration VI 2020; Goto et al. 2021; Maity & Choudhury 2022). A robust tracer of the IGM ionization state is the presence or absence of the H I sensitive Lyman α (Ly α) emission line in detected galaxy spectra. The number density, fraction and spatial distribution of galaxies with observable Ly α emission, so-called Ly α emitters (LAEs), track the mean H I fraction ($\langle\chi_{\rm H\,I}\rangle$) in the IGM (e.g. Mesinger & Furlanetto 2008; Dayal, Maselli & Ferrara 2011; Dijkstra et al. 2014; Hutter et al. 2014; Pentericci et al. 2014, 2018; Schenker et al. 2014; Fuller et al. 2020) and spatial distribution of ionized regions (e.g. Jensen

★ E-mail: anne.hutter@nbi.ku.dk (AH); heneka@thphys.uni-heidelberg.de (CH)

et al. 2013; Mesinger et al. 2015; Castellano et al. 2016, 2018; Hutter et al. 2017, 2020; Qin et al. 2022). However, the fraction of Ly α radiation transmitted through the IGM is sensitive to the shape of the Ly α line emerging from a galaxy, which again is subject to the gas density and velocity distribution of its interstellar and circumgalactic media (e.g. Verhamme et al. 2015; Dijkstra, Gronke & Venkatesan 2016; Gronke et al. 2017; Kimm et al. 2019; Kakiichi & Gronke 2021).

Fortunately, current and forthcoming radio interferometers, such as the Square Kilometre Array[1] (SKA; Carilli & Rawlings 2004), Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017), Murchison Widefield Array[2] (MWA; Barry et al. 2019; Li et al. 2019), and Low Frequency Array[3] (LOFAR; Patil et al. 2017; Mertens et al. 2020), will detect the cosmic H I 21-cm signal that traces the topology of the ionized regions in the IGM, with SKA being expected to have sufficient angular resolution and sensitivity to provide us with real-space 21-cm maps. Statistical analyses applied to the 21-cm signal measured in reciprocal space alone, such as the 21-cm auto power spectra, will constrain our models of reionization and the underlying galaxy population driving this phase transition. However, they rely on the accurate removal of various 21-cm signal foregrounds interfering with the EoR signal (e.g. Shaver et al. 1999; Barry et al. 2016; Patil et al. 2016, 2017; Trott & Wayth 2016; Mertens, Ghosh & Koopmans 2018; Mertens et al. 2020). Theoretically, cross-correlating the 21-cm signal with galaxy surveys eases the removal of bright 21-cm foregrounds, as the only foregrounds that survive the cross-correlation are those arising from the cosmological volume of the galaxy survey,[4] confirming the reality of the cosmological 21-cm signal (Furlanetto & Lidz 2007; Beane, Villaescusa-Navarro & Lidz 2019). In practice, however, 21-cm foregrounds will inflate the variance of 21-cm-galaxy cross-correlations compared to a hypothetical 21-cm foreground-free survey. For this reason, 21-cm foreground mitigation is still desirable and its quality increases for larger survey areas (Liu & Shaw 2020). Thus, efforts have concentrated on investigating the power of cross-correlations between the 21-cm signal and galaxies (Furlanetto & Lidz 2007; Wyithe, Loeb & Schmidt 2007; Park et al. 2014). Furthermore, as the Ly α line detected in spectroscopic or narrow-band surveys allows for more precise redshift estimates of the selected galaxies than broad-band Lyman break galaxy surveys, a strong focus has been on exploring the constraining power of 21-cm-LAE cross-correlations, either in terms of cross-power spectra or cross-correlation functions (Wiersma et al. 2013; Sobacchi, Mesinger & Greig 2016; Vrbanec et al. 2016, 2020; Heneka, Cooray & Feng 2017; Hutter et al. 2017; Hutter, Trott & Dayal 2018; Kubota et al. 2018; Heneka & Mesinger 2020; Weinberger, Kulkarni & Haehnelt 2020). Indeed, for various reionization scenarios and LAE models, 21-cm-LAE cross-correlations exhibit $\langle\chi_{\rm H\,I}\rangle$-sensitive signatures, such as the cross-correlation or cross-power amplitude and the scale where the cross-power spectrum switches signs or the cross-correlation function changes its curvature. These signatures stem from the large-scale anticorrelation (correlation) between the 21-cm signal in emission (absorption) and the LAEs located in ionized regions (see e.g. Heneka & Mesinger 2020) as well as the corresponding cross-correlations tracing the size of ionized regions around LAEs. However, despite these fundamental relations, the values and signs for the small-scale cross-correlation function and power spectra differ among different works. While the change in sign reflects whether the 21-cm signal is predominantly in absorption or emission, it remains unclear whether the remaining differences are signatures of different reionization scenarios and LAE models or arise from limited simulated volumes or the chosen normalizations for the underlying 21-cm and LAE number density fluctuations. Only a thorough understanding of the 21-cm-LAE cross-correlations will allow us to tighten constraints on the reionization history and topology as well as the nature of Ly α emitting galaxies and assess which supplementary statistics and/or data might be required further.

We address this question in this paper. For this purpose, we derive the small-scale analytic limit of the 21-cm-LAE cross-correlation function and propose an analytic fitting function. We compare the analytic predictions with results from different simulations with ASTRAEUS (Hutter et al. 2023) and 21CMFAST (Mesinger, Greig & Sobacchi 2016), and analyse: What reionization characteristics (e.g. ionization history and topology) does the small-scale amplitude of the 21-cm-LAE cross-correlation function trace? Which feature in the 21-cm-LAE cross-correlation function tracks the typical size of the ionized regions? What are the effects of self-shielded regions around LAEs and limited simulation volumes?

This paper is organized as follows. In Section 2, we derive the analytic limits and model for the 21-cm-LAE cross-correlation function during reionization. We then compare the results from different ASTRAEUS and 21CMFAST reionization simulations to the analytic predictions derived in Section 2 and assess the dependence of the 21-cm-LAE cross-correlation function amplitude on the reionization topology in Section 3. In Section 4, we investigate the effects of limited simulation volumes on the 21-cm-LAE cross-correlation function and discuss the results from existing literature in Section 5. We conclude in Section 6.

2 21-CM-LAE CROSS-CORRELATIONS

The 21-cm line is emitted when a neutral hydrogen atom in its electronic ground state transitions from the triplet to the singlet hyperfine state. The spin temperature T_s describes the ratio of atoms in the triplet to singlet state. It shapes the intensity of the emitted or absorbed 21-cm radiation characterized by the brightness temperature δT_b. Importantly, we can only measure this 21-cm radiation relative to the background radiation, the CMB, with a temperature $T_{\rm CMB}$. Thus the measurable differential 21-cm brightness temperature δT_b is given by (e.g. Furlanetto, Oh & Briggs 2006)

$$\delta T_b(\mathbf{x}) = \frac{T_s(\mathbf{x}) - T_{\rm CMB}}{1+z}\left(1 - e^{-\tau(\mathbf{x})}\right) \tag{1}$$

$$\simeq \frac{3c\lambda_{21}^2 h A_{10}}{32\pi k_{\rm B} H_0} n_{H,0}\left(\frac{1+z}{\Omega_m}\right)^{1/2}\left(1 - \frac{T_{\rm CMB}}{T_{\rm s}(\mathbf{x})}\right)\chi_{\rm HI}(\mathbf{x})$$

$$= T_0\left(1 - \frac{T_{\rm CMB}}{T_{\rm s}(\mathbf{x})}\right)\chi_{\rm HI}(\mathbf{x})\left(1+\delta(\mathbf{x})\right). \tag{2}$$

Here τ describes the corresponding 21-cm optical depth, which we assume to be small in equation (2). A_{10} represents the Einstein coefficient for spontaneous emission of a photon with an energy of hc/λ_{21}, corresponding to the energy difference of the hydrogen singlet and triplet hyperfine levels. $n_{\rm H,\,0}$ is the neutral hydrogen density today and $\chi_{\rm HI}(\mathbf{x})$ and $1 + \delta(\mathbf{x}) = \rho(\mathbf{x})/\langle\rho\rangle$ describe the neutral hydrogen fraction and overdensity at position $\mathbf{x}$, respectively.

[1]Square Kilometre Array, https://www.skatelescope.com.
[2]Murchison Widefield Array, http://www.mwatelescope.org.
[3]Low Frequency Array, http://www.lofar.org.
[4]For example, low-redshift interlopers in high-redshift galaxy surveys could correlate with point sources that are part of the 21-cm foregrounds and generate false correlation signatures.

In this paper, we use the following definitions for the 21-cm signal,

$$\delta_{21}(\mathbf{x}) = \frac{\delta T_b(\mathbf{x})}{T_0} = \left(1 - \frac{T_{\rm CMB}}{T_s(\mathbf{x})}\right) \chi_{\rm HI}(\mathbf{x})\,(1+\delta(\mathbf{x}))\,, \tag{3}$$

the number density of LAEs,

$$\delta_{\rm LAE}(\mathbf{x}) = \frac{n_{\rm LAE}(\mathbf{x})}{\langle n_{\rm LAE}\rangle} - 1, \tag{4}$$

and the 21-cm cross-correlation function,

$$\xi_{21,\rm LAE}(\mathbf{r}) = \frac{1}{V}\int {\rm d}^3x\ \ \delta_{21}(\mathbf{x}+\mathbf{r})\,\delta_{\rm LAE}(\mathbf{x}), \tag{5}$$

to derive the cross-correlations between the 21-cm signal fluctuations and the LAE distribution. In the following, we will phrase our calculations under the assumption that the simulation volume V is gridded on N cells, i.e. $\frac{1}{V}\int {\rm d}^3x \rightarrow \frac{1}{N}\sum_1^N$. While our calculations remain valid in the limit of $N\rightarrow\infty$, we choose this gridding approach to better reflect the typical outputs of simulations used to compute the 21-cm-LAE cross-correlation functions.

2.1 The cross-correlation amplitude at LAE positions

To derive an analytic expression for the 21-cm-LAE cross-correlation function, we first evaluate the 21-cm-LAE cross-correlation function at the positions of LAEs,

$$\xi_{21,\rm LAE}(r=0) = \frac{1}{N}\sum_{\mathbf{x}} \delta_{21}(\mathbf{x})\,\delta_{\rm LAE}(\mathbf{x}), \tag{6}$$

From equation (4) we see that $\delta_{\rm LAE}$ adopts only positive values at LAE locations and remains negative with a value of -1 otherwise, while the 21-cm signal vanishes in ionized regions ($\chi_{\rm HI}=0$). LAEs are preferentially located in sufficiently large ionized regions with residual H I fractions up to 10^{-4}, allowing the Ly α line to redshift out of absorption and traverse the IGM. We note that sufficiently strong gas outflows from LAEs can relax this criterion, such that some LAEs could be located in neutral regions. However, as LAE surveys detect relatively bright LAEs (e.g. Ouchi et al. 2018) likely to be located in overdense and ionized regions, we assume here that $\delta_{21}=0$ at LAE locations for LAEs with $L_\alpha \gtrsim 10^{42}$ erg s^{-1}. For these assumptions, the only regions contributing to $\xi_{21,\rm LAE}(\mathbf{r}=0)$ are the neutral regions where no LAEs are found. Moreover, since the ionization fronts are sharp, most of the N cells will be either neutral or highly ionized ($\chi_{\rm HI} \lesssim 10^{-4}$). We thus consider the ionization field to be binary and neglect partially ionized cells at the ionization fronts or around galaxies with ionized regions smaller than the cell size. The 21-cm-LAE cross-correlation function values at very small scales are then given by

$$\begin{aligned}
\xi_{21,\rm LAE}(r=0) &= \frac{1}{N}\left[\sum_{\mathbf{x}_{\rm HI}} \delta_{21}(\mathbf{x})\ \underbrace{\delta_{\rm LAE}(\mathbf{x})}_{=-1} + \sum_{\mathbf{x}_{\rm HII}} \underbrace{\delta_{21}(\mathbf{x})}_{\simeq 0}\ \delta_{\rm LAE}(\mathbf{x})\right] \\
&= -\frac{1}{N}\sum_{\mathbf{x}_{\rm HI}} \left(1 - \frac{T_{\rm CMB}}{T_s(\mathbf{x})}\right) \underbrace{\chi_{\rm HI}(\mathbf{x})}_{\simeq 1}\,(1+\delta(\mathbf{x})) \\
&\simeq -\langle\chi_{\rm HI}\rangle \left\langle \left(1 - \frac{T_{\rm CMB}}{T_s}\right)\,(1+\delta)\right\rangle_{\rm HI}.
\end{aligned} \tag{7}$$

Here $\langle\rangle_{\rm HI}$ denotes the mean value across neutral regions. We note that this expression for $\xi_{21,\rm LAE}(r=0)$ represents a lower limit, as LAEs in partially or complete neutral regions will contribute positively (based on the reasonable assumption that $T_s \gg T_{\rm CMB}$ at LAE locations). As the Universe becomes ionized, the IGM is heated by the energetic photons from the first stars and galaxies, the spin temperature rises and exceeds the CMB temperature during the early phases of reionization. Assuming the post-heating regime $T_{\rm s} \gg T_{\rm CMB}$ to be valid in neutral patches, the 21-cm-LAE cross-correlation at very small scales becomes

$$\xi_{21,\rm LAE}(r=0) \simeq -\langle\chi_{\rm HI}\rangle\,\langle 1+\delta\rangle_{\rm HI}. \tag{8}$$

We note that this limit also applies for any representation of δ_{21} that solely shifts the zero-point, e.g. $\delta_{21}(\mathbf{x}) = (\delta T_b(\mathbf{x}) - \langle\delta T_b\rangle)/T_0$.

2.2 The cross-correlation amplitude profile around LAEs

Next we derive the 21-cm-LAE cross-correlation profile depending on the size distribution of the ionized bubbles around LAEs. Here we limit our calculations to the post-heating regime of the EoR. Separating the 21-cm-LAE cross-correlation functions into $N_{\rm LAE}$ pixels containing ($\delta_{\rm LAE} > -1$) and $N - N_{\rm LAE}$ pixels devoid ($\delta_{\rm LAE} = -1$) of LAEs, we yield for the 21-cm-LAE cross-correlation as a function of radial distance from an LAE

$$\begin{aligned}
\xi_{21,\rm LAE}(r) &= \frac{1}{N}\sum_{n=0}^{N} \delta_{21}(\mathbf{x}+\mathbf{r})\,\delta_{\rm LAE}(\mathbf{x}) \\
&= \frac{1}{N}\sum_{n=0}^{N-N_{\rm LAE}} -\delta_{21}(\mathbf{x}+\mathbf{r})|_{\mathbf{x}\neq\mathbf{x}_{\rm LAE}} \\
&\quad + \frac{1}{N}\sum_{n=0}^{N_{\rm LAE}} \frac{N}{N_{\rm LAE}}\delta_{21}(\mathbf{x}+\mathbf{r})|_{\mathbf{x}=\mathbf{x}_{\rm LAE}} \\
&\simeq -\langle\delta_{21}(\mathbf{x})\rangle_{\mathbf{x}} + \frac{1}{N_{\rm LAE}}\sum_{n=0}^{N_{\rm LAE}} \delta_{21}(\mathbf{x}+\mathbf{r})|_{\mathbf{x}=\mathbf{x}_{\rm LAE}} \\
&= -\langle\delta_{21}\rangle + \langle\delta_{21}\rangle^{\rm LAE}(r) \\
&\simeq -\langle\chi_{\rm HI}\rangle\langle 1+\delta\rangle_{\rm HI} + \chi_{\rm HI}^{\rm LAE}(r)\left(1+\delta^{\rm LAE}\right)(r).
\end{aligned} \tag{9}$$

Here we have assumed that pixels are either small enough to contain only one LAE or that LAEs are sparse enough that not more than one LAE is found in a pixel. $\langle\delta_{21}\rangle^{\rm LAE}$ is the average 21-cm signal profile around LAEs, while $\langle\delta_{21}\rangle$ is the average overall 21-cm signal. Correspondingly, $\chi_{\rm HI}^{\rm LAE}(r)$ and $(1+\delta^{\rm LAE})(r)$ are the average neutral fraction and density profiles around LAEs.

$$\xi_{21,\rm LAE}(r) = -\langle\chi_{\rm HI}\rangle\langle 1+\delta\rangle_{\rm HI}\left[1 - \frac{\chi_{\rm HI}^{\rm LAE}(r)}{\langle\chi_{\rm HI}\rangle}\frac{\left(1+\delta^{\rm LAE}\right)(r)}{\langle 1+\delta\rangle_{\rm HI}}\right]. \tag{10}$$

The main factor determining $\xi_{21,\rm LAE}(r)$ is the average neutral hydrogen profile around LAEs beyond the halo scale, $\chi_{\rm HI}^{\rm LAE}(r)$. While $\chi_{\rm HI}^{\rm LAE}(r)$ is determined by the sizes of the ionized regions around LAEs at small r values, it converges to the average neutral hydrogen fraction $\langle\chi_{\rm HI}\rangle$ as r increases beyond the typical sizes of the ionized regions around LAEs.

To obtain an analytic form for $\xi_{21,\rm LAE}(r)$, we assume $(1+\delta^{\rm LAE})(r) \simeq \langle 1+\delta\rangle_{\rm HI}$ and derive $\chi_{\rm HI}^{\rm LAE}(r)$ as follows: we assume the distribution of the radii of the ionized regions around LAEs to follow a lognormal distribution (McQuinn et al. 2007; Zahn et al. 2007; Meerburg, Dvorkin & Spergel 2013). With the probability density function

$${\rm PDF}(r) = \frac{1}{r\sqrt{2\pi\sigma_{\ln r}^2}}\exp\left[-\frac{\left[\ln\frac{r}{r_{\rm ion}}\right]^2}{2\sigma_{\rm ion}^2}\right], \tag{11}$$

describing the probability of the size of an ionized region around an LAE, the cumulative density function

$$\mathrm{CDF}(r) = \int_0^r \mathrm{d}r' \, \mathrm{PDF}(r') \tag{12}$$

$$= \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left[\frac{\ln \frac{r}{r_{\rm ion}}}{\sqrt{2}\sigma_{\rm ion}}\right] \tag{13}$$

describes then the average profile of the neutral hydrogen fraction around LAEs, where each ionized region containing an LAE lies in an overall neutral medium, i.e. $\langle \chi_{\rm H\,I} \rangle \simeq 1$. Inversely, $1 - \mathrm{CDF}(r)$ depicts the average profile of the ionization fraction around LAEs in an effectively neutral IGM. However, as the Universe becomes ionized, the distances between ionized regions reduces, and the probability to encounter a neutral or an ionized regions at large distances r scales with the average neutral hydrogen fraction, $\langle \chi_{\rm H\,I} \rangle$. Therefore, we approximate the average neutral hydrogen fraction profile as

$$\chi_{\rm HI}^{\rm LAE}(r) = \langle \chi_{\rm HI} \rangle \, \mathrm{CDF}(r). \tag{14}$$

As the 21-cm–LAE cross-correlation function depicts the probability of detecting a 21-cm signal of a given strength at a distance r from an LAE and thus traces the mean ionization profile around LAEs, we propose the following ansatz for $\xi_{21,\,\rm LAE}$:

$$\xi_{21,\rm LAE}(r) = -\langle \chi_{\rm HI} \rangle \, \langle 1+\delta \rangle_{\rm HI} \, [1 - \mathrm{CDF}(r)] \, . \tag{15}$$

We will show in the following that this ansatz provides an excellent fit for the numerically derived results.

We note that $\chi_{\rm HI}^{\rm LAE}(r)$ may not be entirely dominated by the size distribution of the ionized regions around LAEs. At $r \lesssim 5$ cMpc self-shielding systems can increase the neutral hydrogen fraction around LAEs, leading to $\chi_{\rm HI}^{\rm LAE}(r \lesssim 5 \text{ cMpc}) > 0$. At these distances, the average density profile around LAEs increases towards smaller distances. For example, in the ASTRAEUS simulations, it rises approximately as $(1+\delta)(r) \simeq 1 + \frac{4}{3} r^{-4/3}$. The increase in the neutral hydrogen density towards LAEs causes then the 21-cm-LAE cross-correlation function to reduce its negative amplitude towards smaller distances r [cf. equation (10) and Kubota et al. (2018) and Weinberger et al. (2020)]. The presence of this feature depends strongly on the modelling of the self-shielding systems in simulations (see e.g. appendix D in Weinberger et al. 2020), which is a complex function of the ionizing radiation and feedback processes from the stellar populations as well as the temperature and metallicity of the IGM gas.

3 RESULTS FROM SIMULATIONS

In this section we describe the different reionization scenarios and simulations that we use then to analyse the dependency of the 21-cm-LAE cross-correlation functions on reionization and its topology.

3.1 Simulations

We analyse results from two different reionization simulation frameworks, (1) the seminumerical galaxy evolution and reionization model ASTRAEUS and (2) the seminumerical 21CMFAST code, which we describe in the following. Both use cosmological parameters consistent with the results from the Planck mission; the exact values used can be found in Klypin et al. (2016) for the underlying VSMDPL simulation that ASTRAEUS uses and Mesinger et al. (2016) for the Evolution of 21-cm Structure (EOS) simulations.

3.1.1 *ASTRAEUS simulations (Hutter et al. 2023)*

ASTRAEUS[5] couples a semi-analytical galaxy evolution model (an enhanced version of DELPHI; Dayal et al. 2014, 2022; Mauerhofer & Dayal 2023) to a seminumerical reionization scheme (CIFOG; Hutter 2018) and runs on the outputs of a dark-matter (DM) only N-body simulation (merger trees and density fields). It includes not only models for all key processes of galaxy evolution thought to be relevant during the EoR, such as gas accretion, mergers, star formation, supernovae feedback, metal and dust enrichment, radiative feedback from reionization, but also follows the spatially inhomogeneous reionization process accounting for recombinations and tracking the residual H I fraction in ionized regions [see Hutter et al. (2021, 2023) and Ucci et al. (2023) for modelling details].

The ASTRAEUS simulations exploited for this analysis are based on the high-resolution VERY SMALL MULTIDARK PLANCK (VSMDPL) DM-only simulation from the MULTIDARK simulation project and has been run with the GADGET-2 Tree+PM N-body code (Springel 2005). The VSMDPL simulation follows the trajectories of 3840^3 DM particles in a box with a side length of 160 h^{-1} comoving Mpc (cMpc), and each DM particle has a mass of $6.2 \times 10^6 h^{-1}$ $\mathrm{M}_\odot$. Halos and subhalos down to 20 particles or a minimum halo mass of 1.24×10^8 h^{-1} $\mathrm{M}_\odot$ have been identified with the phase space ROCKSTAR halo finder (Behroozi, Wechsler & Wu 2013a) for all 150 snapshots ranging from $z = 25$ to $z = 0$. To generate the necessary input files for ASTRAEUS, we have used the pipeline internal CUTNRESORT scheme to cut and resort the vertical merger trees for $z = 0$ galaxies [sorted on a tree-branch by tree-branch basis within a tree and generated by CONSISTENT TREES; Behroozi et al. (2013b)] to local horizontal merger trees (sorted on a redshift-by-redshift basis within a tree) for galaxies at $z = 4.5$. Moreover, for all snapshots at $z \geq 4.5$, we have mapped the DM particles onto 2048^3 grids and re-sampling these to 512^3 grids to generate the DM density fields with cells with a side length of 312.5 h^{-1} ckpc.

ASTRAEUS has recently been extended to also include a model for LAEs (see Hutter et al. 2023), where the latter are defined as all galaxies exceeding a Ly α luminosity of $L_\alpha \geq 10^{42}$ erg s^{-1}. This model describes the Ly α line profile emerging from a galaxy as a function of the ISM gas and dust distribution as well as the escape fraction of H I ionizing photons ($f_{\rm esc}$), and follows the line-of-sight dependent Ly α attenuation by the H I in the IGM during reionization. In Hutter et al. (2023), we explored three different Ly α line profile models and underlying relations between $f_{\rm esc}$ and halo mass. Here we will consider the two physically plausible bracketing reionization scenarios of $f_{\rm esc}$ increasing (MHINC) and decreasing (MHDEC) with rising halo mass, and the *Gaussian* Ly α line profile model also used in Hutter et al. (2014) and Dayal et al. (2011). We note that these ASTRAEUS simulations reproduce all available observational star-forming galaxy data sets at $z = 5$–10, such as the UV luminosity functions, stellar mass functions, star formation rate, and stellar mass densities. Moreover, the $f_{\rm esc}$ relations are normalized such that they reproduce the constraints on the reionization history from gamma-ray burst (GRB) optical afterglow spectrum analyses, quasar sightlines, Ly α luminosity functions, LAE clustering, and fraction as well as the CMB optical depth from Planck Collaboration VI (2020).

For this combination of Ly α line profile model and reionization scenarios, we compute the 21-cm-LAE cross-correlation functions following the approach outlined in Hutter et al. (2017, 2018). In brief, we derive the 21-cm signal fields from the simulated ionization

[5] https://github.com/annehutter/astraeus

$[\chi_{\rm H\,I}(\mathbf{x})]$ and density grids $[1+\delta(\mathbf{x})]$ by applying equation (2), assuming $T_x(\mathbf{x}) >> T_{\rm CMB}$ and

$$T_0 = 28.5{\rm mK}\left(\frac{1+z}{10}\right)^{1/2}\frac{\Omega_b}{0.042}\frac{h}{0.073}\left(\frac{\Omega_m}{0.24}\right)^{-1/2} \quad (16)$$

to each grid cell. We then obtain the dimensionless 21-cm-LAE cross-correlation function as

$$\xi_{21,{\rm LAE}}(r) = \int P_{21,{\rm LAE}}(k)\,\frac{\sin(kr)}{kr}\,4\pi k^2\,{\rm d}k. \quad (17)$$

The cross-power spectrum $P_{21,{\rm LAE}}(k) = V\langle\widetilde{\delta}_{21}(\mathbf{k})\,\widetilde{\delta}_{\rm LAE}(-\mathbf{k})\rangle$ is in units of cMpc^3 for a volume V and derived from the product of the Fourier transformation of the fractional fluctuation fields $\delta_{21}(\mathbf{x})$ and $\delta_{\rm LAE}(\mathbf{x})$ as defined in equations (3) and (4).[6] We note that the 21-cm-LAE cross-correlation results for the *Clumpy* and *Porous* Ly α line profile models also explored in Hutter et al. (2023) are identical to those of the *Gaussian* model, as the galaxies identified as observable LAEs, i.e. after accounting for the attenuation by the IGM, are effectively the same. While the *Gaussian* Ly α line profile model describes the Ly α line emerging from galaxies as a Gaussian centred around the Ly α resonance, the *Clumpy* and *Porous* models consider the gas and dust to be clumpy and in case of the *Porous* model also dispersed with gas-free channels, resulting in double-peak profiles with varying emission at the Ly α resonance depending on the assumed clump size and $f_{\rm esc}$.

3.1.2 EOS 21CMFAST simulations (Mesinger et al. 2016)

21CMFAST combines the excursion-set formalism and perturbation theory to follow the evolving density, velocity, ionization, and spin temperature fields. The EOS project[7] provides public 21-cm simulations of the EoR of 1.6 Gpc box length, computed on a 1024^3 grid using 21CMFASTV2 (Sobacchi & Mesinger 2014), and cell sizes of $\sim$1 h^{-1} Mpc. The EOS simulations include a sub-grid prescription for inhomogeneous recombinations, photo-heating suppression of the gas fraction in small halos, and a calibration of the X-ray emissivity of galaxies with high-mass X-ray binary observations in local star forming galaxies (Mineo, Gilfanov & Sunyaev 2012). Also, the Lyman series radiation background is self-consistently computed, determining how closely the spin temperature tracks the kinetic gas temperature through the Wouthuysen–Field effect (Wouthuysen 1952; Field 1958). The EOS simulations explore two models for the EoR morphology: (1) the faint galaxy model characterized by many small ionized H II regions (SmallH II), and (2) the bright galaxy model of fewer, larger H II regions (LargeH II). These two models are based on different star-formation scenarios, corresponding to efficient star formation in either faint or bright galaxies and, thus, different typical masses for the underlying DM halos. In both cases, the (constant) ionizing escape fraction is matched to yield similar Thompson scattering optical depths, consistent with estimates from *Planck* (Planck Collaboration XIII 2016). To assign LAEs to host halos, we connect LAE intrinsic luminosity and host halo mass via a minimum halo mass (corresponding to a minimum observed luminosity) as well as a duty cycle that accounts for the stochasticity of Ly α emission. This relation is calibrated to match the observed $z = 6.6$ LAE number density and luminosity function of the Subaru Suprime-Cam ultra-deep field (Ouchi et al. 2010), taking into account the IGM attenuation along the line-of-sight for a typical velocity shift of $\sim$230 km s^{-1} redward of the line centre. We define LAEs as galaxies with a Ly α luminosity of $L_\alpha \geq 2.5\times10^{42}$ erg s^{-1} and have checked that the resulting LAEs also match the observed angular clustering signal.

We calculate the cross-correlation function directly from our real-space 21-cm and LAE boxes using the estimator from Croft et al. (2016). We do not directly Fourier transform from the cross-power spectrum to the cross-correlation function, as we found this to be less stable in the presence of 21-cm noise in mock realizations. Our noise model assumes an SKA1-low tracked scanning strategy with 1000 h on-sky integration and is calculated using the 21CMSENSE code (Pober et al. 2013, 2014). Specifically, we assume modes in the so-called foreground wedge to be lost, a frequency-dependent scaling for the sky temperature, and a compact antennae core of a maximum baseline of 1.7 km for the antennae configuration from the SKA1-low baseline design. We sum over the visible, narrow-band projected LAE–21-cm cell pairs at distance r,

$$r_{21,{\rm LAE}}(r) = \frac{1}{N_{\rm LAE}N(r)}\sum_i^{N_{\rm LAE}}\sum_j^{N(r)}\delta_{21}\left(\mathbf{r}_{\rm i}+\mathbf{r}_{\rm j}\right), \quad (18)$$

where $\mathbf{r}_{\rm i}$ is the position of the i-th LAE and $|\mathbf{r}_{\rm j}| = r$; $N_{\rm LAE}$ is the number of LAEs in the observed volume and $N(r)$ is the number of 21-cm cells at distance r from the i-th LAE.

3.2 Understanding the 21-cm-LAE cross-correlation dependencies

To understand how the 21-cm-LAE cross-correlation function, particularly its amplitude, depends on the ionization state of the IGM and the reionization topology, we compare the 21-cm-LAE cross-correlation functions derived from our simulations to the analytic limits and profiles outlined in Section 2. We show the respective 21-cm-LAE cross-correlation functions at different stages of reionization and for different scenarios in Fig. 1 for the EOS simulations (SmallH II, LargeH II) and in Fig. 2 for the ASTRAEUS simulations (MHDEC, MHINC).

Firstly, from these figures, we see that the negative 21-cm-LAE cross-correlation amplitude at small scales, $|\xi_{21,{\rm LAE}}(r=0)|$, decreases in all reionization scenarios as the Universe becomes more ionized. This trend has been found in a number of works (e.g. Sobacchi et al. 2016; Heneka et al. 2017; Hutter et al. 2017, 2018; Heneka & Mesinger 2020; Weinberger et al. 2020) and agrees with the $\langle\chi_{\rm H\,I}\rangle$-scaling of $\xi_{21,{\rm LAE}}$ derived in equations (8) and (10). The latter echos the fact that LAEs are located in ionized regions, and thus the difference between the average ionization level and that at LAE positions is $\langle 1-\chi_{\rm H\,II}\rangle = \langle\chi_{\rm H\,I}\rangle$.

However, as we expect from equations (8) and (10) and can see from the simulated $\xi_{21,{\rm LAE}}$ values and the corresponding ionization levels (cf. Tables 1 and 2), the overall ionization state of the IGM is not the only quantity that defines $|\xi_{21,{\rm LAE}}(r=0)|$. As the average 21-cm differential brightness temperature depends on the ionization state *and* the gas density in neutral regions, $|\xi_{21,{\rm LAE}}(r=0)|$ is also proportional to the gas overdensity in neutral regions, $\langle 1+\delta\rangle_{\rm H\,I}$. In Figs 1 and 2, we see that the $|\xi_{21,{\rm LAE}}(r=0)|$ expectations according to equation (8) (depicted as diamonds in Fig. 1 and dotted lines in Fig. 2) match well with the simulation-derived cross-correlations at small r (solid lines) for the LargeH II, MHINC, and MHDEC models. We note that for the SmallH II model the mean neutral density and thus the $|\xi_{21,{\rm LAE}}(r=0)|$ expectation is up to $\sim$ 10 per cent lower in absolute value depending on $\langle\chi_{\rm H\,I}\rangle$ as compared to the LargeH II expectation; due to the on average smaller size of ionized regions

[6]The Fourier transformation is computed as $\widetilde{\delta}(\mathbf{k}) = V^{-1}\int\delta(\mathbf{x})\,e^{-2\pi i\mathbf{kx}}\,{\rm d}^3x$.
[7]http://homepage.sns.it/mesinger/EOS.html

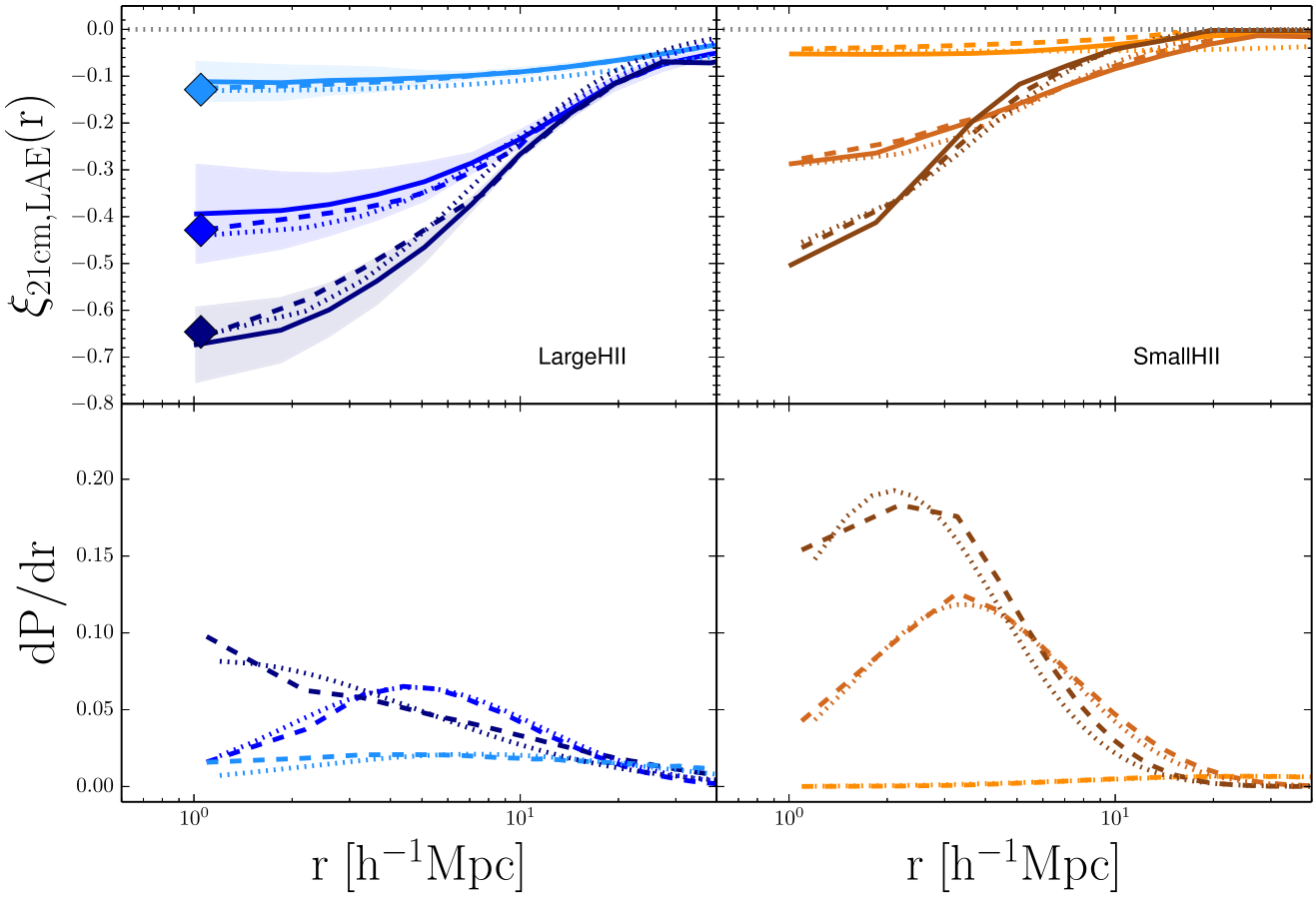

Figure 1. Upper panels: 21-cm-LAE cross-correlation functions for the EOS LargeH II (left) and SmallH II (right) simulations for varied hydrogen neutral fraction, for neutral hydrogen fraction ∼0.15, ∼0.52, and ∼0.74 (top to bottom line, light to dark). Shaded regions depict 2σ scatter computed for each 10 mock Monte-Carlo SKA1-low and Subaru HSC realizations. The diamonds at small r (left) depict the $r = 0$ (equation 8) expectation as derived from the corresponding simulations. Bottom: Probability density distribution of ionized regions of the LargeH II (left) and SmallH II (right) simulations. Solid lines show the results from the simulations in the top panels. Dashed and dotted lines show our analytical fit using the (1) ionization profiles, and (2) the lognormal distribution, respectively.

in the SmallH II model we probably need to resolve smaller scales, such as in the ASTRAEUS simulations ($\lesssim 0.5\ h^{-1}$ cMpc), for a better extrapolation to $r = 0$. The limits given are thus representing results for an upper limit on the bubble sizes.

The shaded regions in Fig. 1 (as in Fig. 3) show the 2σ uncertainty from 10 Monte-Carlo mock realizations of the 21-cm signal assuming 1000h of SKA-Low observations and of a narrow-band LAE survey with Subaru HSC characteristics. For the narrow-band LAE survey we assumed a systemic redshift uncertainty of $\Delta z = 0.1$, a survey area of 3.5 deg^2, and a limiting narrow-band luminosity of $L_{\alpha,\min} = 2.5 \times 10^{42}$ erg s^{-1}. As can be seen in this figure, we can expect the cross-correlation signals at different neutral hydrogen fractions (0.15, 0.52, 0.74) depicted to be distinguishable with such experiments. We would like to draw attention here mostly to the finding, that our analytical expectation and the simulation-derived cross-correlations agree well within the uncertainty bands depicted.

Secondly, we note that equation (8) is only valid in the post-heating regime where $T_s \gg T_{CMB}$. During Cosmic Dawn when $T_s \lesssim T_{CMB}$, $\xi_{21,LAE}(r = 0)$ depends also on the average spin temperature in neutral regions (as $1 - T_s/T_{CMB}$) as predicted by equation (7). Fig. 3 depicts the 21-cm-LAE cross-correlation function at $\langle \chi_{HI} \rangle \simeq 0.5$ for different average spin temperature values T_s in the neutral IGM derived from the EOS simulations (solid lines). These simulations track the spatially inhomogeneous evolution of the IGM temperature and Lyman series radiation background relevant for determining the coupling between the kinetic gas and spin temperatures, and follow the spin temperature fluctuations. In Fig. 3, the spatial fluctuations of the spin temperature were considered when calculating the 21-cm brightness temperature and respective cross-correlations. We refer the reader to Heneka & Mesinger (2020) for a detailed discussion of how the spin temperature fluctuations shape the 21-cm-LAE cross-correlation functions when the IGM or parts of it remain cold, revealing that assuming an average spin temperature would not yield the same results. The comparison to the analytical prediction at $r \simeq 0$ (coloured diamonds) shows again that these are in good agreement with the results from the simulations. This underlines that $\xi_{21,LAE}$ is sensitive to the spin temperature and gas densities in neutral regions and not to their full-box averages. We explore this power of the 21-cm signal to probe the density in neutral regions in the next section.

Next, we analyse how the size distribution of the ionized regions around LAEs is imprinted in $\xi_{21,LAE}(r)$. For this purpose, we derive the size distribution of ionized regions around LAEs in both ASTRAEUS and EOS simulations by shooting rays from each simulated LAEs along all major axes of the simulation box and measure the sizes of the surrounding ionized regions. Taking the resulting size distribution of ionized regions as PDF(r) (dashed lines in bottom

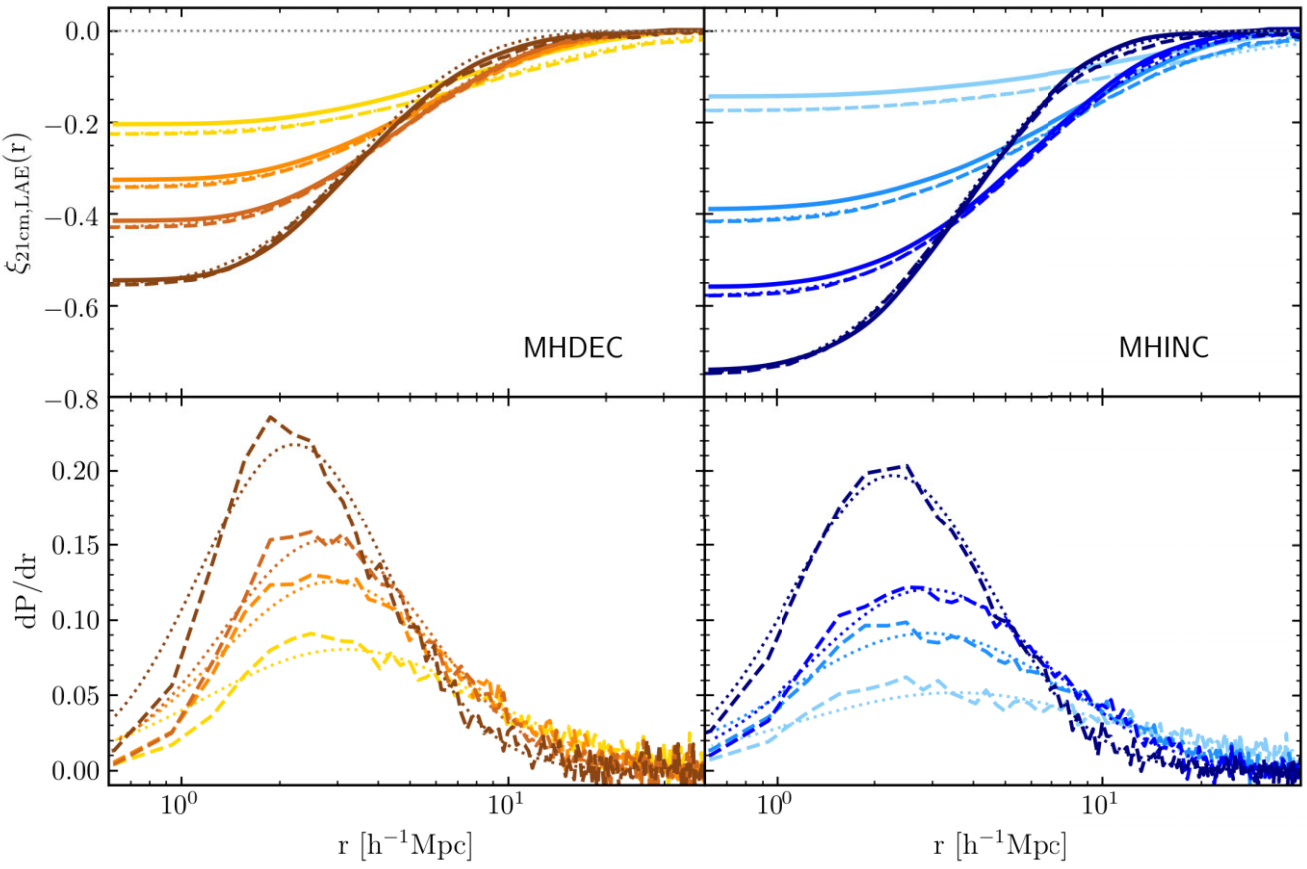

Figure 2. 21-cm-LAE cross-correlation functions (top) and probability density distribution of ionized regions (bottom) of the ASTRAEUS MHDEC and MHINC simulations for varied global HI fractions at z= 8.0, 7.3, 7.0, 6.7 from dark to bright colours. For these redshifts the HI fractions are $\langle\chi_{\mathrm{H\,I}}\rangle = 0.84, 0.69, 0.52, 0.23$ for MHINC and 0.71, 0.59, 0.49, 0.34 for the MHDEC simulations, respectively (see Table 1). Solid lines show the results from the simulations in the top panels. Dashed and dotted lines show our analytical fit using the (1) the ionization profiles along the six lines of sights (along major axes), and (2) the lognormal distribution that fits best to the line-of-sight averaged ionization profile, respectively.

Table 1. Best-fitting values for fitting the lognormal distribution to the size distribution of ionized regions around LAEs derived from the ASTRAEUS MHINC and MHDEC simulations.

z	MHINC $\langle\chi_{\mathrm{H\,I}}\rangle$	MHINC $\frac{r_{\mathrm{ion}}}{h^{-1}\mathrm{cMpc}}$	MHINC $\frac{\sigma_{\mathrm{ion}}}{h^{-1}\mathrm{cMpc}}$	MHDEC $\langle\chi_{\mathrm{H\,I}}\rangle$	MHDEC $\frac{r_{\mathrm{ion}}}{h^{-1}\mathrm{cMpc}}$	MHDEC $\frac{\sigma_{\mathrm{ion}}}{h^{-1}\mathrm{cMpc}}$
8.0	0.84	3.70	0.70	0.71	3.45	0.66
7.3	0.69	5.68	0.81	0.59	4.68	0.72
7.0	0.52	7.24	0.96	0.49	5.48	0.80
6.6	0.23	13.00	1.14	0.34	8.19	0.97

Note. In the MHINC simulation, f_{esc} increases with rising halo mass, while it decreases in the MHDEC simulation.

Table 2. Best-fitting values for fitting the lognormal distribution to the size distribution of ionized regions around LAEs derived from the LargeH II and SmallH II model of the EOS simulations.

$\langle\chi_{\mathrm{H\,I}}\rangle$	LargeH II $\frac{r_{\mathrm{ion}}}{h^{-1}\mathrm{cMpc}}$	LargeH II $\frac{\sigma_{\mathrm{ion}}}{h^{-1}\mathrm{cMpc}}$	SmallH II $\frac{r_{\mathrm{ion}}}{h^{-1}\mathrm{cMpc}}$	SmallH II $\frac{\sigma_{\mathrm{ion}}}{h^{-1}\mathrm{cMpc}}$
0.74	<9.77*	<1.47*	3.65	0.75
0.52	10.33	0.89	5.97	0.74
0.16/0.15[1]	32.81	1.23	100.61	1.16

Notes. *Upper limit due to limited spatial resolution of the simulations.
[1]During late states of reionization, or at comparably low neutral hydrogen fraction, the size distribution of ionized regions in the EOS simulations has a broad peak at a few tens of Mpc with a considerable tail towards larger radii, traceable due to the large simulated volume of 1 Gpc^3. We therefore caution the best-fitting values in this row to have a large uncertainty.

panels of Fig. 1 and Fig. 2), we derive $\xi_{21,\,\mathrm{LAE}}(r)$ with equations (13) and (15) (dashed lines in top panels of Figs 1 and 2). The derived $\xi_{21,\,\mathrm{LAE}}(r)$ values agree very well with the numerically derived ones. Various works have pointed out that the 21-cm-LAE cross-correlation function can measure the typical sizes of ionized regions around LAEs (e.g. Lidz et al. 2009; Wiersma et al. 2013; Vrbanec et al. 2020); however, they do not agree on which characteristic point in $\xi_{21,\,\mathrm{LAE}}(r)$ traces the scale of the average or typical size of ionized regions. Here we confirm that the peak of the size distribution of the ionized regions coincides with the inflection point of $\xi_{21,\,\mathrm{LAE}}(r)$. We also note that at the same $\langle\chi_{\mathrm{H\,I}}\rangle$ values the ionized regions around LAEs in the EOS simulations have on average larger sizes than in the ASTRAEUS simulations (cf. Tables 1 and 2 at $\langle\chi_{\mathrm{H\,I}}\rangle \simeq 0.5$). This might be due to the ASTRAEUS simulations assuming a lower Ly α luminosity required for a galaxy to be an LAE and an ionizing emissivity biased more towards lower-mass halos through the scaling of f_{esc} with halo mass.

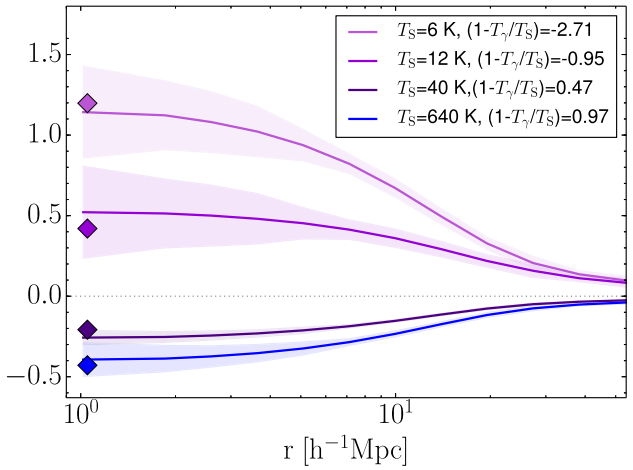

Figure 3. 21-cm-LAE cross-correlation functions for the EOS LargeH II and different average spin temperature values T_s in the neutral IGM at fixed neutral fraction of $\sim$ 50 per cent. Shaded regions depict 2σ scatter computed for each 10 mock Monte-Carlo SKA1-Low (1000 h) and Subaru HSC realizations.

Finally, we test whether a lognormal distribution adequately describes the size distribution of the ionized regions around LAEs and can be used to quickly forecast $\xi_{21,\,\mathrm{LAE}}(r)$ from a given set of parameters ($\langle\chi_{\mathrm{H\,I}}\rangle$, $\langle 1+\delta\rangle_{\mathrm{H\,I}}$, r_{ion}, σ_{ion}). We test this hypothesis by finding the parameters of the lognormal distribution (r_{ion}, σ_{ion}) that best-fitting the PDF(r) derived from the measured ionization profiles around LAEs in the simulations. The bottom panels of Fig. 2 show that the lognormal distribution (dotted lines) provides indeed a good fit to the measured size distributions of ionized regions (dashed lines). Most notably the lognormal distribution only tends to overpredict the number of smaller sized ionized regions, leading to the corresponding $\xi_{21,\,\mathrm{LAE}}(r)$ values [derived following equation (15) and shown as dashed lines in top panels of Fig. 2] shifting to higher values or smaller scales.

In summary, we find our analytic limits and profiles to match the $\xi_{21,\,\mathrm{LAE}}(r)$ values derived from the EOS and ASTRAEUS simulations very well as long as LAEs reside in highly ionized cells. Assuming a lognormal distribution is an adequate approximation for the size distribution of ionized regions and can be used to fit future 21-cm-LAE cross-correlation functions derived from observations.

3.3 Tracing the reionization topology

The different reionization scenarios covered in the EOS and ASTRAEUS simulations allow us to analyse the signatures of the reionization topology in their 21-cm-LAE cross-correlation function. We focus on two signatures: the small-scale amplitude and the inflection point of the 21-cm-LAE cross-correlation function $\xi_{21,\,\mathrm{LAE}}(r)$.

Firstly, the small-scale amplitude $|\xi_{21,\,\mathrm{LAE}}|(r \simeq 0)$ depends on the reionization topology, i.e. the propagation of the ionization fronts through the cosmic web, as it traces the average hydrogen gas density in neutral regions, $\langle 1+\delta\rangle_{\mathrm{H\,I}}$: the stronger the correlation between the underlying gas distribution and ionizing emissivity distribution emerging from galaxies or the redshift when a region became ionized, the lower is $\langle 1+\delta\rangle_{\mathrm{H\,I}}$ at any time during the EoR. We see this relation when comparing the SmallH II and LargeH II reionization scenarios in the EOS simulations in Table 3 and the MHDEC and MHINC reionization scenarios in the ASTRAEUS simulations in Fig. 4. In both, the LargeH II and MHINC scenario, the majority of the ionizing photons are produced and escape from more massive galaxies with

Table 3. Neutral gas overdensities at given global neutral hydrogen fractions for the LargeH II and SmallH II simulations.

$\langle\chi_{\mathrm{H\,I}}\rangle$	$\langle 1+\delta\rangle_{\mathrm{H\,I}}$ for LargeH II	$\langle 1+\delta\rangle_{\mathrm{H\,I}}$ for SmallH II
0.74	0.90	0.89
0.52	0.85	0.83
0.16/0.15	0.83	0.78

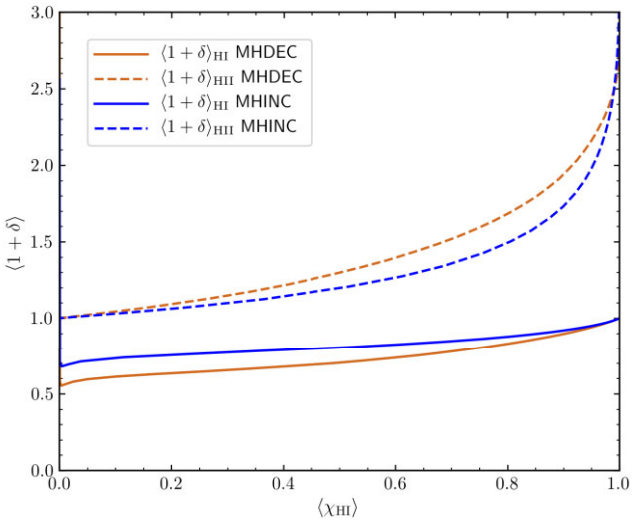

Figure 4. Relation between the neutral (ionized) gas overdensity and the global neutral hydrogen fraction in the ASTRAEUS MHDEC and MHINC simulations. Solid lines show the results for the neutral hydrogen gas and dashed lines for the ionized hydrogen gas.

$T_{\mathrm{vir}} > 2 \times 10^5$ K and $M_h \gtrsim 10^{9.5}$ M$_\odot$, respectively. Located in significantly overdense regions, the ionized regions originating from these galaxies trace indeed these significantly overdense regions but not the less dense regions where lower mass halos are located; $\langle 1+\delta\rangle_{\mathrm{H\,I}}$ drops only slightly as reionization progresses. In contrast, ionized regions in the SmallH II and MHDEC scenarios follow the underlying DM and gas density distribution closely as the low-mass halos located in intermediate to dense regions are the dominant sources of ionizing photons; as a consequence $\langle 1+\delta\rangle_{\mathrm{H\,I}}$ traces increasingly the least dense regions as the Universe becomes ionized.

Secondly, the shape of $\xi_{21,\,\mathrm{LAE}}$ directly maps the size distribution of the ionized regions; in particular, the peak of the size distribution of ionized regions coincides with the inflection point of $\xi_{21,\,\mathrm{LAE}}$.[8] Importantly, the peak of the size distribution is highly sensitive to the distribution of the ionizing emissivity within the galaxy population, e.g. the more ionizing radiation escapes from lower mass halos, the more similar sized are the ionized regions and the smaller is the average ionized region. Indeed, these trends can be seen in Fig. 2 and Table 1 when going from the MHINC (bottom right-hand panel) to the MHDEC scenario (bottom left-hand panel): the size distribution of ionized regions becomes more peaked and shifts to smaller scales. Hence, the inflection point of $\xi_{21,\,\mathrm{LAE}}$ provides an estimate of the typical size of ionized regions around LAEs.

We note that the ASTRAEUS simulations show lower $\langle 1+\delta\rangle_{\mathrm{H\,I}}$ values than the EOS simulations due to the following reasons: (1)

[8]The inflection point is given by $\frac{\partial^2\xi_{21,\mathrm{LAE}}(r)}{\partial r^2} \simeq \frac{\partial^2\mathrm{CDF}(r)}{\partial r^2} = \frac{\partial\mathrm{PDF}(r)}{\partial r} = 0$.

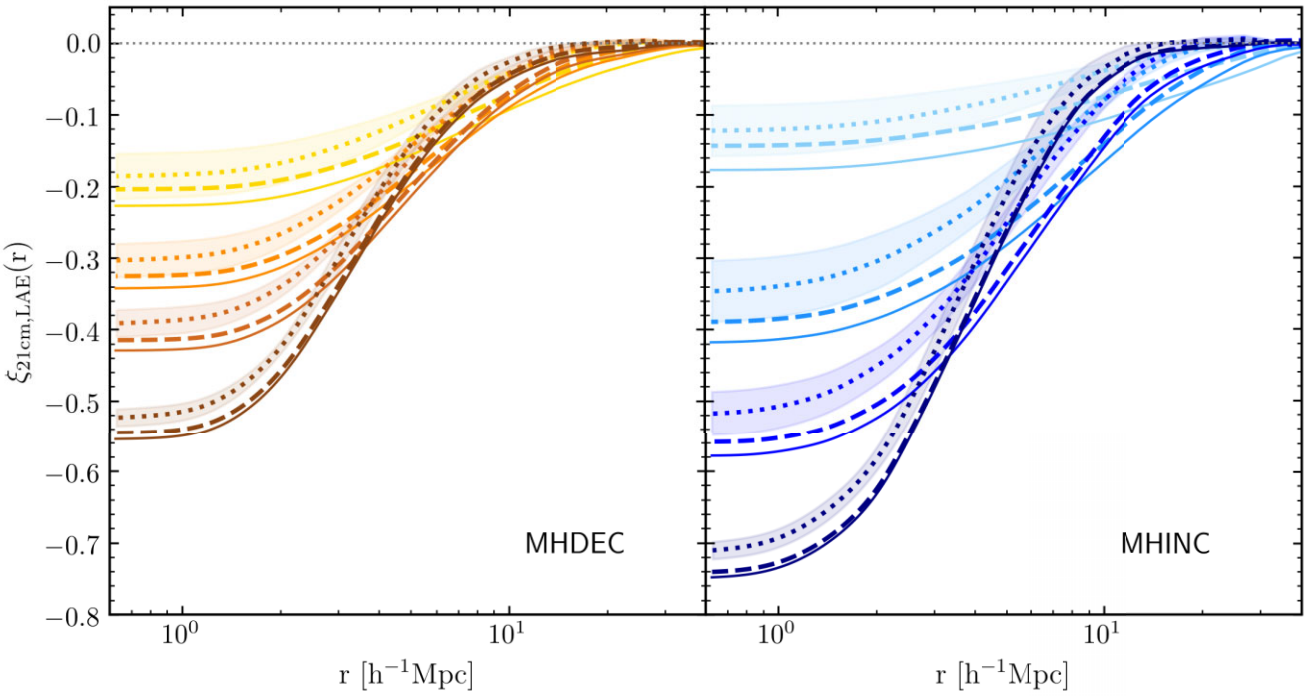

Figure 5. 21-cm-LAE cross-correlation functions of the ASTRAEUS MHDEC and MHINC simulations for simulation box sizes of 80 h^{-1} cMpc (dotted), 160 h^{-1} cMpc (dashed), and 320 h^{-1} cMpc (solid). The shaded region shows the standard deviation of $\xi_{21,\,\mathrm{LAE}}(r)$ across the eight subboxes. Results are shown for varied global H I fractions at $z = 8.0, 7.3, 7.0, 6.7$ from dark to bright colours. For these redshifts the H I fractions are $\langle\chi_{\mathrm{H\,I}}\rangle = 0.84, 0.69, 0.52, 0.23$ for MHINC, and 0.71, 0.59, 0.49, 0.34 for the MHDEC simulations, respectively (see Table 1).

while both the SmallH II and MHDEC scenarios consider halos that exceed virial temperatures of $T_{\mathrm{vir}} = 10^4$K and are not star-formation suppressed by radiative feedback from reionization, the contribution of low-mass halos ($M_h \lesssim 10^{9.5}$ M$_\odot$) to reionization is higher in the MHDEC simulation, as it includes also an ionizing escape fraction (f_{esc}) that decreases with rising halo mass. (2) While the LargeHII simulation considers only halos with $T_{\mathrm{vir}} > 2 \times 10^5$ K ($M_h \gtrsim 10^{9.5}$ M$_\odot$ at $z = 7$) to contribute to the ionizing budget, the MHINC scenario includes the same halos as the MHDEC scenario but an f_{esc} that increases with halo mass and thus has also minor contribution from low-mass halos.

In summary, as the reionization topology depends sensitively on the trends of galactic properties shaping the emerging ionizing emissivity with galaxy mass (e.g. f_{esc}, stellar populations, initial mass function), not only the inflection point of the 21-cm-LAE cross-correlation function traces the ionizing properties of LAEs but also their small-scale amplitude during the EoR. The more ionizing radiation emerges from low-mass objects that follow the underlying cosmic web structure more closely than more massive objects, the stronger is the correlation between the underlying density and ionization fields,[9] and thus the weaker is the 21-cm-LAE anticorrelation amplitude at small scales.

4 IMPACT OF LIMITED VOLUMES

With the analytic limits for the 21-cm-LAE cross-correlation function $\xi_{21,\,\mathrm{LAE}}$ at hand, we can investigate the impact of the simulation box size on the $\xi_{21,\,\mathrm{LAE}}(r)$ values derived from the gridded simulation boxes via the cross-power spectra during different stages of reionization. We note that while cosmic variance affects the 21-cm-LAE cross-correlation amplitudes, our analytic estimates (equation 15) should remain valid for the corresponding $\langle\chi_{\mathrm{H\,I}}\rangle$ and $\langle 1 + \delta\rangle_{\mathrm{H\,I}}$ values in the selected volume. For this reason, in this section, convergence refers to the deviation of the via the cross-power spectra derived $\xi_{21,\,\mathrm{LAE}}(r)$ values from our analytic estimates. Therefore, in Fig. 5 we show the $\xi_{21,\,\mathrm{LAE}}(r)$ for different simulation box sizes, ranging from 80 to 320 h^{-1}Mpc for the ASTRAEUS simulations. We note that since the ASTRAEUS simulations have only been run on a periodic 160 h^{-1}Mpc box, we derive the results for a 320 h^{-1}Mpc box by concatenating the 160 h^{-1} box at the same redshift in all three directions. While this will not recover the large-scale power missed due to cosmic variance (impacting the large-scale reionization topology), it provides a rough estimate of the large scale modes required to derive converged $\xi_{21,\,\mathrm{LAE}}(r)$ values.

To obtain $\xi_{21,\,\mathrm{LAE}}(r)$ for the 80 h^{-1}Mpc box, we divide the 160 h^{-1}Mpc box into eight subboxes. For each of the subboxes, we compute $\xi_{21,\,\mathrm{LAE}}(r)$. We show their mean value (dotted lines) and standard deviation (shaded areas) across the subboxes in Fig. 5. Here the standard deviation measures the cosmic variance of such a volume, while the deviation of the mean value to our analytic estimate for the 21-cm-LAE cross-correlation functions (essentially represented by $\xi_{21,\,\mathrm{LAE}}(r)$ for the 320 h^{-1}Mpc box in Fig. 5) estimates the convergence. We briefly digress to discuss the effects of cosmic variance. Listing average ionization fraction $\langle\chi_{\mathrm{H\,II}}\rangle$ and the overdensity of neutral regions $\langle 1 + \delta\rangle_{\mathrm{H\,I}}$ for each subbox, we can see from Tables A1 and A2 in Appendix A that the ionization history has not converged in a volume of a 80 h^{-1}Mpc

[9]We note that a stronger correlation between the underlying density and ionization fields results in a lower average overdensity in neutral regions $\langle 1 + \delta\rangle_{\mathrm{H\,I}}$ at fixed $\langle\chi_{\mathrm{H\,I}}\rangle$.

box: $\langle\chi_{\rm H\,II}\rangle$ varies around $\sim 2-7$ per cent, with the variation amplitude rising as reionization progresses. As the ionization fronts propagate from dense to less dense regions, simulation boxes with lower ionization levels show higher average densities of the neutral regions. As expected, we find these values to predict by how much $\xi_{21,\,\rm LAE}(r)$ of a subbox exceeds or subceeds the main value across all subboxes, with the deviation being proportional to $\langle\chi_{\rm H\,I}\rangle\langle 1+\delta\rangle_{\rm H\,I|subbox} - \langle\chi_{\rm H\,I}\rangle\langle 1+\delta\rangle_{\rm H\,I}$, as expected from our analytic estimate (equation 8).

Next we discuss the convergence by comparing the $\xi_{21,\,\rm LAE}(r)$ results across the different simulation volumes. In Fig. 5 we see that the strength of the anticorrelation between the 21-cm signal and LAEs drops the more large-scale cross-power is missed due to a decreasing box size (going from solid to dashed to dotted lines). How much large-scale power is missed depends on (1) the global ionization fraction $\langle\chi_{\rm H\,I}\rangle$ and (2) the reionization topology, i.e. the correlation between the reionization redshift $z_{\rm reion}$ and the underlying gas density field:

Dependence on $\langle\chi_{\rm H\,I}\rangle$: For both ASTRAEUS simulations the difference between the 160 or 80 h^{-1} cMpc to the 320 h^{-1} cMpc box increases as the Universe becomes more ionized and $\langle\chi_{\rm H\,I}\rangle$ decreases. It shifts from $\lesssim 1$ per cent ($\lesssim 3$ per cent) at $z = 8$ to $3-4$ per cent ($5-6$ per cent) at $z = 6.6$ for the 160 h^{-1} cMpc (80 h^{-1} cMpc) box. The growth of the ionized regions in size enhances the importance of large-scale power. The closer their sizes become to those of the simulation box, the less volume is left to map the background accurately.

Dependence on $1+\delta$-$z_{\rm reion}$ cross-correlation: The MHINC scenario where brighter galaxies are the main drivers of reionization shows larger differences in the small-scale 21-cm-LAE cross-correlation amplitude among different box sizes than the MHDEC scenario [cf. values at $\langle\chi_{\rm H\,I}\rangle \simeq 0.5$ in MHDEC (orange lines) and MHINC (medium blue lines)]. The reason is similar to that for the dependence on $\langle\chi_{\rm H\,I}\rangle$: in the MHINC scenario, the spatial variance of the ionizing emissivity is higher as more massive galaxies have higher $f_{\rm esc}$ values therefore the ionized regions around LAEs are larger and their sizes get closer to that of the simulation box at higher $\langle\chi_{\rm H\,I}\rangle$ values.

Finally, for full convergence, i.e. the $\xi_{21,\,\rm LAE}(r)$ values derived via the cross-power spectra numerically agree with our analytical estimates, we find the ASTRAEUS simulation box to be at the limit. Ideally, larger simulation box of $\sim$300 h^{-1}Mpc on the side would be required to obtain converged results. Interestingly, these volumes are in good agreement with those found necessary for the 21-cm power spectrum to converge due to cosmic variance in previous studies (Iliev et al. 2014; Kaur, Gillet & Mesinger 2020). However, it should be also noted that the combined SKA-Subaru HSC observational uncertainties depicted in Fig. 1 are of similar order than the deviation of the 80 h^{-1} cMpc box $\xi_{21,\,\rm LAE}(r)$ values from the 320 h^{-1} cMpc box results. Finally, we note that computing the 21-cm-LAE cross-correlation functions not via the cross-power spectra but directly in real-space may be an avenue to avoid the convergence issues for smaller volumes, as obtaining the small-scale amplitudes does not rely then on capturing the large-scale fluctuations. We also find them more stable when deriving the 21-cm-LAE cross-correlation functions from mock realizations that include the thermal noise in the 21-cm signal maps. However, computing the 21-cm-LAE cross-correlation functions in real-space is significantly slower than via the cross-power spectra (approximately hours versus minutes for the 1024^3 grids of the EOS simulations) but should remain feasible for similar sized grids and probably more robust when cross-correlating future noisy 21-cm maps with volume-limited LAE data.

5 COMPARISON TO PREVIOUS WORK

As we have seen in previous sections, the 21-cm-LAE cross-correlation functions are not only sensitive to the overall ionization state of the IGM but also to the reionization topology and the simulation box size. In the following, we compare our results to previous works and highlight why their predictions agree or differ from our analytic model.

Firstly, all works where the field of the 21-cm signal fluctuations, δ_{21}, scales with δT_b depict the $\langle\chi_{\rm H\,I}\rangle$-dependency of $\xi_{21,\,\rm LAE}(r)$ (see Sobacchi et al. 2016; Vrbanec et al. 2016; Hutter et al. 2017, 2018; Heneka & Mesinger 2020; Vrbanec et al. 2020; Weinberger et al. 2020). This scaling is not seen in Kubota et al. (2018) as their 21-cm signal fluctuation field is normalized by $\langle\delta T_b\rangle$ and thus not sensitive to $\langle\chi_{\rm H\,I}\rangle\langle 1+\delta\rangle_{\rm H\,I}$. Although $\xi_{21,\,\rm LAE}(r)$ in Weinberger et al. (2020) is shown in units of mK, dividing their $\lim_{r\to 0}\xi_{21,\,\rm LAE}(r)$ values by T_0 (as given by their equation 2) yields anticorrelation amplitudes that are in rough agreement with our analytic limits, with $\langle 1+\delta\rangle_{\rm H\,I}$ decreasing from $\lesssim 1$ at $z \simeq 7.4$ to $\sim$0.8 at $z \simeq 6.6$ for their Very Late model. Similarly, the 21-cm-LAE cross-correlation results in Heneka & Mesinger (2020) confirm our analytic limits for both the post-heating as well as heating epoch when $T_s \sim T_{\rm CMB}$. The results in Vrbanec et al. (2016, 2020) are also in line with our predictions; however, dividing their $\xi_{21,\,\rm LAE}$ values at the smallest scales shown yields a constant $\langle 1+\delta\rangle_{\rm H\,I}$ value during the second half of reionization ($\langle\chi_{\rm H\,I}\rangle \lesssim 0.5$). Secondly, the 21-cm-LAE cross-correlation predictions in some works (Hutter et al. 2017, 2018; Kubota et al. 2018) show lower anticorrelation amplitudes due to their box sizes of $\lesssim$ 200 cMpc and deriving the cross-correlations from the cross-power spectra. As a result, the survey parameters predicted in these works to distinguish between different stages of reionization represent conservative limits. Thirdly, Sobacchi et al. (2016) finds also lower anticorrelation amplitudes despite a sufficient volume of $\gtrsim 500^3$ cMpc^3. This might be due to actually showing the 2D 21-cm-LAE cross-correlation functions or their method of connecting the intrinsic Ly α luminosity to halos or their LAEs being located in partially neutral regions (as their LAE model might allow sufficient IGM transmission of Ly α because of the redshifted Ly α line emerging from galaxies). The latter is likely to be also the main reason for the weaker 21-cm-galaxy anticorrelation amplitudes in Park et al. (2014). Their galaxy sample extends down to halo masses of $M_h \simeq 2\times10^8$ M$_\odot$, which are most abundant but not able to ionize a cell of $\sim$0.5 cMpc length alone.

6 CONCLUSIONS

We have computed the 21-cm-LAE cross-correlation functions, $\xi_{21,\,\rm LAE}(r)$, for different reionization scenarios simulated with two different seminumerical schemes following galaxy evolution and reionization. While ASTRAEUS derives the galaxy properties from the simulated DM mass assembly histories like semi-analytical galaxy evolution models, 21CMFAST follows a more semi-empirical approach to infer galaxy properties. The scenarios differ in the large-scale distribution of the ionizing emissivity and cover the physically plausible range of (1) the escape fraction of ionizing photons $f_{\rm esc}$ decreasing to increasing with halo mass and (2) different galaxy mass (virial temperature) thresholds for star formation. This diverse data set has allowed us to verify the analytic limit of the small-scale 21-cm-LAE cross-correlation amplitude we derived and to propose a physically motivated analytic fitting function for the 21-cm-LAE cross-correlation function during the EoR. Our fitting function assumes the sizes of ionized regions around LAEs to follow

a lognormal distribution and fits the numerical results derived from the different simulations well. The analytic limit and fitting function for the 21-cm-LAE cross-correlation function allow us to draw the following conclusions:

(i) The small-scale 21-cm-LAE cross-correlation amplitude, $\xi_{21,\,\mathrm{LAE}}(r \simeq 0)$, is directly proportional to the mean neutral hydrogen fraction and the average spin-temperature weighted overdensity in neutral regions. In the post-heating regime ($T_\mathrm{s} \gg T_\mathrm{CMB}$) the dependence on the spin temperature becomes negligible.

(ii) Assuming a lognormal distribution for the sizes of the ionized regions provides a good approximation for the ionized regions around LAEs and allows us to analytically derive the shape of $\xi_{21,\,\mathrm{LAE}}(r)$ across all scales. The peak of the size distribution of the ionized regions and thus typical size of ionized regions around LAEs corresponds to the inversion point in $\xi_{21,\,\mathrm{LAE}}(r)$.

(iii) Scaling with the average overdensity in neutral regions, the 21-cm-LAE cross-correlation amplitude is also sensitive to the reionization topology, i.e. the propagation of the ionization fronts within the cosmic web structure. The stronger the emerging ionizing emissivity is correlated to the underlying gas distribution, i.e. the more the ionizing emissivity is biased to low-mass galaxies, the weaker is the 21-cm-LAE anticorrelation amplitude.

(iv) The smaller the simulation box is, the more large-scale modes are not contributing to the large-scale anticorrelation between the 21-cm signal and LAEs, relevant when the 21-cm-LAE cross-correlation function is derived via the 21-cm-LAE cross-power spectrum and leading to the 21-cm-LAE anticorrelation amplitude being underestimated. This effect increases with the size of the ionized regions and their size and distribution being sensitive to cosmic variance. We find that $\sim$300 h^{-1} cMpc boxes provide large enough volumes for the numerically derived 21-cm-LAE cross-correlation functions to reproduce our small-scale analytic limit.

(v) Our analytic predictions and volume studies can explain the different 21-cm-LAE (21-cm-galaxy) cross-correlation predictions to date. Given the information provided in previous works, we find them due to different normalizations of the 21-cm fluctuation field δ_{21}, too small simulation volumes, or galaxies/LAEs being located in partially neutral simulation cells (due to a too large cell size for the given galaxy population).

The functional form of the 21-cm-LAE cross-correlation function that we derived in this paper provides not only a test for future cross-correlation predictions from simulations, e.g. whether the simulated volume is sufficient, but also a fitting function for the 21-cm-LAE cross-correlation functions derived from future observations. The latter could provide a quick way to derive constraints for reionization. However, the problem remains that the 21-cm-LAE anticorrelation amplitude is sensitive to both the ionization history and topology. Breaking this degeneracy would require at least a tight relation between the size distribution of the ionized regions and the average ionization level of the IGM. Future work, however, needs to show whether such a relation is sufficient or whether the non-Gaussian nature of the ionized large-scale structure needs to be accounted for. We already see that the size distributions of ionized regions do not differ very strongly for opposing f_esc scenarios. For this reason, 21-cm LAE cross-correlations are unlikely to provide tighter constraints on reionization unless they are complemented by analyses that trace the non-Gaussianity of the 21-cm signal, such as the bispectrum (Majumdar et al. 2018, 2020; Hutter et al. 2020; Tiwari et al. 2022) and other shape-sensitive statistics (Gazagnes, Koopmans & Wilkinson 2021).

Despite these shortcomings, our analytic representation of the 21-cm-LAE cross-correlation function offers a computationally cheap way to predict which combinations and designs of 21-cm and LAE surveys with forthcoming telescopes (e.g. SKA, Roman, Subaru's Hyper Suprime-Cam) would provide the best constrained 21-cm-LAE cross-correlation functions and thus constraints on reionization.

ACKNOWLEDGEMENTS

AH acknowledges support by the VILLUM FONDEN under grant 37459. The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. CH acknowledges funding by Volkswagen Foundation and is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1—390900948 (the Heidelberg STRUCTURES Excellence Cluster). PD and MT acknowledge support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) grant 016.VIDI.189.162 ('ODIN'). PD warmly thanks the European Commission's and University of Groningen's CO-FUND Rosalind Franklin program. GY acknowledges Ministerio de Ciencia e Innovación (Spain) for partial financial support under research grant PID2021-122603NB-C21. The VSMDPL simulation has been performed at LRZ Munich within the project *pr87yi*. The CosmoSim data base (www.cosmosim.org) provides access to the simulation and the Rockstar data. The data base is a service by the Leibniz Institute for Astrophysics Potsdam (AIP). This research made use of MATPLOTLIB, a PYTHON library for publication quality graphics (Hunter 2007); and the PYTHON library NUMPY (Harris et al. 2020).

DATA AVAILABILITY

The ASTRAEUS code is publicly available on GitHub (https://github.com/annehutter/astraeus; Hutter, Legrand & Ucci 2020) as well as the EOS simulations on http://homepage.sns.it/mesinger/EOS.html. The full ASTRAEUS and EOS simulations and the data derived from these data sets will be shared on reasonable request to the corresponding authors.

REFERENCES

Barry N. et al., 2019, ApJ, 884, 1
Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, MNRAS, 461, 3135
Beane A., Villaescusa-Navarro F., Lidz A., 2019, ApJ, 874, 133
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, ApJ, 763, 18
Bosman S. E. I. et al., 2022, MNRAS, 514, 55
Carilli C. L., Rawlings S., 2004, New Astron. Rev., 48, 979
Castellano M. et al., 2016, ApJ, 818, L3
Castellano M. et al., 2018, ApJ, 863, L3
Croft R. A. C. et al., 2016, MNRAS, 457, 3541
Dayal P. et al., 2022, MNRAS, 512, 989
Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1
Dayal P., Ferrara A., Dunlop J. S., Pacucci F., 2014, MNRAS, 445, 2545
Dayal P., Maselli A., Ferrara A., 2011, MNRAS, 410, 830
DeBoer D. R. et al., 2017, PASP, 129, 045001
Dijkstra M., Gronke M., Venkatesan A., 2016, ApJ, 828, 71
Dijkstra M., Wyithe S., Haiman Z., Mesinger A., Pentericci L., 2014, MNRAS, 440, 3309
Field G. B., 1958, Proc. IRE, 46, 240
Fuller S. et al., 2020, ApJ, 896, 156
Furlanetto S. R., Lidz A., 2007, ApJ, 660, 1030

Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181
Gazagnes S., Koopmans L. V. E., Wilkinson M. H. F., 2021, MNRAS, 502, 1816
Goto H. et al., 2021, ApJ, 923, 229
Gronke M., Dijkstra M., McCourt M., Oh S. P., 2017, A&A, 607, A71
Harris C. R. et al., 2020, Nature, 585, 357
Heneka C., Cooray A., Feng C., 2017, ApJ, 848, 52
Heneka C., Mesinger A., 2020, MNRAS, 496, 581
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Hutter A., 2018, MNRAS, 477, 1549
Hutter A., Dayal P., Müller V., Trott C. M., 2017, ApJ, 836, 176
Hutter A., Dayal P., Partl A. M., Müller V., 2014, MNRAS, 441, 2861
Hutter A., Dayal P., Yepes G., Gottlöber S., Legrand L., Ucci G., 2021, MNRAS, 503, 3698
Hutter A., Legrand L., Ucci G., 2020, Astrophysics Source Code Library, record ascl:2004.006
Hutter A., Trebitsch M., Dayal P., Gottlöber S., Yepes G., Legrand L., 2023, MNRAS, 524, 6124
Hutter A., Trott C. M., Dayal P., 2018, MNRAS, 479, L129
Hutter A., Watkinson C. A., Seiler J., Dayal P., Sinha M., Croton D. J., 2020, MNRAS, 492, 653
Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725
Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2013, MNRAS, 428, 1366
Kakiichi K., Gronke M., 2021, ApJ, 908, 30
Kaur H. D., Gillet N., Mesinger A., 2020, MNRAS, 495, 2354
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, MNRAS, 491, 1736
Kimm T., Blaizot J., Garel T., Michel-Dansac L., Katz H., Rosdahl J., Verhamme A., Haehnelt M., 2019, MNRAS, 486, 2215
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Kubota K., Yoshiura S., Takahashi K., Hasegawa K., Yajima H., Ouchi M., Pindor B., Webster R. L., 2018, MNRAS, 479, 2754
Li W. et al., 2019, ApJ, 887, 141
Lidz A., Zahn O., Furlanetto S. R., McQuinn M., Hernquist L., Zaldarriaga M., 2009, ApJ, 690, 252
Liu A., Shaw J. R., 2020, PASP, 132, 062001
Maity B., Choudhury T. R., 2022, MNRAS, 515, 617
Majumdar S., Kamran M., Pritchard J. R., Mondal R., Mazumdar A., Bharadwaj S., Mellema G., 2020, MNRAS, 499, 5090
Majumdar S., Pritchard J. R., Mondal R., Watkinson C. A., Bharadwaj S., Mellema G., 2018, MNRAS, 476, 4007
Mauerhofer V., Dayal P., 2023, preprint (arXiv:2305.01681)
McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043
Meerburg P. D., Dvorkin C., Spergel D. N., 2013, ApJ, 779, 124
Mertens F. G. et al., 2020, MNRAS, 493, 1662
Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, MNRAS, 478, 3640
Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, MNRAS, 446, 566
Mesinger A., Furlanetto S. R., 2008, MNRAS, 386, 1990
Mesinger A., Greig B., Sobacchi E., 2016, MNRAS, 459, 2342
Mineo S., Gilfanov M., Sunyaev R., 2012, MNRAS, 426, 1870
Ouchi M. et al., 2010, ApJ, 723, 869
Ouchi M. et al., 2018, PASJ, 70, S13
Park J., Kim H.-S., Wyithe J. S. B., Lacey C. G., 2014, MNRAS, 438, 2474
Patil A. H. et al., 2016, MNRAS, 463, 4317
Patil A. H. et al., 2017, ApJ, 838, 65
Pentericci L. et al., 2014, ApJ, 793, 113
Pentericci L. et al., 2018, A&A, 619, A147
Planck Collaboration XIII, 2016, A&A, 594, A13
Planck Collaboration VI, 2020, A&A, 641, A6
Pober J. C. et al., 2013, AJ, 145, 65
Pober J. C. et al., 2014, ApJ, 782, 66
Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021, MNRAS, 506, 2390
Qin Y., Wyithe J. S. B., Oesch P. A., Illingworth G. D., Leonova E., Mutch S. J., Naidu R. P., 2022, MNRAS, 510, 3858
Schenker M. A., Ellis R. S., Konidaris N. P., Stark D. P., 2014, ApJ, 795, 20
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Sobacchi E., Mesinger A., 2014, MNRAS, 440, 1662
Sobacchi E., Mesinger A., Greig B., 2016, MNRAS, 459, 2741
Springel V., 2005, MNRAS, 364, 1105
Tiwari H., Shaw A. K., Majumdar S., Kamran M., Choudhury M., 2022, J. Cosmol. Astropart. Phys., 2022, 045
Trott C. M., Wayth R. B., 2016, PASA, 33, e019
Ucci G. et al., 2023, MNRAS, 518, 3557
Verhamme A., Orlitová I., Schaerer D., Hayes M., 2015, A&A, 578, A7
Vrbanec D. et al., 2016, MNRAS, 457, 666
Vrbanec D., Ciardi B., Jelić V., Jensen H., Iliev I. T., Mellema G., Zaroubi S., 2020, MNRAS, 492, 4952
Weinberger L. H., Kulkarni G., Haehnelt M. G., 2020, MNRAS, 494, 703
Wiersma R. P. C. et al., 2013, MNRAS, 432, 2615
Wouthuysen S. A., 1952, AJ, 57, 31
Wyithe J. S. B., Loeb A., Schmidt B. P., 2007, MNRAS, 380, 1087
Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, ApJ, 654, 12
Zhu Y. et al., 2021, ApJ, 923, 223

APPENDIX: CROSS-CORRELATION AMPLITUDE FOR SUBVOLUMES

In A1 and A2 we list the average ionization fraction $\langle \chi_{\rm H\,II} \rangle$ and neutral overdensity in each of the eight 80 h^{-1} cMpc subboxes at redshifts $z = 8.0$, 7.3, 7.0, and 6.7 for the MHDEC and MHINC scenarios, respectively. Values exceeding the average values of the entire simulation box (160 h^{-1} cMpc, second column) are marked in red, while those falling short are marked in blue. The more a subbox is ionized, the lower its average overdensity in neutral regions, leading to a lower 21-cm-LAE cross-correlation amplitude. The resulting variance in the 21-cm-LAE cross-correlation amplitude across the eight subboxes is shown as shaded regions in Fig. 5.

Table A1. Global hydrogen ionization fraction $\langle \chi_{\rm H\,II} \rangle$ and mean overdensity in neutral regions $\langle 1+\delta \rangle_{\rm H\,I}$ at different redshifts in the MHDEC ASTRAEUS simulation (column 2).

z	$\langle \chi_{\rm H\,II} \rangle$	$\langle \chi_{\rm H\,II} \rangle^{0,0,0}$	$\langle \chi_{\rm H\,II} \rangle^{1,0,0}$	$\langle \chi_{\rm H\,II} \rangle^{0,1,0}$	$\langle \chi_{\rm H\,II} \rangle^{0,0,1}$	$\langle \chi_{\rm H\,II} \rangle^{1,1,0}$	$\langle \chi_{\rm H\,II} \rangle^{1,0,1}$	$\langle \chi_{\rm H\,II} \rangle^{0,1,1}$	$\langle \chi_{\rm H\,II} \rangle^{1,1,1}$	$\sigma_{\langle \chi_{\rm HII} \rangle}$
8.0	0.290	0.281	0.313	0.268	0.308	0.302	0.261	0.287	0.296	0.019
7.3	0.413	0.400	0.446	0.381	0.445	0.433	0.371	0.409	0.420	0.028
7.0	0.511	0.497	0.548	0.468	0.558	0.541	0.455	0.501	0.521	0.038
6.6	0.660	0.653	0.715	0.602	0.723	0.702	0.577	0.635	0.673	0.053
z	$\langle 1+\delta \rangle_{\rm H\,I}$	$\langle 1+\delta \rangle_{\rm HI}^{0,0,0}$	$\langle 1+\delta \rangle_{\rm HI}^{1,0,0}$	$\langle 1+\delta \rangle_{\rm HI}^{0,1,0}$	$\langle 1+\delta \rangle_{\rm HI}^{0,0,1}$	$\langle 1+\delta \rangle_{\rm HI}^{1,1,0}$	$\langle 1+\delta \rangle_{\rm HI}^{1,0,1}$	$\langle 1+\delta \rangle_{\rm HI}^{0,1,1}$	$\langle 1+\delta \rangle_{\rm HI}^{1,1,1}$	$\sigma_{\langle 1+\delta \rangle_{\rm HI}}$
8.0	0.781	0.788	0.773	0.795	0.771	0.774	0.793	0.783	0.775	0.0096
7.3	0.730	0.737	0.721	0.746	0.719	0.722	0.744	0.732	0.723	0.0108
7.0	0.698	0.705	0.688	0.714	0.686	0.689	0.713	0.700	0.690	0.0115
6.6	0.661	0.669	0.654	0.679	0.649	0.651	0.677	0.664	0.652	0.0121

Note. Columns 3–10 show the respective values for the eight subboxes with each having a length of 80 h^{-1}Mpc. Column 11 depicts the standard deviation across the eight subboxes of the respective values.

Table A2. Global hydrogen ionization fraction $\langle \chi_{\rm H\,II} \rangle$ and mean overdensity in neutral regions $\langle 1+\delta \rangle_{\rm H\,I}$ at different redshifts in the MHINC ASTRAEUS simulation (column 2).

z	$\langle \chi_{\rm H\,II} \rangle$	$\langle \chi_{\rm H\,II} \rangle^{0,0,0}$	$\langle \chi_{\rm H\,II} \rangle^{1,0,0}$	$\langle \chi_{\rm H\,II} \rangle^{0,1,0}$	$\langle \chi_{\rm H\,II} \rangle^{0,0,1}$	$\langle \chi_{\rm H\,II} \rangle^{1,1,0}$	$\langle \chi_{\rm H\,II} \rangle^{1,0,1}$	$\langle \chi_{\rm H\,II} \rangle^{0,1,1}$	$\langle \chi_{\rm H\,II} \rangle^{1,1,1}$	$\sigma_{\langle \chi_{\rm HII} \rangle}$
8.0	0.157	0.156	0.179	0.137	0.166	0.166	0.140	0.153	0.161	0.014
7.3	0.313	0.313	0.359	0.267	0.339	0.339	0.278	0.297	0.314	0.032
7.0	0.482	0.497	0.545	0.405	0.522	0.523	0.427	0.453	0.484	0.062
6.6	0.770	0.804	0.839	0.671	0.795	0.820	0.686	0.739	0.808	0.064
z	$\langle 1+\delta \rangle_{\rm H\,I}$	$\langle 1+\delta \rangle_{\rm HI}^{0,0,0}$	$\langle 1+\delta \rangle_{\rm HI}^{1,0,0}$	$\langle 1+\delta \rangle_{\rm HI}^{0,1,0}$	$\langle 1+\delta \rangle_{\rm HI}^{0,0,1}$	$\langle 1+\delta \rangle_{\rm HI}^{1,1,0}$	$\langle 1+\delta \rangle_{\rm HI}^{1,0,1}$	$\langle 1+\delta \rangle_{\rm HI}^{0,1,1}$	$\langle 1+\delta \rangle_{\rm HI}^{1,1,1}$	$\sigma_{\langle 1+\delta \rangle_{\rm HI}}$
8.0	0.892	0.895	0.885	0.902	0.886	0.887	0.900	0.894	0.889	0.0065
7.3	0.843	0.847	0.834	0.856	0.833	0.835	0.852	0.847	0.839	0.0081
7.0	0.806	0.810	0.796	0.821	0.794	0.797	0.817	0.812	0.802	0.0103
6.6	0.757	0.757	0.750	0.774	0.746	0.743	0.775	0.773	0.740	0.0148

Note. Columns 3–10 show the respective values for the eight subboxes with each having a length of 80 h^{-1}Mpc. Column 11 depicts the standard deviation across the eight subboxes of the respective values.

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