astraeus - VIII. A new framework for Lyman-α emitters applied to different reionization scenarios

Hutter, Anne; Trebitsch, Maxime; Dayal, Pratika; Gottlöber, Stefan; Yepes, Gustavo; Legrand, Laurent

Published in:
Monthly Notices of the Royal Astronomical Society

DOI:
10.1093/mnras/stad2230

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **524,** 6124–6148 (2023) https://doi.org/10.1093/mnras/stad2230

Advance Access publication 2023 July 28

ASTRAEUS – VIII. A new framework for Lyman-α emitters applied to different reionization scenarios

Anne Hutter,[1,2,3]★ Maxime Trebitsch,[1] Pratika Dayal,[1] Stefan Gottlöber,[4] Gustavo Yepes[5,6] and Laurent Legrand[1]

[1] Kapteyn Astronomical Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

[2] Cosmic Dawn Center (DAWN)

[3] Niels Bohr Institute, University of Copenhagen, Jagtvej 128, DK-2200 Copenhagen N, Denmark

[4] Leibniz-Institut für Astrophysik, An der Sternwarte 16, D-14482 Potsdam, Germany

[5] Departamento de Fısica Teorica, Modulo 8, Facultad de Ciencias, Universidad Autonoma de Madrid, E-28049 Madrid, Spain

[6] CIAFF, Facultad de Ciencias, Universidad Autonoma de Madrid, E-28049 Madrid, Spain

Accepted 2023 July 18. Received 2023 July 18; in original form 2022 September 30

ABSTRACT

We use the ASTRAEUS framework to investigate how the visibility and spatial distribution of Lyman-α (Lyα) emitters (LAEs) during reionization is sensitive to a halo mass-dependent fraction of ionizing radiation escaping from the galactic environment ($f_{\rm esc}$) and the ionization topology. To this end, we consider the two physically plausible bracketing scenarios of $f_{\rm esc}$ increasing and decreasing with rising halo mass. We derive the corresponding observed Lyα luminosities of galaxies for three different analytic Lyα line profiles and associated Lyα escape fraction ($f_{\rm esc}^{\rm Ly\alpha}$) models: importantly, we introduce two novel analytic Lyα line profile models that describe the surrounding interstellar medium (ISM) as dusty gas clumps. They are based on parameterizing results from radiative transfer simulations, with one of them relating $f_{\rm esc}^{\rm Ly\alpha}$ to $f_{\rm esc}$ by assuming the ISM of being interspersed with low-density tunnels. Our key findings are: (i) for dusty gas clumps, the Lyα line profile develops from a central to double peak dominated profile as a galaxy's halo mass increases; (ii) LAEs are galaxies with $M_h \gtrsim 10^{10}\,{\rm M}_\odot$ located in overdense and highly ionized regions; (iii) for this reason, the spatial distribution of LAEs is primarily sensitive to the global ionization fraction and only weakly in second-order to the ionization topology or a halo mass-dependent $f_{\rm esc}$; (iv) furthermore, as the observed Lyα luminosity functions reflect the Lyα emission from more massive galaxies, there is a degeneracy between the $f_{\rm esc}$-dependent intrinsic Lyα luminosity and the Lyα attenuation by dust in the ISM if $f_{\rm esc}$ does not exceed ~ 50 per cent.

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

1 INTRODUCTION

The Epoch of Reionization (EoR) marks the second major phase transition in the Universe. With the emergence of the first galaxies, ultraviolet (UV) radiation gradually ionizes the neutral hydrogen (H I) in the intergalactic medium (IGM) until the Universe is reionized by $z \simeq 5.3$ (Fan et al. 2006; Keating et al. 2020; Zhu et al. 2021; Bosman et al. 2022). However, as only the brighter galaxies during the EoR are observed to date, key questions detailing the reionization process remain outstanding: Did the few bright and more massive or the numerous faint and low-mass galaxies contribute more to reionization? Feedback processes, such as heating by supernovae (SN) and photoionization, suppress star formation in low-mass galaxies (Gnedin 2000; Gnedin & Kaurov 2014; Ocvirk et al. 2016, 2020; Hutter et al. 2021a), and reduce the contribution of very low-mass galaxies to reionization. An even more critical quantity that regulates the ionizing radiation (with energies $E > 13.6$ eV) escaping from galaxies and thus the galaxy population driving the reionization of the IGM is the fraction of ionizing photons $f_{\rm esc}$ that escape from galaxies into the IGM (e.g. Kim et al. 2013; Seiler et al. 2019; Hutter et al. 2021b; Garaldi et al. 2022).

While the presence of H I in the IGM during the EoR impedes direct measurements of $f_{\rm esc}$, different theoretical models and simulations have investigated the physical processes determining and dependencies of $f_{\rm esc}$ (e.g. Ferrara & Loeb 2013; Kimm & Cen 2014; Wise et al. 2014; Kimm et al. 2019). Cosmological radiation hydrodynamical simulations suggest that $f_{\rm esc}$ decreases towards deeper gravitational potential (e.g. Kimm & Cen 2014; Wise et al. 2014; Xu et al. 2016; Anderson et al. 2017; Kimm et al. 2017, 2019; Lewis et al. 2020). High-resolution simulations of the ISM indicate that $f_{\rm esc}$ is dominated by the escape from star-forming clouds. The ionizing radiation of massive stars and their explosions as SN ionize, heat, and destroy the star-forming clouds clearing the way for the ionizing radiation to escape (Howard et al. 2018; Kim, Kim & Ostriker 2019; He, Ricotti & Geen 2020; Kimm et al. 2022). The complex dependency of $f_{\rm esc}$ on the underlying gravitational potential, the gas distribution and stellar populations in the ISM leaves marks not only in the radiation emitted by galaxies but also in the ionization topology, the time, and spatial distribution of the ionized regions around galaxies.

Current and forthcoming observations of galaxies and the ionization state of the IGM have the potential to constrain galactic

★ E-mail: anne.hutter@nbi.ku.dk

properties, such as f_{esc}, and the reionization process. On the one hand, detecting the 21-cm signal from H I in the IGM with forthcoming large radio interferometers (e.g. Square Kilometre Array) will measure the ionization topology, which provides constraints on the dependence of f_{esc} on galaxy mass (Kim et al. 2013; Seiler et al. 2019; Hutter et al. 2020b). On the other hand, being extremely sensitive to the attenuation by H I in the IGM, the observable Lyman-α (Lyα) radiation at 1216Åfrom high-redshift galaxies has gained popularity in probing reionization for the following reason: A $z \gtrsim 6$ galaxy only exhibits detectable Lyα emission when: (i) it is surrounded by an ionized region that is large and ionized (i.e. low residual H I fraction) enough to allow a sufficient fraction of its emerging Lyα line to traverse the IGM, or (ii) it is gas-rich enough (corresponding to a high H I column density) such that the red part of the Lyα line emerging from the galaxy is redshifted out of absorption, or (iii) it has strong outflows that redshift the emerging Lyα line out of absorption, or it is a combination of all three. The first criterion suggest that more massive galaxies able to retain more gas might be the most likely to show observable Lyα emission during the EoR: their higher rates of forming stars emitting ionizing photons lead to an increased production of Lyα radiation in the ISM and the growth of large ionized regions around them. The latter is accelerated by their ionized regions merging earlier with those of the surrounding lower mass objects attracted by their deeper gravitational potentials (Chardin, Aubert & Ocvirk 2012; Furlanetto & Oh 2016; Chen et al. 2019). As reionization progresses and the ionized regions grow, increasingly lower mass galaxies become visible as Lyα emitters (LAEs), which leads not only to a higher fraction of galaxies showing Lyα emission but also to a reduced clustering of LAEs (McQuinn et al. 2007; Jensen et al. 2013; Hutter, Dayal & Müller 2015; Sobacchi & Mesinger 2015).

This picture is increasingly supported by observations of $z > 6$ LAEs. Not only the fraction of Lyman Break Galaxies (LBGs) showing Lyα emission rises from $z \simeq 8$ to $z \simeq 6$ (Pentericci et al. 2014, 2018; Schenker et al. 2014; Fuller et al. 2020), but also the majority of Lyα emission at $z \gtrsim 6.5$ is detected in galaxies with a bright UV continuum (Oesch et al. 2015; Zitrin et al. 2015; Roberts-Borsani et al. 2016; Endsley & Stark 2022; Endsley et al. 2022). Moreover, the close proximity of UV-bright LAEs suggests that LAEs are located in over-dense regions (Vanzella et al. 2011; Castellano et al. 2016, 2018; Jung et al. 2020; Tilvi et al. 2020; Hu et al. 2021; Endsley & Stark 2022) that exhibit the first and largest ionized regions during the EoR. This hypothesis is also in line with the observed double-peaked Lyα profiles in $z \gtrsim 6.5$ galaxies (Hu et al. 2016; Matthee et al. 2018; Songaila et al. 2018; Meyer et al. 2021), indicating that the ionized regions surrounding them are so large that even the part bluewards the Lyα resonance redshifts out of resonance. Current theoretical predictions of the large-scale LAE distribution confirm this picture, suggesting that the LAEs we see during the EoR are more massive galaxies naturally located in over-dense regions (c.f. Dayal, Maselli & Ferrara 2011; Jensen et al. 2013; Hutter et al. 2014; Mesinger et al. 2015; Weinberger et al. 2018; Qin et al. 2022).

Yet, all these LAE models effectively assume a constant f_{esc} value across the entire galaxy population at a given redshift. This assumption remains highly uncertain as f_{esc} is very sensitive to the ISM and the circumgalactic medium (CGM) of galaxies that again depend on the underlying gravitational potential of a galaxy. However, it is essential, since f_{esc} defines the critical processes that shape the Lyα luminosities observed from galaxies. An f_{esc} varying with galactic properties and the underlying gravitational potential might alter the galaxy population seen as LAEs for the following reasons: first, within a galaxy, most Lyα radiation is produced by recombining hydrogen atoms [see e.g. Laursen et al. (2019) and Faucher-Giguère et al. (2010) for an estimate showing that Lyα cooling radiation is subdominant] and scales with the number of H I ionizing photons absorbed within the galaxy ($\propto 1 - f_{esc}$); secondly, a fraction of these Lyα photons undergoes only a few scattering events when they escape through the same low-density tunnels that facilitate the escape of H I ionizing photons. In contrast, the other fraction that traverses optically thick clouds upon its escape is scattered and absorbed by hydrogen and dust, respectively (see, e.g. Verhamme et al. 2015; Dijkstra, Gronke & Venkatesan 2016; Kimm et al. 2019; Kakiichi & Gronke 2021). These different escape mechanisms result not only in f_{esc} posing a lower limit to the fraction of Lyα photons escaping from a galaxy but also determining the Lyα line profile that emerges from a galaxy. Detailed low-redshift galaxy observations increasingly supported the f_{esc}-sensitivity of these Lyα properties (Verhamme et al. 2017; Jaskot et al. 2019; Gazagnes et al. 2020). Thirdly, f_{esc} shapes the IGM ionization topology by determining the number of ionizing photons available to ionize the IGM surrounding a galaxy. While a higher f_{esc} value enlarges the ionized region surrounding a galaxy and enhances the transmission of Lyα radiation through the IGM (Dayal et al. 2011; Hutter et al. 2014), the corresponding Lyα line emerging from a galaxy will be more peaked around the Lyα resonance and raise the absorption by H I in the IGM. Given this complex f_{esc}-dependency of the observed Lyα luminosity, it remains unclear whether different dependencies of f_{esc} with galaxy properties (e.g. increasing or decreasing with rising halo mass) would (i) identify the same galaxies as LAEs (exceeding a threshold Lyα luminosity) and/or (ii) lead to different spatial large-scale distribution of the LAEs' Lyα luminosities. In other words, which of these f_{esc}-dependent Lyα processes dominates the observed Lyα luminosities? For example, is the f_{esc}-dependency of the intrinsic Lyα luminosity dominant, and we yield a weaker clustering of LAEs when f_{esc} value decreases with rising halo mass? Or do they compensate each other once we reproduce the observed Lyα luminosity functions (Lyα LFs)?

To address these questions, we use our ASTRAEUS framework that models galaxy evolution and reionization self-consistently (Hutter et al. 2021a; Ucci et al. 2023), and simulate different reionization scenarios that gauge the physically plausible range of f_{esc} dependencies, i.e. f_{esc} decreasing and increasing with rising halo mass. Moreover, we parameterize results from numerical Lyα radiative transfer (RT) simulations of clumpy media (Gronke 2017) and build an analytic model for the fraction of Lyα photons escaping and the corresponding Lyα line profile emerging from high-redshift galaxies. Importantly, we explore three different Lyα line profile models, including (i) a Gaussian profile around the Lyα resonance where the Lyα escape fraction is directly related to the dust attenuation of the UV continuum (used in previous LAE models outlined in Dayal et al. 2011; Hutter et al. 2014), (ii) a Lyα line profile emerging from a shell of dusty gas clumps, which we model by using the different Lyα escape regimes identified in Gronke (2017), and (iii) a Lyα line profile emerging from a shell of gas clumps with a fraction f_{esc} of the solid angle interspersed by gas-free tunnels. The latter two give rise to various combinations of a central peak around the Lyα resonance (Lyα photons hardly scatter in an optically thin medium) and two peaks in the red and blue wings (Lyα photons are scattered in an optically thick medium). By deriving the observed Lyα luminosities of all simulated galaxies for all combinations of reionization scenarios and Lyα line models, we address the following questions: Which f_{esc}-dependent Lyα process, i.e. intrinsic production, escape, or transmission through the IGM of Lyα radiation, dominates the observed Lyα luminosity? Can the observed Lyα luminosities of galaxies inform us on their emerging Lyα line profile? Given the ionization topology depends sensitively on the assumed dependency of f_{esc} with halo mass, are the same or different galaxies identified as LAEs and do they differ in the spatial distribution of their Lyα luminosities?

This paper is organized as follows. In Section 2 we briefly describe the ASTRAEUS model, its implementation of dust, and the different reionization simulations. In Section 3 we introduce the different Lyα line profile models and their corresponding attenuation by dust. We then (Section 4) discuss how the Lyα line profiles depend on halo mass in our different reionization scenarios, how free model parameters, such as the ISM clumpiness or size of the dust gas clumps, need to be adjusted to fit the observed Lyα LFs, and how the galaxy properties determining the observed Lyα luminosities depend on the halo mass of a galaxy. In Section 5 we identify the location of LAEs in the large-scale density and ionization structure and assess whether the spatial distribution of LAEs differs for different f_{esc}-dependencies on halo mass/ionization topologies. Finally, we briefly discuss which Lyman Break galaxies are preferentially identified as LAEs (Section 6) and conclude in Section 7. In this paper we assume a Λ cold dark matter Universe with cosmological parameter values of $\Omega_\Lambda = 0.69$, $\Omega_{\rm m} = 0.31$, $\Omega_b = 0.048$, $H_0 = 100h = 67.8$ km s^{-1} Mpc^{-1}, $n_s = 0.96$, and $\sigma_8 = 0.83$, and a Salpeter initial mass function (IMF; Salpeter 1955) between 0.1 M$_\odot$ and 100 M$_\odot$.

2 THE MODEL AND SIMULATIONS

In this paper, we use the ASTRAEUS framework. This framework couples a semi-analytic galaxy evolution model (an enhanced version of DELPHI; Dayal et al. 2014) with a semi-numerical reionization scheme (CIFOG; Hutter 2018) and runs the resulting model on the outputs of a dark matter (DM) only N-body simulation. In this section, we provide a brief description of the physical processes implemented in ASTRAEUS (for more details, see Hutter et al. 2021a) and introduce the different reionization simulations.

2.1 N-body simulation

As part of the Multidark simulation project, the underlying DM N-body simulation (VERY SMALL MULTIDARK PLANCK; VSMDPL) has been run with the GADGET-2 TREE + PM code (Springel 2005). In a box with a side length of 160 h^{-1}Mpc, it follows the trajectories of 3840^3 DM particles. Each DM particle has a mass of $6 \times 10^6 h^{-1}$ M$_\odot$. For a total of 150 snapshots ranging from $z = 25$ to $z = 0$, the phase space ROCKSTAR halo finder (Behroozi, Wechsler & Wu 2013a) has been used to identify all halos and subhalos down to 20 particles or a minimum halo mass of $1.24 \times 10^8 h^{-1}$ M$_\odot$. To obtain the local horizontal merger trees (sorted on a redshift-by-redshift basis within a tree) for galaxies at $z = 4.5$ that ASTRAEUS requires as input, we have used the pipeline internal CUTNRESORT scheme to cut and resort the vertical merger trees (sorted on a tree-branch by tree-branch basis within a tree) generated by CONSISTENT TREES (Behroozi et al. 2013b). For the first 74 snapshots that range from $z = 25$ to $z = 4.5$, we have generated the DM density fields by mapping the DM particles onto 2048^3 grids and re-sampling these to 512^3 grids used as input for the ASTRAEUS pipeline.

2.2 Galaxy evolution

ASTRAEUS tracks key processes of early galaxy formation and reionization by post-processing the DM merger trees extracted from the VSMDPL simulation. At each time step (i.e. snapshot of the N-body simulation) and for each galaxy, it tracks the amount of gas that is accreted, the gas and stellar mass merging, star formation and associated feedback from SNII and metal enrichment, as well as the large-scale reionization process and its associated feedback on the gas content of early galaxies.

2.2.1 *Gas and stars*

In the beginning, when a galaxy starts forming stars in a halo with mass $M_{\rm h}$, it has a gas mass of $M_g^i(z) = f_g(\Omega_b/\Omega_m)M_h(z)$, with f_g being the gas fraction not evaporated by reionization, i.e. $f_g = 1$ and $f_g < 1$ as the galaxy forms in a neutral and ionized region, respectively. In subsequent time steps a galaxy gains gas from its progenitors ($M_g^{\rm mer}(z)$) and smooth accretion ($M_g^{\rm acc}$), while its total gas mass never exceeds the limit given by reionization feedback,

$$M_g^i(z) = \min\left(M_g^{\rm mer}(z) + M_g^{\rm acc}(z),\ f_g(\Omega_m/\Omega_b)M_h\right) \qquad (1)$$

with

$$M_g^{\rm acc}(z) = M_h(z) - \sum_{p=1}^{\rm N_p} M_{h,p}(z+\Delta z) \qquad (2)$$

$$M_g^{\rm mer}(z) = \sum_{p=1}^{\rm N_p} M_{h,p}(z+\Delta z), \qquad (3)$$

where N_p is the galaxy's number of progenitors and $M_{{\rm h},p}$ the halo mass of each progenitor.

At each time step, a fraction of the merged and accreted (initial) gas mass is transformed into stellar mass, $M_\star^{\rm new}(z) = (f_\star^{\rm eff}/\Delta t)M_g^i(z)$.[1] Here $f_\star^{\rm eff}$ represents the fraction of gas that forms stars over a time span Δt and is limited by the minimum amount of stars that need to form to eject all gas from the galaxy, $f_\star^{\rm ej}$, and an upper limit, $f_\star$. $f_\star^{\rm eff}$ depends on the gravitational potential: more massive galaxies form stars at the constant rate $f_\star$, while low-mass galaxies form stars at the limited rate $f_\star^{\rm ej}$ due to SN and radiative feedback. While we account for radiative feedback from reionization by modifying the initial gas mass reservoir with the factor f_g, $f_\star^{\rm eff}$ incorporates the suppression of star formation in low-mass halos as gas is heated and ejected by SNII explosions. Our model incorporates a delayed SN feedback scheme, i.e. at each time step the effective star formation efficiency accounts for the SNII energy released from stars formed in the current and previous time steps, following the mass-dependent stellar lifetimes (Padovani & Matteucci 1993). In contrast to Hutter et al. (2021a), we have updated our model and do not assume stars to form in bursts to calculate the number of SNII exploding within a time step but $M_\star^{\rm new}(z)$ to form at a constant star formation over the entire time step (see Appendix B for a detailed calculation). The star formation efficiency in the SN feedback-limited regime is given by

$$f_\star^{\rm ej}(z) = \frac{v_c^2}{v_c^2 + f_w E_{51} \nu_z}\left[1 - \frac{f_w E_{51} \sum_j \nu_j M_{\star,j}^{\rm new}(z_j)}{M_g^i(z)\ v_c^2}\right], \qquad (4)$$

with v_c being the rotational velocity of the halo, E_{51} the energy released by a SNII, f_w the fraction of SNII energy injected into the winds driving gas outflows, $M_{\star,j}^{\rm new}(z_j)$ the stellar mass formed during previous time steps j, and ν_j the fraction of stellar mass formed in previous time step j that explodes in the current time step given the assumed IMF.

ASTRAEUS incorporates multiple models for radiative feedback from reionization, ranging from a weak and time-delayed (*Weak Heating*) to a strong instantaneous feedback (*Jeans mass*). In this

[1] We note that this definition has been altered compared to the first version of ASTRAEUS in Hutter et al. (2021a).

Table 1. ASTRAEUS model parameters and chosen values in this work.

Parameter	Value or reference	Description
$f_\star$	0.025	Maximum star-formation efficiency
f_w	0.2	SN coupling efficiency
-	Photoionization	Radiative feedback model
IMF	Salpeter (1955)	For stellar evolution, enrichment, SED
SED	STARBURST99	ionizing SED model

work, we use the intermediate and time-delayed *Photoionization* model, where the characteristic mass defining the gas fraction not evaporated by reionization grows on a dynamical time-scale to the respective Jeans mass (for a detailed description see Hutter et al. 2021a). We list the ASTRAEUS model parameters and their assumed values in Table 1. $f_\star$ and f_w have been adjusted to reproduce the observed UV LFs, stellar mass functions, global star formation rate density, and global stellar mass density at $z = 10 - 5$.

2.2.2 *Metals and dust*

The current ASTRAEUS model also incorporates the metal enrichment by stellar winds, SNII and SNIa explosions (for a detailed description see Ucci et al. 2023). At each time step, we assume that gas smoothly accreted has the average metallicity of the gas in the IGM, $Z_{\rm IGM}$. Metals are produced through stellar winds, SNII and SNIa explosions. The amount of newly forming metals depends on the number of massive stars exploding as SN in the current time step according to Padovani & Matteucci (1993), Yates et al. (2013), and Maoz, Mannucci & Brandt (2012). For the corresponding stellar metal yields, ASTRAEUS uses the latest yield tables from Kobayashi, Karakas & Lugaro (2020). We assume that gas and metals are perfectly mixed. Thus, the metals ejected from the galaxy are proportional to the ejected gas mass and the metallicity of the gas in the galaxy. This ejected metal mass contributes to $Z_{\rm IGM}$.

In this work, we have extended the ASTRAEUS model (Hutter et al. 2021a; Ucci et al. 2023) to follow the formation, growth, destruction, astration, and destruction of dust in each galaxy (c.f. Dayal et al. 2022 for details). We note that we consider dust to be part of our metal reservoir (i.e. $M_{\rm dust} \leq M_{\rm m}$). At each time step, ASTRAEUS computes the evolution of the dust mass $M_{\rm dust}$ in a galaxy by solving the following differential equation

$$\frac{{\rm d}M_{\rm dust}}{{\rm d}t} = \dot{M}_{\rm dust}^{\rm prod} + \dot{M}_{\rm dust}^{\rm grow} - \dot{M}_{\rm dust}^{\rm dest} - \dot{M}_{\rm dust}^{\rm astr} - \dot{M}_{\rm dust}^{\rm ej}. \tag{5}$$

The first term on the right-hand side (RHS) of equation (5) denotes the production of dust in SNII and asymptotic giant branch (AGB) stars through condensation of metals in stellar ejecta,

$$\dot{M}_{\rm dust}^{\rm prod} = y_{\rm SNII}\gamma_{\rm SNII} + \dot{M}_{\rm dust}^{\rm AGB}, \tag{6}$$

with $y_{\rm SNII} = 0.45\,{\rm M}_\odot$ being the dust mass formed per SNII,

$$\gamma_{\rm SN}(t) = \int_{8\,{\rm M}_\odot}^{40\,{\rm M}_\odot} {\rm SFR}(t - \tau_m)\phi(m){\rm d}m. \tag{7}$$

the number of SNII events,

$$\dot{M}_{\rm dust}^{\rm AGB}(t) = \int_{0.85\,{\rm M}_\odot}^{50\,{\rm M}_\odot} y_{\rm AGB}(m){\rm SFR}(t - \tau_m)\phi(m){\rm d}m \tag{8}$$

the contribution from AGB stars and $y_{\rm AGB}$ the dust yields from AGB stars. In agreement with (Ucci et al. 2023), we adopt the latest yield tables from Kobayashi et al. (2020) for $y_{\rm AGB}$. The second term on the RHS of equation (5) describes the dust grain growth through the accretion of heavy elements in dense molecular clouds in the ISM,

$$\dot{M}_{\rm dust}^{\rm grow} = \left(Z' - \frac{M_{\rm dust}}{M_{\rm g}^i}\right) f_{\rm cold\ gas}\frac{M_{\rm dust}}{\tau_{{\rm gg},0}Z_\odot}, \tag{9}$$

where Z' is the metallicity after accretion and star formation, $M_{\rm dust}$ is the dust mass, $f_{\rm cold\ gas}$ the fraction of cold and molecular gas, and $\tau_{\rm gg} = \tau_{0,\,{\rm gg}}/Z$ the accretion time-scale adopted from Asano et al. (2013; see also Triani et al. 2020). We assume $f_{\rm cold\ gas} = 0.5$ and $\tau_{{\rm gg},\,0} = 30$ Myr. The third term in equation (5) describes the destruction of dust by SN blastwaves, for which we adopt the analytic description outlined in McKee (1989),

$$\dot{M}_{\rm dust}^{\rm dest} = \left(1 - f_{\rm cold\ gas}\right)\frac{M_{\rm dust}}{M_{\rm g}^i}\,\gamma_{\rm SN}\epsilon\,M_{\rm SN,bw}, \tag{10}$$

with ϵ being the effifiency of dust destruction in a SN-shocked ISM and $M_{\rm SN,\,bw}$ the mass accelerated to 100 km s^{-1} by the SN blast wave. In line with McKee (1989) and Lisenfeld & Ferrara (1998), we adopt $\epsilon = 0.03$ and $M_{\rm SN,bw} = 6.8 \times 10^3\,{\rm M}_\odot$. Finally, equation (5) accounts also for the destruction of dust by astration as new stars form from the metal-enriched gas,

$$\dot{M}_{\rm dust}^{\rm astr} = Z^{\rm i}\,\frac{M_\star^{\rm new}}{\Delta t}, \tag{11}$$

and the ejection of metals through winds powered by the energy injected by SN,

$$\dot{M}_{\rm dust}^{\rm ej} = Z'\,\frac{M_{\rm g}^{\rm ej}}{\Delta t}. \tag{12}$$

The parameter values ($y_{\rm SNII}$, $\tau_{{\rm gg},\,0}$, ϵ, $M_{\rm SN,\,bw}$) quoted reasonably reproduce the observed UV LFs when the UV is attenuated by dust as follows [please see Hutter et al. (2021a) for observational UV LFs data points included]: from the dust mass, M_d, we obtain the total optical depth to UV continuum photons as (see e.g. Dayal et al. 2011)

$$\tau_{\rm UV,c} = \frac{3\Sigma_d}{4as}, \tag{13}$$

with $\Sigma = M_d/(\pi r_d^2)$ being the dust surface mass density, r_d the dust distribution radius, and $a = 0.03\ \mu$m and $s = 2.25$ g cm^{-3} the radius and material density of graphite/carbonaceous grains (Todini & Ferrara 2001). Since we assume that dust and gas are perfectly mixed, we equate the dust distribution radius, r_d, with the radius of the gas, $r_g = 4.5\lambda r_{\rm vir}[(1 + z)/6]^{1.8}$. Here λ is the spin parameter of the simulated halo, $r_{\rm vir}$ the virial radius, and the third factor accounts for the redshift evolution of the compactness of galaxies and ensures that the observed UV LFs at $z = 5$–10 are well reproduced. For a slab-like geometry, the escape fraction of UV continuum photons of a galaxy is then given by

$$f_{\rm esc}^{\rm c} = \frac{1 - \exp(-\tau_{\rm UV,c})}{\tau_{\rm UV,c}}, \tag{14}$$

and its observed UV luminosity by

$$L_{\rm c}^{\rm obs} = f_{\rm esc}^{\rm c} L_{\rm c}, \tag{15}$$

with the intrinsic UV luminosity, $L_{\rm c}$, being computed as outlined in section 2.2.4 in Hutter et al. (2021a).

2.3 Reionization

At each time step ASTRAEUS follows the time and spatial evolution of the ionized regions in the IGM. For this purpose, it derives the number of ionizing photons produced in each galaxy, $\dot{Q}$, by convolving the

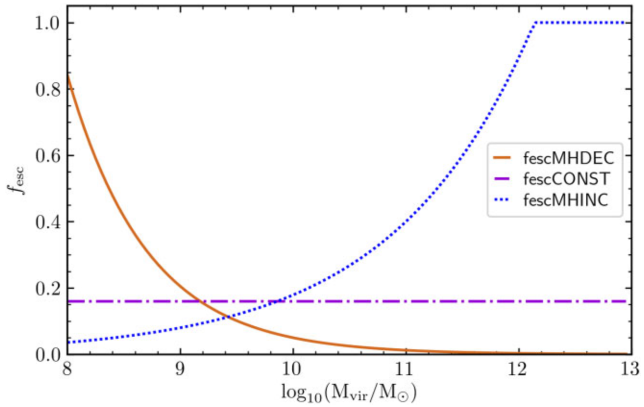

Figure 1. The ionizing escape fraction $f_{\rm esc}$ for the three models, decreasing (solid orange line), being constant (dash–dotted magenta line), and increasing (dotted blue line) with halo mass $M_{\rm h}$.

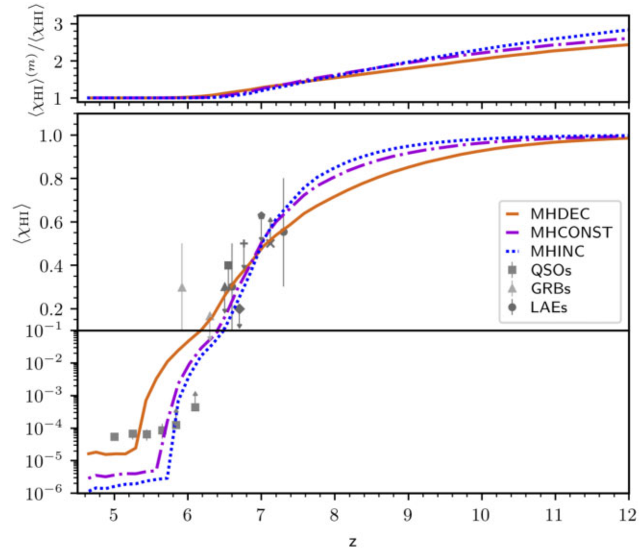

Figure 2. Ratio of the mass- and volume-averaged neutral hydrogen fraction (top panel) and volume averaged neutral hydrogen fraction (bottom panel) as a function of redshift. In each panel, we show results for our three $f_{\rm esc}$ models: decreasing (solid orange line), being constant (dash–dotted magenta line), and increasing (dotted blue line) with halo mass $M_{\rm h}$. In the lower panel, grey points indicate observational constraints from: gamma-ray burst (GRB) optical afterglow spectrum analyses (light triangles, Totani et al. 2006, 2014), quasar sightlines (medium squares, Fan et al. 2006), Lyα LFs (dark circles, Konno et al. 2018; dark squares, Kashikawa et al. 2011; dark diamonds, Ouchi et al. 2010; dark pentagons, Ota et al. 2010; and dark triangles, Malhotra & Rhoads 2004), LAE clustering (dark plus signs, Ouchi et al. 2010), and the Lyα emitting galaxy fraction (dark crosses, Pentericci et al. 2011, 2014; Caruana et al. 2012, 2014; Ono et al. 2012; Schenker et al. 2012; Treu et al. 2012).

galaxy's star formation rate history with the spectra of a metal-poor ($Z = 0.05 Z_{\odot}$) stellar population. Spectra have been obtained from the stellar population synthesis model STARBURST99 (Leitherer et al. 1999). Again we assume that stars form continuously over a time step. Then the number of ionizing photons that contribute to the ionization of the IGM is given by

$$\dot{N}_{\rm ion} = f_{\rm esc} \dot{Q}, \quad (16)$$

where $f_{\rm esc}$ is the fraction of ionizing photons that escape from the galaxy into the IGM. From the resulting ionizing emissivity and gas density distributions ASTRAEUS derives the spatial distribution of the ionized regions by comparing the cumulative number of ionizing photons with the number of absorption events (CIFOG; see Hutter 2018 for details). Within ionized regions, it also derives the photoionization rate and residual H I fraction in each grid cell. The ionization and photoionization fields obtained allow us then to determine on the fly whether the environment of a galaxy has been reionized and account for the corresponding radiative feedback by computing the gas mass the galaxy can hold on to ($f_g M_g^i$).

2.4 Simulations

In the following we consider three different reionization scenarios that explore the physically plausible space of the ionizing escape fraction $f_{\rm esc}$ (c.f. Fig. 1):

(i) MHDEC: $f_{\rm esc}$ decreases with rising halo mass of a galaxy (red solid line)

$$f_{\rm esc} = f_{\rm esc,low} \left(\frac{f_{\rm esc,high}}{f_{\rm esc,low}} \right)^{\frac{\log_{10}(M_h/M_{h,\rm low})}{\log_{10}(M_{h,\rm high}/M_{h,\rm low})}} \quad (17)$$

with $f_{\rm esc,\,low} = 0.55$, $f_{\rm esc,\,high} = 0.05$, $M_{h,\rm low} = 2 \times 10^8 h^{-1}\,{\rm M}_{\odot}$, and $M_{h,\rm high} = 10^{10} h^{-1}\,{\rm M}_{\odot}$.

(ii) MHCONST: $f_{\rm esc} = 0.16$ for each galaxy (magenta dash–dotted line).

(iii) MHINC: $f_{\rm esc}$ increases with rising halo mass of a galaxy (blue dotted line) following equation (17) with $f_{\rm esc,\,low} = 0.08$, $f_{\rm esc,\,high} = 0.4$, $M_{h,\rm low} = 10^9 h^{-1}\,{\rm M}_{\odot}$, and $M_{h,\rm high} = 10^{11} h^{-1}\,{\rm M}_{\odot}$.

These three $f_{\rm esc}$ prescriptions have been adjusted to reproduce the electron optical depth measured by Planck (Planck Collaboration 2020) and fit the observational constraints from LAEs, quasar absorption spectra, and gamma ray bursts (as depicted in the lower panel of Fig. 2). In addition, for MHINC the maximum $f_{\rm esc}$ value of more massive galaxies is also limited by the observed Lyα LFs. Despite having very similar electron optical depths, these three $f_{\rm esc}$ prescriptions lead to different ionization histories and topologies (see Figs 2 and 6). As $f_{\rm esc}$ decreases with rising halo mass, reionization is dominated by the low-mass galaxies ($M_h \lesssim 10^{10}\,{\rm M}_{\odot}$), leading to on average smaller ionized regions and lower photoionization rates. Since these low-mass galaxies appear earlier, reionization begins earlier (see solid red line in Fig. 2); however, as shown in Hutter et al. (2021a) for the *Photoionization* model their overall star formation rate decreases around $z \simeq 7$, resulting in the Universe being reionized at a later time and exhibiting a higher average residual H I fraction in ionized regions. In contrast, as $f_{\rm esc}$ increases with rising halo mass, more massive galaxies ($M_h \gtrsim 10^{10}\,{\rm M}_{\odot}$) drive reionization. On average, ionized regions are larger and more clustered around more massive galaxies, and photoionization rates within these ionized regions are higher. Reionization begins later with the appearance of more massive galaxies and ends earlier as the abundance of these massive galaxies increases.

3 MODELLING LYα EMITTERS

In this section, we introduce the different models for the emergent Lyα line profiles (Section 3.1) and fractions of Lyα radiation escaping from a galaxy (Section 3.3), describe the attenuation of Lyα radiation by H I in the IGM, and the derivation of the observed Lyα luminosity of a galaxy (Section 3.4). We summarize the combinations of emerging Lyα line profile and dust attenuation models investigated in this paper in Section 3.3.7.

3.1 Emerging Lyα line profiles

We investigate three Lyα line profiles $J(x)$: (1) a thermally Doppler-broadened Gaussian centred at the Lyα resonance; (2) a single, double, or triple-peaked profile that depends on the clumpiness and H I column density of the gas in a galaxy; (3) a single, double, or triple-peaked profile that depends both on the ionizing escape fraction $f_{\rm esc}$ and the clumpiness and H I column density of the gas in a galaxy. While the first model represents a simple assumption used in previous works (e.g. Dayal et al. 2011; Hutter et al. 2014), the latter two models are inspired by observations and detailed Lyα RT simulations (e.g. Dijkstra et al. 2016; Gronke 2017). The Lyα line emerging from a galaxy is given by the intrinsic Lyα luminosity, $L_\alpha^{\rm intr} = \frac{2}{3} Q(1-f_{\rm esc})\, h\nu_\alpha$, the escape fraction Ly$\alpha$ photons from the galaxy, $f_{\rm esc}^{\rm Ly\alpha}$, and the line profile $J(x)$.

$$L_\alpha^{\rm gal}(x) = L_\alpha^{\rm intr} f_{\rm esc}^{\rm Ly\alpha} J(x). \quad (18)$$

Here we have expressed the frequency deviation from the Lyα resonance ν_α in terms of the thermal line broadening $\sigma_{\rm th} = (v_{\rm th}/c)\nu_\alpha$ with $v_{\rm th} = \sqrt{2k_B T/m_{\rm H}}$, yielding $x = \frac{\nu-\nu_\alpha}{\sigma_{\rm th}}$. k_B is the Boltzmann constant, $m_{\rm H}$ the mass of a hydrogen atom, and T the temperature of the H I gas. In the remainder of this section, we detail our different models for the Lyα line profiles and escape fractions.

3.1.1 Central Gaussian

This model assumes that the emission sites of Lyα radiation, the hydrogen atoms within a galaxy, move at velocities that reflect the galaxy's rotation. The corresponding Doppler-broadened Lyα line profile is then given by

$$J_{\rm centre}(x) = \frac{1}{\sqrt{\pi}} \frac{\sigma_{\rm th}}{\sigma_r} \exp\left[-x^2 \frac{\sigma_{\rm th}^2}{\sigma_r^2}\right], \quad (19)$$

We note that since the $\sigma_{\rm th}$-dependence of x cancels any dependency of $J_{\rm Gaussian}(\nu)$ on $\sigma_{\rm th}$, the assumed gas temperature has no effect on the emerging Lyα line profile (we use $T = 10^4$ K in Fig. 3). $\sigma_r \simeq (v_r/c)\nu_\alpha$ describes the Doppler broadening of the line due to the rotation of the galaxy. The rotation velocity of the galaxy v_r is closely linked to the halo rotational velocity $v_c = (3\pi G H_0)^{1/3}\Omega_m^{1/6}(1+z)^{1/2}M_h^{1/3}$, ranging between $v_r = v_c$ and $v_r = 2v_c$ (Mo, Mao & White 1998; Cole et al. 2000). We assume $v_r = 1.5v_c$.

3.1.2 Single, double, or triple-peaked in a clumpy/homogeneous medium

This model describes the Lyα line profile emerging from a clumpy medium. It implements the regimes and characteristic escape frequencies identified in Gronke (2017). We consider a slab with a thickness of $2B$ and a total optical depth of $2\tau_0$. The source is located at the slab's midplane and injects photons at the Lyα resonance $x = 0$. If the slab medium is homogeneous, Neufeld (1990) derived the emergent Lyα profile as

$$J_{\rm slab}(T, \tau_0, x) = 4\pi \frac{\sqrt{6}}{24} \frac{x^2}{a(T)\,\tau_0} \frac{1}{\cosh\left(\sqrt{\frac{\pi^4}{54}}\frac{|x^3|}{a(T)\,\tau_0}\right)}, \quad (20)$$

for $a(T)\tau_0 \gtrsim 10^3$ with $a(T) = \frac{A_\alpha}{4\pi\sigma_{\rm th}(T)}$ and $\int_{-\infty}^{\infty} J_{\rm slab}(x, x_i)\,{\rm d}x = 1$. A_α is the Einstein for the spontaneous emission of Lyα photons.

In the following we will revisit the regimes for Lyα escape in a clumpy medium that have been identified in Gronke (2017). The clumpy medium is characterized by the total optical depth of the clumps and the average number of clumps each Lyα photon escaping the slab scatters with. For a slab consisting of clumps with each having an optical depth $\tau_{0,\rm cl}$ at the line centre, Lyα photons escaping the slab will encounter on average f_c clumps and have a total optical depth of $\tau_0 = \frac{4}{3} f_c \tau_{0,\rm cl}$[2] at the line centre. The emerging Lyα line profile depends sensitively on the total and clump optical depth at line centre, τ_0 and $\tau_{0,cl}$, respectively, and the number of clumps the Lyα photons scatter with. Gronke (2017) identified the following regimes:

(i) *Free-streaming regime:* The clumpy medium is optically thin ($\tau_0 < 1$), and Lyα photons can stream through. The emerging line profile peaks around $x = 0$.

(ii) *Porous regime:* The clumps are optically thick to Lyα photons ($\tau_{\rm cl} > 1$), but only a fraction $1 - \exp(-\tau_{0,cl})$ of the Lyα photons scatter with a clump. The emerging line profile is again peaked around $x = 0$.

(iii) *Random walk regime:* The clumps are optically thick to Lyα ($\tau_{\rm cl} > 1$), and each Lyα photon encounters $N_{\rm sct,rw} \propto f_c^2$ scattering events (Hansen & Oh 2006). However, the number of scattering events is too low for the Lyα photons to scatter in frequency space far enough into the wings to escape through excursion. Hence, the emerging line profile peaks also around $x = 0$.

(iv) *Homogeneous regime:* The clumps are optically thin ($\tau_{\rm cl} \leq 1$) and Lyα photons scatter $\sim\tau_0$ times ($N_{\rm sct,exc} \propto f_c$) and escape via excursion: they follow a random walk in space *and* frequency and escape as they are scattered into the wings where the clumps become optically thin. The emerging line profile is a double-peaked with the two peaks being located at

$$x_{\rm esc}(\tau_0) \simeq \begin{cases} \pm\left(ka\tau_0/\sqrt{\pi}\right)^{1/3}, & \frac{\sqrt{\pi}x_\star^2}{ka} \leq \tau_0 \\ \pm x_\star, & \tau_0 < \frac{\sqrt{\pi}x_\star^2}{ka} \end{cases} \quad (21)$$

for an injection frequency $x = 0$ (c.f. Adams 1975; Gronke 2017). Here $x_\star$ is the frequency where the Lyα absorption profile transitions from the Gaussian core to the Lorentzian wings, and $\tau_0 = \frac{\sqrt{\pi}x_\star^2}{ka}$ marks the transition where the slab becomes optically thin at the escape frequency $x_{\rm esc}$.

Gronke (2017) derived the boundary criteria between these regimes for a static clumpy medium, which we briefly revisit here. To derive the critical number of clumps, f_c, separating the regimes, we first consider the time and distances covered that it takes a Lyα to traverse the slab.

3.1.3 Excursion

As Lyα photons traverse or escape the slab, they scatter with H I many times. This alters their direction and frequency x, and they essentially perform a random walk. However, as the Lyα cross section is higher close to the line centre, most scatterings will occur close to the line centre and remain spatially close. Only as the Lyα photons are scattered into the wings of the Lyα absorption profile their mean free paths become larger, allowing them to escape the slab (Adams 1975). The series of these so-called wing scatterings that allow Lyα photons to escape are referred to as excursion. We can estimate the mean displacement and time spent in such an excursion event: a Lyα photon with frequency x will scatter on average $N_{\rm sct,exc} \sim x^2$ times before it returns to the core. For a slab of thickness B, its average mean free path is $\lambda_{\rm mfp,exc}(x) = B\sigma_0/(k\tau_0\sigma_{\rm HI}(x)) = B/(k\tau_0 H(a,x))$ using the wing approximation of the Lyα cross section and $k \simeq \sqrt{3}$ being a geometrical factor that accounts for slant paths in a plane-parallel

[2]The factor 4/3 arises from the mean path length through a sphere.

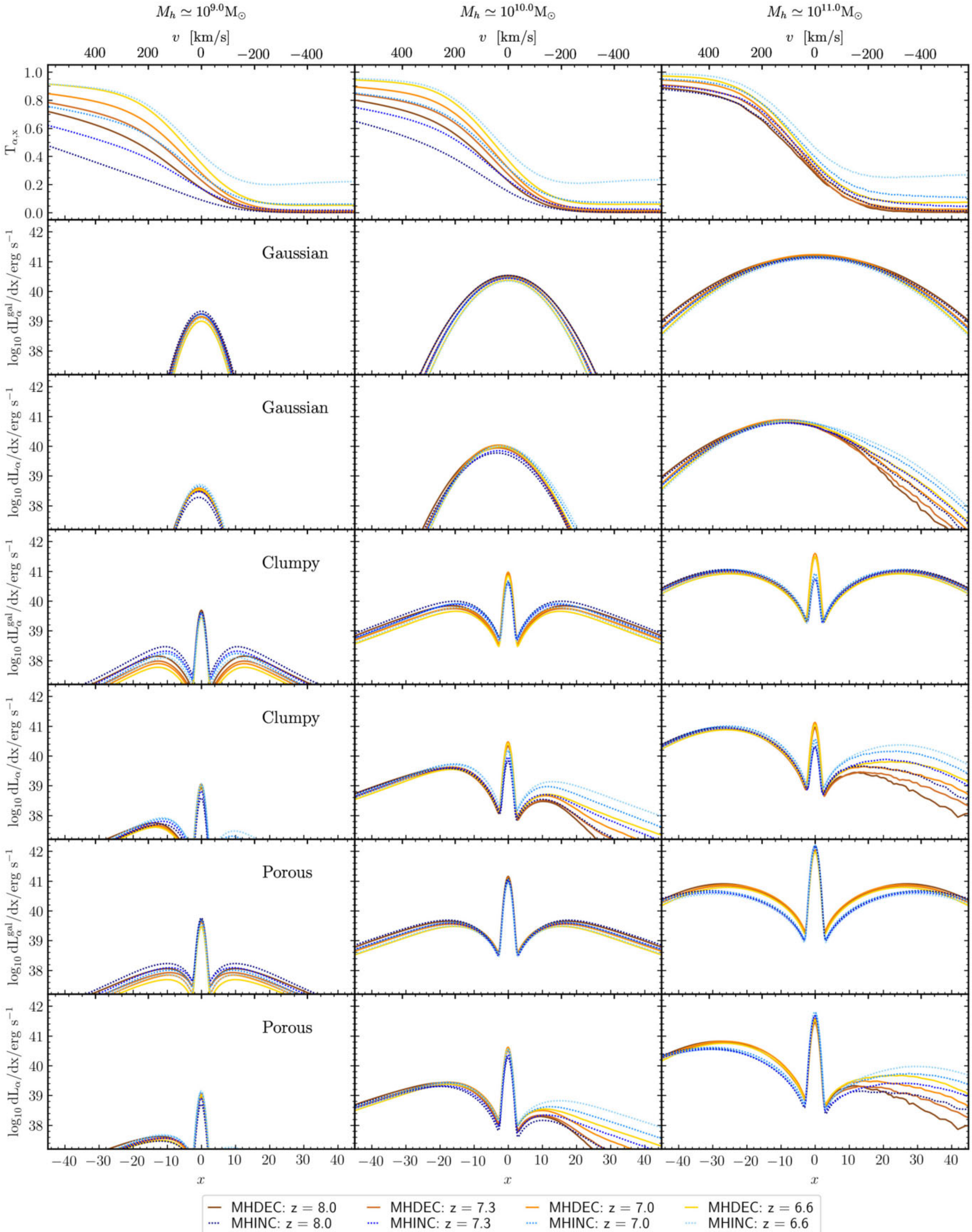

Figure 3. Intrinsic (top) and observed (bottom) Lyα line profile and IGM transmission (centre) at $z = 8.0, 7.3, 7.0, 6.6$ for a homogeneous static gas shell (left), a clumpy gas shell (centre), and a clumpy gas shell with holes through which Lyα radiation escapes without scattering. Solid (dash–dotted) lines show results for the reionization scenario where f_{esc} decreases (increases) with halo mass M_{h}.

medium and was determined in Adams (1975).[3] This and the random walk nature of the Lyα escape imply an average displacement of

$$d_{\rm exc} = \sqrt{N_{\rm sct,exc}}\lambda_{\rm mfp,exc}(x) = \frac{\sqrt{N_{\rm sct,exc}}B}{k\tau_0 H_v(a,x)} = \frac{xB}{k\tau_0 H_v(a,x)} \quad (22)$$

and time spent in the excursion of

$$t_{\rm exc} = N_{\rm sct,exc}\frac{\lambda_{\rm mfp,exc}(x)}{c} = \frac{N_{\rm sct,exc}B}{ck\tau_0 H_v(a,x)} = \frac{x^2 B}{ck\tau_0 H_v(a,x)}, \quad (23)$$

with $H_v(a,x) = \frac{a}{\sqrt{\pi}(x)^2}$ being the effective line absorption profile in the wings.

3.1.4 *Random walk*

As the clumps become optically thick at corresponding escape frequencies $x_{\rm esc}(\tau_0)$, the Lyα photons do not escape the slab via excursion anymore but by random walking: the number of scattering events is smaller than required for excursion and scale with the square of the number of clumps, $N_{\rm sct,rw} \propto f_c^2$ (Hansen & Oh 2006). With the mean free path given by the average clump separation $\lambda_{\rm mfp,\,rw} = kB/f_c$, the average displacement and time are then

$$d_{\rm rw} = \sqrt{N_{\rm sct,rw}}\lambda_{\rm mfp,rw} = \frac{\sqrt{N_{\rm sct,rw}}kB}{f_c} = kB \quad (24)$$

and

$$t_{\rm rw} = \sqrt{N_{\rm sct,rw}}\frac{\lambda_{\rm mfp,rw}}{c} = \frac{\sqrt{N_{\rm sct,rw}}kB}{cf_c} = \frac{kBf_c}{c}. \quad (25)$$

(i) *Division between random-walk and homogeneous regime in optically thick medium:* For a given total optical depth at the line centre, τ_0, we can derive the critical number of clumps along a line of sight that marks the transition from the random (clumps are optically thick) to the homogeneous regime (clumps become optically thick at the excursion frequency). We estimate this transition to arise when both regimes contribute equally to the flux of escaping Lyα photons.

$$\frac{F_{\rm rw}}{F_{\rm exc}} = \frac{t_{\rm exc}}{t_{\rm rw}} = \frac{x^2}{k^2\tau_0 H(x) f_c} = \frac{\sqrt{\pi}x^4}{k^2 a\tau_0 f_c} = 1 \quad (26)$$

With $\tau_0 = 4/3 f_c \tau_{0,\,\rm cl}$, the critical number of clumps for Lyα photons escaping at frequency x yields then as

$$f_{c,\rm crit} = \frac{\sqrt{3}\pi^{1/4}}{2k}\frac{x^2}{\sqrt{a\tau_{0,\rm cl}}}. \quad (27)$$

As long as the wings remain optically thick, the majority of Lyα photons (with injection frequency $x = 0$) will escape at $x_{\rm esc} \simeq \left(\frac{ka\tau_0}{\sqrt{\pi}}\right)^{1/3}$ leading to

$$f_{c,\rm crit} = \frac{2}{\sqrt{3}k\pi^{1/4}}\sqrt{a\tau_{0,\rm cl}}. \quad (28)$$

This $f_{c,\,\rm crit}$ value marks the transition from the random walk to the excursion regime. We can understand its increase with the clump optical depth $\tau_{0,\,\rm cl}$ as follows: for optically thicker clumps to become optically thin at the escape frequency $x_{\rm esc}$, a higher escaping frequency and thus a higher effective total optical depth are required. This can only be achieved by interacting with more clumps (higher $f_{c,\,\rm crit}$).
Because the transition described by $f_{c,\,\rm crit}$ is not sharp, we model the Lyα line profile emerging from the moving slab by superposing the Lyα radiation escaping in the homogeneous ($J_{\rm slab}$) and random walk regimes ($J_{\rm centre}$).

$$J_{\rm rh}(\tau_0, x) = (1 - f_{\rm rw})J_{\rm slab}(T, \tau_0, x) + f_{\rm rw}J_{\rm centre}(T, x) \quad (29)$$

Here we assume $J_{\rm centre}$ is given by equation (19) with $\sigma_r = \sigma_{\rm th}$ and

$$J_{\rm slab}(T,\tau_0,x) = \begin{cases} 4\pi\frac{\sqrt{6}}{24}\frac{x^2}{a(T)\;\tau_0}\frac{1}{\cosh\left(\sqrt{\frac{\pi^4}{54}\frac{|x^3|}{a(T)\;\tau_0}}\right)}, & \frac{\sqrt{\pi}x_\star^2}{ka(T)} \le \tau_0 \\ 4\pi\frac{\sqrt{6}}{24}\frac{kx^2}{\sqrt{\pi}x_\star^3}\frac{1}{\cosh\left(\sqrt{\frac{\pi^3}{54}\frac{k|x^3|}{x_\star^3}}\right)}, & \tau_0 < \frac{\sqrt{\pi}x_\star^2}{ka(T)} \end{cases}. \quad (30)$$

We derive the corresponding ratio $f_{\rm fw}$ by assuming that the Lyα flux escapes predominantly where the Lyα profiles peak,

$$f_{\rm rw} = \frac{F_{\rm rw}/F_{\rm exc}}{F_{\rm rw}/F_{\rm exc} + \frac{J_{\rm centre}(0)}{2(J_{\rm slab}(\tau_0,x_{\rm esc})+J_{\rm slab}(\tau_0,-x_{\rm esc}))}}. \quad (31)$$

We note that $J_{\rm slab}$ reproduces the results of Lyα RT simulations down to $\tau_0 \simeq 10^5$. While the line profiles for lower τ_0 values start deviating, we will see in Section 3.2 that the galaxies considered here exceed this threshold. Importantly, we find that the assumed $J_{\rm slab}$, $J_{\rm centre}$, and $f_{\rm rw}$ reproduce the Lyα line profiles for resting clumps, fixed τ values, and varying f_c values in Gronke (2017).

(ii) *Division between porous and homogeneous regime in optically thin medium:* As the medium becomes optically thinner, Lyα photons that scatter into the wings can escape the slab before completing their excursion. This transition occurs as the wings become optically thin, i.e. $k\tau(x_\star) \le 1$ translating to $ka\tau_0 \le \sqrt{\pi}x_\star^2$, and Ly$\alpha$ photons escape at $x_{\rm esc} = x_\star$. While the slab is optically thin at $x_\star$, depending on whether the clumps are optically thin or thick at $x_{\rm esc}$, the escape of Lyα photons is described by the homogeneous and porous regime, respectively. Again we estimate the transition to arise when both regimes contribute equally to the flux of escaping Lyα photons,

$$\frac{F_{\rm por}}{F_{\rm hom}} = \frac{t_{\rm exc}}{t_{\rm rw}} = \frac{\sqrt{\pi}x_\star^4}{k^2 a\tau_0 f_c} = 1. \quad (32)$$

We yield the critical number of clumps that mark the transition from the porous to the homogeneous regime as

$$f_{c,\rm crit} = \frac{x_\star}{k(1 - e^{-\tau_{0,\rm cl}})}. \quad (33)$$

We note that if clumps are optically thin at line centre ($\tau_{0,\,\rm cl} < 1$), not every clump encounter leads to a scattering event; this reduces the number of clumps encountered by a factor $1 - e^{-\tau_{0,\rm cl}}$. The emerging Lyα line profile accounts again for Lyα photons escaping in homogeneous ($J_{\rm slab}$; see equation 30) and porous regime ($J_{\rm centre}$, see equation 19),

$$J_{\rm ph}(\tau_0, x) = (1 - f_{\rm por})J_{\rm slab}(T, \tau_0, x) + f_{\rm por}J_{\rm centre}(T, x). \quad (34)$$

The ratio between the two different escape regimes is then again given by assuming that most Lyα photons escape at the peak frequencies,

$$f_{\rm por} = \frac{F_{\rm por}/F_{\rm hom}}{F_{\rm por}/F_{\rm hom} + \frac{J_{\rm centre}(0)}{2(J_{\rm slab}(\tau_0,x_{\rm esc})+J_{\rm slab}(\tau_0,-x_{\rm esc})}}. \quad (35)$$

(iii) *Division between porous and random-walk regime:* As the optical depth of the already optically thick clumps, $\tau_{0,\,\rm cl}$, exceeds the optical depth of the slab, τ_0, a fraction of the Lyα photons traverse the slab without scattering, leaving the random and entering the porous regime. This transition occurs as the number of clumps encountered by the Lyα photons becomes less than unity, $f_c < 1$.

[3] The geometrical factor k is often not explicitly included in the literature when displacements and excursion times are discussed.

3.1.5 Ionizing escape fraction dependent in a clumpy/homogeneous medium

For this Lyα line profile model, we assume a model similar to the so-called picket fence model (Heckman et al. 2011). Here a fraction f_{esc} of the ionizing radiation escapes through low-density channels, while the other fraction of ionizing photons is absorbed by the dense shell. Correspondingly, the Lyα photons escaping through the channels scatter only a few times, while those escaping through the shell encounter many scattering events. For optically thin channels, the former gives rise to a single-peaked Lyα line centred around $x = 0$, while the latter creates a broader double-peaked Lyα line (assuming a homogeneous slab model with peaks at x_{esc}).

Here we assume the channels to be fully ionized and a fraction f_{esc} of the Lyα photons to escape through them without scattering. As a consequence, the other fraction of Lyα photons encounter the shell with an optical depth of

$$\tau_{\rm shell} = \frac{\tau_0}{1 - f_{\rm esc}}, \tag{36}$$

at their first scattering. τ_0 is the optical depth as derived in Section 3.2. We make the simplifying assumption that those photons traverse the shell without being scattered into the channels. In reality a fraction of these photons encounter a lower optical depth when traversing the empty channels on their scattering path out of the slab, leading to a profile closer centered around $x = 0$. Thus, the emerging Lyα line profile, which we assume to be

$$J(\tau, x) = f_{\rm esc}\ J_{\rm slab}(T, 0, x) + (1 - f_{\rm esc})\ J_{\rm shell}(T, \tau_{\rm shell}, x) \tag{37}$$

$$J_{\rm shell} = f_{\rm shell} J_{\rm slab}(T, 0, x) + (1 - f_{\rm shell})\ J_{\rm slab}(T, \tau_{\rm shell}, x) \tag{38}$$

represents a lower limit to the fraction of Lyα photons escaping close to the resonance. We derive f_{shell} by choosing a clump optical depth $\tau_{0,\,cl}$ and use equations (31) and (35) for the random-homogeneous ($ka\tau_0 > \sqrt{\pi}x_\star^3$) or porous-homogeneous ($ka\tau_0 < \sqrt{\pi}x_\star^3$) transitions.

3.2 Optical depth and H I column density

To derive the Lyα line profile emerging from a simulated galaxy for the models described in Sections 3.1.2 (*Clumpy* model) and 3.1.5 (*Porous* model), we yield the optical depth at the Lyα line centre from our simulated galaxies as

$$\tau_0 = \frac{4}{3} f_c \tau_{0,\rm cl} = N_{\rm HI}\sigma_{\rm HI}. \tag{39}$$

$\tau_{0,\,cl}$ is a free parameter in our model and reflects the optical depth of a dense clump in the ISM. We will use this parameter to calibrate our model to the observed Lyα LFs at $z = 6.6-8$ in Section 4. To obtain a rough estimate of $\tau_{0,\,cl}$, we consider the median mass (M_{cl}) and size (r_{cl}) of molecular clouds ($M_{cl} \simeq 10^5\ M_\odot$ and $r_{cl} = 20$ pc) resembling the mass and size of possible dense structures in the ISM,

$$\tau_{0,\rm cl} = \sigma_{\rm HI}\frac{M_{\rm cl}}{r_{\rm cl}}. \tag{40}$$

To obtain the optical depth τ_0, we derive the neutral hydrogen column density N_{HI} from the initial gas mass, M_g^{i} as

$$N_{\rm HI} = \xi\frac{3M_{\rm HI}}{4\pi r_g^2 m_{\rm H}} = \xi\frac{3X_c(1-Y)M_g^{\rm i}}{4\pi\,(4.5\lambda r_{\rm vir})^2 m_{\rm H}} = \xi\frac{3 f_m M_{\rm vir}}{81\pi\lambda^2 r_{\rm vir}^2 m_{\rm H}}$$
$$= \xi\left(\frac{9\pi^2 H_0^2\Omega_m}{G}\right)^{2/3}(1+z)^2 M_{\rm vir}^{1/3}\frac{3 f_m}{81\pi\lambda^2 m_{\rm H}}. \tag{41}$$

Here r_g describes the gas radius, for which we assume $r_g = 4.5\lambda r_{\rm rvir}$. X_c and Y are the cold gas and helium mass fractions, respectively. Gas accretion and SN feedback processes determine the relation between the initial gas mass and the halo mass, f_m, which ranges typically between $\sim 10^{-3}$ for low-mass galaxies to $\sim 10^{-1}$ for more massive galaxies. ξ is a geometrical correction factor that depends on τ_0 and the dust optical depth at the Lyα resonance τ_d. Its maximum values is 0.35 and we describe its derivation and dependencies in Appendix A. For the cosmological parameters assumed in this paper, we yield

$$N_{\rm HI} = 6.5\times 10^{17}{\rm cm}^2\ (1+z)^2\frac{\xi f_m}{\lambda^2}\left(\frac{M_{\rm vir}}{10^8 {\rm M}_\odot}\right)^{1/3}. \tag{42}$$

3.3 Dust attenuation

We employ two different dust models. The first one links the Lyα escape fraction to the escape fraction of UV continuum photons, f_{esc}^{c}. The second one is more complex. It assumes a clumpy medium where the attenuation of Lyα by dust follows different relations in the regimes identified in Gronke (2017). Both models assume a slab-like geometry and we describe their details in the following.

3.3.1 Simple attenuation model

In this model, we assume that (i) dust and gas are perfectly mixed, (ii) the dust distribution is slab-like, and (iii) the dust attenuation of Lyα photons is proportional to the dust attenuation of UV continuum photons. The escape fraction of Lyα photons, $f_{esc}^{Ly\alpha}$, is then directly related to the escape of UV continuum photons, f_{esc}^{c}, derived in Section 2.2.2.

$$f_{\rm esc}^{\rm Ly\alpha} = p\ f_{\rm esc}^{\rm c}. \tag{43}$$

We use p as a free parameter to obtain the observed Lyα LFs at $z = 6.6$–7.3.

3.3.2 Refined attenuation model

This model assumes that dust and gas are perfectly mixed and distributed in clumps. The dust attenuation of Lyα photons depends on the total optical depth of the dust, $\tau_{d,\,total}$, the optical depth of a clump, $\tau_{d,\,cl}$, and the number of clumps, f_c, encountered along the sightline from the midplane to the surface of the slab. We derive its value by estimating the dust absorption cross section. Following Galliano (2022) and assuming the radius and density of graphite/carbonaceous grains (see Section 2.2.2), we assume $\kappa_{abs} \simeq \frac{Q_{abs}}{as} \simeq 2\times 10^5$ cm^2 g^{-1} with $Q_{abs} \simeq 1$ being the absorption efficiency.[4]

$$\tau_{\rm d,total} = \frac{4}{3} f_c \tau_{\rm d,cl} = \xi\ \frac{3}{4\pi}\frac{M_{\rm d}}{r_{\rm d}^2}\kappa = \frac{M_{\rm d}}{M_{\rm HI}}\frac{\kappa_{\rm abs} m_{\rm H}}{\sigma_{\rm HI}}\tau_0. \tag{44}$$

The resulting estimates for $\tau_{d,\,total}$ and $\tau_{d,\,cl}$ allow us to compute the Lyα escape fractions in the different escape regimes as follows.

3.3.3 Free-streaming regime

In an optically thin slab ($\tau_0 < 1$), the Lyα photons stream through $\sim f_c$ clumps. On their way, they are attenuated by the dust in clumps

[4]We note that this is in rough agreement with the dust extinction cross sections of the Small and Large Magellanic clouds $\kappa_{ext} = \sigma_d/m_H = \sigma_{d,ref}\frac{M_z}{M_d\ Z_{ref} m_H} \simeq 4\times 10^5$ cm^2 g^{-1}, with the extinction efficiency $Q_{ext} = Q_{abs} + Q_{sca}$ being given by the similar sized absorption (Q_{abs}) and scattering efficiencies (Q_{sca}) at Lyα, a dust-to-metal mass ratio $M_d/M_Z \simeq 0.25$, $\sigma_{ref} \simeq 4\times 10^{-22}$cm^2, and $Z_{ref} \simeq 0.0025$ for SMC and $\sigma_{ref} \simeq 7\times 10^{-22}$cm^2 and $Z_{ref} \simeq 0.005$ for LMC (for further explanations, see Laursen 2010).

and hence, the total dust optical depth determines the Lyα escape fraction, $\tau_{\rm d,\,total}$, as

$$f_{\rm esc}^{\rm Ly\alpha,fs} = \exp\left(-\tau_{\rm d,total}\right) = \exp\left(-\frac{4}{3} f_c \tau_{\rm d,cl}\right). \quad (45)$$

We note that in this regime, the number of clumps along the sightline f_c and clump optical depth $\tau_{0,\,\rm cl}$ are degenerate.

3.3.4 *Random walk regime*

In the random walk regime, both the slab and individual clumps are optically thick ($\tau_{\rm cl} \geq 1$). As a result, Lyα photons escape by mostly being scattered by the clumps, and their escape fraction is determined by the number of clumps encountered along their random walk, $N_{\rm cl}(f_c)$, and the absorption probability per clump interaction ϵ. According to Hansen & Oh (2006), it is then given by

$$f_{\rm esc}^{\rm Ly\alpha,rw} = f_{\rm HO06} = \frac{1}{\cosh(\sqrt{2N_{\rm cl}(f_c)\,\epsilon})} \quad (46)$$

We assume $N_{\rm cl}(f_c) \simeq \frac{3}{2} f_c^2 + 2 f_c$ as found in Gronke (2017). The scaling of N_c with f_c also agrees with the findings in Hansen & Oh (2006) and prefactors vary slightly due to different geometries of the scattering surface. However, since ϵ is sensitive to how deep the photons permeate the clump, it depends non-trivially on the clump optical depth and movement. For simplicity, we assume $\epsilon = 1 - \exp(-\tau_{\rm d,cl}^{2/3}/10)$, which we have found to be in rough agreement with the fits and results shown in Gronke (2017).

3.3.5 *Homogeneous regime*

In the homogeneous regime, the slab is optically thick ($\tau_0 \geq 1$), while the individual clumps are optically thin at the escape frequencies ($\tau_{\rm cl}(x_{\rm esc}) < 0$). During their initial random walk, the Lyα photons scatter with $N_{\rm cl}(f_{c,\,\rm crit})$ clumps before they diffuse into the wings and escape by free-stream through f_c clumps. The resulting Lyα escape fraction,

$$f_{\rm esc}^{\rm Ly\alpha,hom} = f_{\rm HO06}(f_{c,\rm crit}) \exp(-\tau_{\rm d,total}), \quad (47)$$

depends on $f_{c,\,\rm crit}$ and $\tau_{\rm d,\,total}$, with $f_{c,\,\rm crit}$ being determined by τ_0 and $\tau_{0,\,\rm cl}$,

$$f_{c,\rm crit} = \begin{cases} \frac{2}{\sqrt{3k\pi^{1/4}}}\sqrt{a\tau_{0,\rm cl}} & \text{for } ka\tau_0 \geq \sqrt{\pi}x_{\rm max}^2 \\ \frac{x_{\rm max}}{k\left(1-e^{-\tau_{0,\rm cl}}\right)} & \text{for } ka\tau_0 < \sqrt{\pi}x_{\rm max}^2 \end{cases}. \quad (48)$$

3.3.6 *Porous regime*

In the porous regime, the individual clumps are optically thick ($\tau_{0,\,\rm cl} \geq 1$), but only a fraction $1 - e^{-f_c}$ of the Lyα photons will encounter a clump along their sightlines. The other fraction of Lyα photons does not interact with any clumps and is thus not attenuated by dust as they escape the slab.[5]

$$f_{\rm esc}^{\rm Ly\alpha,por} = e^{-f_c} + \left[1 - e^{-f_c}\right] \exp\left(-\frac{4}{3} f_c \tau_{\rm d,cl}\right). \quad (49)$$

[5]We note that our expression is here a lower limit of $f_{\rm esc}^{\rm Ly\alpha,por}$ as we assume the Lyα radiation interacting with clumps to experience attenuation as if they streamed through the clump. It might be more appropriate to consider these Lyα photons to be absorbed as in the random walk regime, $f_{\rm esc}^{\rm Ly\alpha,por} = e^{-f_c} + \left[1 - e^{-f_c}\right] \frac{1}{\cosh(\sqrt{2N_{\rm cl}(f_c)\,\epsilon})}$; however in practice, galaxies in the porous regime have not much, if any, dust.

3.3.7 *Emerging Lyα line profile models*

We briefly summarize the combinations of Lyα line and dust attenuation models that we will investigate in this paper.

3.3.8 *Gaussian*

The Lyα line profile emerging from a galaxy is given by the central Gaussian Lyα line profile (Section 3.1.1). To account for the attenuation by dust, we apply the Lyα escape fraction, $f_{\rm esc}^{\rm Ly\alpha}$, derived in our simple dust model (Section 3.3.1) to all frequencies x.

3.3.9 *Clumpy*

This model assumes an shell of dusty gas clumps, whereas gas and dust are perfectly mixed. It combines the Lyα line model described in Section 3.1.2 with the refined dust model depicted in Section 3.3.2. The gas in the galaxies is assumed to have a temperature of $T = 10^4$ K.[6] In contrast to the *Gaussian* model, we dust-attenuate the Lyα line of each escape regime (homogeneous, random, porous) by its corresponding escape fraction $f_{\rm esc}^{\rm Ly\alpha}$. The emerging Ly$\alpha$ line profile is then the superposition of the line profiles of all relevant escape regimes,

$$L_\alpha^{\rm gal}(x) = f_{\rm esc,slab}^{\rm Ly\alpha}(1 - f) J_{\rm slab}(x) + f_{\rm esc,centre}^{\rm Ly\alpha} f\, J_{\rm centre}(x), \quad (50)$$

with $f_{\rm esc,slab}^{\rm Ly\alpha} = f_{\rm esc,hom}^{\rm Ly\alpha}$, and $(f,\, f_{\rm esc,centre}^{\rm Ly\alpha})$ given by $(1,\, f_{\rm esc,fs}^{\rm Ly\alpha})$, $(f_{\rm rw},\, f_{\rm esc,rw}^{\rm Ly\alpha})$ or $(f_{\rm por},\, f_{\rm esc,por}^{\rm Ly\alpha})$ depending on the total and clump optical depths τ_0 and $\tau_{0,\,\rm cl}$.

3.3.10 *Porous*

This model is very similar to the *Clumpy* model. However, it considers the shell of clumps to be pierced with gas and dust-free channels through which a fraction $f_{\rm esc}$ of the Lyα photons escape without scattering. It combines the Lyα line model described in Section 3.1.5 and assuming $\tau_{\rm channel}^{\rm LyC} = 0$ with the refined dust model depicted in Section 3.3.2. Again we assume the gas in the galaxy to be heated to a temperature of $T = 10^4$ K, and the Lyα line of each escape regime (homogeneous, random, porous) to be dust-attenuated by its corresponding escape fraction $f_{\rm esc}^{\rm Ly\alpha}$. The emerging Ly$\alpha$ line profile is again a superposition of the Lyα photons escaping through the channels and the clumpy shell,

$$L_\alpha^{\rm gal}(x) = f_{\rm esc}\, J_{\rm channel}(x) + (1 - f_{\rm esc})\, J_{\rm shell}(x) \quad (51)$$

$$J_{\rm channel}(x) = J_{\rm centre}(x) \quad (52)$$

$$J_{\rm shell}(x) = f_{\rm esc,slab}^{\rm Ly\alpha}\, (1 - f)\, J_{\rm slab}(x) + f_{\rm esc,centre}^{\rm Ly\alpha}\, f\, J_{\rm centre}(x) \quad (53)$$

with $f_{\rm esc,slab}^{\rm Ly\alpha} = f_{\rm esc,hom}^{\rm Ly\alpha}$, and $(f,\, f_{\rm esc,centre}^{\rm Ly\alpha})$ given by $(1,\, f_{\rm esc,fs}^{\rm Ly\alpha})$, $(f_{\rm rw},\, f_{\rm esc,rw}^{\rm Ly\alpha})$, or $(f_{\rm por},\, f_{\rm esc,por}^{\rm Ly\alpha})$ depending on the total and clump optical depths τ_0 and $\tau_{0,\,\rm cl}$. We note that τ_0 exceeds the τ_0 value in the *Clumpy* model when $f_{\rm esc} > 0$ (see equation 36), as the same amount of gas and dust is distributed over a smaller solid angle.

[6]We have chosen $T = 10^4$ K for simplicity. If we were to assume the virial temperature ($T_{\rm vir}$), the double-peak line profile would narrow as $T_{\rm vir}$ increases.

3.4 IGM attenuation

The Lyα radiation escaping from a galaxy is attenuated by the H I it encounters along the line of sight from the location of emission, $r(z_{\rm em})$, to the location of absorption, $r(z_{\rm obs})$. Expressing the frequency ν of a photon in terms of its rest-frame velocity $x = v/b = (\nu_\alpha/\nu - 1)c/b$ relative to the Lyα line centre, the transmitted fraction of radiation at frequency x is given by

$$T_{\alpha,x}(x) = \exp\left[-\tau_\alpha(x)\right] \tag{54}$$

$$\tau_\alpha(x) = c\int_{z_{\rm em}}^{z_{\rm obs}} \sigma_0\ \phi(x + x_{\rm p}(r(z)))\ \frac{n_{\rm HI}(r(z))}{(1+z)H(z)}\ {\rm d}z. \tag{55}$$

Here τ_α describes the optical depth to Lyα, while $n_{\rm HI}(r)$ and $v_{\rm p}(r) = bx_{\rm p}(r)$ the H I density and peculiar velocity (in the rest-frame of the emitted Lyα radiation) at a physical distance r from the emitter, respectively. σ_0 is the specific absorption cross section, described in the cgs system as

$$\sigma_0 = \frac{\pi e^2 f}{m_e c^2} = \frac{3\lambda_\alpha^2 A_{21}}{8\pi}, \tag{56}$$

where $f = 0.4162$ is the oscillator strength, e the electron charge, m_e the electron mass, $\lambda_\alpha = 1216$ Å the wavelength of a photon at the Lyα line centre, and $A_{21} = 6.265\times10^8\ {\rm s}^{-1}$ the Einstein coefficient for spontaneous emission of Lyα photons. $\phi(x)$ depicts the Lyα profile for absorption and is given by a Voigt profile consisting of a Gaussian core

$$\phi_{\rm Gauss}(x) = \frac{\lambda_\alpha}{\sqrt{\pi}b}\ \exp\left(-x^2\right) \tag{57}$$

$$b = \sqrt{\frac{2k_B T_{\rm IGM}}{m_H}}, \tag{58}$$

and Lorentzian damping wings

$$\phi_{\rm Lorentz}(x) = \frac{A_{21}\lambda_\alpha^2}{4\pi^2(x\ b)^2 + \frac{1}{4}A_{21}^2\lambda_\alpha^2}. \tag{59}$$

Here b is the Doppler parameter, $T_{\rm IGM}$ the temperature of the IGM, k_B the Boltzmann constant, and $m_{\rm H}$ the mass of a hydrogen atom. While pressure line broadening is unimportant in regions of low H I density and the profile can be approximated by the Gaussian core, the absorption in the Lorentzian damping wings is non-negligible in regions of high H I density. In practice, we mimic the Voigt profile by assuming the Gaussian core profile $\phi(x) = \phi_{\rm Gauss}(x)$ for $|x| < x_\star$ and the Lorentzian wing profile $\phi(x) = \phi_{\rm Lorentz}(x)$ otherwise. Fitting to numerical results yields the transition frequency as

$$x_\star = 0.54\log_{10}\left(\frac{b}{\rm cm\,s^{-1}}\right) \tag{60}$$

for temperatures between $T = 0.01$ K and 10^8 K.

Our calculations of T_α include the Hubble flow and peculiar velocities $v_{\rm p}$: outflows (inflows) of gas from a galaxy that correspond to positive (negative) $v_{\rm p}$ values will redshift (blueshift) the Lyα photons and lead to an increase (decrease) in T_α. For each galaxy in a simulation snapshot we derive T_α along all directions along the major axes (i.e. along and against the x, y and z axes). By stepping through the simulation box that is divided into 512 cells on the side (and each cell having a size of 461ckpc), we derive the $n_{\rm HI}(r)$ and $v_{\rm p}(r)$ profiles from the ASTRAEUS ionization and VSMDPL density and velocity grids. For any galaxy, we start the profiles at the galaxy position $r_{\rm em} = 0$ and end them once the highest frequency $x_{\rm max} = v_{\rm max}/b = 40$ tracked in our Lyα line profiles has redshifted out of absorption at $r \simeq v_{\rm max}/[H_0\Omega_m^{1/2}(1+z)^{1/2}] \simeq 13.6/(1+z)^{1/2}$ cMpc.

Table 2. Parameters for our three different Lyα line profile models.

Parameter	Scenario	*Gaussian*	*Clumpy*	*Porous*
$\tau_{0,\,\rm cl}$	MHDEC	–	1.2×10^6	2.4×10^6
$\tau_{0,\,\rm cl}$	MHINC	–	5×10^5	1.8×10^6
p	MHDEC	1.0	–	–
p	MHINC	1.4	–	–
T	All	10^4 K	10^4 K	10^4 K

We assume $T_{\rm IGM} = 10^4$K in ionized and $T_{\rm IGM} = 10^2$ K in neutral regions. Since the Lyα line redshifts out of resonance very quickly (the light travel time for distance r at $z = 7$ is less than 2 Myr, shorter than the simulation time steps), a single simulation snapshot suffices for computing the T_α values of the galaxies in that snapshot. We also assume periodic boundary conditions when computing T_α.

Finally, we derive the observed, i.e. dust and IGM attenuated, Lyα luminosity and line profile along each major axes (resulting in six lines of sight) as

$$L_{\alpha,x}(x) = L_\alpha^{\rm gal}(x)\ T_{\alpha,x}(x) = L_\alpha^{\rm intr}\ f_{\rm esc}^{{\rm Ly}\alpha}\ J(x), \tag{61}$$

where $f_{\rm esc}^{{\rm Ly}\alpha}$ and $J(x)$ are the respective Lyα escape fraction and line profile for a one of the models as outlined in Section 3.3.7. The total observed Lyα luminosity L_α and total fraction of Lyα radiation transmitted through the IGM are yielded when integrating the respective quantity over the frequency x.

$$L_\alpha = \int_{-\infty}^{\infty} L_{\alpha,x}(x)\ {\rm d}x \tag{62}$$

$$T_\alpha = \int_{-\infty}^{\infty} T_{\alpha,x}(x)\ {\rm d}x. \tag{63}$$

In the following, we use all lines of sight as independent probes when line-of-sight-sensitive Lyα quantities are analysed.

We derive the observed Lyα luminosities (L_α) for all galaxies at z = 20, 15, 12, 10, 9, 8, 7.3, 7, and 6.6 for any combination of emerging Lyα line model (*Gaussian*, *Clumpy*, *Porous*) and reionization scenario (MHDEC, MHCONST, MHINC). Free model parameters (p for the *Gaussian* model, $\tau_{0,\,\rm cl}$ for the *Clumpy* and *Porous* models) have been chosen to visually best-fit the observed Lyα LFs at $z \simeq$ 6.7, 7.0, and 7.3 (see Table 2). For simplicity and better comparison we assume in all models the gas in galaxies to have the temperature of photo-ionized gas, $T = 10^4$ K. Moreover, we note that since the MHCONST scenario represents an intermediate case and provides no further insights, we limit our discussion to the MHDEC and MHINC scenarios in the remainder of this paper.

4 NUMBERS AND PROPERTIES OF LYα EMITTING GALAXIES

In this section, we aim to identify which physical process – the intrinsic Lyα production ($L_\alpha^{\rm intr}$), the absorption by dust within the galaxies ($f_{\rm esc}^{{\rm Ly}\alpha}$), or the scattering by H I in the IGM (T_α) – dominates the observed Lyα emission. To this end, we analyse (i) how the IGM attenuation profile $T_\alpha(x)$ depends on galaxy mass and the $f_{\rm esc}$-sensitive ionization topology, (ii) how the Lyα line profiles emerging from a galaxy depend on the density and velocity distributions of gas and dust within a galaxy and $f_{\rm esc}$, and how much it affects the fraction of Lyα radiation that is transmitted through the IGM, and (iii) to which degree the $f_{\rm esc}$ dependency of $L_\alpha^{\rm intr}$, $f_{\rm esc}^{{\rm Ly}\alpha}$, and T_α leave characteristic imprints in the Lyα LFs and the population emitting visible Lyα emission.

4.1 The transmission through the IGM

We start by discussing the frequency-dependent IGM transmission $T_{\alpha,x}$ shown in the top row of Fig. 3. These profiles depend solely on the underlying ionization topology and density distribution of the IGM. From the different panels depicting the average $T_{\alpha,x}$ in different halo mass bins of width $\Delta\log_{10}M_\mathrm{h} = 0.125$, we see that all $T_{\alpha,x}$ profiles follow a common trend: $T_{\alpha,x}$ decreases towards higher frequencies with an stronger decline around the Lyα resonance ($x = 0$). Photons bluewards the Lyα resonance redshift into the Lyα resonance as they propagate through the IGM and have the largest likelihood to be absorbed by the H I present. Photons redwards the Lyα resonance are also redshifted, but their likelihood of being absorbed by H I decreases significantly as their energy drops.

In each panel in the top row of Fig. 3 we show $T_{\alpha,x}$ for the two reionization scenarios MHDEC (yellow/orange/brown lines) and MHINC (blue lines) and redshifts $z = 8.0$, 7.3, 7.0, 6.6 (bright to dark lines as redshift decreases). In general, i.e. for both reionization scenarios and all halo masses, $T_{\alpha,x}$ increases as the ionized regions grow around galaxies and the IGM is increasingly ionized (bright to dark lines). First, a larger ionized region shifts not only the point of strongest Lyα absorption to higher frequencies x but also reduces the absorption in the damping wings of the Lyα absorption profile. Secondly, lower H I fractions in ionized regions diminish the number of H I atoms absorbing Lyα photons.

These two mechanisms shape $T_{\alpha,x}$ redwards and bluewards the Lyα resonance. As Lyα photons travel through the IGM and redshift, photons emitted at frequencies $x \gtrsim 0$ see the Gaussian core of the Lyα absorption profile $\phi(x)$ and are absorbed by H I abundances as low as $\chi_\mathrm{H\,I} \gtrsim 10^{-4}$; thus they are sensitive to the residual H I fraction in ionized regions. Correspondingly, we see in Fig. 3 that $T_{\alpha,x}$ increases for $x \gtrsim 0$ with decreasing redshift as the photoionization rate around galaxies increases and lowers the residual H I fraction in ionized regions. However, photons emitted at frequencies $x \lesssim 0$ are absorbed by the damping wings of the Lyα absorption profile $\phi(x)$. Since the Lyα absorption cross section is lower in the damping wings, the abundance of H I needs to be significantly higher for Lyα photons to be absorbed; thus, as the sizes of ionized regions decrease, photons emitted at these frequencies are increasingly absorbed by the neutral regions located beyond the ionized regions around the emission sites. For this reason, we find $T_{\alpha,x}$ for $x \lesssim 0$ to increase as the sizes of the ionized regions around galaxies rise with increasing halo mass and decreasing redshift. The rising sizes of ionized regions also become manifest in the shift of the frequency at which $T_{\alpha,x}$ has a value of 0.5 to higher frequencies.

Its dependence on the size of the ionized regions around galaxies makes $T_{\alpha,x}$ a tracer of the ionization topology: our two extreme reionization scenarios where f_esc either increases (MHINC, blue dotted lines) or decreases (MHDEC, yellow to brown solid lines) with rising halo mass M_h exhibit very different ionization topologies (see Fig. 6). These differences are imprinted in $T_{\alpha,x}$ as follows. First, since in the MHINC scenario the higher f_esc values of more massive galaxies ($M_h \gtrsim 10^{10}\,\mathrm{M}_\odot$) raise the photoionization rate within ionized regions (leading to lower $\chi_\mathrm{H\,I}$ values, also seen in Fig. 2 at $z \lesssim 6$), the corresponding $T_{\alpha,x}$ values are higher bluewards the Lyα resonance than in the MHDEC scenario. Moreover, in the MHINC scenario, reionization proceeds faster, leading to the Universe being more ionized at $z < 7$, and the bias of the ionizing emissivity towards more massive galaxies grows with time, raising the photoionization rate in the ionized regions. Both effects contribute to the relative increase in $T_{\alpha,x}$ from MHDEC to MHINC to rise towards lower redshifts bluewards the Lyα resonance. Secondly, as the size of the ionized regions around galaxies is imprinted in $T_{\alpha,x}$ redwards the Lyα resonance, MHINC shows lower (higher) $T_{\alpha,x}$ values at $z \gtrsim 7$ ($z \lesssim 7$) than the MHDEC scenario for galaxies with $M_h < 10^{11}\,\mathrm{M}_\odot$: At $z \gtrsim 7$, ionized regions become increasingly smaller towards lower mass halos ($M_h \lesssim 10^{9.5}\,\mathrm{M}_\odot$) and higher redshifts as the corresponding f_esc values and global ionization fraction decrease. However, at $z \lesssim 7$, this trend reverses as the ionized regions become large enough for the red wing of the Lyα to be redshifted out of the absorption resonance of the Gaussian core. Towards more massive halos and higher global ionization fractions, $T_{\alpha,x}$ becomes sensitive to the residual H I fraction in ionized regions (c.f. $T_{\alpha,x}$ in MHINC (light blue dotted line) exceeds $T_{\alpha,x}$ in MHDEC (yellow solid line) at $z = 6.6$). It is interesting to note that the respective $T_{\alpha,x}$ values are very similar in both reionization scenarios, despite the f_esc values of more massive halos ($M_h > 10^{10}\,\mathrm{M}_\odot$) differing by about one order of magnitude or more.

4.2 The Lyα line profiles and LFs

The Lyα line profile emerging from a galaxy represents a quantity that (i) is shaped by the density and velocity distribution of gas and dust within the galaxy and (ii) affects which fraction of the Lyα radiation escaping from a galaxy is transmitted through the IGM. In this section, for our three models of the emerging Lyα line profiles, we discuss the following: (i) How do the assumed gas and dust distributions affect the attenuation of Lyα by dust in a galaxy and the emerging Lyα line profile? (ii) How does the Lyα line profile affect the Lyα transmission through the IGM? And, since the LF of the intrinsic Lyα luminosity (L_α^intr) will be steeper for the scenario where f_esc increases (MHINC) than when it decreases (MHDEC) with rising halo mass, (iii) which characteristics are required for the Lyα line profiles of the simulated galaxy population to reproduce the observed Lyα LFs?

4.2.1 The Gaussian model

The *Gaussian* line model centers the Lyα line at the Lyα resonance. The second row in Fig. 3 shows that its width increases as the rotational velocity of a galaxy increases with rising halo mass. Both the increase in the line width and the size of the ionized region surrounding the galaxy lead to higher IGM transmission values of Lyα radiation as galaxies become more massive (c.f. third row in Fig. 3). At the same time, the fraction of Lyα photons that escape from the galaxies drops as the abundance of dust increases. We use the ratio between the Lyα and UV continuum escape fractions to adjust the Lyα luminosities emerging from the galaxies and fit the observed Lyα LFs in each of our reionization scenarios. In the MHINC scenario the more massive galaxies – that dominate the observed Lyα LF – have higher f_esc values than in the MHDEC scenario; to compensate the corresponding lower L_α^intr values (and steeper slope of the intrinsic Lyα LF), we need a higher $f_\mathrm{esc}^\mathrm{Ly\alpha}/f_\mathrm{esc}^\mathrm{UV}$ ratio (1.4) than in the MHDEC scenario (1.0). Despite this compensation, the slopes of the observed Lyα LFs at $z \lesssim 8$ (c.f. left-hand panel in Fig. 4) is still steeper for the MHINC than for the MHDEC scenario.

4.2.2 The Clumpy model

In the *Clumpy* model, the clumpiness of the gas in the shell and the attenuation by dust molecules in these clumps determine the shape of the Lyα line profile. We note that in the following clumpiness describes the number of clumps in the dusty gas shell, i.e. a higher clumpiness corresponds to fewer clumps and thus a higher ratio between the clump ($\tau_{0,\mathrm{cl}}$) and total line optical depth (τ_0). We find the following characteristic trends for the Lyα line profile: first, the clumpier the gas in the shell is, the more Lyα radiation escapes around the Lyα resonance (profile showing a central peak), and the

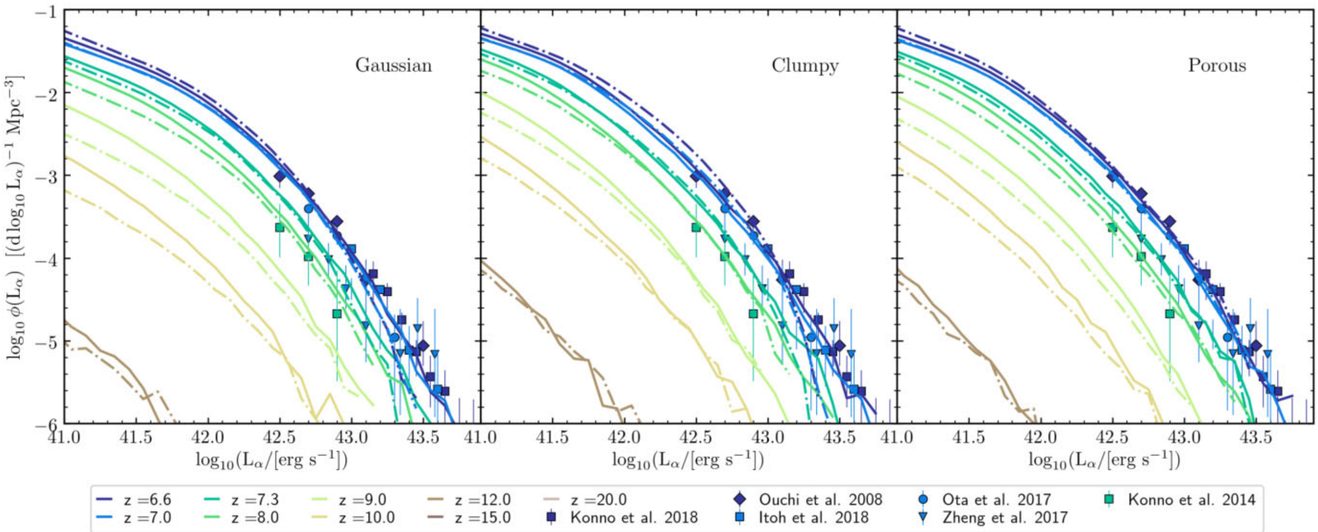

Figure 4. Observed Lyα LFs at $z = 20, 15, 12, 10, 9, 8, 7.3, 7, 6.6$ for a homogeneous static gas shell (left), a clumpy gas shell (centre), and a clumpy gas shell with holes through which Lyα radiation escapes without scattering. Solid (dash–dotted) lines show results for the reionization scenario where $f_{\rm esc}$ decreases (increases) with halo mass $M_{\rm h}$. Observational data points are from Ouchi et al. (2010), Konno et al. (2014, 2018), Ota et al. (2017), Zheng et al. (2017), and Itoh et al. (2018).

fewer Lyα photons escape through excursion or via the wings (double peak profile); secondly, when assuming the same clump size for all galaxies – as we do in this paper – the gas clumpiness decreases as galaxies become more massive and contain more gas. Thus, from low-mass to more massive galaxies, we find the Lyα line profile to shift from a central peak dominated to a double-peak domintaed profile (see the fourth row in Fig. 3 from left to right), reflecting the transition from the random to the homogeneous regime (see Section 3.1.2). This transition also goes in hand with an increased transmission through the IGM, which we can see when comparing the Lyα profiles emerging from galaxies (fourth row) with those after having traversed the IGM (fifth row in Fig. 3). The Lyα luminosity at $x = 0$ decreases by $\sim$0.5 orders of magnitude for all halo masses (from $10^{41.6}$ erg s^{-1} to $10^{41.1}$ erg s^{-1} for $M_h \simeq 10^{11}\,\mathrm{M}_\odot$ and from $10^{39.7}$ erg s^{-1} to $10^{39.2}$ erg s^{-1} for $M_h \simeq 10^{9}\,\mathrm{M}_\odot$ for e.g. MHDEC model), while the peak Lyα luminosity of the red wing decreases only about $\lesssim 0.3$ orders of magnitude at all halo masses. While the blue wing is similarly or more attenuated than the central peak in the IGM, the total fraction of Lyα radiation transmitted through the IGM for a fully double-peaked profile exceeds that of profiles with a central peak component. Furthermore, as the galaxies' gravitational potentials flatten with decreasing redshift, τ_0 decreases and leads to (i) a narrower double-peak profile (following the dependence of the peak position on $\tau_0^{1/3}$) and (ii) a stronger central peak (the gas becomes clumpier as the ratio $\tau_{0,\,\rm cl}/\tau_0$ increases).

A change in the clumpiness of the gas and dust shell (or clump optical depth $\tau_{0,\,\rm cl}$ and $\tau_{\rm d,\,cl}$) goes not only in hand with a change in the Lyα profile affecting T_α but also an altered attenuation of the escaping Lyα radiation by dust. Thus, adjusting the clump optical depth allows us to enhance and reduce the Lyα luminosities and reproduce the observed Lyα LFs: as we increase the size of the clumps, i.e. increase $\tau_{0,\,\rm cl}$, Lyα photons will scatter with fewer clumps, leading to (i) a higher fraction $f_{\rm esc}^{\rm Ly\alpha}$ escaping, and (ii) a higher fraction escaping at the Lyα resonance, which again leads to stronger attenuation by H I in the IGM. However, we note that once the emerging Lyα profile is fully double peaked, the attenuation by H I in the IGM can not be further decreased (by changing the injected Lyα line profile). The observable Lyα emission can only be enhanced by decreasing $\tau_{0,\,\rm cl}$ as long as the $f_{\rm OH06}(f_{\rm c,\,crit})$ factor in $f_{\rm esc}^{\rm Ly\alpha,hom}$ remains significantly below unity. With the observed Lyα LF being dominated by the more massive galaxies ($M_h \gtrsim 10^{10}\,\mathrm{M}_\odot$, as we will discuss in the next section), we find the $\tau_{0,\,\rm cl}$ value to reflect the factor by which the bright end of the intrinsic Lyα LF needs to be reduced to reproduce the observational Lyα LF data points (filled points in Fig. 4). As the intrinsic Lyα LFs is lower at the bright end in the MHINC scenario, a lower $\tau_{0,\,\rm cl}$ value (5×10^5) is required than for the MHDEC scenario (1.2×10^6). Nevertheless, the slopes of the resulting observed Lyα LFs at $z \lesssim 8$ keep the trends of the intrinsic Lyα LFs, with the bright ends of the Lyα LFs being steeper in the MHINC than in the MHDEC scenario.

4.2.3 The Porous model

The *Porous* model represents a refinement of the *Clumpy* model. It adds gas-free channels through which Lyα and ionizing photons escape freely. This explains why, to first order, we find the trends in the last two rows of Fig. 3 to be similar to those in the fourth and fifth rows: a lower clumpiness of gas and dust in the shell induces a stronger prevalence of the double-peak component in the Lyα line profile emerging from a galaxy, enhancing the IGM transmission T_α and absorption by dust within the galaxy, and causing the corresponding Lyα LFs to shift to lower values. On the other hand, it differs from the *Clumpy* model substantially, as $f_{\rm esc}$ determines the minimum fraction of Lyα radiation that escapes at the Lyα resonance and contributes to the central peak in our modelling. Hence, as long as $\tau_{0,\,\rm cl}$ remains above the $\tau_{0,\,\rm cl}$ value that leads to the same fraction of Lyα escaping in the central peak than given by $f_{\rm esc}$ (referred to as $\tau_{0,\rm cl}^{\rm crit}$ in the following), the *Porous* model inherits the trend of the *Clumpy* model. As $\tau_{0,\,\rm cl}$ drops below $\tau_{0,\rm cl}^{\rm crit}$, a further decrease in $\tau_{0,\,\rm cl}$ affects the Lyα line profile emerging from a galaxy hardly, and once the $f_{\rm HO06}(f_{\rm c,\,crit})$ factor of $f_{\rm esc}^{\rm Ly\alpha,hom}$ approaches unity, the observed Lyα LFs remain 'fixed'. The resulting upper limit of $f_{\rm esc}^{\rm Ly\alpha}$ (determined by the total dust optical depth $\tau_{\rm d,\,total}$) is essential, as together with $L_\alpha^{\rm intr}$ it provides an upper limit to $f_{\rm esc}$ values that fit the observed Lyα LFs. We find this upper limit to be about $f_{\rm esc} \sim 0.5$ in our ASTRAEUS model.

Due to their opposing dependencies of $f_{\rm esc}$ with halo mass, the Lyα profiles in the *Porous* model show the most noticeable differences between the MHDEC and MHINC scenarios among our three Lyα line profile models. While the double-peak component is more prominent in the most massive galaxies ($M_h \simeq 10^{11}$ M$_\odot$) in the MHDEC scenario, the central peak is slightly stronger in the MHINC scenario. To fit the observed Lyα LFs, we find that we require for both reionization scenarios a more clumpy gas and dust distribution than in the *Clumpy* model, i.e. a (higher) $\tau_{0,\,\rm cl}$ value of $1.8 - 2.4 \times 10^6$. These increased $\tau_{0,\,\rm cl}$ values enhance the corresponding $f_{\rm c,\,crit}$ values and thus the dust attenuation in the homogeneous regime giving rise to the double-peak components and counteract the increased escape close to the Lyα resonance. This model-integrated correlation between $f_{\rm esc}^{\rm Ly\alpha}$ and $f_{\rm esc}$ counteracts the trend of flattening (steepening) the slope of the intrinsic Lyα LFs due to $f_{\rm esc}$ decreasing (increasing) with rising halo mass: If $f_{\rm esc}$ is low (high), more (less) Lyα radiation is subject to dust attenuation. This model feature explains why the observed Lyα LFs of the MHINC simulation are shallower than in the *Clumpy* model and hardly changes for the MHDEC simulation due to its low $f_{\rm esc}$ values for more massive galaxies.

As the dust composition and absorption cross section for Lyα remain highly uncertain during the EoR, we note that a lower (higher) dust absorption cross section $\kappa_{\rm abs}$ could still reproduce the observed Lyα LFs in our *Clumpy* and *Porous* models by raising (decreasing) the clump optical depth $\tau_{0,\,\rm cl}$. However, this would go along with an enhanced (reduced) double-peak and reduced (enhanced) central-peak component in the average Lyα line profile emerging from galaxies.

Finally, we briefly comment on how our emerging and IGM-attenuated Lyα profiles compare to those obtained from radiative hydrodynamical simulations of clouds and small cosmological volumes ($\sim 10^3$cMpc^3). While the *Clumpy* and *Porous* reproduce the double- and triple-peak profiles and their dependence on $N_{\rm HI}$ and $f_{\rm esc}$ found in cloud simulations (Kimm et al. 2019; Kakiichi & Gronke 2021; Kimm et al. 2022) by construction, their Lyα line profiles differ from those obtained from the SPHINX simulation (Garel et al. 2021). In SPHINX the median angle-averaged Lyα line profile has been found to be less double-peaked towards brighter galaxies, with the blue peak being seemingly increasingly suppressed. This is the opposite trend of our findings. The discrepancy lies in the differently assumed or simulated ISM structures: While our LAE models assume an idealized scenario of same-sized dusty gas clumps, the SPHINX simulation follows the formation of star-forming clouds within galaxies. With rising galaxy mass, we expect the simulated SPHINX galaxies to contain a higher number of star-forming clouds with various velocity and size distributions. A single or very few star-forming clouds – as found in low-mass galaxies – will give rise to a double-peaked Lyα line profile. Adding the profiles of multiple/many star-forming clouds at different velocities will give rise to increasingly more complex Lyα line profiles as galaxies become more massive. Adjusting our current Lyα line profile models to the complex structure of the ISM will be the subject of future work.

4.3 The dependence of Lyα properties on halo mass

In this section, we provide a more detailed discussion of how the intrinsic Lyα luminosity ($L_\alpha^{\rm intr}$), the Lyα escape fraction, the Lyα transmission through the IGM, the observed Lyα luminosity, and Lyα equivalent width depend on halo mass and evolve with redshift for the different reionization scenarios. To this end, we show these quantities as a function of halo mass for both reionization scenarios (MHDEC: yellow/orange/brown lines; MHINC: blue lines) and redshifts $z \simeq 8$, 7.3, 7, 6.6 in Fig. 5 and list the corresponding average H I fractions in Table 3. Solid and dot–dashed lines in Fig. 5 depict the median value for galaxies in the given halo mass bin, and shaded regions indicate the range spanned by 68 per cent of the values. For line-of-sight-dependent Lyα properties (T_α, L_α, EW$_\alpha$), we include all six lines of sight.

4.3.1 *Intrinsic Lyα luminosity $L_\alpha^{\rm intr}$*

As the most recent star formation dominates the production of ionizing photons within galaxies, we find $L_\alpha^{\rm intr}$ to follow the SFR–$M_{\rm h}$ relation (for a detailed discussion, see Hutter et al. 2021a). While the range of SFR values is broad for low-mass halos ($M_h \lesssim 10^{9.5}$ M$_\odot$) where SN feedback drives stochastic star formation, the SFR–$M_{\rm h}$ relation becomes tighter towards more massive galaxies as SN feedback ejects an increasingly lower fraction of gas from the galaxy. Being mainly produced by recombining hydrogen atoms within a galaxy, the Lyα radiation produced within the galaxy correlates with the escape fraction of ionizing photons as $1 - f_{\rm esc}$. As we can see from the first row in Fig. 5, this dependency on $f_{\rm esc}$ leads to higher (lower) Lyα luminosities for more massive galaxies, lower (higher) Lyα luminosities for low-mass galaxies, and thus a shallower (steeper) LFs in the MHDEC (MHINC) scenario.

4.3.2 *Lyα escape fraction $f_{\rm esc}^{\rm Ly\alpha}$*

As the dust content in galaxies increases with their mass, we find $f_{\rm esc}^{\rm Ly\alpha}$ to decrease with rising halo mass at all redshifts and for all Lyα line models. However, the different assumed distributions of dust and their resulting attenuation of Lyα radiation lead to differences in the details of this global trend: first, the *Gaussian* model shows a steeper decline in $f_{\rm esc}^{\rm Ly\alpha}$ for galaxies with $M_h \gtrsim 10^{10.5}$ M$_\odot$ than the *Clumpy* and *Porous* models. Secondly, $f_{\rm esc}^{\rm Ly\alpha}$ is always higher in the MHINC than in the MHDEC scenario. This is necessary to reproduce the observed Lyα LFs by compensating the lower intrinsic Lyα luminosities with a more clumpy gas-dust distribution in the MHINC scenario. In case of the *Clumpy* model, it also highlights how a decrease in the clump optical depth by a factor $\sim$2 can increase $f_{\rm esc}^{\rm Ly\alpha}$ by reducing the fraction of Lyα photons escaping in the homogeneous regime (i.e. a decrease in $f_{\rm c,\,crit}$ and $\tau_{0,\,\rm cl}$ leads to a reduced number of clumps encounters $N_{\rm cl}$ and the clump albedo ϵ). Thirdly, for the MHDEC (MHINC) scenario, the $f_{\rm esc}^{\rm Ly\alpha}$ values show higher (lower) values in the *Porous* model than in the *Clumpy* model for $M_h \lesssim 10^{10}$ M$_\odot$. The reason for this difference is as follows. In both scenarios the higher $\tau_{0,\,\rm cl}$ values in the *Porous* model increase the dust attenuation of Lyα escaping in the homogeneous regime. But only a fraction $1 - f_{\rm esc}$ of the Lyα photons is subject to dust attenuation. This unattenuated escape of Lyα radiation imprints the mass-dependency of $f_{\rm esc}$ in the $f_{\rm esc}^{\rm Ly\alpha}$ values. However, for galaxies with $M_h \gtrsim 10^{10}$ M$_\odot$, this imprint ($f_{\rm esc}^{\rm Ly\alpha}$ enhancement in *Porous* model) is only visible in the MHINC scenario where $f_{\rm esc}$ values are sufficiently large (>0.1); in the MHDEC scenario $f_{\rm esc}$ values are too small.

4.3.3 *Lyα IGM transmission* T$_\alpha$

As outlined in Section 4.1, the surrounding ionized region (in particular its size and residual H I fraction) and the Lyα line profile emerging from a galaxy determine how much of a galaxy's escaping Lyα radiation is transmitted through the IGM.

For more massive galaxies with $M_h \gtrsim 10^{10}$ M$_\odot$, T_α is mainly shaped by the Lyα profile. This is because the ionized regions surrounding them are sufficiently large – due to their enhanced ionizing

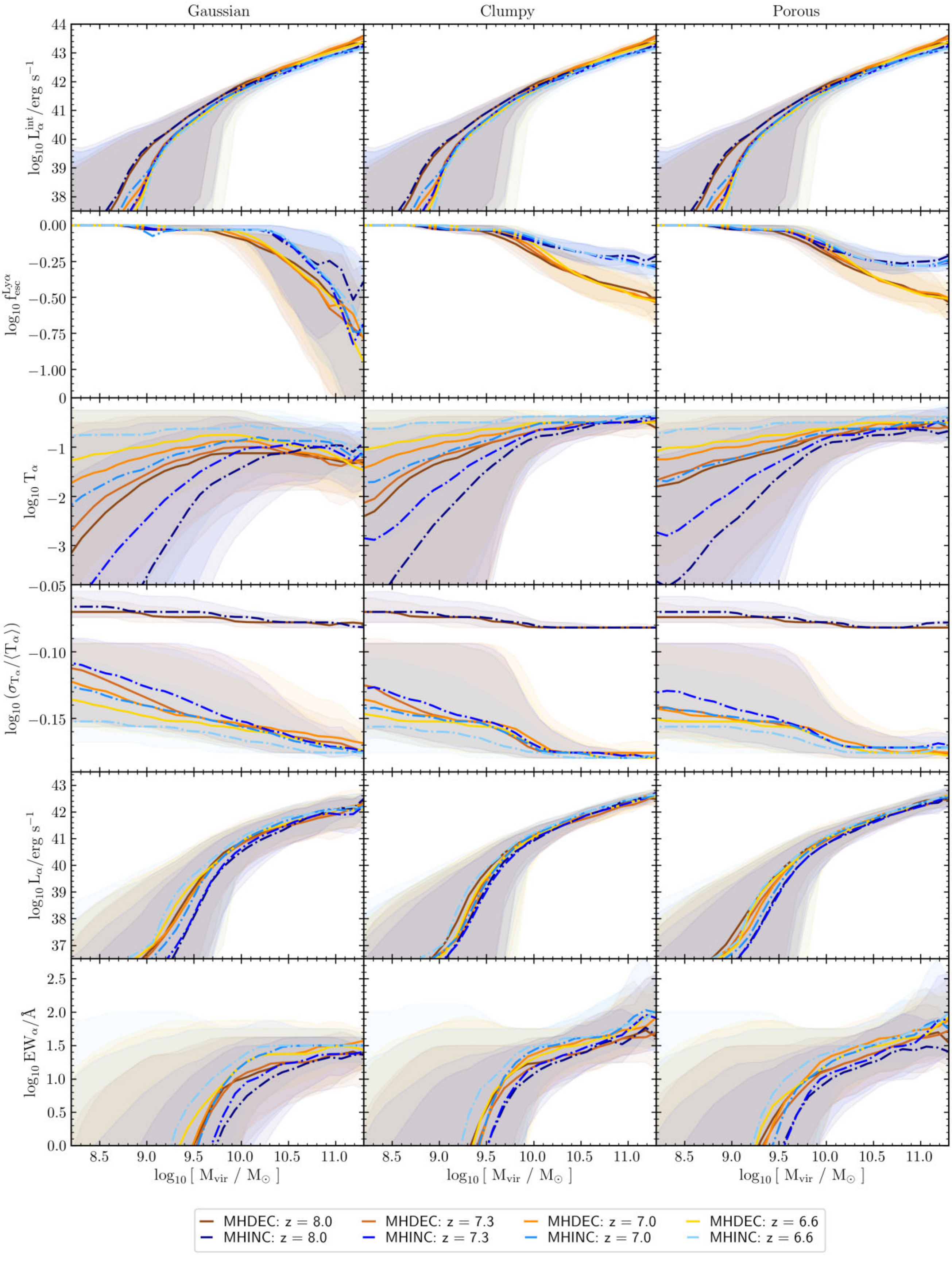

Figure 5. Median of indicated galactic properties (lines) and their $\sim 1.3\sigma$ uncertainties (shaded regions) as a function of halo mass M_h at $z = 8.0, 7.3, 7.0, 6.6$ for a homogeneous static gas shell. Solid (dash–dotted) lines show results for the reionization scenario where f_{esc} decreases (increases) with halo mass M_h.

Table 3. The evolution of the global H I fractions of the IGM for our reionization scenarios.

z	$\langle \chi_{\rm H\,I} \rangle^{\rm MHINC}$	$\langle \chi_{\rm H\,I} \rangle^{\rm MHDEC}$
8.0	0.84	0.71
7.3	0.69	0.59
7.0	0.52	0.49
6.6	0.23	0.34

emissivity and their clustered neighbourhood – for the Lyα radiation to redshift out of absorption. Hence, at these high halo masses, any trends in T_α reflect the ratio between the Lyα radiation escaping around the Lyα resonance and escaping through the wings: the more Lyα escapes in the central peak, the lower is the T_α value. Indeed, as can be seen in Fig. 5, the *Gaussian* model concentrating the emerging Lyα radiation around the Lyα resonance shows the lowest median T_α values at $M_h \gtrsim 10^{10}\,{\rm M}_\odot$ among all Lyα line profile models. In the *Clumpy* model, where the fraction of Lyα escaping through the wings increases with rising halo mass, we find the median T_α value to increase accordingly. This effect is more evident for the MHINC scenario as it transitions from a Lyα line profile with a dominating central peak at $M_h \simeq 10^{10}\,{\rm M}_\odot$ to one with a prevailing double peak component at $M_h \simeq 10^{11}\,{\rm M}_\odot$. The *Porous* model also confirms that T_α is highly sensitive to the Lyα line profile. In the MHINC scenario, the double peak component is weaker and increases less with halo mass, leading to slightly lower T_α values than in the *Clumpy* model for $M_h \simeq 10^{10-11}\,{\rm M}_\odot$ and T_α hardly changing with halo mass. In the MHDEC scenario, we see the same effect but to a lower degree.

However, for less massive galaxies ($M_h \lesssim 10^{10}\,{\rm M}_\odot$), T_α is more sensitive to the properties of their surrounding ionized regions. Since the ionized regions around less massive galaxies can differ significantly depending on their environment and phase in their stochastic star formation cycle [see Hutter et al. (2021b) and Legrand et al. (2023) for environment dependence], their T_α values span across an extensive range from as low as effectively zero to as high as $\simeq 70$ per cent. Nevertheless, the median T_α value shows a definite trend. It increases with rising halos mass for all models and at all stages of reionization. With increasing halo mass, galaxies are surrounded by larger ionized regions as they form more stars emitting ionizing photons and are more likely to be located in clustered regions that are reionized earlier. The larger the surrounding ionized regions are, the higher the transmission of Lyα radiation through the IGM. We can see this relationship when comparing the median T_α values of the MHDEC and MHINC simulations. In the MHDEC scenario low-mass galaxies are surrounded by larger ionized regions at $z \gtrsim 7$ than in the MHINC, causing their corresponding T_α values to be raised (c.f. orange/brown solid lines vs dark blue/blue lines in the third row of Fig. 5). At $z \lesssim 7$, however, reionization progresses faster and the photoionization rate in clustered ionized regions yields a lower residual H I fraction in the MHINC simulation, both leading to a higher median T_α value for the MHINC than MHDEC scenario at $z \simeq 6.6$. Finally, we briefly discuss how the Lyα line profile emerging from a galaxy affects T_α for less massive galaxies. From Fig. 5 we see that the T_α values differ between our three different Lyα line profile models: While at all stages of reionization the T_α values for $M_h \lesssim 10^{10}\,{\rm M}_\odot$ are very similar in the *Porous* and *Clumpy* model, the *Porous* model shows lower T_α values for $M_h \gtrsim 10^{10}\,{\rm M}_\odot$ at $z \lesssim 7$ than the *Clumpy* model in the MHINC scenario. This drop goes in hand with the increased central peak component in these more massive galaxies (c.f. Fig. 3 and the previous section). The median T_α values of the *Gaussian* model always lie below those of the *Clumpy* and *Porous* models; a larger fraction of Lyα radiation escapes closer to the Lyα resonance and is thus subject to stronger attenuation by the IGM.

4.3.4 *Variance of the IGM transmission along different lines of sight*

To investigate how strongly the transmission of Lyα radiation through the IGM depends on the direction, we show the standard deviation of T_α values over the six lines of sight aligning with the major axes in relation to the corresponding mean value, $\sigma_{T_\alpha}/\langle T_\alpha \rangle = \sqrt{\langle T_\alpha^2 \rangle - \langle T_\alpha \rangle^2}/\langle T_\alpha \rangle$, in the fourth row of Fig. 5. At all redshifts and for all models, $\sigma_{T_\alpha}/\langle T_\alpha \rangle$ decreases with rising halo mass and decreasing redshift for the following reason. As galaxies grow in mass, they produce more ionizing photons that can ionize larger regions around them and are also more likely to be located in more strongly clustered ionized regions, both enhancing and homogenizing Lyα transmission through the IGM along different lines of sight. However, we note that parts of the decrease of $\sigma_{T_\alpha}/\langle T_\alpha \rangle$ with decreasing redshift is also due to $\langle T_\alpha \rangle$ rising. Since it is hard to disentangle these two effects, we will focus on relative differences between the different reionization scenarios and Lyα line profile models. First, the more the emerging Lyα line profile is concentrated around the Lyα resonance, the more sensitive is T_α to the varying H I abundance around a galaxy, and the larger is the variance across different lines of sight (c.f. the higher $\sigma_{T_\alpha}/\langle T_\alpha \rangle$ values in the *Gaussian* compared to the other two models, and in the *Porous* compared to the *Clumpy* model for $M_h \gtrsim 10^{10.5}\,{\rm M}_\odot$ when central peak component dominates). Secondly, we focus on Lyα line profiles more sensitive to the environmental H I abundance of a galaxy (*Gaussian* model). When accounting for the $\langle T_\alpha \rangle$ values to be lower in the MHINC than in the MHDEC scenario at $z \simeq 7$ (see median T_α values in the third row of Fig. 5), we can deduce that the variance of T_α across different lines of sight is higher in the MHDEC than in the MHINC scenario. Indeed in the MHINC scenario, the shape of ionized regions is closer to spheres and less filamentary, which results in more 'homogeneous' T_α values.

4.3.5 *Observed Lyα luminosity* L_α

For any model and reionization scenario, the trend of L_α with rising halo mass depends on the respective trends of $L_\alpha^{\rm intr}$, $f_{\rm esc}^{\rm Ly\alpha}$, and T_α. Being surrounded by smaller ionized regions, the low T_α values of less massive galaxies ($M_h \lesssim 10^{10}\,{\rm M}_\odot$) strongly suppress and shape their emerging Lyα radiation. In contrast, the T_α values of more massive galaxies ($M_h \gtrsim 10^{10}\,{\rm M}_\odot$) show only weak trends with halo mass and similar values throughout reionization. For this reason, the trends of their L_α values with halo mass are predominantly shaped by the corresponding trends of $L_\alpha^{\rm intr}$ and $f_{\rm esc}^{\rm Ly\alpha}$. Though, for model parameters that reproduce the observed Lyα LFs, a relative increase (decrease) of $L_\alpha^{\rm intr}$ towards higher halo masses, such as in the MHINC (MHDEC) scenario, is compensated by an $f_{\rm esc}^{\rm Ly\alpha}$ that decreases more (less) strongly with halo mass. Nevertheless, the resulting relation between L_α and halo mass does not significantly change. It shows that only more massive galaxies where SN and radiative feedback do not considerably suppress star formation exhibit observable Lyα emission of $L_\alpha \gtrsim 10^{41}\,{\rm erg\,s^{-1}}$.

4.3.6 *Observed Lyα equivalent width EW$_\alpha$*

We compute the Lyα equivalent width EW$_\alpha$ from L_α and the observed UV continuum luminosity at 1500 Å (L_c). The trend of the median EW$_\alpha$ with halo mass follows that of L_α, with median

EW_α values ranging from ∼5–30 Å for galaxies in $M_h \simeq 10^{10}\,\mathrm{M}_\odot$ halos to ∼25–100 Å for galaxies in $M_h \simeq 10^{11.3}\,\mathrm{M}_\odot$ halos. More massive galaxies with a strongly attenuated UV continuum – the fraction of these galaxies increases towards higher halo masses due to the higher abundance of dust – and high L_α values show EW_α values up to ∼300 Å in the *Clumpy* and *Porous* models. However, these high EW_α values are not present in the *Gaussian* model for the following reason: in this model, the escape of Lyα and UV continuum radiation differs just by a constant factor, while the dust attenuation of Lyα and UV continuum photons within a galaxy are not only linked via the dust mass in the *Clumpy* and *Porous* models.

In summary, we find that only more massive galaxies ($M_h \gtrsim 10^{10}\,\mathrm{M}_\odot$) where star formation is not substantially suppressed by SN and radiative feedback from reionization show significant Lyα emission of $L_\alpha \gtrsim 10^{41}$ erg s^{-1}. This limitation of observable Lyα emission to more massive galaxies allows the $f_{\rm esc}$-dependency of the intrinsic Lyα luminosity to be compensated by a weaker or stronger attenuation of Lyα by dust within a galaxy. If less massive galaxies were visible in Lyα, they would break this degeneracy as they would not contain enough dust to attenuate the Lyα radiation in all scenarios sufficiently.

5 THE SPATIAL DISTRIBUTION OF LYα EMITTING GALAXIES

In this section, we analyse where galaxies with observable Lyα emission are located in the large-scale structure and how their environment and Lyα luminosity distributions differ in our reionization scenarios (MHDEC and MHINC). For this purpose, we discuss the environment of Lyα emitting galaxies in terms of their large-scale spatial distribution (Fig. 6), their surrounding over-density $(1+\delta)$ and H I fraction ($\chi_{\rm H\,I}$) (Fig. 7), and their 3D autocorrelation functions (Fig. 8). As we yield very similar results for our three Lyα lines profile, we use the *Porous* model as a representative case.

5.1 The environment

Before detailing the location of Lyα emitting galaxies in the large-scale matter distribution, we briefly discuss the ionization structure of the IGM using Figs 6 and 7. Fig. 6 shows the ionization fields at $z = 8$, 7, and 6.7 for the MHDEC (top) and MHINC scenarios (bottom). As can be seen in this figure, if $f_{\rm esc}$ decreases with halo mass (MHDEC scenario), reionization is not only more extended but also ionized regions are on average smaller, follow more the large-scale density distribution and thus have less bubble-like shapes than if $f_{\rm esc}$ increases with halo mass (MHINC scenario). The grey contours in Fig. 7, showing the two-dimensional probability density distribution of the H I fraction ($\chi_{\rm H\,I}$) and over-density of the IGM at $z = 8$, 7.3, 7, and 6.7 (derived from all cells of the 512^3 ionization and density grids output by ASTRAEUS), complement the picture. These contours indicate that not only an increasing fraction of the volume becomes ionized as reionization progresses (from right to left) but also the $\chi_{\rm H\,I}$ values in ionized regions decrease [e.g. from $\chi_{\rm H\,I} \simeq 10^{-4}$ ($10^{-4.3}$) in average dense regions with $\log_{10}(1+\delta)+1$ at $z = 8$ to $\chi_{\rm H\,I} \simeq 10^{-4.7}$ ($10^{-5.2}$) at $z = 6.7$ for the MHDEC (MHINC) scenario]. The latter is because as galaxies grow in mass with decreasing redshift, their emission of ionizing photons increases, leading to a rise of the photoionization rates within ionized regions and thus lower $\chi_{\rm H\,I}$ values. Moreover, at the same time, as the photoionization rate within ionized regions becomes increasingly homogeneous, the enhanced number of recombinations in denser regions (for the detailed modelling description, see Hutter 2018) leads to a positive correlation between the H I fraction and density in ionized regions. However, the exact value of the photoionization rate within ionized regions and its spatial distribution depends strongly on the ionizing emissivities escaping from the galaxies into the IGM. If less clustered low-mass galaxies drive reionization – as in the MHDEC scenario (top row in Fig. 7) – the resulting photoionization rate is more homogeneous and lower than if the more strongly clustered massive galaxies are the main drivers of reionization (c.f. MHINC scenario in the bottom row of Fig. 7). The difference in the photoionization rate's magnitude explains the shift of the $\chi_{\rm H\,I}$ values by an order of magnitude to lower values in under-dense to moderately over-dense regions ($\log_{10}(1+\delta) \lesssim 1.2$) when going from the MHDEC to the MHINC scenario. In contrast, the more inhomogeneous distribution of the photoionization rate's values enhances this drop in $\chi_{\rm H\,I}$ in over-dense regions where the most massive galaxies are located.

As we can see from the red stars in Fig. 6 and coloured contours in Fig. 7, galaxies emitting Lyα luminosities of $L_\alpha \geq 10^{42}$ erg s^{-1} always lie in ionized regions for our Lyα line profile models. Although these galaxies trace the ionization topology, their populations (and thus locations) hardly differ for our two opposing reionization scenarios. This absence of a significant difference is due to their massive nature (see also e.g. Kusakabe et al. 2018); hence, all Lyα emitting galaxies lie in over-dense regions, with the ones brighter in Lyα located in denser regions (c.f. green to blue to red contours). The latter trend is mainly because more massive galaxies, which exhibit higher star formation rates and produce more ionizing and Lyα radiation, are located in denser regions.

5.2 The clustering

In this section, we address the question whether the Lyα luminosity-dependent distribution of LAEs could differ for reionization scenarios with opposing trends of $f_{\rm esc}$ with halo mass. For this purpose, we analyse the 3D autocorrelation function for LAE samples with different minimum Lyα luminosities (Fig. 8). We define a galaxy to be an LAE if it has an observed Lyα luminosity of $L_\alpha \geq 10^{42}$ erg s^{-1}.

Before we discuss the differences between our opposing $f_{\rm esc}$ descriptions, we give a brief overview of the global trends and their origins. First, as predicted by hierarchical structure formation, all autocorrelation functions in Fig. 8 decrease from small to large scales, implying stronger clustering of galaxies on small scales than on large scales. Secondly, the dropping amplitude of the LAE autocorrelation functions with decreasing redshift (from ochre to blue lines) reflects the growth and increasing ionization of ionized regions. Thirdly, since the L_α value of a galaxy is strongly correlated to its halo mass in our galaxy evolution model, selecting galaxies with increasingly brighter Lyα luminosities (left to right in Fig. 8) corresponds to selecting more massive galaxies. The latter explains the increasing amplitude and stronger clustering. Comparing the correlation functions of the L_α selected galaxies with those of LBGs (galaxies with $M_{\rm UV} \geq -17$) shows that the Lyα selected galaxies are more massive than our LBGs (solid grey lines). It also shows that the decrease in the clustering of LAEs is partially due to galaxies of a given mass becoming a less biased tracer of the underlying density field as the density of the Universe drops with decreasing redshift.

Comparing the autocorrelation functions of our two opposing $f_{\rm esc}$ descriptions, we find that the MHINC scenario (dotted lines) has higher autocorrelation amplitudes than the MHDEC scenario (solid lines) throughout reionization and for all minimum Lyα luminosities studied. This difference decreases towards larger scales. The reason

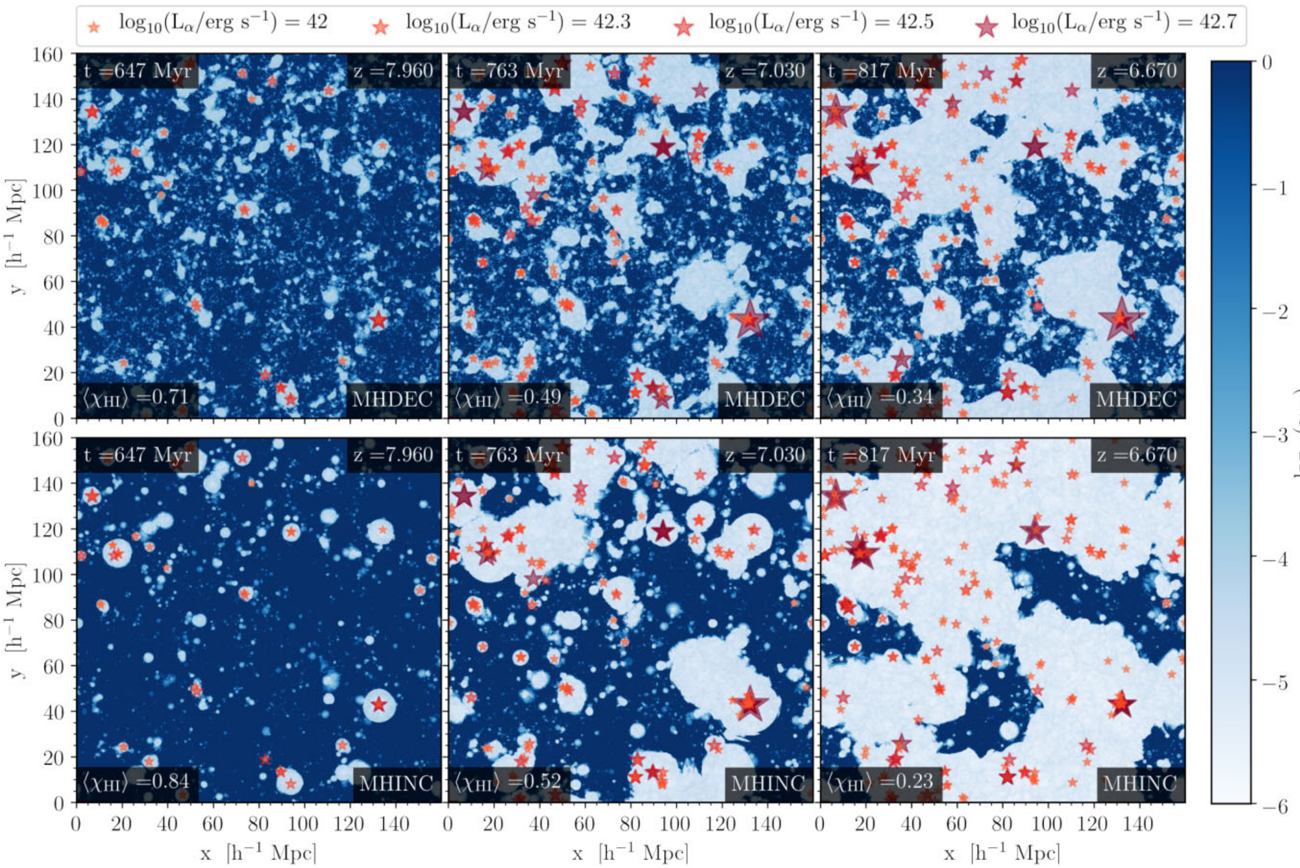

Figure 6. Neutral hydrogen fraction fields at $z = 8.0$ (left), $z = 7.0$ (centre), and $z = 6.6$ (right) for the MHDEC (top) and MHINC *Porous* models (bottom). We show a $1.6h^{-1}$cMpc-thick (5 cells) slice through the centre of the simulation box. The blue color scale depicts the volume-averaged value of the neutral fraction in each cell. Red stars show the location of LAEs, with their sizes and colour scale encoding the observed Lyα luminosity along the z-direction.

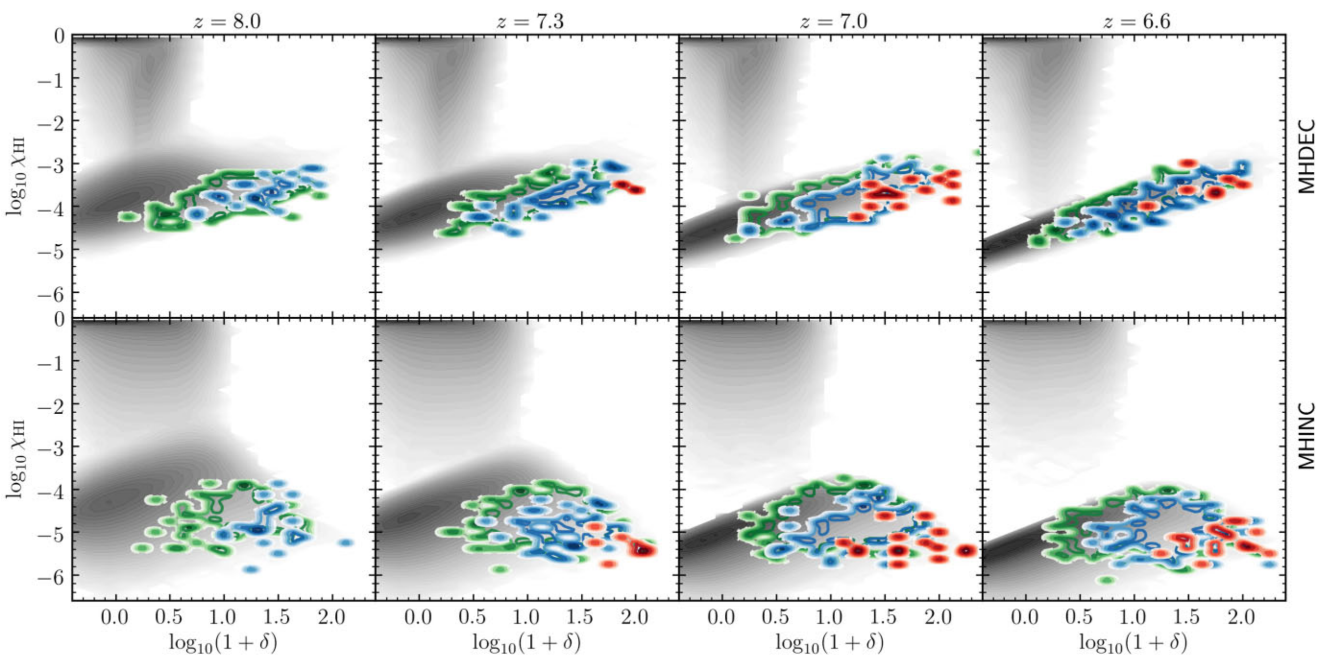

Figure 7. 2D probability distribution in $\chi_{\rm H\,I}$ and overdensity for all simulation cells (grey) and galaxies with $L_\alpha \geq 10^{42}$ erg s^{-1} (green), $L_\alpha \geq 10^{42.5}$ erg s^{-1} (blue), and $L_\alpha \geq 10^{43}$ erg s^{-1} (red) in the *Porous* model. The top (bottom) row shows results for the reionization scenario where $f_{\rm esc}$ decreases (increases) with halo mass $M_{\rm h}$.

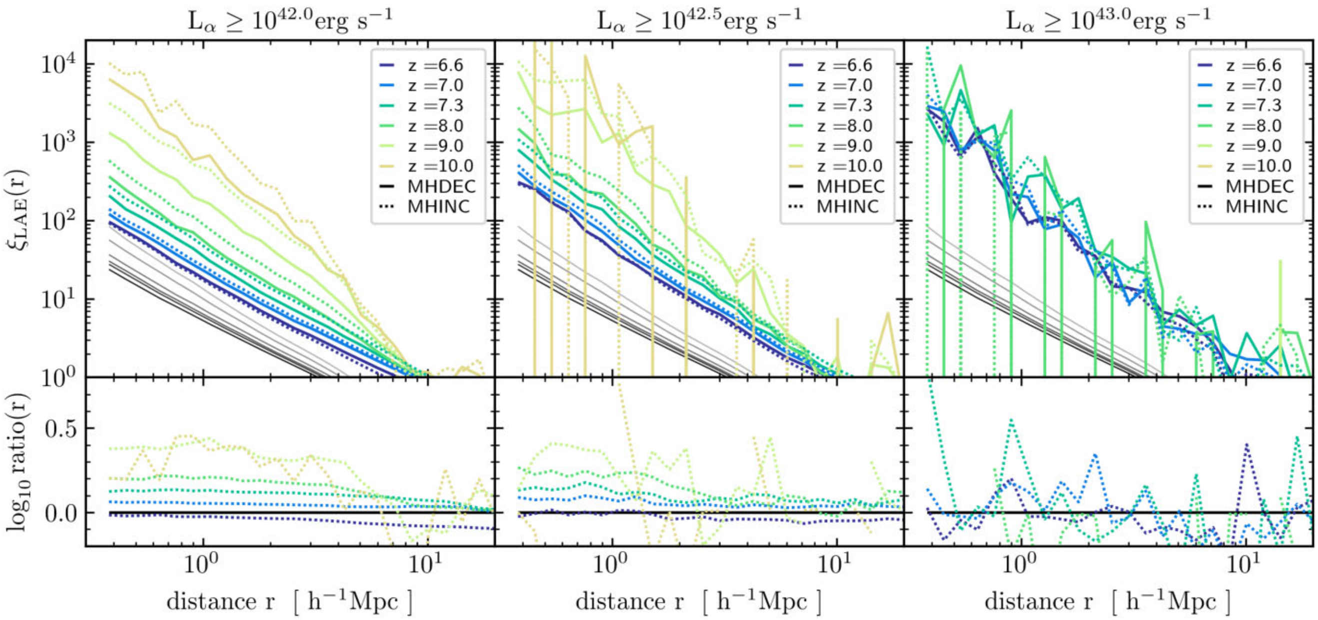

Figure 8. Top panels: 3D correlation function of galaxies that exceed an observed Lyα luminosity of $L_\alpha > 10^{42}$ erg s^{-1} (left), $L_\alpha > 10^{42.5}$ erg s^{-1} (centre), and $L_\alpha > 10^{43}$ erg s^{-1} (right) at $z = 10, 9, 8, 7.3, 7, 6.6$ for the *Porous* model. Solid (dash–dotted) lines show results for the reionization scenario where $f_{\rm esc}$ decreases (increases) with halo mass $M_{\rm h}$ and assumes $\tau_{0,\,\rm cl} = 4 \times 10^5$ (2×10^5). The grey to black lines indicate the corresponding LBG ($M_{\rm UV} < -17$) 3D correlation functions from $z = 10$ to 6.6. Bottom panels: Ratio between the 3D LAE correlation functions of the MHINC and the MHDEC scenario at fixed redshifts.

for these higher amplitudes is twofold: On the one hand, the MHINC scenario has a lower global average ionization fraction at $z \gtrsim 7$ than the MHDEC scenario (see Fig. 2). Its ionized regions are located around more biased tracers of matter, i.e. more massive galaxies, leading to a stronger clustering. While the scenarios' difference in $\langle \chi_{\rm H\,I} \rangle$ reaches its maximum with $\sim$0.13 around $z \simeq 8$, the difference in the autocorrelation amplitudes rises even towards higher redshifts. This is because, with increasing redshift, galaxies of the same mass become more biased tracers of the underlying matter distribution. Thus, the same difference in $\langle \chi_{\rm H\,I} \rangle$ at higher $\langle \chi_{\rm H\,I} \rangle$ values leads to a larger difference in the clustering of LAEs, since the Lyα luminosity of a galaxy correlates strongly with its halo mass. We note that selecting LAEs with a higher minimum Lyα luminosity also corresponds to selecting more biased tracers and yields higher correlation amplitudes (c.f. different panels in Fig. 8). On the other hand, during the early stages of reionization, ionized regions grow preferentially around the most biased tracers of the underlying matter field (most massive galaxies) in the MHINC scenario. Thus, we would expect that, at the same $\langle \chi_{\rm H\,I} \rangle$ value, LAEs in this scenario are more clustered than LAEs in the MHDEC scenario where the $f_{\rm esc}$ decreasing with rising halo mass counteracts the biased growth of ionized regions. Indeed, at $z \lesssim 7$, the correlation amplitude in the MHINC scenario is higher or similar than in the MHDEC scenario, although the Universe is similarly or more ionized in the former, respectively. This difference becomes more apparent as we consider higher minimum Lyα luminosities of $L_\alpha > 10^{42.5}$ erg s^{-1}. It is driven by the higher photoionization rates in the ionized regions around massive galaxies.

We conclude that, since LAEs coincide with the most massive galaxies located in dense and ionized regions, their clustering is primarily a tracer of the global ionization state of the IGM. While the exact ionization topology at fixed $\langle \chi_{\rm H\,I} \rangle$ values has only a secondary effect on the clustering of LAEs during the second half of reionization, the spatial distribution of LAEs provides a relatively robust tool to map the detailed ionization history at early times.

6 THE RELATION OF LAES TO LBGS

In this section, we address the question of what defines whether an LBG shows Lyα emission and why the fraction of LBGs with observable Lyα emission changes as the observed UV continuum luminosity (at 1500 Å) or the minimum Lyα equivalent width, EW$_\alpha$, rise. For this purpose, we show both the fraction of LBGs with a Lyα equivalent width of at least EW$_\alpha \geq 25$ Å (top row) and EW$_\alpha \geq 50$ Å (central row) and the median EW$_\alpha$ value (bottom row) as a function of the UV continuum luminosity in Fig. 9.

For our three different Lyα line profile models, we find the median EW$_\alpha$ to exhibit similar values of $\sim$4–40 Å at all redshifts shown. Furthermore, the EW$_\alpha$ values range to lower values as galaxies become UV fainter. As galaxies become less massive, this spread in EW$_\alpha$ values reflects the increasingly broader range of star formation rate values to lower values, which traces back to the larger variety of mass assembly histories that increasingly include progenitors with particularly SN feedback-suppressed star formation. The $M_{\rm UV}$-dependency of the fraction of LBGs with Lyα emission ($f_{\rm LAE}$) also reflects this shift towards lower EW$_\alpha$ values (c.f. top and central row of Fig. 9): first, $f_{\rm LAE}$ decreases towards lower UV luminosities, and secondly, this decrease is stronger for lower than higher EW$_\alpha$ cuts. These trends imply that UV bright galaxies are more likely to show higher EW$_\alpha$ values for all our Lyα line profile models and reionization scenarios. For example, while only <20 per cent of galaxies with $M_{\rm UV} \simeq -18$ exceed EW$_\alpha > 25$ Å, >40 per cent of galaxies with $M_{\rm UV} \gtrsim -20$ exceed EW$_\alpha > 25$ Å and >5 per cent even EW$_\alpha > 50$ Å.

Moreover, both the slight rise of EW$_\alpha$ and $f_{\rm LAE}$ values with decreasing redshift and their variation among our different reionization scenarios can be attributed to the increasing fraction of Lyα radiation that is transmitted through the IGM as the Universe becomes more ionized (see T_α in Fig. 5). For example, at a given redshift $z > 7$ ($z < 7$), the EW$_\alpha$ and $f_{\rm LAE}$ values are on average higher (lower) in the MHDEC than in the MHINC scenario, which is due to a more

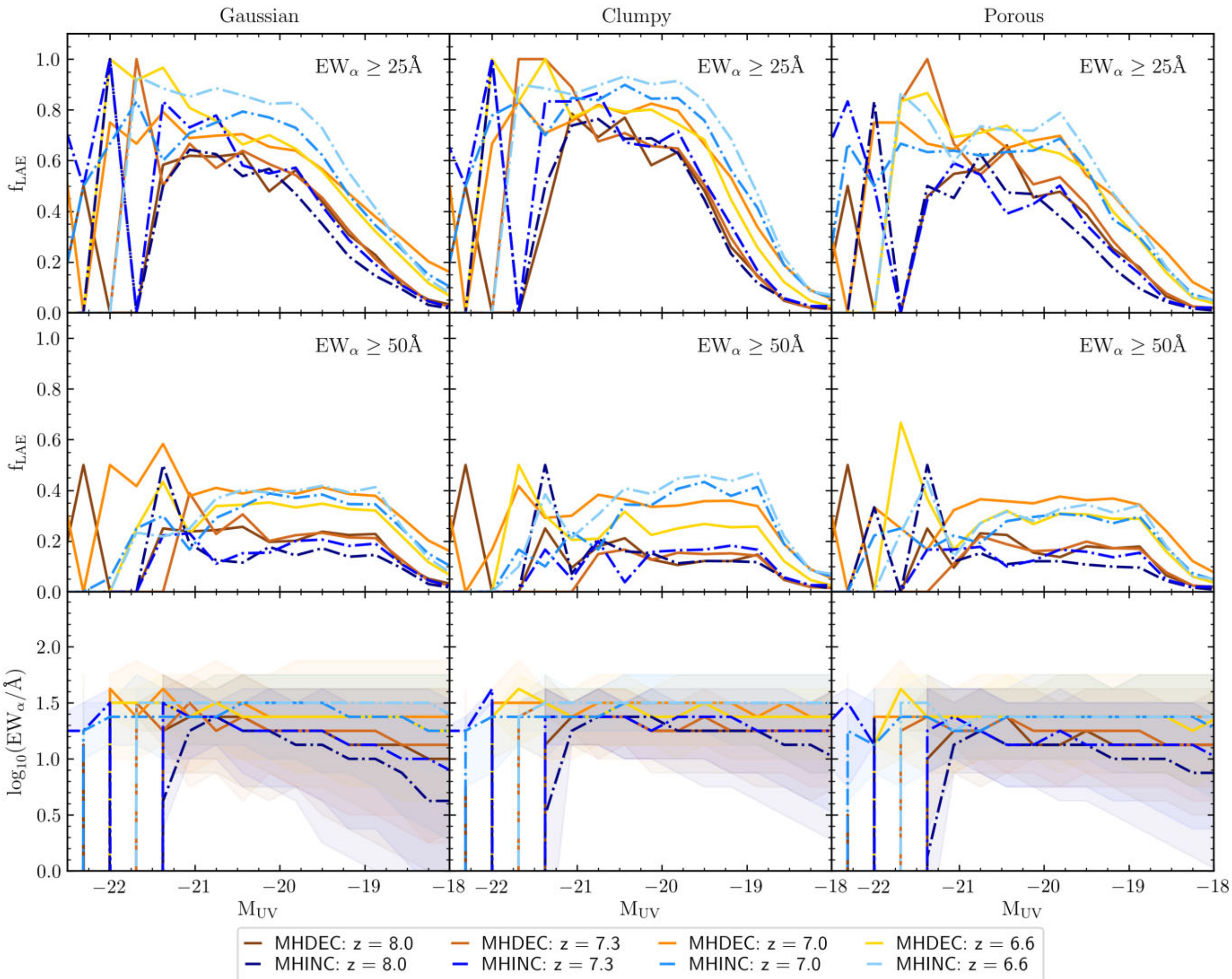

Figure 9. The two top rows depict the fraction of LBGs showing Lyα emissions with $L_\alpha \geq 10^{42}$ erg s^{-1} and EW$_\alpha$ exceeding the value marked in the panels for the different Lyα line profile models as marked. The bottom row shows the corresponding medians of the EW$_\alpha$ values (lines) and their $\sim 1.3\sigma$ uncertainties (shaded regions). Solid (dash–dotted) lines show results for the reionization scenario where $f_{\rm esc}$ decreases (increases) with halo mass $M_{\rm h}$.

(less) ionized IGM. Similarly, the lower EW$_\alpha$ values reached in the *Gaussian* model for UV fainter galaxies are due to the stronger absorption of Lyα radiation by H I in the IGM. Finally, we note that since in the *Clumpy* and *Porous* models the attenuation of the UV continuum and Lyα by dust do not necessarily correlate with each other (as e.g. parts of Lyα can escape via random walk), a few galaxies that are attenuated strongly in the UV but less in Lyα show high EW$_\alpha$ values of $\sim$1000 Å. Thus, the main driver of these high EW$_\alpha$ values is the dust attenuation of the UV continuum assumed in our models.

Comparing our fraction of LBGs showing Lyα emission with those obtained in observations (e.g. Schenker et al. 2012, 2014; Caruana et al. 2014; Pentericci et al. 2014, 2018; Mason et al. 2019), we find that (i) the observed trend of $f_{\rm LAE}$ decreasing towards higher UV luminosity agrees roughly with our results for EW$_\alpha$ > 50 Å but not for EW$_\alpha$ > 25 Å, and (ii) our $f_{\rm LAE}$ values are higher than those inferred from observations (again more so for EW$_\alpha$ > 25 Å than EW$_\alpha$ > 50 Å). These discrepancies hint either at our model predicting too high Lyα or too low UV luminosities (particularly for more massive galaxies) despite reproducing the observed Lyα and UV LFs, or observations missing bright LAEs. Interestingly, we find that the fraction of LBGs with high EW$_\alpha$ values of $f_{\rm LAE}(\mathrm{EW}_\alpha > 100\ \mathrm{Å}) \simeq 1-12$ per cent and $f_{\rm LAE}(\mathrm{EW}_\alpha > 240\ \mathrm{Å}) \lesssim 1$ per cent in the *Clumpy* and *Porous* models are in rough agreement with the results from deep MUSE observations at z = 3–6 that consider only LAEs with detected UV continuum (Kerutt et al. 2022). A higher abundance of high EW$_\alpha$ values has been found in various high-redshift LAE observations (e.g. Malhotra & Rhoads 2002; Shimasaku et al. 2006; Shibuya et al. 2018). Nevertheless, our $f_{\rm LAE}$ values agree roughly with the results from radiative hydrodynamical simulations post-processed with Lyα RT, such as SPHINX (c.f. fig. B1 in Garel et al. 2021).

7 CONCLUSIONS

We apply our new framework for LAEs to different reionization scenarios, and analyse how the escape fraction of H I ionizing photons, $f_{\rm esc}$, and its dependence on halo mass affect the luminosity-dependent number and spatial distributions of LAEs. Besides $f_{\rm esc}$ affecting the IGM ionization topology and the strength of the Lyα line produced in the ISM, its sensitivity to the density and velocity structure of ISM gas and dust has been found to correlate with the Lyα line profile emerging from a galaxy and the fraction of Lyα radiation escaping into the IGM. Notably, the emerging Lyα line profile reflects the attenuation by dust in the ISM and can also change the fraction of Lyα radiation that traverses the IGM unattenuated by H I. For this reason, we build an analytical model for Lyα line profiles that emerge from a Lyα source surrounded by a shell of dusty gas clumps interspersed with low-density channels. Our model reproduces the numerical RT results of a shell with dust gas clumps of different sizes as presented in Gronke (2017). By coupling this model to ASTRAEUS, a semi-numerical model coupling galaxy evolution and reionization self-consistently, we derive the Lyα line profiles emerging from the simulated galaxy population and explore the resulting large-scale distribution of LAEs for different dependencies of $f_{\rm esc}$ on halo mass (decreasing, constant, increasing) and Lyα line profiles (Gaussian profile, shell of dusty clumps interspersed with low-density channels or not). For this parameter space, we analyse the resultant ionization topologies, the dependencies of Lyα line profiles and Lyα properties on halo mass, and the location of galaxies with observable Lyα emission in the large-scale structure. Our main results are the following:

(i) For a shell consisting of clumps of the same size, the Lyα line profile emerging from a galaxy develops from a central peak at the Lyα resonance dominated to a double peak dominated profile as it becomes more massive. Adding low-density channels results in either a weakening of this trend, particularly as $f_{\rm esc}$ increases with rising halo mass.

(ii) In all reionization scenarios and Lyα line profile models, LAEs (galaxies with $L_\alpha \geq 10^{42}$ erg s^{-1}) are more massive galaxies with $M_h \gtrsim 10^{10}\,{\rm M}_\odot$. These galaxies exhibit continuous star formation and are biased tracers of the underlying mass density distribution. Both allow efficient transmission of the Lyα line through the IGM by facilitating the build-up of ionized regions around them. In contrast, less massive galaxies are surrounded by smaller ionized regions, which results in their Lyα radiation being significantly attenuated by H I in the IGM.

(iii) As LAEs are more massive galaxies and the most biased tracers of the underlying mass density distribution, they are located in the densest and most highly ionized regions. This finding holds for any inside-out reionization scenario where dense regions containing massive galaxies are ionized before under-dense voids and for Lyα line profiles exhibiting emission around/close to the Lyα resonance (see also Hutter et al. 2014, 2017). In such scenarios, the spatial distribution of LAEs is primarily sensitive to the global ionization fraction and only in second-order to the ionization topology or the trend of $f_{\rm esc}$ with halo mass.

(iv) As the observable Lyα LFs are composed of the Lyα emission from more massive galaxies, a decrease in their intrinsic Lyα luminosities (Lyα produced in the ISM) due to higher $f_{\rm esc}$ values can be compensated by reducing the attenuation by dust in the ISM (echoing the degeneracy found in Hutter et al. 2014). However, if $f_{\rm esc}$ exceeds $\sim$0.5 for the most massive galaxies ($M_h \gtrsim 10^{11}\,{\rm M}_\odot$), their intrinsic Ly$\alpha$ luminosity is too low to reproduce the observed Lyα LFs (see also Hutter et al. 2014).

All combinations of our reionization scenarios and Lyα line profile models result in Lyα and UV luminosities in reasonable agreement with observational constraints. However, although two of the three Lyα line profile models investigated use parameterizations of numerical Lyα RT simulation results, they represent idealized scenarios where the gas in each galaxy is distributed in clumps of the same mass. In reality, the density and velocity distributions of gas and dust in the ISM are more complex. First, the dusty gas clumps will have different masses, with a distribution close to that of a scale-free one at the massive end. Such a mass distribution would result in more massive galaxies having larger clumps than less massive galaxies, which again would lead to a homogenization of their Lyα line profiles where more massive (less massive) galaxies have an enhanced (weakened) central peak component and a weakened (enhanced) double-peak component. This change in the Lyα line profiles would result in the Lyα radiation being less (more) attenuated by dust in the ISM and traversing the IGM more (less) efficiently. Secondly, the medium between the clumps as well as the low-density channels might not be fully ionized (and very unlikely to be gas-free), causing the Lyα radiation escaping close to its resonance (central peak used in this work) to contribute to a narrower double-peak profile. Additionally, the gas may exhibit a turbulent velocity structure that could broaden the double-peak component. Both partially neutral low-density channels and an inhomogeneous velocity structure are likely to enhance the transmission of Lyα through the IGM. Thirdly, the attenuation of Lyα radiation by dust in the ISM depends on the distribution of dust in clumps. While our model assumes that gas and dust are perfectly mixed, a scenario where dust condensates in the centre surrounded by a shell of hydrogen gas would lower the absorption probability per clump and enhance the escape fraction of Lyα photons from a galaxy. Finally, simulations and observations of local analogues of high-redshift galaxies (i.e. regarding their extreme metallicity and ionization continuum properties) indicate that stellar feedback, especially that of SN, heat the gas and drive gas outflows (e.g. Gronke & Oh 2020; Carr et al. 2021; Kakiichi & Gronke 2021; Fielding & Bryan 2022; Hu et al. 2023; Xu et al. 2023). Indeed expanding homogeneous shell models have been used to fit observed Lyα profiles from high-redshift analogues (e.g. Gronke 2017; Orlitová et al. 2018); however, the inferred outflow velocities are on average lower than those inferred from UV absorption lines of low-ionization-state elements (Orlitová et al. 2018; Xu et al. 2023), hinting at more complex outflow geometries and kinematics of the neutral gas (see e.g. Carr et al. 2021; Blaizot et al. 2023). In general, outflowing neutral gas causes the Lyα photons to redshift, enabling easier escape from the galaxy and transmission through the IGM. While in principle outflows could enhance the observed Lyα emission, particularly from low-mass ($M_h \lesssim 10^{10}\,{\rm M}_\odot$) galaxies, and make the large-scale LAE distribution more sensitive to the ionization topology, their velocities or neutral gas fraction might be not sufficient to redshift the Lyα radiation out of absorption. In future work, we will extend our analytical Lyα line models towards more realistic outflow geometries and kinematics and explore whether this will affect the large-scale LAE distribution during reionization significantly.

Our Lyα line profile models, despite being limited by the simplified structure assumed for the ISM, represent a first step towards more complex analytical models for the Lyα line emerging from galaxies that are computationally efficient enough to derive the LAE populations in large cosmological simulations. To date, many models deriving the large-scale distribution of LAEs assume Lyα line profiles that arise from outflowing gas, consisting of a dominant red and a negligible blue peak (Mesinger et al. 2015; Mason et al. 2018; Weinberger, Haehnelt & Kulkarni 2019). However, such profiles are hardly seen in Lyα RT simulations of simulated galaxies (see e.g. Laursen, Sommer-Larsen & Razoumov 2011; Garel et al. 2021; Blaizot et al. 2023).

Finally, our finding that the spatial distribution of LAEs is not sensitive to the dependence of $f_{\rm esc}$ with halo mass suggests that

LAEs alone can not help to constrain any gradual dependence of $f_{\rm esc}$ with galactic properties. Any dependency introduced in the intrinsic Lyα luminosity can be compensated by the opposed trend of the Lyα escape fraction, achieved by changing the ISM gas and dust distribution. This insensitivity to $f_{\rm esc}$ dependencies makes LAEs relatively robust tracers of the underlying density field that we can use to pin down the ionization topology. Constraining $f_{\rm esc}$ during the EoR will require a combination of ionization topology measurements through the H I 21-cm signal and measurements of other emission lines.

ACKNOWLEDGEMENTS

We thank Max Gronke and Peter Laursen for useful discussions and the anonymous referee for their comments. AH, GY, LL, PD, and SG acknowledge support from the European Research Council's starting grant ERC StG-717001 ('DELPHI'). AH, MT, and PD also acknowledge support from the NWO grant 016.VIDI.189.162 ('ODIN'). AH acknowledges support by the Villum Fonden under grant 37459. PD acknowledges support from the European Commission's and University of Groningen's CO-FUND Rosalind Franklin program. PD thanks the Institute for Advanced Study (IAS) Princeton, where a part of this work was carried out, for their generous hospitality and support through the Bershadsky Fund. GY acknowledges financial support from MICIU/FEDER under project grant PGC2018-094975-C21. We thank Peter Behroozi for creating and providing the ROCKSTAR merger trees of the VSMDPL and ESMDPL simulations. The authors wish to thank V Springel for allowing us to use the L-Gadget2 code to run the different Multidark simulation boxes, including the VSMDPL and ESMDPL used in this work. The VSMDPL and ESMDPL simulations have been performed at LRZ Munich within the project pr87yi. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SUPERMUC-NG at Leibniz Supercomputing Centre (www.lrz.de). The CosmoSim database (www.cosmosim.org) provides access to the simulations and the Rockstar data. The database is a service by the Leibniz Institute for Astrophysics Potsdam (AIP). The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. This research made use of MATPLOTLIB, a PYTHON library for publication quality graphics (Hunter 2007); and the PYTHON library NUMPY (Harris 2020).

DATA AVAILABILITY

The source code of the semi-numerical galaxy evolution and reionization model within the ASTRAEUS framework is available on GitHub (https://github.com/annehutter/astraeus; Hutter 2020a). The underlying N-body DM simulation, the ASTRAEUS simulations and derived data in this research will be shared on reasonable request to the corresponding author.

REFERENCES

Adams T. F., 1975, ApJ, 201, 350
Anderson L., Governato F., Karcher M., Quinn T., Wadsley J., 2017, MNRAS, 468, 4077
Asano R. S., Takeuchi T. T., Hirashita H., Inoue A. K., 2013, Earth Planets Space, 65, 213
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
Blaizot J. et al., 2023, MNRAS, 523, 3749
Bosman S. E. I. et al., 2022, MNRAS, 514, 55
Carr C., Scarlata C., Henry A., Panagia N., 2021, ApJ, 906, 104
Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lacy M., Jarvis M. J., Lorenzoni S., Hickey S., 2012, MNRAS, 427, 3055
Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lorenzoni S., Jarvis M. J., Ebert H., 2014, MNRAS, 443, 2831
Castellano M. et al., 2016, ApJ, 818, L3
Castellano M. et al., 2018, ApJ, 863, L3
Chardin J., Aubert D., Ocvirk P., 2012, A&A, 548, A9
Chen Z., Xu Y., Wang Y., Chen X., 2019, ApJ, 885, 23
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Dayal P., Maselli A., Ferrara A., 2011, MNRAS, 410, 830
Dayal P., Ferrara A., Dunlop J. S., Pacucci F., 2014, MNRAS, 445, 2545
Dayal P. et al., 2022, MNRAS, 512, 989
Dijkstra M., Gronke M., Venkatesan A., 2016, ApJ, 828, 71
Endsley R., Stark D. P., 2022, MNRAS, 511, 6042
Endsley R. et al., 2022, MNRAS, 517, 5642
Fan X. et al., 2006, AJ, 132, 117
Faucher-Giguère C.-A., Kereš D., Dijkstra M., Hernquist L., Zaldarriaga M., 2010, ApJ, 725, 633
Ferrara A., Loeb A., 2013, MNRAS, 431, 2826
Fielding D. B., Bryan G. L., 2022, ApJ, 924, 82
Forero-Romero J. E., Yepes G., Gottlöber S., Knollmann S. R., Cuesta A. J., Prada F., 2011, MNRAS, 415, 3666
Fuller S. et al., 2020, ApJ, 896, 156
Furlanetto S. R., Oh S. P., 2016, MNRAS, 457, 1813
Galliano F., 2022, Habilitation Thesis, Université Paris-Saclay
Garaldi E., Kannan R., Smith A., Springel V., Pakmor R., Vogelsberger M., Hernquist L., 2022, MNRAS, 512, 4909
Garel T., Blaizot J., Rosdahl J., Michel-Dansac L., Haehnelt M. G., Katz H., Kimm T., Verhamme A., 2021, MNRAS, 504, 1902
Gazagnes S., Chisholm J., Schaerer D., Verhamme A., Izotov Y., 2020, A&A, 639, A85
Gnedin N. Y., 2000, ApJ, 542, 535
Gnedin N. Y., Kaurov A. A., 2014, ApJ, 793, 30
Gronke M., 2017, A&A, 608, A139
Gronke M., Oh S. P., 2020, MNRAS, 494, L27
Hansen M., Oh S. P., 2006, MNRAS, 367, 979
Harris C. et al., 2020, Springer Science and Business Media, 585, 357, available at: https://numpy.org/
He C.-C., Ricotti M., Geen S., 2020, MNRAS, 492, 4858
Heckman T. M. et al., 2011, ApJ, 730, 5
Howard C. S., Pudritz R. E., Harris W. E., Klessen R. S., 2018, MNRAS, 475, 3121
Hu E. M., Cowie L. L., Songaila A., Barger A. J., Rosenwasser B., Wold I. G. B., 2016, ApJ, 825, L7
Hu W. et al., 2021, Nat. Astron., 5, 485
Hu W. et al., 2023, preprint (arXiv:2307.04911)
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Hutter A., 2018, MNRAS, 477, 1549
Hutter A., Dayal P., Partl A. M., Müller V., 2014, MNRAS, 441, 2861
Hutter A., Dayal P., Müller V., 2015, MNRAS, 450, 4025
Hutter A., Dayal P., Müller V., Trott C. M., 2017, ApJ, 836, 176
Hutter A., Legrand L., Ucci G., 2020a, Astrophysics Source Code Library, record ascl:2004.006, available at: https://github.com/annehutter/astraeus
Hutter A., Watkinson C. A., Seiler J., Dayal P., Sinha M., Croton D. J., 2020b, MNRAS, 492, 653
Hutter A., Dayal P., Yepes G., Gottlöber S., Legrand L., Ucci G., 2021a, MNRAS, 503, 3698
Hutter A., Dayal P., Legrand L., Gottlöber S., Yepes G., 2021b, MNRAS, 506, 215
Itoh R. et al., 2018, ApJ, 867, 46
Jaskot A. E., Dowd T., Oey M. S., Scarlata C., McKinney J., 2019, ApJ, 885, 96

Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2013, MNRAS, 428, 1366
Jung I. et al., 2020, ApJ, 904, 144
Kakiichi K., Gronke M., 2021, ApJ, 908, 30
Kashikawa N. et al., 2011, ApJ, 734, 119
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, MNRAS, 491, 1736
Kerutt J. et al., 2022, A&A, 659, A183
Kim H.-S., Wyithe J. S. B., Park J., Lacey C. G., 2013, MNRAS, 433, 2476
Kim J.-G., Kim W.-T., Ostriker E. C., 2019, ApJ, 883, 102
Kimm T., Cen R., 2014, ApJ, 788, 121
Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 466, 4826
Kimm T., Blaizot J., Garel T., Michel-Dansac L., Katz H., Rosdahl J., Verhamme A., Haehnelt M., 2019, MNRAS, 486, 2215
Kimm T., Bieri R., Geen S., Rosdahl J., Blaizot J., Michel-Dansac L., Garel T., 2022, ApJS, 259, 21
Kobayashi C., Karakas A. I., Lugaro M., 2020, ApJ, 900, 179
Konno A. et al., 2014, ApJ, 797, 16
Konno A. et al., 2018, PASJ, 70, S16
Kusakabe H. et al., 2018, PASJ, 70, 4
Laursen P., 2010, PhD thesis, Niels Bohr Institute for Astronomy, Physics and Geophysics
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
Legrand L., Dayal P., Hutter A., Gottlöber S., Yepes G., Trebitsch M., 2023, MNRAS, 519, 4564
Leitherer C. et al., 1999, ApJS, 123, 3
Lewis J. S. W. et al., 2020, MNRAS, 496, 4342
Lisenfeld U., Ferrara A., 1998, ApJ, 496, 145
Malhotra S., Rhoads J. E., 2002, ApJ, 565, L71
Malhotra S., Rhoads J. E., 2004, ApJ, 617, L5
Maoz D., Mannucci F., Brandt T. D., 2012, MNRAS, 426, 3282
Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, ApJ, 856, 2
Mason C. A. et al., 2019, MNRAS, 485, 3947
Matthee J., Sobral D., Gronke M., Paulino-Afonso A., Stefanon M., Röttgering H., 2018, A&A, 619, A136
McKee C., 1989, in Allamandola L. J., Tielens A. G. G. M.eds, IAU Symp. 135, Interstellar Dust. Kluwer, Dordrecht. p. 431
McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, MNRAS, 381, 75
Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, MNRAS, 446, 566
Meyer R. A., Laporte N., Ellis R. S., Verhamme A., Garel T., 2021, MNRAS, 500, 558
Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
Neufeld D. A., 1990, ApJ, 350, 216
Ocvirk P. et al., 2016, MNRAS, 463, 1462
Ocvirk P. et al., 2020, MNRAS, 496, 4087
Oesch P. A. et al., 2015, ApJ, 804, L30
Ono Y. et al., 2012, ApJ, 744, 83
Orlitová I., Verhamme A., Henry A., Scarlata C., Jaskot A., Oey M. S., Schaerer D., 2018, A&A, 616, A60
Ota K. et al., 2010, ApJ, 722, 803
Ota K. et al., 2017, ApJ, 844, 85
Ouchi M. et al., 2010, ApJ, 723, 869
Padovani P., Matteucci F., 1993, ApJ, 416, 26
Pentericci L. et al., 2011, ApJ, 743, 132
Pentericci L. et al., 2014, ApJ, 793, 113
Pentericci L. et al., 2018, A&A, 619, A147
Planck Collaboration, 2020, A&A, 641, A6
Qin Y., Wyithe J. S. B., Oesch P. A., Illingworth G. D., Leonova E., Mutch S. J., Naidu R. P., 2022, MNRAS, 510, 3858
Roberts-Borsani G. W. et al., 2016, ApJ, 823, 143
Salpeter E. E., 1955, ApJ, 121, 161
Schenker M. A., Stark D. P., Ellis R. S., Robertson B. E., Dunlop J. S., McLure R. J., Kneib J.-P., Richard J., 2012, ApJ, 744, 179
Schenker M. A., Ellis R. S., Konidaris N. P., Stark D. P., 2014, ApJ, 795, 20
Seiler J., Hutter A., Sinha M., Croton D., 2019, MNRAS, 487, 5739
Shibuya T. et al., 2018, PASJ, 70, S14
Shimasaku K. et al., 2006, PASJ, 58, 313
Sobacchi E., Mesinger A., 2015, MNRAS, 453, 1843
Songaila A., Hu E. M., Barger A. J., Cowie L. L., Hasinger G., Rosenwasser B., Waters C., 2018, ApJ, 859, 91
Springel V., 2005, MNRAS, 364, 1105
Tilvi V. et al., 2020, ApJ, 891, L10
Todini P., Ferrara A., 2001, MNRAS, 325, 726
Totani T., Kawai N., Kosugi G., Aoki K., Yamada T., Iye M., Ohta K., Hattori T., 2006, PASJ, 58, 485
Totani T. et al., 2014, PASJ, 66, 63
Treu T., Trenti M., Stiavelli M., Auger M. W., Bradley L. D., 2012, ApJ, 747, 27
Triani D. P., Sinha M., Croton D. J., Pacifici C., Dwek E., 2020, MNRAS, 493, 2490
Ucci G. et al., 2023, MNRAS, 518, 3557
Vanzella E. et al., 2011, ApJ, 730, L35
Verhamme A., Orlitová I., Schaerer D., Hayes M., 2015, A&A, 578, A7
Verhamme A., Orlitová I., Schaerer D., Izotov Y., Worseck G., Thuan T. X., Guseva N., 2017, A&A, 597, A13
Weinberger L. H., Kulkarni G., Haehnelt M. G., Choudhury T. R., Puchwein E., 2018, MNRAS, 479, 2564
Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, MNRAS, 485, 1350
Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, MNRAS, 442, 2560
Xu H., Wise J. H., Norman M. L., Ahn K., O'Shea B. W., 2016, ApJ, 833, 84
Xu X. et al., 2023, preprint (arXiv:2301.11498)
Yates R. M., Henriques B., Thomas P. A., Kauffmann G., Johansson J., White S. D. M., 2013, MNRAS, 435, 3500
Zheng Z.-Y. et al., 2017, ApJ, 842, L22
Zhu Y. et al., 2021, ApJ, 923, 223
Zitrin A. et al., 2015, ApJ, 810, L12

APPENDIX A: GEOMETRICAL CORRECTION FACTOR ξ

When deriving the attenuation of the UV continuum by dust, we have assumed the light sources and dust to be homogeneously distributed within a slab. However, the numerical Lyα RT simulations in Gronke (2017) assume a screen of dust and gas between the light sources and the observer. To make the escape fractions of the UV continuum and Lyα radiation consistent, we introduce a geometrical correction factor that adjusts the Lyα escape fraction for sources behind a screen to that for sources distributed in a dusty gas slab. According to Forero-Romero et al. (2011), the Lyα escape fraction relations for these two geometries are given by

$$f_{\rm esc}^{\rm screen}(\tau_0) = \frac{1}{\cosh\left(\xi_0\sqrt{(a\tau_0)^{1/3}\tau_a}\right)} \quad \text{(A1)}$$

$$f_{\rm esc}^{\rm slab}(\tau_0) = \frac{1-\exp(-P)}{P} \quad \text{(A2)}$$

$$P = \epsilon_0\left((a\tau_0)^{1/3}\tau_a\right)^{3/4} \quad \text{(A3)}$$

$$\tau_a = \tau_{\rm d,total}(1-A), \quad \text{(A4)}$$

where A is the albedo. By equating the expressions for the Lyα escape fractions,

$$f_{\rm esc}^{\rm screen}(\tau_0^{\rm eff}) = f_{\rm esc}^{\rm slab}(\tau_0), \tag{A5}$$

we derive a correction factor ξ that reduces the H I column density in the screen geometry to the slab geometry,

$$\xi = \frac{\tau_0^{\rm eff}}{\tau_0} = \min\left(\xi_{\rm max}, \frac{\epsilon_0}{\xi_0^{3/2}} \frac{\left({\rm arcosh}\left(\frac{P}{1-e^{-P}}\right)\right)^{3/2}}{P}\right) \tag{A6}$$

$$P = \epsilon_0\, a^{1/4} \left((1-A)\frac{M_{\rm d}}{M_{\rm H\,I}}\frac{\kappa_{\rm abs} m_{\rm H}}{\sigma_{\rm H\,I}}\right)^{3/4} \tau_0. \tag{A7}$$

Here we assume $A = 0.5$ and $\xi_0 = 2.48$, with ξ_0 being adjusted to reproduce $f_{\rm esc}^{\rm screen}$ shown in fig. 1 in Forero-Romero et al. (2011). $\xi_{\rm max} = 0.35$ represents an upper limit for a dust free homogeneous distribution of gas and sources. We derived its value as follows: first, we add the Neufeld solutions (equation 30) for an equidistant set of τ_0 values between $[0, \tau_0]$; secondly, from the resulting Lyα line profile we estimate the effective $\tau_0^{\rm eff}$ value by measuring the peak positions ($x_p^{\rm eff}$). Relating these peak positions to those obtained for the single Neufeld solution for τ_0 (x_p), we obtain the ratio $x_p^{\rm eff}/x = \left(\tau_0^{\rm eff}/\tau_0\right)^{1/3} = \left(N_{\rm HI}^{\rm eff}/N_{\rm HI}\right)^{1/3}$, and the correction factor $\xi_{\rm max} = (x_p^{\rm eff}/x_p)^{1/3}$. We have also checked that applying the correction factor ξ to $N_{\rm HI}$ reproduces the correct shift in the Lyα peak positions x_p shown in fig. A5 in Forero-Romero et al. (2011).

APPENDIX B: DELAYED NON-BURSTY SUPERNOVA FEEDBACK SCHEME

We briefly describe our new formalism for the number of SN exploding if the stellar mass formed in a time step is assumed to form at a continuous rate across that time step. For a given star formation history SFR(t), the differential number of SN after a time t is given by

$$\frac{{\rm d}N_{\rm SN}}{{\rm d}t}(t) = \int_0^\infty {\rm d}t'\ {\rm SFR}(t')\ \nu(t-t'). \tag{B1}$$

$\nu(t)$ is the differential number of SN per stellar mass formed at $t' = 0$ and exploding at $t' = t$, and hence yields as

$$\nu(t) = M_{\rm SN}^{-\gamma}(t)\,\frac{{\rm d}M_{\rm SN}}{{\rm d}t}\Theta(t - t_{\star,\rm high})\,\Theta(t_{\rm SN,low} - t), \tag{B2}$$

with γ being the slope of the assumed IMF, $t_{\star,\,\rm high}$ being the time after which the most massive stars sampled by the IMF, $M_{\star,\,\rm high}$, explode as SN, and $t_{\rm SN,\,low}$ the time that it takes a star with the lowest stellar mass to explode as SN ($M_{\rm SN,low} = 8{\rm M}_\odot$). Stars of mass $M_{\rm SN}$ explode after a time t and the corresponding relation is described by

$$\frac{M_{\rm SN}}{{\rm M}_\odot} = \left(\frac{t/{\rm Myr}-3}{1.2\times10^3}\right)^{-1/1.85} = a^{-c}(t-3)^{-c}. \tag{B3}$$

For constant star formation with

$${\rm SFR}(t) = \begin{cases} 0 & t < t_i \\ s_0 & t_i \le t \le t_f \\ 0 & t_f < t \end{cases} \tag{B4}$$

we yield after inserting equation (B2) and B3 into equation (B1)

$$\begin{aligned}\frac{{\rm d}N_{\rm SN}}{{\rm d}t}(t) &= \int_{t_i}^{t_f} {\rm d}t'\ s_0\ \nu(t-t') \\ &= \int_{t_{\rm min}}^{t_{\rm max}} {\rm d}t'\ s_0\ M_{\rm SN}^{-\gamma}(t-t')\ \frac{{\rm d}M_{\rm SN}}{{\rm d}(t-t')} \\ &= -\int_{t_{\rm min}}^{t_{\rm max}} {\rm d}t'\ s_0\ c a^{c(\gamma-1)}\left(t-t'-3\right)^{c(\gamma-1)-1} \\ &= s_0\ \frac{a^{c(\gamma-1)}}{1-\gamma} \\ &\quad\times\left[(t-t_{\rm min}-3)^{c(\gamma-1)} - (t-t_{\rm max}-3)^{c(\gamma-1)}\right] \\ &= s_0\ \frac{a^{c(\gamma-1)}}{1-\gamma}\left[f_{\rm min}(t) - f_{\rm max}(t)\right]\end{aligned} \tag{B5}$$

with

$$t_{\rm min} = \max[t_i, t - t_{\rm SN,low}] \tag{B6}$$

$$t_{\rm max} = \min[t_f, t - t_{\star,\rm high}] \tag{B7}$$

and

$$t_{\rm max} \ge t_{\rm min}. \tag{B8}$$

These relations result in the following additional criteria

$$t \ge t_i + t_{\star,\rm high} \tag{B9}$$

$$t \le t_f + t_{\rm SN,low}. \tag{B10}$$

equation (B6) describes the differential number of SN exploding between the onset of star formation (t_i) and time t assuming constant star formation from t_i to t_f. However, to obtain the total number of SN exploding in a given time step, i.e. between t_{j-1} and t_j, we need to integrate over all contributions from $t_{j-1} \le t \le t_j$ (i.e. integrating equation (B6) over time t),

$$\begin{aligned}N_{\rm SN}(t_i, t_f, t_{j-1}, t_j) &= \int_{t_{j-1}}^{t_j} {\rm d}t\ \frac{{\rm d}N_{\rm SN}}{{\rm d}t} \\ &= \int_{t_{j-1}}^{t_j} {\rm d}t\ s_0\ \frac{a^{c(\gamma-1)}}{1-\gamma}\left[f_{\rm min}(t) - f_{\rm max}(t)\right].\end{aligned} \tag{B11}$$

We solve the different summands in the integral separately, yielding

$$\begin{aligned}F_{\rm max} &= \int_{t_{j-1}}^{t_j} {\rm d}t\ f_{\rm max}(t) \\ &= \int_{t_{j-1}}^{t_j} {\rm d}t\ (t-t_{\rm max}-3)^{c(\gamma-1)} \\ &\quad\times\Theta(t_f + t_{\rm SN,low} - t)\ \Theta(t - t_i + t_{\star,\rm high}) \\ &= \int_{\max(t_{j-1}, t_i+t_{\star,\rm high})}^{\min(t_j, t_f+t_{\star,\rm high})} {\rm d}t\ \left(t_{\star,\rm high}-3\right)^{c(\gamma-1)} \\ &\quad+ \int_{\max(t_{j-1}, t_f+t_{\star,\rm high})}^{\min(t_j, t_f+t_{\rm SN,low})} {\rm d}t\ \left(t-t_f-3\right)^{c(\gamma-1)} \\ &= \left[\left(t_{\star,\rm high}-3\right)^{c(\gamma-1)}\ t\right]_{\max(t_{j-1}, t_i+t_{\star,\rm high})}^{\min(t_j, t_f+t_{\star,\rm high})} \\ &\quad+ \left[\frac{\left(t-t_f-3\right)^{c(\gamma-1)+1}}{c(\gamma-1)+1}\right]_{\max(t_{j-1}, t_f+t_{\star,\rm high})}^{\min(t_j, t_f+t_{\rm SN,low})}\end{aligned} \tag{B12}$$

and

$$F_{\min} = \int_{t_{j-1}}^{t_j} \mathrm{d}t\ f_{\min}(t)$$
$$= \int_{t_{j-1}}^{t_j} \mathrm{d}t\ (t - t_{\min} - 3)^{c(\gamma-1)} \times \Theta(t_f + t_{\mathrm{SN,low}} - t)\ \Theta(t - t_i + t_{\star,\mathrm{high}})$$
$$= \int_{\max(t_{j-1}, t_i + t_{\star,\mathrm{high}})}^{\min(t_j, t_i + t_{\mathrm{SN,low}})} \mathrm{d}t\ (t - t_i - 3)^{c(\gamma-1)} + \int_{\max(t_{j-1}, t_i + t_{\mathrm{SN,low}})}^{\min(t_j, t_f + t_{\mathrm{SN,low}})} \mathrm{d}t\ \left(t_{\mathrm{SN,low}} - 3\right)^{c(\gamma-1)}$$
$$= \left[\left(t_{\star,\mathrm{high}} - 3\right)^{c(\gamma-1)}\ t\right]_{\max(t_{j-1}, t_i + t_{\star,\mathrm{high}})}^{\min(t_j, t_i + t_{\mathrm{SN,low}})} + \left[\frac{\left(t - t_f - 3\right)^{c(\gamma-1)+1}}{c(\gamma-1)+1}\right]_{\max(t_{j-1}, t_i + t_{\mathrm{SN,low}})}^{\min(t_j, t_f + t_{\mathrm{SN,low}})}. \quad \text{(B13)}$$

Inserting equation (B13) and B14 into equation (B12), we obtain

$$N_{\mathrm{SN}}(t_i, t_f, t_{j-1}, t_j) = s_0\ \frac{a^{c(\gamma-1)}}{1-\gamma}\ [F_{\min} - F_{\max}]\,. \quad \text{(B14)}$$

For a given star formation law SFR(t), the total stellar mass formed across all time is

$$M_\star^{\mathrm{tot}} = \int_0^\infty \mathrm{d}t\ \mathrm{SFR}(t)\ \left[\int_{M_{\star,\mathrm{high}}}^{M_{\star,\mathrm{low}}} \mathrm{d}m\ m^{1-\gamma}\right]$$
$$= \int_0^\infty \mathrm{d}t\ \mathrm{SFR}(t)\ \frac{M_{\star,\mathrm{low}}^{2-\gamma} - M_{\star,\mathrm{high}}^{2-\gamma}}{2-\gamma}. \quad \text{(B15)}$$

Hence, for a constant star formation between t_i and t_f, we finally obtain

$$M_\star^{\mathrm{tot}}(t_i, t_f) = \frac{s_0}{t_f - t_i}\ \frac{M_{\star,\mathrm{low}}^{2-\gamma} - M_{\star,\mathrm{high}}^{2-\gamma}}{2-\gamma}. \quad \text{(B16)}$$

Finally, from equation (B14) and (B16), we derive the number of SN exploding between times t_{j-1} and t_j from stars formed between t_i and t_j per stellar mass as

$$\frac{N_{\mathrm{SN}}(t_i, t_f, t_{j-1}, t_j)}{M_\star^{\mathrm{tot}}(t_i, t_f)} = \frac{2-\gamma}{1-\gamma}\ \frac{a^{c(\gamma-1)}}{M_{\star,\mathrm{low}}^{2-\gamma} - M_{\star,\mathrm{high}}^{2-\gamma}}\ \frac{F_{\min} - F_{\max}}{t_f - t_i}. \quad \text{(B17)}$$

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