



**HAL**  
open science

## Galactic ionizing photon budget during the epoch of reionization in the Cosmic Dawn II simulation

Joseph Lewis, Pierre Ocvirk, Laurence Voutquenne-Nazabadioko, Jenny Sorce, Paul Shapiro, Nicolas Deparis, Taha Dawoodbhoy, Romain Teyssier, Gustavo Yepes, Stefan Gottlöber, et al.

► **To cite this version:**

Joseph Lewis, Pierre Ocvirk, Laurence Voutquenne-Nazabadioko, Jenny Sorce, Paul Shapiro, et al.. Galactic ionizing photon budget during the epoch of reionization in the Cosmic Dawn II simulation. Monthly Notices of the Royal Astronomical Society, 2020, 496 (4), pp.4342-4357. 10.1093/mnras/staa1748 . hal-03142985

**HAL Id: hal-03142985**

**<https://hal.science/hal-03142985v1>**

Submitted on 17 Nov 2021

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

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



Distributed under a Creative Commons Attribution 4.0 International License

# Galactic ionizing photon budget during the epoch of reionization in the Cosmic Dawn II simulation

Joseph S. W. Lewis<sup>1</sup>,<sup>1</sup>★ Pierre Ocvirk,<sup>1</sup> Dominique Aubert,<sup>1</sup> Jenny G. Sorce,<sup>2,3,4</sup> Paul R. Shapiro,<sup>5</sup> Nicolas Deparis,<sup>1</sup> Taha Dawoodbhoj,<sup>5</sup> Romain Teyssier,<sup>6</sup> Gustavo Yepes,<sup>7,8</sup> Stefan Gottlöber,<sup>4</sup> Kyungjin Ahn,<sup>9</sup> Ilian T. Iliev<sup>10</sup> and Jonathan Chardin<sup>1</sup>

<sup>1</sup>Observatoire Astronomique de Strasbourg, Université de Strasbourg, CNRS UMR 7550, 11 rue de l'Université, F-67000 Strasbourg, France

<sup>2</sup>Univ. Lyon, ENS de Lyon, Univ. Lyon I, CNRS, Centre de Recherche Astrophysique de Lyon, UMR5574, F-69007 Lyon, France

<sup>3</sup>Univ. Lyon, Univ. Lyon I, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon, UMR5574, F-69230 Saint-Genis-Laval, France

<sup>4</sup>Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany

<sup>5</sup>Department of Astronomy, University Texas, Austin, TX 78712-1083, USA

<sup>6</sup>Institute for Theoretical Physics, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

<sup>7</sup>Departamento de Física Teórica M-8, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain

<sup>8</sup>Centro de Investigación Avanzada en Física Fundamental (CIAFF), Universidad Autónoma de Madrid, E-28049 Madrid, Spain

<sup>9</sup>Chosun University, 375 Seosuk-dong, Dong-gu, Gwangju 501-759, Korea

<sup>10</sup>Astronomy Center, Department of Physics & Astronomy, Pevensey II Building, University of Sussex, Falmer, Brighton BN1 9QH, United Kingdom

Accepted 2020 June 11. Received 2020 June 9; in original form 2020 January 17

## ABSTRACT

Cosmic Dawn II yields the first statistically meaningful determination of the relative contribution to reionization by galaxies of different halo mass, from a fully coupled radiation-hydrodynamics simulation of the epoch of reionization large enough ( $\sim 100$  Mpc) to model global reionization while resolving the formation of all galactic haloes above  $\sim 10^8 M_\odot$ . Cell transmission inside haloes is bi-modal – ionized cells are transparent, while neutral cells absorb the photons their stars produce – and the halo escape fraction  $f_{\text{esc}}$  reflects the balance of star formation rate (SFR) between these modes. The latter is increasingly prevalent at higher halo mass, driving down  $f_{\text{esc}}$  (we provide analytical fits to our results), whereas halo escape luminosity, proportional to  $f_{\text{esc}} \times \text{SFR}$ , increases with mass. Haloes with dark matter masses within  $6 \times 10^8 M_\odot < M_{\text{halo}} < 3 \times 10^{10} M_\odot$  produce  $\sim 80$  per cent of the escaping photons at  $z = 7$ , when the universe is 50 per cent ionized, making them the main drivers of cosmic reionization. Less massive haloes, though more numerous, have low SFRs and contribute less than 10 per cent of the photon budget then, despite their high  $f_{\text{esc}}$ . High-mass haloes are too few and too opaque, contributing  $< 10$  per cent despite their high SFRs. The dominant mass range is lower (higher) at higher (lower) redshift, as mass function and reionization advance together (e.g. at  $z = 8.5$ ,  $x_{\text{H I}} = 0.9$ ,  $M_{\text{halo}} < 5 \times 10^9 M_\odot$  haloes contributed  $\sim 80$  per cent). Galaxies with UV magnitudes  $M_{\text{AB}1600}$  between  $-12$  and  $-19$  dominated reionization between  $z = 6$  and 8.

**Key words:** Galaxy: formation – galaxies: high-redshift – dark ages, reionization, first stars.

## 1 INTRODUCTION

Current observations are consistent with the hypothesis that the Universe was reionized when UV starlight from massive stars escaped from the early galaxies in which they formed, creating intergalactic H II regions that grew in size and number until they overlapped to fully ionize the intergalactic medium (hereafter IGM) by  $z = 6$ . [For reviews and references, see, for instance, Dayal & Ferrara 2018 and Barkana & Loeb 2007]. During the epoch of reionization (EOR), the globally averaged ionized fraction was equivalent to the volume filling factor of these H II regions, which increased monotonically in an evolving patchwork of fully ionized and fully neutral zones. How fast this volume filling factor grew was determined primarily

by the average balance between the rate of release of ionizing photons by galaxies and the recombination rate of H atoms in their surrounding IGM. The release rate in a given patch depended upon the galaxy formation rate there, the star formation rates (hereafter ‘SFRs’) inside each galaxy, the spectra and luminosities of those stars, and the galactic escape fractions  $f_{\text{esc}}$  of their ionizing photons. The recombination rate in the surrounding IGM depended upon its evolving inhomogeneous density field. All these ingredients varied in space and time in a complex way. First, since structure formation was inhomogeneous. Secondly, because Reionization and the energy release associated with the star formation that drives it exerted hydrodynamical feedback therefore the ingredients were interdependent.

To predict their coupled evolution in the context of  $\Lambda$ CDM cosmology, to test the latter against observations, we must model the gravitational and gas dynamics of dark and baryonic matter and

\* E-mail: j.lewis@astro.unistra.fr

the radiative transfer of ionizing radiation in and between galaxies as they form. To capture the large-scale structure of inhomogeneous reionization, this must be done in a representative volume large enough ( $\sim 100$  Mpc), and with enough resolving power to form all the galaxies in that volume which contribute to reionization. As the dominant contributors are thought to be the ‘atomic-cooling haloes’ (hereafter ‘ACHs’) – those of virial temperatures above  $10^4$  K and masses above  $10^8 M_\odot$  – which means we must be able to resolve the formation of all the millions of haloes above  $10^8 M_\odot$  in that large volume.<sup>1</sup> In principle, to capture the full details of star formation within each galaxy, we would also have to resolve the interstellar medium (ISM) of each galaxy down to the sub-parsec scale on which molecular clouds fragment into protostars that then collapse into stars. The latter is currently out-of-reach computationally, however, even in the highest resolution simulations to date of a single galaxy. This means star formation and its local energy release must generally be treated as a ‘sub-grid’ process.

We have developed the Cosmic Dawn (hereafter ‘CoDa’) Project, to simulate reionization and galaxy formation together, self-consistently, with fully coupled, radiation-hydrodynamics, on a large-enough scale and with sufficient mass resolution to satisfy these requirements (Ocvirk et al. 2016; Aubert et al. 2018; Dawoodbhoy et al. 2018; Ocvirk et al. 2020). CoDa I (91 Mpc box), described in Ocvirk et al. (2016) and Dawoodbhoy et al. (2018), and CoDa II (94.5 Mpc box), described in Ocvirk et al. (2020), both used the massively parallel, hybrid CPU-GPU code RAMSES-CUDATON on a uniform grid of  $4096^3$  cells for the baryons and the radiation field, with  $4096^3$  N-body particles for the dark matter. CoDa I-AMR (91 Mpc), on the other hand, used another massively parallel, hybrid CPU-GPU code EMMA (Aubert, Deparis & Ocvirk 2015), with adaptive mesh refinement (AMR), with  $2048^3$  particles on a grid of  $2048^3$  coarse cells from which AMR increased the resolution locally, by up to a factor of 8, to follow the increasing local overdensity, leading to 18 billion cells after refinement. All three CoDa simulations were in volumes large enough to model the inhomogeneity and globally averaged time history of reionization, while also serving to model reionization and galaxy formation in the Local Universe, by adopting ‘constrained realizations’ of the Gaussian random noise initial conditions that were derived from galaxy survey data so as to reproduce the familiar structures of the Local Universe, such as the MW, M31, and the Virgo cluster, when evolved forward to  $z = 0$  (Sorice et al. 2016). We refer the reader to the papers cited above to describe our CoDa simulations and their relative differences in more detail. Our purpose here is to use the most recent of them, CoDa II, to find the ionizing luminosities of all the galaxies that formed in it during the EOR, to make the first statistically meaningful determination of the relative contribution to reionization by galaxies of different halo mass, over the full range of masses that contribute significantly, in a fully coupled radiation hydrodynamical numerical simulation.

The escape fraction  $f_{\text{esc}}$  of galaxies is difficult to observe directly. Indeed, the individual galaxies must be bright enough to detect, but also to compare their fluxes and spectral information at different

wavelengths (above and below the H Lyman limit). This must then be interpreted in terms of a model in which stars are assumed to have some initial mass function (hereafter ‘IMF’) and an SFR, which determines their spectral energy distribution (hereafter ‘SED’) over time. The radiation that emerges from the galaxy at wavelengths longwards of the Lyman limit is then assumed to be a combination of this SED and the nebular emission that results from re-processing the absorbed fraction of ionizing starlight by the interstellar gas, and may be partially attenuated by internal dust. Starlight emitted bluewards of the H Lyman limit is attenuated by photoionizing H atoms in the ISM of the galaxy and possibly attenuated further by dust. These processes are reflected in the net absorbed fraction  $(1 - f_{\text{esc}})$ , which may also include attenuation by the bound-free opacity of foreground Lyman limit absorbers along the line of sight. Observations of galaxies at different redshifts face different challenges, as they involve different spectral regions depending on  $z$ . Moreover, the foreground opacity is also a strong function of increasing redshift. As a consequence, observational determinations of  $f_{\text{esc}}$  are few and still uncertain. A review of this subject is well beyond the scope of this paper; the reader is referred to, e.g. Izotov et al. (2016), and the review by Dayal & Ferrara (2018) and references therein for a summary.

On the theory side, results are also rather limited. Some attempt to derive an empirical  $f_{\text{esc}}$ , one-size-fits-all. For an assumed form and amplitude of the UV luminosity function of galaxies above redshift 6, they proceed by adjusting  $f_{\text{esc}}$  to release enough ionizing photons to finish reionization in time to satisfy various observational constraints (we refer to these efforts as ‘one-zone’ models, e.g. Robertson et al. 2015). In doing so values such as  $f_{\text{esc}} = 10$  or 20 percent are sometimes reported, but this depends strongly on the underlying assumptions that led to it. These models often require some redshift evolution (for instance, in Haardt & Salvaterra 2015; Puchwein et al. 2019). Other attempts to determine a global value for  $f_{\text{esc}}$  employ semi-analytical or semi-analytical models of reionization. Again, these are similarly adjusted to match observational constraints, but use the model’s own statistical determinations of the rate of formation of galactic haloes from cosmological initial conditions, and some assumption about the SFR per halo (e.g. Ferrara & Loeb 2013; Dayal et al. 2020). A variety of galaxy formation simulations also exist that attempt to predict the  $f_{\text{esc}}$  and SFR from their simulated galaxies. However, these are mostly without radiative transfer or only post-processed with radiative transfer (such as in Razoumov & Sommer-Larsen 2010; Yajima, Choi & Nagamine 2011; Paardekooper, Khochfar & Dalla Vecchia 2015; Ma et al. 2015; Anderson et al. 2017). Recent studies do account for fully coupled radiation-hydrodynamics, with a focus on spatial resolution, considering a single galaxy or a fraction thereof (Kimm & Cen 2014; Trebitsch et al. 2017, 2018; Kimm et al. 2019; Trebitsch et al. 2020; Yoo, Kimm & Rosdahl 2020). This focus on resolving internal galactic structure precludes the follow-up of ionizing radiation propagation at cosmological scales.

A review of this subject, too, is well beyond the scope of this paper; the reader is again referred to Dayal & Ferrara (2018) for a summary and references, but we will describe some of these results in what follows, as we compare with our own.

The relative contribution of different mass haloes to the total ionizing photon budget released into the IGM during the EOR depends, not only upon the values of  $f_{\text{esc}}$  for each galaxy, but on their SFRs and the evolution of their population, as well. These aspects combine to determine the ionizing luminosity function of galaxies. In this work, we shall investigate this galactic ionizing photon budget. Since star formation efficiency typically rises with halo mass within the range of masses we represent (not only in

<sup>1</sup>Although the first stars form in mini-haloes (hereafter ‘MHs’; i.e. those with halo mass  $M < 10^8 M_\odot$  and virial temperatures  $T < 10^4$  K) that can cool gas by  $\text{H}_2$  molecular cooling (some stars may even form in metal-cooling MHs, as shown in Wise et al. 2014), the rising UV background during the EOR limits their contribution to the earliest stages of reionization. As a result, their relative contribution when compared to the more massive ACHs appears small ( $\sim 10$ – $20$  percent), (Ahn et al. 2012; Kimm et al. 2017).

CoDa II, but as generally expected, see Moster, Naab & White 2013; Legrand et al. 2019), while the abundance of haloes decreases with halo mass (e.g. see the halo mass functions of Sheth, Mo & Tormen 2001; Watson et al. 2013), the mass range of contributing haloes may, in principle, be broad, with a maximum contribution from haloes that, at different redshifts, may be anywhere within the broad range  $10^8 M_{\odot} < M_{\text{halo}} < 10^{12} M_{\odot}$ .

Previous work on the role of simulated galaxies in ionizing the IGM has sometimes been difficult to reconcile. However, it seems that with the recent advent of higher resolution and of fully coupled simulations, a few elements of consensus have begun to emerge. Studies such as Anderson et al. (2017) and Yajima et al. (2011) find that reionization is driven by the more numerous, low-mass galaxies ( $M_{\text{halo}} < 10^{9.5} M_{\odot}$ ), which broadly agrees with the conclusions of Kimm & Cen (2014), who find that the photon budget is dominated by masses  $10^8 M_{\odot} < M_{\text{halo}} < 10^9 M_{\odot}$  before  $z = 8$ , after which more massive haloes take over. Similarly, a more recent effort by Katz et al. (2018) seems to favour haloes within the range  $10^9 M_{\odot} < M_{\text{halo}} < 10^{10} M_{\odot}$  during the EOR, and those of higher mass at  $z = 6$ .

Most of the previous simulations are in volumes that are not large enough (most are boxes smaller than 25 Mpc across) to fully represent the halo mass function above  $\sim 10^{10} M_{\odot}$ . They may therefore be missing some of the largest haloes and galaxies, the ones, in fact, that form the most stars. The contributions of these highest mass haloes ( $> 10^{10} M_{\odot}$ ) to the photon budget in these studies is therefore partially absent. This could have a further, profound effect, on the rate LyC photons are released from the lowest mass galaxies, the ones that are reionized and suppressed by external sources, as well as dramatically alter the geometry of ionized regions throughout the EOR. Moreover, when reionization is simulated in too small a box, the duration of reionization is too small compared with that found in a volume large enough to capture the globally averaged history. In turn, this can affect the relative importance of haloes of different mass as their relative abundances evolve with redshift (Iliev et al. 2006, 2014), alongside their importance to reionization.

To overcome these limitations, we will address the photon budget of galaxies during the EOR using the CoDa II simulation (Ocvirk et al. 2020), produced with the RAMSES-CUDATON code (Ocvirk et al. 2016), which couples RAMSES (Teyssier 2002), the code for baryonic hydrodynamics and dark matter N-body dynamics, to ATON (Aubert & Teyssier 2008), the code for radiative transfer of ionizing radiation and non-equilibrium ionization rate equations. The CoDa II simulation ran from  $z = 150$  to  $z = 5.8$  in a comoving cubic box 94.533 Mpc on a side, with a high-enough mass resolution to form every galaxy in that volume with halo mass above  $10^8 M_{\odot}$ . This is sufficient to satisfy the requirement for large enough volume to simulate the EOR and its inhomogeneity, with a statistically meaningful halo mass function over the full mass range that may contribute significantly to reionization. Further featuring the fully coupled radiation-hydrodynamics (including radiation transport at the full speed of light) necessary to study the release of ionizing starlight into the IGM by galaxies during the EOR and its transport between galaxies involving highly supersonic ionization fronts. This combination of very large volume with complete sampling of the galactic sources within it represents a necessary compromise. The focus on large scales comes at a cost; we do not attempt to achieve the higher resolution inside galaxies that some other recent simulations do (e.g. Trebitsch et al. 2017; Katz et al. 2018, 2019; Rosdahl et al. 2018; Kimm et al. 2019; Trebitsch et al. 2020), which may affect some of the internal halo physics that are important for our problem, such as the escape of ionizing photons. One of the goals of this study,

however, is to demonstrate that, despite our relatively coarser spatial resolution internal to individual galaxies in CoDa II, the ‘global’ halo quantities relevant for describing the radiative properties of high-redshift galaxies, such as their escape fraction and total escape luminosity in ionizing photons, are meaningful and well captured, thereby validating the CoDa II-like approach, and paving the way towards even larger numerical simulations of the EOR with it in the future.

In this paper, we first, in Section 2, outline our numerical approach and computations. Then, in Section 3, we lay out our escape fraction results. Then, in Section 4, we present the ionizing galactic photon budget. Finally, in Section 5, we summarize our findings and propose some directions in which to take our subsequent efforts.

## 2 METHODS

### 2.1 Cosmic Dawn simulations

CoDa I and CoDa II are the largest coupled radiation hydrodynamics cosmological grid-based simulations aimed at studying the EoR. In the simulation code RAMSES-CUDATON, the RAMSES hydrodynamics+N-body code (Teyssier 2002) and the ATON radiative transfer code (Levermore 1984; Aubert & Teyssier 2008) are coupled, forming a hybrid code: hydrodynamics, gravitation, star formation, and supernova (SN) feedback are managed by the central processing units (CPUs), while the more computationally intensive mono-group<sup>2</sup> radiative transfer, Hydrogen photochemistry and cooling are managed by the graphics processing units (GPUs). The resulting acceleration allows us to perform simulations using the full speed of light, thereby circumventing possible artefacts arising from the use of a reduced speed of light (see Gnedin 2016, Deparis et al. 2019, Ocvirk et al. 2019, and Wu et al. 2019 for details on the impact of reduced or variable speed of light approximations).

We do not account for chemical enrichment, nor dust. The simulation focuses on the stellar component’s ability to drive cosmic reionization, and therefore active galactic nuclei (AGN) formation, feedback, and ionizing emissivity are not taken into account.

The large box size (94.44 cMpc box) and relatively high resolution for a simulation of cosmic reionization ( $4096^3$  dark matter particles and  $4096^3$  cells) yield a comoving cell size of 23.06 ckpc (i.e. 3.3 kpc physical at  $z = 6$ ) and a dark matter (stellar) particle mass of  $4.07 \times 10^5 M_{\odot}$  (11 732  $M_{\odot}$ ). With these specifications, CoDa II can represent the large-scale spatial variance in the reionization process whilst self-consistently dealing with the physics of the haloes that interests us (i.e. those within  $10^8 M_{\odot} \lesssim M_{\text{halo}} \lesssim 10^{12} M_{\odot}$ ), and providing us a huge sample of galactic haloes: there are around 13 million dark matter haloes at the end of the EoR in CoDa II.

CoDa II is compatible with a number of observational constraints related to the EoR, most notably the reionization history of the Universe, the Thomson optical depth measured from the cosmic microwave background, and the UV luminosity function of galaxies, as shown in Ocvirk et al. (2020). For further information relating to the code’s design, set-up, and runs, we refer the reader to Ocvirk et al. (2020).

### 2.2 Halo detection and boundaries

Dark matter haloes (haloes throughout the text) are detected using a Friends-of-Friends algorithm described in Roy, Bouillot & Rasera (2014), which produces a catalogue of haloes with their positions

<sup>2</sup>Effective photon energy 20.28 eV.



and masses  $M_{\text{halo}}$ . We can define a halo's virial radius, based on its mass  $M_{\text{halo}}$ , as  $r_{200}$ , as in Ocvirk et al. (2016, 2020):

$$M_{\text{halo}} = \frac{4}{3}\pi r_{200}^3 \times 200 \bar{\rho}_{\text{DM}}, \quad (1)$$

where  $\bar{\rho}_{\text{DM}}$  is the average cosmic dark matter density.

As in Ocvirk et al. (2016, 2020), we assume that one galaxy resides in each dark matter halo, and that the  $r_{200}$  limit of the halo is the limit of the galaxy. This assumption is valid in the vast majority of cases, in which the star-forming region of each halo has one clear stellar mass peak within  $r_{200}$ .

### 2.3 Halo escape fraction: ray-tracing, sub-grid, and net

Using CoDa II's gas density and ionization fields, we can compute the optical depths encountered by photons emitted in the halo along paths, from their injection to  $r_{200}$ . For a given halo, and for a given halo cell, we use the *healpix* PYTHON module *healpy* to sample the  $r_{200}$  sphere with 768 evenly distributed end points (we pick this number so as to adequately resolve our largest haloes. Increasing the number of rays by a factor of 2 only yields a difference of the order of  $10^{-4}$  in  $f_{\text{esc}}^{\text{ray}}$  for the most error susceptible computation). We then compute the optical depth along the path from the source cell centre to each end point as

$$\tau_{\text{path}} = \int_{\text{path}} \sigma_E \times \rho_{\text{H1}} \times dl, \quad (2)$$

where  $\sigma_E = 2.493 \times 10^{-22} \text{m}^2$  is the effective Hydrogen ionization cross-section of the photon group considered here in CoDa II,  $\rho_{\text{H1}}$  is the neutral Hydrogen physical density of the cell, and  $dl$  is an element of length.

The fraction of photons reaching  $r_{200}$  from a cell,  $\text{Tr}_{200}$ , is then obtained as the average of the transmissions along all 768 paths connecting the cell to the  $r_{200}$  sphere:  $\text{Tr}_{200} = \langle \exp(-\tau_{\text{path}}) \rangle_{\text{paths}}$  where the bracket denotes the average over all 768 paths. Fig. C1 is provided as an explanatory figure in Appendix C.

Finally, the halo escape fraction  $f_{\text{esc}}^{\text{ray}}$  is obtained as the SFR-weighted average of the transmissions of all the emitting cells of the halo, i.e. for a halo containing  $N$  cells indexed by the integer  $i$ :

$$f_{\text{esc}}^{\text{ray}} = \frac{\sum_i \text{SFR}_i \text{Tr}_{200,i}}{\sum_i \text{SFR}_i}. \quad (3)$$

The SFR of a cell is obtained as the stellar mass formed in the last 10 Myr,  $M_{\text{stars}}(< 10 \text{ Myr})$  divided by a 10 Myr duration:

$$\text{SFR} = M_{\text{stars}}(< 10 \text{ Myr}) / 10^7 M_{\odot} \text{ yr}^{-1}, \quad (4)$$

In the rest of the paper, 'star forming' (halo or cell) always means that some stellar mass has been formed in the last 10 Myr, i.e.  $M_{\text{stars}}(< 10 \text{ Myr}) > 0$ . The 'ray' superscript is used to clarify at all times that  $f_{\text{esc}}^{\text{ray}}$  is obtained via ray-tracing. We will refrain from discussing the interstellar medium (ISM) or circum-galactic medium (CGM) nature of the absorbing material. Since these are difficult to separate in CoDa II, we prefer the more general term 'halo escape fraction' for  $f_{\text{esc}}^{\text{ray}}$ .

Furthermore, since CoDa II does not resolve the ISM of our galaxies, CoDa II also uses a sub-grid escape fraction  $f_{\text{esc}}^{\text{sub}} = 0.42$  (chosen to obtain an EoR ending around  $z = 6$ ), which accounts for the photons lost to the star's birth cloud and the ISM, i.e. only 0.42 of the ionizing photons produced by a stellar particle is deposited in the cell containing it. We can now also define the net halo escape fraction  $f_{\text{esc}}^{\text{net}}$  as

$$f_{\text{esc}}^{\text{net}} = f_{\text{esc}}^{\text{sub}} \times f_{\text{esc}}^{\text{ray}}. \quad (5)$$

This is the fraction of a halo's stellar population photon production that manages to reach  $r_{200}$ . Since  $f_{\text{esc}}^{\text{sub}}$  is constant by construction in CoDa II, we focus mostly in the rest of the paper on the determination and behaviour of  $f_{\text{esc}}^{\text{ray}}$ .

### 2.4 Escape luminosity

To obtain the instantaneous amount of produced ionizing photons within a given halo  $\dot{\eta}_{\star}$ , we sum the ionizing photon production of all emitting star particles (i.e. younger than 10 Myr) within  $r_{200}$ , i.e.

$$\dot{\eta}_{\star} = M_{\text{stars}}(< 10 \text{ Myr}) \times 4.32 \times 10^{46} \text{ ph s}^{-1}, \quad (6)$$

where the factor  $4.32 \times 10^{46}$  is the stellar ionizing emissivity in  $\text{ph s}^{-1} M_{\odot}^{-13}$  (taken from table 1 of Ocvirk et al. 2020). We then define the escape luminosity ( $L_{\text{esc}}$ ) of a halo as the product of the intrinsic luminosity and the net halo escape fraction. This gives the number of ionizing photons that reach  $r_{200}$  per second:

$$L_{\text{esc}} = \dot{\eta}_{\star} \times f_{\text{esc}}^{\text{sub}} \times f_{\text{esc}}^{\text{ray}} \text{ ph s}^{-1}, \quad (7)$$

Note that, with these definitions,  $L_{\text{esc}}$  is, as expected, proportional to SFR and  $f_{\text{esc}}^{\text{ray}}$ . If we interpret  $r_{200}$  as the limit between the halo and the IGM,  $L_{\text{esc}}$  is effectively an estimate of the halo luminosity exiting the halo and entering the IGM. Similarly, we can define the cell escape luminosity, by considering  $\dot{\eta}_{\star}$  as the photon production within that cell and the average transmission  $\text{Tr}_{200}$  of the paths from that cell to the  $r_{200}$  sphere of its host halo.

## 3 RESULTS: HALO ESCAPE FRACTIONS

### 3.1 Halo escape fraction as a function of mass and redshift

Fig. 1 shows the distribution of halo escape fraction  $f_{\text{esc}}^{\text{ray}}$  as a function of halo mass for four redshifts:  $z = 6, 7, 8, 10$ .

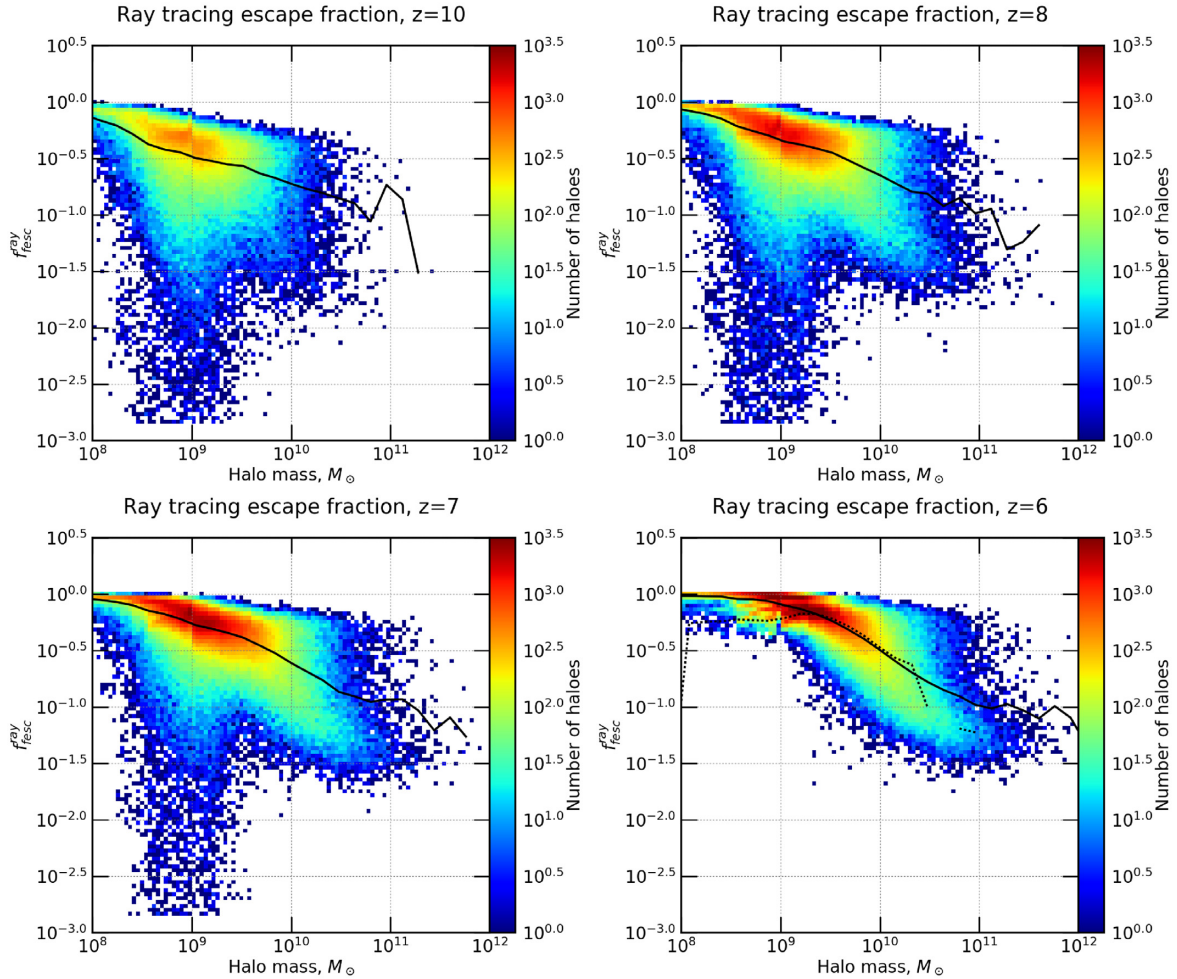
For masses  $\gtrsim 10^{9.5} M_{\odot}$ , the average  $f_{\text{esc}}^{\text{ray}}$  decreases, reaching values of  $< 10$  per cent for  $M_{\text{halo}} > 10^{11} M_{\odot}$ . These smaller values are observed despite massive haloes appearing to heat their  $r_{200}$  sphere surroundings with SN activity, bringing their neutral fraction down to  $10^{-6}$  and below. Indeed, massive haloes tend to feature dense, neutral, and opaque cores that trap a large fraction of their ionizing photon production as seen in Fig. 2. A more detailed investigation of the internal properties of these haloes is performed in Section 3.3.

Halo escape fractions are generally higher at low masses, saturating at 1 for all redshifts. The elongated vertical feature of the distribution of haloes at  $M_{\text{halo}} \approx 9 \times 10^8 M_{\odot}$ , with  $f_{\text{esc}}$  values lower than  $10^{-1}$ , and at  $z = 10, 8, 7$  (top left, top right, bottom left-hand panels of Fig. 1) is populated by haloes in which a star has formed recently, and in which the haloes' expanding H II regions have not yet fully reached the halo boundary at distance  $r_{200}$ . Fig. 3 shows the neutral fraction in a plane containing such a halo, illustrating this case.

Comparing the panels of Fig. 1 shows that the extended vertical distribution attributed to haloes in which stars have recently formed progressively disappears between  $z = 10$  and  $z = 6$ , owing to the ionization of the material of the lower mass haloes by the combination of local UV production, SNe energy injection, and outside radiation affecting these haloes, thereby rendering them more and more UV transparent within  $r_{200}$ .

There is an intrinsically high scatter in the escape fractions. It is due to the wide range spanned by the properties of the halo population

<sup>3</sup>This value is computed from the number of ionizing photons per stellar baryon produced by a binary stellar population of metallicity  $Z = 0.001$  using the BPASS (Eldridge et al. 2017) models, and with a Kroupa IMF (Kroupa 2001).



**Figure 1.** Distributions of halo escape fraction  $f_{\text{esc}}^{\text{ray}}$  of star-forming haloes as a function of  $M_{\text{halo}}$  at epochs  $z = 10, 8, 7, 6$ . The full lines indicate the averages of the distributions, i.e. for the star-forming haloes, whereas the dotted line indicates averages computed for the whole population, for the  $z = 6$  epoch only. The net halo escape fraction can be obtained by multiplying by the sub-grid escape fraction  $f_{\text{esc}}^{\text{net}} = 0.42 f_{\text{esc}}^{\text{ray}}$ .

such as the maximum central density, the gas density profile, and the individual accretion history of the halo. These different properties are well sampled because of CoDa II very large size and therefore abundant halo population.

Close examination of the density maps reveals discontinuities around  $5 \times 10^8 M_{\odot}$  and  $10^9 M_{\odot}$  in the highly populated red/orange areas. These are due to resolution effects. In the case of our less massive haloes, the number of cells that represent them can be small therefore increases in  $r_{200}$  can affect the resulting  $f_{\text{esc}}^{\text{ray}}$ .

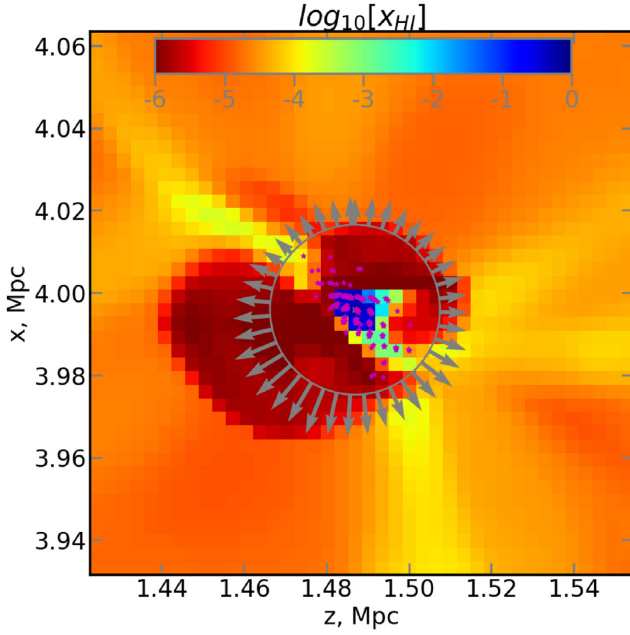
Fig. 4 shows the average escape fractions as a function of mass for five epochs,  $z = 6, 7, 8, 10, 14.9$ . Again, we see high  $f_{\text{esc}}^{\text{ray}}$  values for low-mass haloes, and a negative mass trend from  $\approx 10^9 M_{\odot}$  onwards. There is a clear evolution with redshift for haloes  $M_{\text{halo}} \lesssim 10^{10} M_{\odot}$ : the  $f_{\text{esc}}^{\text{ray}}$  average increases with decreasing redshift, reaching  $f_{\text{esc}}^{\text{ray}} \approx 1$  at  $z = 6$ . Indeed, for a fixed halo mass, higher redshift haloes are denser than their lower redshift counterparts, leading to lower escape fractions. The average behaviour of  $f_{\text{esc}}^{\text{ray}}$  for haloes  $\gtrsim 10^{11} M_{\odot}$  is unclear as the number of such objects is small.

### 3.2 Comparison with the literature

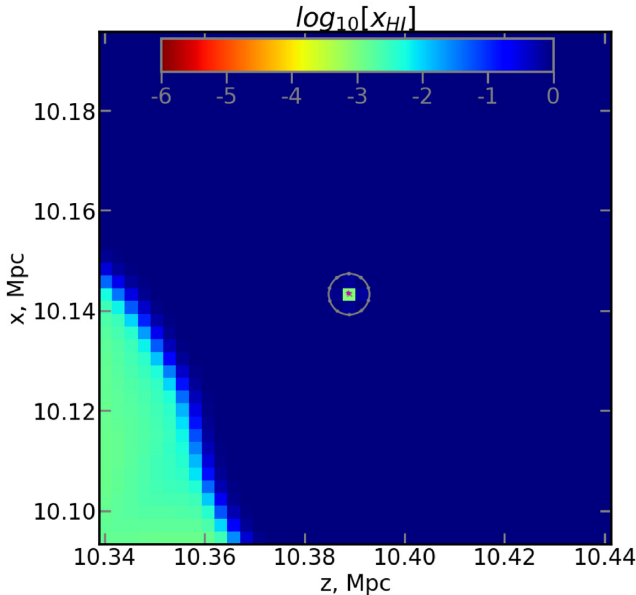
The most prominent feature of Fig. 1 is a decrease of escape fraction with mass, in agreement with the literature investigating the

escape fraction of ionizing radiation in high-redshift galaxies from numerical simulations (Razoumov & Sommer-Larsen 2010; Yajima et al. 2011; Kimm & Cen 2014; Wise et al. 2014; Paardekooper et al. 2015; Katz et al. 2018), although the slope and extent of the decrease may vary, as can be expected given the range in resolution and modelling of these studies. For dark matter haloes of  $10^{11} M_{\odot}$ , the simulations of Yoo et al. (2020) yield an escape fraction of 7 per cent, also in rather good agreement with our results.

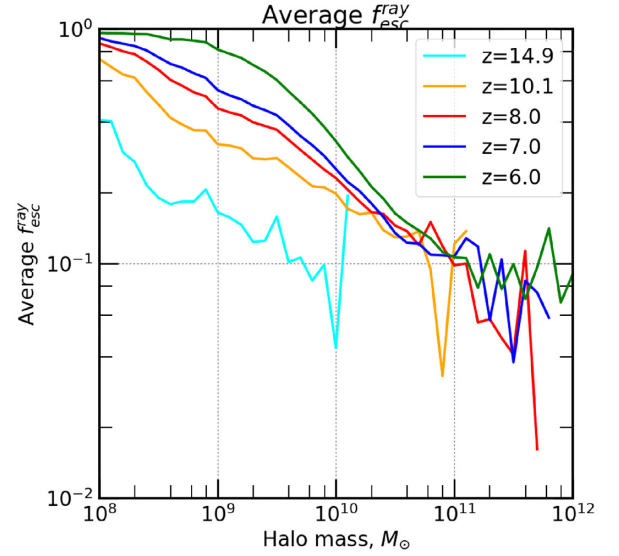
The other striking feature of our results is the evolution of halo escape fraction with time, in particular at masses below  $10^{10} M_{\odot}$ , showing that in CoDa II, such star-forming haloes tend to be more opaque at higher redshifts. For a given halo mass, haloes tend to be denser at higher redshift, and are therefore more likely to yield higher optical depths. Similar evolution, with escape fractions decreasing at higher redshifts, is seen in Kimm & Cen (2014). Razoumov & Sommer-Larsen (2010) report an opposite evolution, perhaps due to their modelling of dust (in their study, dust UV optical depth scales linearly with cell density and metallicity, the latter can be expected to increase on average over time, increasing CGM absorption). This may explain their different result, since dust is not accounted for in our work or in Kimm & Cen (2014).



**Figure 2.**  $X_{\text{H I}}$  (colour) centred in a plane containing an example high-mass halo ( $\approx 5.10^{10} M_{\odot}$ ) at  $z = 6$ . The pink symbols denote active (younger than 10 Myr) stellar particles. Notice the central, neutral region (several cells are close to fully neutral) that exhibits high star formation. The surrounding ionized region is superheated ( $T > 10^5$  K) and is generated by supernova explosions and accretion shocks. The circle represents  $r_{200}$ . The gas in the halo centre appears to be connected to several filament-like structures (with  $10^{-4.5} \gtrsim X_{\text{H I}} \gtrsim 10^{-3}$ ).



**Figure 3.**  $X_{\text{H I}}$  map (coded by colour) centred on a plane containing an example halo that has just formed a stellar particle at  $z = 8$ . Ionizing radiation has not yet reached the surfaces where its flux is sampled. Notice the  $X_{\text{H I}}$  gradient generated by an incoming UV front in the lower left. The symbols are as in Fig. 2.



**Figure 4.** Evolution of the average halo escape fraction  $f_{\text{esc}}^{\text{ray}}$  of star-forming haloes as a function of  $M_{\text{halo}}$  and redshift. The net halo escape fraction is obtained by multiplying by the sub-grid escape fraction  $f_{\text{esc}}^{\text{net}} = 0.42 f_{\text{esc}}^{\text{ray}}$ .

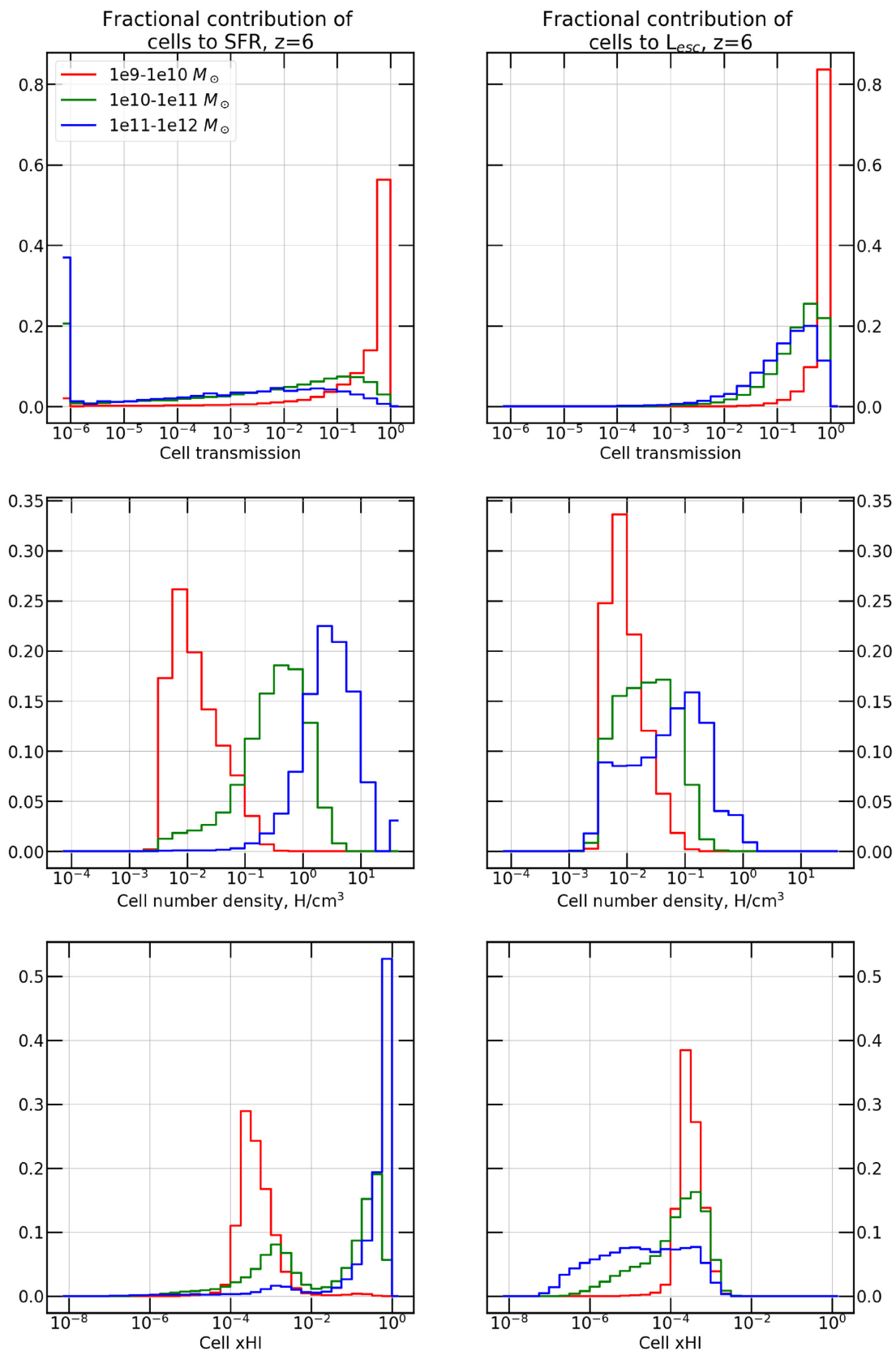
### 3.3 What drives the decrease of escape fractions with mass?

The main trend in all escape fraction plots is a decrease with increasing mass. Since halo escape fraction is determined by density and neutral hydrogen fraction, we want to investigate these particular properties. We show the distributions of the gas properties of star-forming cells in Fig. 5, and their contribution to the SFR and escape luminosity of their halo mass bins.

The top left-hand panel of Fig. 5 shows the distribution of cell transmissions weighted by their SFR for three representative mass bins, normalized by the total SFR in each mass bin (so that the integral of each histogram is 1). The star-forming cells of the lowest mass bin are concentrated around high transmission values, as observed in individual haloes and yielding the low-mass plateaus of  $f_{\text{esc}}^{\text{ray}}$  values close to 1.

There is a stark difference with the intermediate-mass bin, where star-forming cells present a much wider spread of transmission values. The distribution is dominated by cells with intermediate opacity, and peaks at a transmission of  $\sim 0.18$ . For the most massive mass bin, however, the dispersion in transmissions is even higher. Between a few times 0.1 and  $10^{-2}$  the distribution is almost flat, and it extends to extremely opaque cells with transmissions as low as  $10^{-6}$  and below. The high value of the distribution at  $10^{-6}$  indicates that there are close to 40 per cent of star-forming cells with a transmission lower or equal to  $10^{-6}$ . There are very few star-forming cells with a transmission of 1. This implies that most of the absorption of UV photons within our most massive haloes occurs within the cells containing the sources. We remind the reader that this is the distribution of the transmission of halo *cells*, i.e. one value per cell, and not the distribution of the halo-averaged (which would yield one value per halo). This is why the distributions extend much lower than in Fig. 1, which show only halo-averaged values. Please note that the peak at the low transmission end is caused by binning all values  $< 10^{-6}$  together.

We recompute this histogram, weighing this time by the cells' escape luminosity  $L_{\text{esc}}$ . The result is shown in the top right-hand panel of Fig. 5. This allows us to quantify the contribution of cells to the total escape luminosity of haloes. Unsurprisingly, low-mass



**Figure 5.** Distribution of star-forming halo cell properties: transmissions (top row), physical gas densities in  $\text{H cm}^{-3}$  (middle row), neutral fractions (bottom row), for three different mass bins of haloes. The histograms show the contribution of halo cells to (*left:*) total SFR of the mass bin considered, and to (*right:*) the total escape luminosity of the mass bin considered.



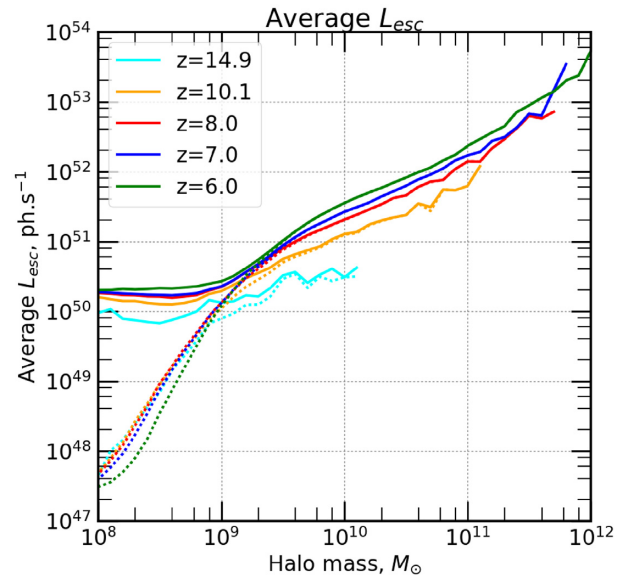
haloes' escape luminosities originate from high transmission cells. For intermediate- and high-mass haloes, the situation is slightly more contrasted: while most of the escaping photons originate from high transmission cells, a small fraction (10–15 per cent) are actually produced by moderately opaque regions with transmission below 0.1.

To gain deeper insight into the properties of UV-bright and UV-dark cells, we further examine their physical properties.

The middle left-hand panel of Fig. 5 shows the distribution of cells' gas density for our three representative mass bins, weighted by their SFR, and normalized by the total SFR in each mass bin. The mode of the distribution shifts to higher densities with increasing halo mass, and the density of star-forming cells in the high mass bin extends up to a few  $10 \text{ H cm}^{-3}$ . This is in stark contrast with the right-hand panel of Fig. 5, which shows the same distribution, but this time weighted by the cells' escape luminosity  $L_{\text{esc}}$ . The mode of the distributions for the highest and intermediate mass bins are located at much lower densities, and extend no further than about  $1 \text{ H cm}^{-3}$  therefore showing that in CoDa II, all stars forming at a density larger than this do not contribute to the escape luminosity of their host halo because their cell transmission is too low.

The bottom left-hand panel of Fig. 5 shows the distribution of cells' neutral hydrogen fractions for our three representative mass bins, weighted by their SFR, and normalized by the total SFR in each mass bin. For the lowest mass bin, the neutral fraction distribution has one strong peak at just over  $10^{-4}$ , which explains their high transmission. The value of  $\approx 10^{-4}$  reflects the ionization equilibrium for a cell of CoDa II resolution at  $z = 6$ , with an overdensity of  $\delta \approx 50$  that contains a single emitting stellar particle. In the case of the  $10^{10-11} M_{\odot}$  mass bin, haloes' cells are split more or less evenly between ionized and neutral, with two peaks, one centred around  $10^{-3}$  and the other one around a few times  $10^{-1}$  in neutral fractions. A possible origin for the binary nature of the distribution may be the rapidity with which dense star-forming cells ionize and recombine. In this context, cells would jump very quickly from one peak to the other, depending on their star formation activity or lack thereof, and spend very little time between the peaks. The star-forming cells of the highest mass bin are neutral in majority. Though the distribution presents a small ionized peak as well, the high neutral fraction peak is hugely dominant (about 90 per cent of the SFR takes place there) when compared to the ionized peak.

The bottom right-hand panel of Fig. 5 shows the same distribution, but this time weighted by the cells' escape luminosity  $L_{\text{esc}}$ . It shows the contribution of the cells in terms of escaping photons. For low-mass haloes, the uni-modal distribution is almost unchanged with respect to the left-hand panel. However, for the two higher mass bins, this different weighing has a strong impact on the relative modes of the distribution: indeed, the high neutral fraction mode of the distribution has completely vanished, so that only the high ionization mode remains ( $X_{\text{H I}} = 10^{-3}$  and below), showing that for these high halo masses, escaping ionizing photons originate predominantly from strongly ionized regions. Another striking aspect of the bottom right-hand panel Fig. 5 is the existence of a tail of the distribution, extending to cells with very high ionization (neutral fractions as low as  $10^{-4}$ – $10^{-7}$ ) and therefore completely transparent. This tail is absent for the low-mass bin. By examining maps of the physical properties of haloes, such as Fig. 2, we found that such high ionization within high-mass haloes is typical of shocked regions. These shocks can either be accretion shocks as seen in Ocvirk, Pichon & Teyssier (2008), Ocvirk et al. (2016), or shocks due to SNe explosions, or a combination of both. Indeed, several studies have shown SN feedback to play an important role in the escape of ionizing photons (Kimm



**Figure 6.** Evolution of the average halo escape luminosity  $L_{\text{esc}}$  of star-forming haloes (the solid line) and all haloes (the dotted line) as a function of  $M_{\text{halo}}$  and redshift.

& Cen 2014; Trebitsch et al. 2017), and we interpret this very high ionization tail in our distribution as another manifestation of this effect. Examining Fig. 2 in the light of these results also allows us to locate the regions contributing to the escape luminosity of the halo. The cells of the low ionization peak, at  $X_{\text{H I}} \geq 0.1$ , belong to the opaque neutral central core of the halo, from which no ionizing photons escape. The high ionization mode at  $X_{\text{H I}} \sim 10^{-3}$  and below, consists of cells located along the accreting gas filaments, up to the virial radius. Finally, the star-forming cells in the very strong ionization tail of the distribution, at  $X_{\text{H I}} < 10^{-5}$ – $10^{-6}$  are located in regions heated by SN feedback.

Although the shock-heated tail is not seen in the low-mass bins, it shows up when increasing spatial resolution, as demonstrated in Appendix A. However, it does not result in an increase in  $f_{\text{esc}}$ , because neutral cells also become more frequent within these haloes, and the net result of increasing resolution is a decrease of  $f_{\text{esc}}$ .

### 3.4 Escape luminosities

Using our determinations of halo escape fractions, we can now compute the halo escape luminosities as in equation (7). We show the resulting  $L_{\text{esc}}$  as a function of halo mass and their evolution with redshift in Fig. 6. The escape luminosity increases with halo mass, despite the decrease of  $f_{\text{esc}}$ . Indeed, at  $z = 6$ , for instance,  $f_{\text{esc}}$  decreases roughly as  $M_{\text{halo}}^{-0.34} \sim M_{\text{halo}}^{-1/3}$ , as shown in Appendix B, whereas halo SFR increases as  $M_{\text{halo}}^{5/3}$ , and therefore the product  $L_{\text{esc}} \propto M_{\text{halo}}^{4/3}$  increases with halo mass. At the low-mass end,  $L_{\text{esc}}$  flattens as we reach the quantization limit of star formation in the simulation: between  $10^8$  and  $10^9 M_{\odot}$ , star-forming haloes host only one emitting star particle of the same elementary mass  $M_{\star} = 11732 M_{\odot}$ . The evolution of  $L_{\text{esc}}$  with redshift reflects that of the escape fraction  $f_{\text{esc}}$ . The impact of star formation suppression by radiation feedback is not readily seen in the solid lines of the figure because they represent the average  $L_{\text{esc}}$  of star-forming haloes only. However, the dotted lines shows the average  $L_{\text{esc}}$  for all haloes, where the suppression-driven decrease of SFR with redshift leads to lower  $L_{\text{esc}}$  at low redshifts below  $10^9 M_{\odot}$ .

## 4 RESULTS: PHOTON BUDGET

We now turn to investigating the contribution of haloes of different masses and different  $M_{\text{AB}1600}$  to cosmic reionization in CoDa II.

### 4.1 Photon budget versus mass

We sort the CoDa II dark matter halo population into 40 logarithmic mass bins between  $10^8$  and  $10^{12} M_{\odot}$ . We define the total escape luminosity of a given mass bin as the sum of the escape luminosity of all haloes within that mass bin. This quantity depends on the total number of haloes in the CoDa II volume, and we wish to conduct our study using a quantity independent of simulation box size, to ease comparison with future semi-analytical models and simulations. Hence, we further define the escape emissivity, as the total escape luminosity of a given mass bin divided by the simulation volume.

The left-hand panel of Fig. 7 shows this escape emissivity as a function of halo mass, at five epochs (the full lines). This represents the cosmic ionizing photon budget for the CoDa II simulation, i.e. the distribution of the contributions of each mass bin to the total rate of photons reaching the IGM and driving reionization, for a  $1 \text{ cMpc}^3$  volume.

The contribution to the cosmic escape emissivity culminates around  $\sim 10^{9.5} M_{\odot}$  between  $z = 6$  and  $z = 8$  and decreases at lower and higher masses. This leads to a photon budget that is dominated by haloes of a few times  $10^8$  up to a few times  $10^{10} M_{\odot}$  between these redshifts.

To quantify the impact of the smaller escape fraction of massive haloes on the photon budget, we also show in the left-hand panel of Fig. 7 the intrinsic photon *production* budget, i.e. proportional to the SFR of haloes (the dotted line), compared to the escaping photon budget. At  $z = 6$ , for instance, intrinsic photon production peaks at  $10^{10.5} M_{\odot}$ . This is the result of two opposite trends: SFR increases with increasing halo mass, whereas the number abundance of haloes decreases, as dictated by the halo mass function. This competition yields  $10^{10.5} M_{\odot}$  haloes as the foremost contributors to the total cosmic star formation at  $z = 6$ . However, their contribution to the escaping photon budget is strongly affected by their low escape fractions, to the point that, even though they dominate all other haloes in terms of SFR, they are outshined in total escape emissivity by the  $10^9$ – $10^{10} M_{\odot}$  halo population.

To further compare the contribution of various mass bins, the right-hand panel of Fig. 7 shows the cumulative version of the photon budget. It allows us to directly read from the plot that at  $z = 6$ , for instance, haloes within  $2 \times 10^9 M_{\odot} \lesssim M_{\text{halo}} \lesssim 10^{10} M_{\odot}$  produce around 50 per cent of all ionizing photons. The shape of the cumulative photon budget is rather similar at all redshifts.

However, the cumulative distributions shift towards lower masses as redshift increases, by about half a decade between  $z = 6$  and  $z = 10$ . This is also seen as a shift of the peak of the distribution (left-hand panel) between  $z = 6$  and  $z = 8$ . This shift reflects the build-up and evolution of the halo mass function towards more abundant and more massive halo populations. At all redshifts but the highest, the cumulative distribution has its 10 per cent and 90 per cent levels separated by about 1.5 decades in mass, meaning this range is responsible for 80 per cent of the ionizing photon budget. In CoDa II, the ionized volume fraction goes from 20 per cent to 100 per cent between  $z = 6$  and 8, and it is  $\sim 50$  per cent at  $z = 7$ . We therefore retain  $z = 7$  as the epoch most representative of ‘ongoing’ reionization and detail the photon budget for this redshift: we read from the cumulative

distribution that the photon budget at this epoch is dominated by galaxies with dark matter masses of  $6 \times 10^8$ – $3 \times 10^{10} M_{\odot}$ , which produce 80 per cent of the ionizing photons reaching the IGM. They are therefore the main drivers of cosmic reionization.

### 4.2 Photon budget by $M_{\text{AB}1600}$

The luminosity function of CoDa II haloes was shown to be in good agreement with observations in Ocvirk et al. (2020). Here, we use the halo magnitudes to recast our photon budget analysis into a  $M_{\text{AB}1600}$  scale instead of the halo mass scale.

Fig. 8 shows the total ionizing photon contribution as a function of  $M_{\text{AB}1600}$  (left), as well as the equivalent cumulative distribution (right). The shape of the distribution and its temporal evolution are similar to that obtained as a function of mass (Fig. 7). This is a direct consequence of the SFR–halo mass relation shown in Ocvirk et al. (2020), which drives an  $M_{\text{AB}1600}$ –halo mass relation.

Combined, these figures illustrate that the main contributors to reionization lie within a magnitude range of around  $-12$  to about  $-19$ . Which is in broad agreement with the semi-analytical results of Liu et al. (2016) for  $z > 7$ . More specifically, the 80 per cent escape luminosity range (reading the 10–90 per cent levels of the cumulative escape luminosity distribution function) is  $M_{\text{AB}1600} = [-13, -19]$  at  $z = 6$ , and  $[-12, -17]$  at  $z = 8$ . This shift with redshift is due to the build-up of the galaxy mass function.

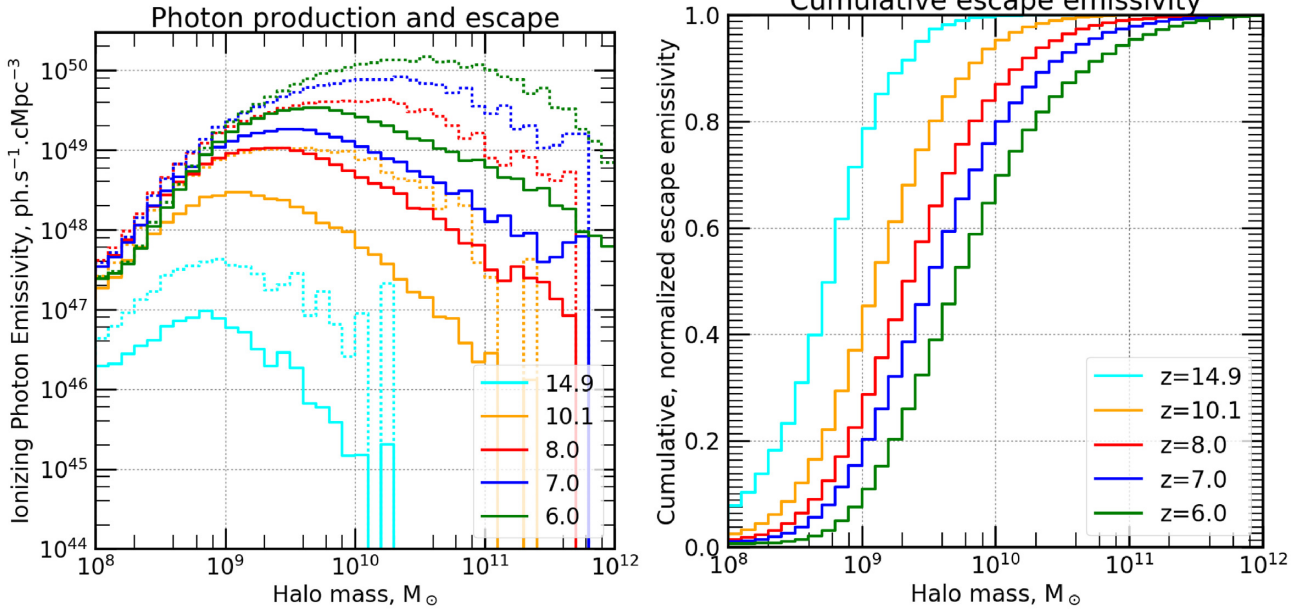
This suggests that very deep surveys in cluster fields, such as Bouwens et al. (2017) and Atek et al. (2018), if reliable down to  $M_{\text{AB}1600} = -13$ , are indeed starting to see the bulk ( $>80$  per cent) of the population driving cosmic reionization at  $z = 6$ . However, at  $z = 6$ , cosmic reionization is already finished in CoDa II, and in the middle of reionization, i.e. at  $z = 7$ , the galaxies seen by Bouwens et al. (2017) are brighter than  $M_{\text{AB}1600} = -17$ , and Fig. 8 shows that these can only account for  $\sim 20$  per cent of the ionizing luminosity. To see the bulk of galaxies driving reionization when it is in full spin, surveys would need to achieve a similar  $M_{\text{AB}1600} = -12$  depth at  $z = 7$ .

Finally, early reionization at  $z = 15$  is driven by galaxies of a very narrow range of halo masses,  $10^8$ – $10^9 M_{\odot}$ , corresponding to magnitudes fainter than  $M_{\text{AB}1600} = -14$ , out of reach of current and future planned observatories at these redshifts.

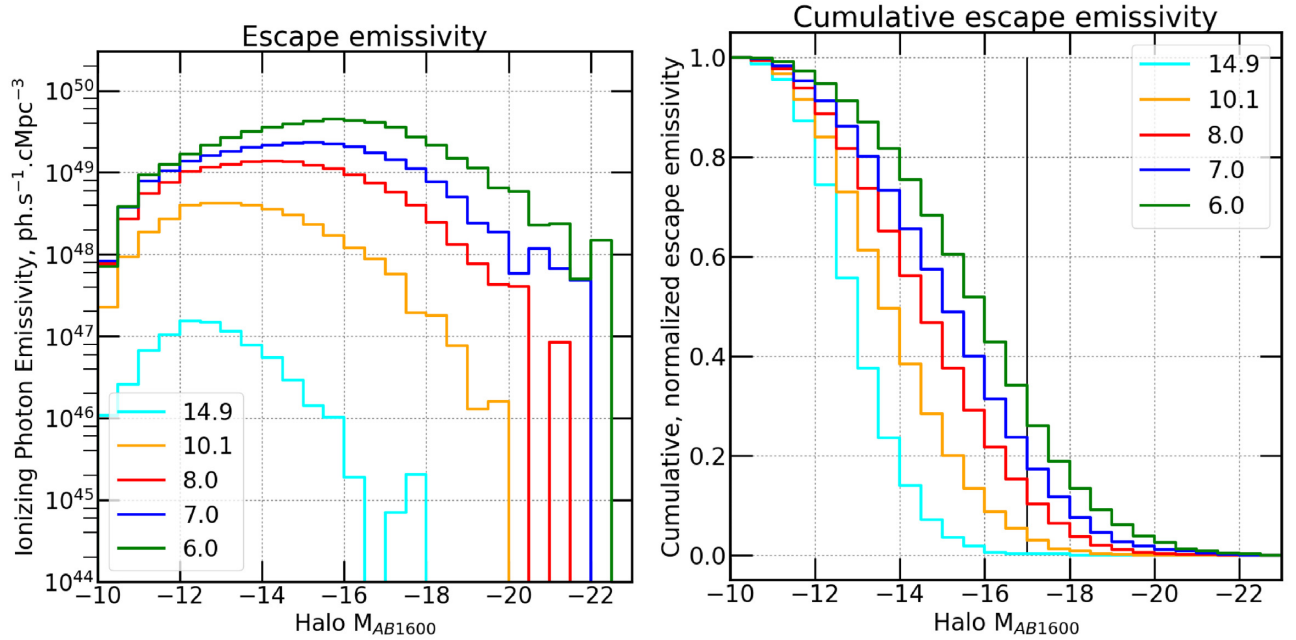
### 4.3 Comparison with the literature

Katz et al. (2018, 2019) use tracers and RAMSES-RT RHD simulations to study the contribution of haloes to the cosmic ionizing luminosity. Although their technique differs in many respects from ours (we do not use such tracers, they use variable speed of light, AMR), their results are in rather good agreement with ours. At  $z = 8$ , they find that 70 per cent of the ionizing luminosity is produced by haloes of  $10^9 M_{\odot} \lesssim M_{\text{halo}} \lesssim 10^{10} M_{\odot}$  (from fig. 6 of Katz et al. 2018). We can read directly from Fig. 7 that at the same redshift, this mass range is responsible for 60 per cent of the ionizing photons in CoDa II. However, at  $z = 6$ , Katz et al. (2018) find that high-mass haloes of  $\gtrsim 10^{10} M_{\odot}$  produce the majority of ionizing photons (60 per cent), while we find that they account for only 40 per cent of the ionizing luminosity, i.e. the largest contribution originates from haloes less massive than  $10^{10} M_{\odot}$  in CoDa II.

At higher redshifts, though ( $z > 12$ ), the Katz et al. (2018) values fluctuate too much for a meaningful comparison. Also,



**Figure 7.** Left : escape emissivity as a function of halo mass for several CoDa II epochs. The displayed emissivity is the sum of all haloes' escape luminosities in the mass bin considered at that redshift, divided by the simulation volume. There are 10 mass bins per mass decade. The full lines correspond to the escape emissivity, whereas the dashed lines correspond to the intrinsic (unabsorbed) emissivities (i.e. derived directly from the star formation rate). Right : Cumulative version of left-hand panel.



**Figure 8.** Left : Total escape emissivity as a function of magnitude for several CoDa II epochs. The displayed emissivity is the sum of all haloes' escape luminosities in the  $M_{AB1600}$  bin considered, divided by the simulation volume. There are four bins per mag. Right : Cumulative version of the total escape emissivity plot on the left. The vertical bar marks the  $M_{AB1600} = -17$  limit.

because of their smaller box size, their sample is devoid of haloes more massive than  $10^{11} M_{\odot}$  even at  $z = 6$ , unlike in CoDa II, where such massive haloes are present already at  $z = 10$ . However, their low number density and escape fractions prevent them from contributing significantly to cosmic reionization: their total escape luminosity is less than 10 per cent at all

redshifts, which is why our results are in fair agreement with Katz et al. (2018), even though they do not include these high-mass haloes.

Yajima et al. (2011) find that haloes below  $M_{\text{halo}} \lesssim 10^{10} M_{\odot}$  contribute about  $\approx 75$  per cent of the ionizing luminosity at  $z = 6$  (summing the two lowest mass bins of the  $z = 6$  panel of their fig. 12).



Again, this is in reasonable agreement with our findings, although at this redshift the largest contribution (45 per cent) comes from haloes in the mass range  $10^9$ – $10^{9.5} M_{\odot}$ , while the contribution of this mass range is about two times smaller in CoDa II. This discrepancy could be due to the lack of radiative feedback on the SFR of low-mass haloes in (Yajima et al. 2011): indeed, their study performs RT as post-processing, whereas radiation and hydrodynamics are fully coupled in CoDa II, which mitigates star formation in low-mass haloes, as shown in Ocvirk et al. (2020), Dawoodbhoy et al. (2018) and similarly in Wu et al. (2019), Ma et al. (2018), and therefore intrinsically reduces their contribution to cosmic reionization.

A possible caveat in our work is the assumption we make in Section 2.2 that each halo contains one galaxy. This could potentially result in the blending of galaxies whose real equivalents would be distinguishable in observations as two separate star-forming objects. A visual inspection of a sample of our haloes reveals that this preferentially occurs in some of the most massive haloes ( $M_{\text{halo}} > 10^{11} M_{\odot}$ ), at a rate of less than 20 per cent. Since this mass range only contributes to a few per cent of the photon budget, tentatively correcting for this effect would impact the photon budget only at the percent level, and would not change our conclusions.

Finally, we address the possible dependence of our results on spatial resolution by performing a follow up higher resolution simulation.

We show in appendix A that indeed, increasing spatial (mass) resolution in RAMSES-CUDATON by a factor of 4 (64) may yield lower halo escape fractions by a factor of 2, while retaining a similar slope. Such a global re-scaling over all masses, without changing the slope of the  $f_{\text{esc}}$ –halo mass relation, should not dramatically affect our results on the photon budget and the predominant halo mass scale driving reionization because it does not alter the *relative* contributions between halo mass bins.

## 5 CONCLUSIONS

We use the CoDa II fully coupled RHD simulation of the EoR to study the photon budget of galaxies during the EoR. To do so, we start out by investigating the escape fractions of CoDa II galaxies. We find that the halo escape fraction (i.e. the fraction of ionizing photons produced by the halo’s stars reaching the virial radius) is a decreasing function of halo mass.

To gain insight into the evolution of the  $f_{\text{esc}}^{\text{ray}}$  with halo mass, we examine the properties of the halo cells as a function of their SFRs and escape luminosity. We find that for intermediate- and high-mass haloes, the neutral fractions of star-forming cells exhibit a strongly bi-modal distribution, with a neutral mode and an ionized mode at  $X_{\text{H I}} \sim 10^{-3}$ . The neutral mode is completely opaque, meaning that escaping ionizing photons originate from the star-forming, ionized regions of the haloes. The halo escape fractions we obtained closely reflect the distribution of the star-forming cells between the neutral opaque mode and the ionized, transparent mode. For instance, CoDa II high-mass haloes ( $10^{11} M_{\odot}$ ) have an average halo escape fraction of  $\sim 10$  per cent because 90 per cent of their young stars reside in central, dense, fully opaque regions, while the remaining 10 per cent of their young stars reside in transparent regions allowing their photons to escape.

Moreover, we find a slow evolution of the halo escape fraction with redshift: haloes of a given mass are more opaque at higher redshift. This is due to the fact that for a fixed mass, haloes at higher redshifts tend to be more concentrated than their low-redshift counterparts.

In Appendix B, we provide a functional form fit to our average halo escape fraction results, so as to allow its use in semi-analytical models of the EoR such as 21cmFAST (Mesinger, Furlanetto & Cen 2011) or Fialkov et al. (2013).

We then use the halo escape fractions of our haloes to investigate the contributions of galaxies of various masses to the total ionizing emissivity during the EoR. We show that CoDa II galaxies within  $3 \times 10^{10} M_{\odot} \gtrsim M_{\text{halo}} \gtrsim 6 \times 10^8 M_{\odot}$  produce about 80 per cent of all the ionizing photons reaching the IGM at  $z = 7$ , which is the middle of reionization in CoDa II ( $x_{\text{H I}} = 50$  per cent). They can therefore be considered as the main drivers of cosmic reionization, although, at  $z = 6$  (8), the mass range accounting for 80 per cent of the photon budget is slightly more (less) massive, by 0.25 dex.

The foremost mass range reionizing the Universe emerges as the result of a competition between the different processes exposed throughout this paper, and can be summarized as follows: the numerous low-mass haloes are too inefficient at forming stars to contribute significantly despite their high halo escape fractions, whereas the high-mass haloes are too few and have escape fractions that are too low to contribute significantly, despite their high SFR.

As a consequence, the low-mass end (below  $5 \times 10^9 M_{\odot}$ ) and the high-mass end (above  $5 \times 10^{10} M_{\odot}$ ) contribute, respectively, only less than 10 per cent each to the total ionizing photon budget between  $z = 8$  and  $z = 6$ .

Our results are in reasonable agreement with the (not exhaustive) literature reviewed, despite a number of differences in numerical treatment, and assumptions on stellar populations and their feedback, which explain the deviations from our results.

Ideally, we would like to follow up on our study by pushing future CoDa II-like simulations to higher spatial resolutions, possibly using AMR to provide a better description of the ISM, and its processes and if possible rely less on a sub-grid escape fraction, as well as improved physics such as chemical enrichment, AGN mechanical and radiative feedbacks, and their possible contribution to reionization. The continued growth of supercomputers, because of hybrid nodes mixing many-core CPUs and GPUs, may allow us to get there in the near future, provided we overcome a number of technical hurdles related to code architecture and optimization.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the judicious comments of the anonymous referee, as well as those of Max Gronke. This work has used PYTHON, and the following packages : PYTHON : MATPLOTLIB (Hunter 2007), NUMPY (Van Der Walt, Colbert & Varoquaux 2011), SCIPY (Virtanen et al. 2020), HEALPIX (Górski et al. 2005). PO acknowledges support from the French ANR funded project ORAGE (ANR-14-CE33-0016). ND and DA acknowledge funding from the French ANR for project ANR-12-JS05-0001 (EMMA). ITI was supported by the Science and Technology Facilities Council (grant numbers ST/F002858/1 and ST/I000976/1) and the Southeast Physics Network (SEPnet). JS acknowledges support from "l'Oréal-UNESCO 'Pour les femmes et la Science'" and the 'Centre National d'études spatiales (CNES)' postdoctoral fellowship programs. KA was supported by NRF (grant number NRF-2016R1D1A1B04935414). GY acknowledges financial support by the MINECO/FEDER under project grant AYA2015-63810-P and MICIU/FEDER under project grant number PGC2018-094975-C21. PRS was supported in part by U.S. NSF grant number AST-1009799, NASA grant number NNX11AE09G, NASA/JPL grant number



RSA Nos. 1492788 and 1515294, and supercomputer resources from NSF XSEDE grant number TG-AST090005 and the Texas Advanced Computing Center (TACC) at the University of Texas at Austin. The CoDa II simulation was performed at Oak Ridge National Laboratory/Oak Ridge Leadership Computing Facility on the Titan supercomputer (INCITE2016 award AST031). Processing was performed on the Eos and Rhea clusters. Auxiliary simulations were performed at pôle HPC de l'Université de Strasbourg (mésocentre). The simulations used for the resolution study were performed on CSCS/Piz Daint (Swiss National Supercomputing Centre), as part of the 'SALT: Shining a light through the dark ages' PRACE allocation obtained via the 16th call for PRACE Project Access (project id pr37). A series of test simulations for the initial conditions of CoDa II were performed at LRZ Munich within the project pr74no. This work has used v2.1 of the Binary Population and Spectral Synthesis (BPASS) models as last described in Eldridge et al. (2017).

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. See Ocvirk et al. (2020) for specifics about the availability of CoDa II.

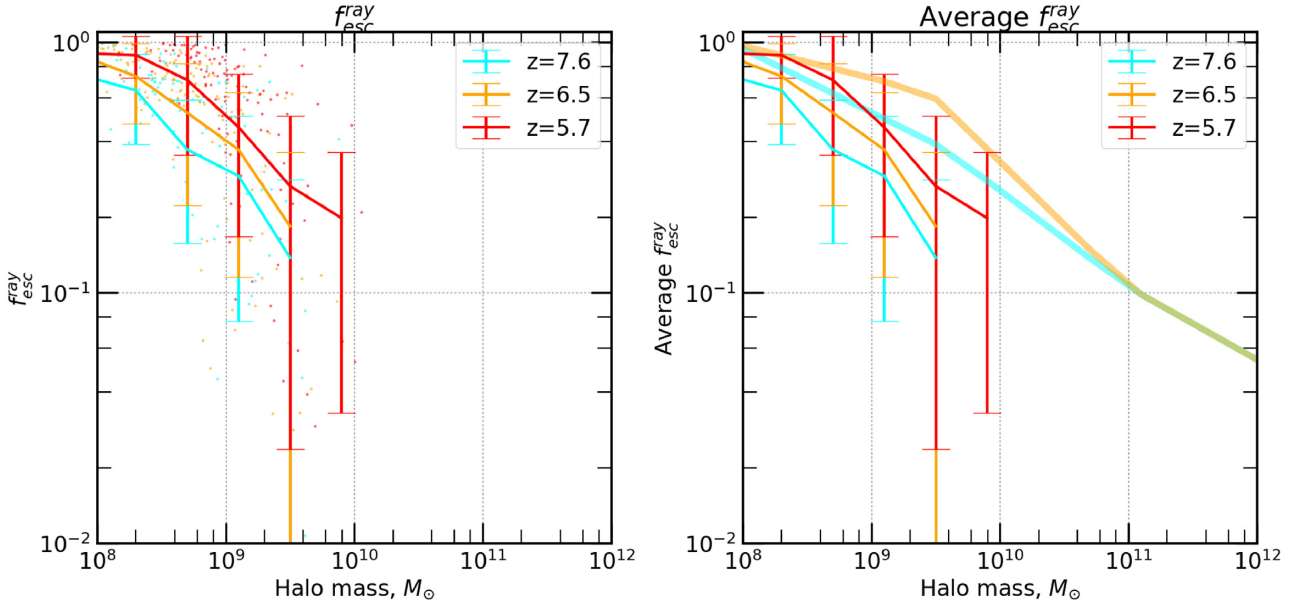
## REFERENCES

- Ahn K., Iliev I. T., Shapiro P. R., Mellema G., Koda J., Mao Y., 2012, *ApJ*, 756, L16
- Anderson L., Governato F., Karcher M., Quinn T., Wadsley J., 2017, *MNRAS*, 468, 4077
- Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, *MNRAS*, 479, 5184
- Aubert D., Teyssier R., 2008, *MNRAS*, 387, 295
- Aubert D., Deparis N., Ocvirk P., 2015, *MNRAS*, 454, 1012
- Aubert D. et al., 2018, *ApJ*, 856, L22
- Barkana R., Loeb A., 2007, *Rep. Prog. Phys.*, 70, 627
- Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, *ApJ*, 843, 129
- Dawoodbhoy T. et al., 2018, *MNRAS*, 480, 1740
- Dayal P., Ferrara A., 2018, *Phys. Rev.*, 780, 1
- Dayal P. et al., 2020, *MNRAS*, 495, 3065
- Deparis N., Aubert D., Ocvirk P., Chardin J., Lewis J., 2019, *A&A*, 622, A142
- Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, *Publ. Astron. Soc. Aust.*, 34, e058
- Ferrara A., Loeb A., 2013, *MNRAS*, 431, 2826
- Fialkov A., Barkana R., Visbal E., Tseliakhovich D., Hirata C. M., 2013, *MNRAS*, 432, 2909
- Gnedin N. Y., 2016, *ApJ*, 833, 66
- Górski K. M., Hivon E., Banday A. J., Wand elt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
- Haardt F., Salvaterra R., 2015, *A&A*, 575, L16
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Iliev I. T., Mellema G., Pen U.-L., Merz H., Shapiro P. R., Alvarez M. A., 2006, *MNRAS*, 369, 1625
- Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, *MNRAS*, 439, 725
- Izotov Y. I., Orlitová I., Schaerer D., Thuan T. X., Verhamme A., Guseva N. G., Worseck G., 2016, *Nature*, 529, 178
- Katz H., Kimm T., Haehnelt M., Sijacki D., Rosdahl J., Blaizot J., 2018, *MNRAS*, 478, 4986
- Katz H., Kimm T., Haehnelt M. G., Sijacki D., Rosdahl J., Blaizot J., 2019, *MNRAS*, 483, 1029
- 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
- Kroupa P., 2001, *MNRAS*, 322, 231
- Legrand L. et al., 2019, *MNRAS*, 486, 5468
- Levermore C. D., 1984, *J. Quant. Spectrosc. Radiat. Transfer*, 31, 149
- Liu C., Mutch S. J., Angel P. W., Duffy A. R., Geil P. M., Poole G. B., Mesinger A., Wyithe J. S. B., 2016, *MNRAS*, 462, 235
- Ma X., Kasen D., Hopkins P. F., Faucher-Giguère C.-A., Quataert E., Kereš D., Murray N., 2015, *MNRAS*, 453, 960
- Ma X. et al., 2018, *MNRAS*, 478, 1694
- Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, 411, 955
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Ocvirk P., Pichon C., Teyssier R., 2008, *MNRAS*, 390, 1326
- Ocvirk P. et al., 2016, *MNRAS*, 463, 1462
- Ocvirk P., Aubert D., Chardin J., Deparis N., Lewis J., 2019, *A&A*, 626, A77
- Ocvirk P. et al., 2020, *MNRAS*
- Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, *MNRAS*, 451, 2544
- Puchwein E., Haardt F., Haehnelt M. G., Madau P., 2019, *MNRAS*, 485, 47
- Razoumov A. O., Sommer-Larsen J., 2010, *ApJ*, 710, 1239
- Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, *ApJ*, 802, L19
- Rosdahl J., et al., 2018, *MNRAS*, 479, 994
- Roy F., Bouillot V. R., Rasera Y., 2014, *A&A*, 564, A13
- Sheth R. K., Mo H. J., Tormen G., 2001, *MNRAS*, 323, 1
- Sorce J. G. et al., 2016, *MNRAS*, 455, 2078
- Teyssier R., 2002, *A&A*, 385, 337
- Trebitsch M., Blaizot J., Rosdahl J., Devriendt J., Slyz A., 2017, *MNRAS*, 470, 224
- Trebitsch M., Volonteri M., Dubois Y., Madau P., 2018, *MNRAS*, 478, 5607
- Trebitsch M. et al., 2020, preprint (arXiv:2002.04045)
- Van Der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22
- Virtanen P. et al., 2020, *Nat. Methods*, 17, 261
- Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, *MNRAS*, 433, 1230
- Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, *MNRAS*, 442, 2560
- Wu X., Kannan R., Marinacci F., Vogelsberger M., Hernquist L., 2019, *MNRAS*, 488, 419
- Yajima H., Choi J.-H., Nagamine K., 2011, *MNRAS*, 412, 411
- Yoo T., Kimm T., Rosdahl J., 2020, preprint (arXiv:2001.05508)

## APPENDIX A: ESCAPE FRACTION RESOLUTION STUDY

We have shown that the main features of our  $f_{\text{esc}}^{\text{ray}}$  values are that they decrease with increasing halo mass above a certain mass scale, with haloes of  $\lesssim 10^9 M_{\odot}$  having high  $f_{\text{esc}}^{\text{ray}}$  values, and increase slightly with the progress of the EoR. Since this first trend is driven by the presence of dense, neutral, and UV opaque cells in haloes, one may expect that increasing the number of resolution elements (and thus within the same volume and same total mass, increasing the maximum possible density) could affect the exact relation between  $f_{\text{esc}}^{\text{ray}}$  and halo mass, as well as the mass scale at which the break between trends occurs.

To investigate this, and to test the robustness of our previously presented results, we proceed to study the resolution convergence of  $f_{\text{esc}}^{\text{ray}}$ . We ran a series of high-resolution simulations using RAMSES-CUDATON: we used a  $4 \text{ cMpc} \cdot \text{h}^{-1}$  sided box, with  $1024^3$  resolution elements (i.e. four times the spatial resolution of CoDa II, i.e. 64 times better mass resolution). We will call it high-res from now on. This set-up yields a comoving cell size of  $5.76 \text{ ckpc}$  ( $0.82 \text{ kpc}$  physical at  $z = 6$ ), and a dark matter particle mass of  $6.4 \times 10^3 M_{\odot}$ .



**Figure A1.** Left : Scatter and mass bin average of  $f_{\text{esc}}^{\text{ray}}$  of star-forming haloes over the full range of masses in the high-res simulation box | Right : average  $f_{\text{esc}}^{\text{ray}}$  of star-forming haloes for the high-res box, compared with our fitting formula (see equation B1) for CoDa II  $f_{\text{esc}}^{\text{ray}}$  (at redshifts that are appropriate,  $10 > z > 6$ ). Here, we plot within the mass range of CoDa II haloes.

All other parameters are kept exactly as in CoDa II, including the minimum stellar mass particle of  $11\,732\,M_{\odot}$ . The initial conditions are necessarily different from CoDa II because the high-res run uses a smaller box. As a consequence, it contains fewer star-forming haloes of a given mass than CoDa II, making  $f_{\text{esc}}^{\text{ray}}$  measurement noisier. The latter must therefore be carefully compared with our former results.

We performed our previous  $f_{\text{esc}}^{\text{ray}}$  measurements on the high-res simulation to test the impact of numerical resolution on  $f_{\text{esc}}^{\text{ray}}$  in CoDa II.

The left-hand panel of Fig. A1 shows  $f_{\text{esc}}^{\text{ray}}$  as a function of halo mass for the high-res simulation. This high-resolution case also presents high  $f_{\text{esc}}^{\text{ray}}$  values for low-mass haloes, as well as decreasing  $f_{\text{esc}}^{\text{ray}}$  with halo mass. Moreover,  $f_{\text{esc}}^{\text{ray}}$  increases on average with time for low-mass haloes, as in CoDa II. The right-hand panel presents a direct comparison of these averages with the fitting formula for CoDa II at the same redshifts where appropriate ( $10 \gtrsim z \gtrsim 6$ , the domain of validity for our fitting formula given in Fig. B1).

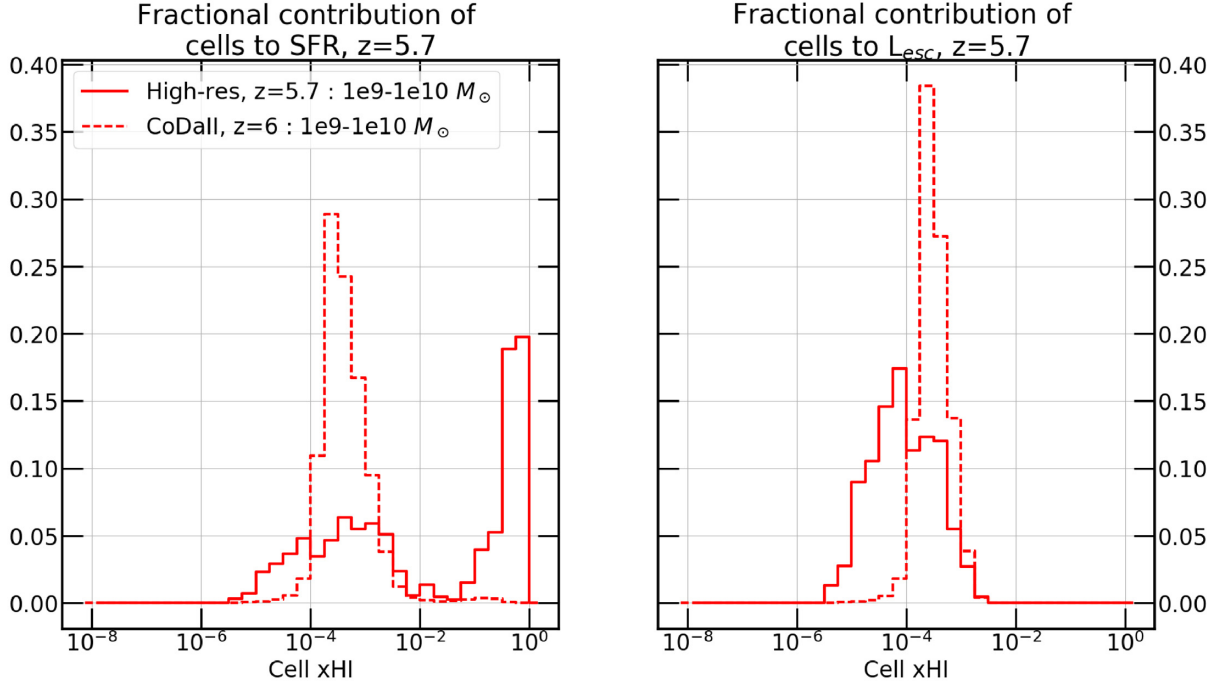
However, as anticipated, there are differences. In the high-res box, the average  $f_{\text{esc}}^{\text{ray}}$  is lower for all masses than in CoDa II, and the slope with mass is slightly more pronounced. However, there is a large scatter around the average of the high-res data set. Indeed, there are only a few hundred star-forming haloes of all masses in the high-res box at  $z = 5.7$ . Reassuringly, the difference between the two boxes is akin to a global re-normalization of the average  $f_{\text{esc}}^{\text{ray}}$ . Therefore, while numerical resolution may change the *absolute*  $L_{\text{esc}}$  of a given halo mass bin, it is not likely to change the *relative* balance between mass bins in the photon budget, which is our main result.

To explain the smaller  $f_{\text{esc}}^{\text{ray}}$  values measured in high-res, we proceed as in 3.3, and examine the properties of the gas of star-forming cells. Fig. A2 shows the distribution of the neutral fractions of star-forming cells in high-res, weighted by their SFR ( $L_{\text{esc}}$ ) in the

left-hand right)-hand panel, at  $z = 5.7$ . In high-res, the neutral fractions of star-forming cells in haloes of  $10^{10}\,M_{\odot} > M_{\text{halo}} > 10^9\,M_{\odot}$  are grouped into two peaks: an ionized mode, centred around  $X_{\text{HI}} \sim 10^{-4}$ , and a neutral/quasi-neutral mode at  $X_{\text{HI}} \sim 0.5$ . The  $L_{\text{esc}}$  weighted distribution shows that only the cells with  $X_{\text{HI}} \leq 10^{-3}$  contribute significantly to the final  $L_{\text{esc}}$  of the haloes in high-res. The cells belonging to the high neutral fraction account for  $\approx 50$  per cent of star formation, and do not contribute to haloes'  $L_{\text{esc}}$ .

Fig. A2 also shows the CoDa II distributions, allowing us to gauge directly the impact of increased resolution. The first striking difference between the two simulations is the shape of the SFR-weighted distribution. In CoDa II, most of the star formation in the  $10^{10}\,M_{\odot} > M_{\text{halo}} > 10^9\,M_{\odot}$  haloes happens in cells with neutral fractions  $< 10^{-3}$ , and there is no equivalent to the high neutral fraction peak seen in high-res. However, this feature is present in the distributions of cell  $x_{\text{HI}}$  in higher mass haloes in CoDa II, which are better resolved. This is a convincing hint that the increase of resolution between CoDa II and high-res allows higher density cells to exist within haloes of the same mass, in turn allowing for higher recombination rates, higher neutral fractions, and finally lower LyC transmissions and escape fractions.

There is, however, a second aspect worthy of discussion. In CoDa II, the star-forming cells of haloes of  $10^{10}\,M_{\odot} > M_{\text{halo}} > 10^9\,M_{\odot}$  do not exhibit neutral fractions smaller than  $10^{-4}$ , whereas in high-res the distributions of neutral fraction stretch all the way to  $10^{-5}$ . Higher mass haloes in CoDa II have similarly low neutral fractions that can only occur when the gas is heated by SN. This suggests that although the higher resolution of high-res allows for higher gas densities that can decrease the escape fraction, it also permits the existence of lower gas density cells that are more susceptible to the heating from SN, plausibly increasing the  $f_{\text{esc}}^{\text{ray}}$  of haloes by allowing higher transmission cells to exist. The fact that the escape fractions of



**Figure A2.** Distribution of star-forming halo cell neutral fractions for haloes of masses  $10^{10} M_{\odot} > M_{\text{halo}} > 10^9 M_{\odot}$ . The full lines show the curves for the high-res simulation, whereas the dashed lines show the curves for the CoDaII simulation. The histograms show the contribution of halo cells to (left:) total SFR of the mass bin considered, and to (right:) the total escape luminosity of the mass bin considered.

the high-res box are lower suggests that this transmission boost does not compensate the opacity boost occurring at higher resolution, and the latter dominates.

## APPENDIX B: ESCAPE FRACTION FITS

In this section, we propose simple functional forms for the both the average of  $f_{\text{esc}}^{\text{ray}}$  and the SFR-weighted average of  $f_{\text{esc}}^{\text{ray}}$  as a function of halo mass and of redshift. These could be useful as a  $f_{\text{esc}}^{\text{ray}}$  model for projects in which the determination of  $f_{\text{esc}}^{\text{ray}}$  is impossible or difficult: either when working with simulated data that was not produced with fully coupled radiation-hydrodynamics simulation codes, performing radiative transfer in post-processing, or with simulations with lower spatial resolution than CoDa II, and also for semi-analytical models of the EoR (Mesinger et al. 2011; Fialkov et al. 2013).

In all cases, we caution the reader that the functions presented here represent the halo escape fraction, i.e. to obtain the net halo escape fraction of haloes,  $f_{\text{esc}}^{\text{net}}$ , one needs to multiply  $f_{\text{esc}}^{\text{ray}}$  by  $f_{\text{esc}}^{\text{sub}}$  (0.42), as show in equation (5).

We provide the reader with PYTHON functions for plotting the fits that follow here (github).

### B1 Average escape fraction

Based on the trends presented by the average values of  $f_{\text{esc}}^{\text{ray}}$  as a function of mass and of redshift between  $z = 6$  and  $z = 10$  shown in Fig. 4, we opted for a double power law of halo mass, with a knee separating both laws, that shifts with redshift. Due to the noisy appearance of the data for  $M_{\text{halo}} > 10^{11} M_{\odot}$ , the time behaviour of  $f_{\text{esc}}^{\text{ray}}$  within this mass range is unclear, which is why

we chose to simply model the evolution of  $f_{\text{esc}}^{\text{ray}}$  with mass as the extension of the  $f_{\text{esc}}^{\text{ray}}(M_{\text{halo}}, z = 10)$  fit, to produce indicative values despite the high amount of noise in the average curves for these masses.

We propose the following functional form (equation B1):

$$f_{\text{esc}}^{\text{ray}} = \begin{cases} f_{\text{esc}}^{\text{ray}, 10.1 \leq z \leq 14.9}(M_{\text{halo}}, z); & \text{if } M_{\text{halo}} > M_{\text{join}}, \\ f_{\text{esc}}^{\text{knee}}(z) f_{\text{esc}}^{\text{knee}}(M_{\text{halo}}, z); & \text{if } M_{\text{join}} \geq M_{\text{halo}} \geq M_{\text{knee}}(z), \\ f_{\text{esc}}^{\text{max}}(z) f_{\text{esc}}^{\text{max}}(M_{\text{halo}}, z); & \text{if } M_{\text{halo}} < M_{\text{knee}}(z), \end{cases} \quad (\text{B1})$$

$$\text{with} = \begin{cases} z_6 & = \left( \frac{1+z}{1+6} \right), \\ f_{\text{esc}}^{\text{max}}(z) & = \min \left( f_{\text{esc}}^{0, \text{max}} z_6^{\beta}, 1.0 \right), \\ f_{\text{esc}}^{\text{knee}}(z) & = \max \left( \min \left( f_{\text{esc}}^{0, \text{knee}} z_6^{\gamma}, 1.0 \right), f_{\text{min}}^{\text{knee}} \right), \\ \Delta^{\text{max}}(z) & = \frac{\log_{10}(f_{\text{esc}}^{\text{knee}}(z)) - \log_{10}(f_{\text{esc}}^{\text{max}}(z))}{\log_{10}(M_{\text{knee}}(z)) - \log_{10}(10^8)}, \\ \Delta^{\text{knee}}(z) & = \frac{\log_{10}(f_{\text{esc}}^{\text{knee}}(z)) - \log_{10}(f_{\text{esc}}^{\text{join}})}{\log_{10}(M_{\text{knee}}(z)) - \log_{10}(M_{\text{join}})}, \\ M_{\text{knee}}(z) & = 2.5 \times 10^9 z_6^{\zeta} M_{\odot}, \\ f_{\text{esc}}^{\text{max}}(M_{\text{halo}}, z) & = \left( \frac{M_{\text{halo}}}{M_{\text{knee}}(z)} \right)^{\Delta^{\text{max}}(z)}, \\ f_{\text{esc}}^{\text{knee}}(M_{\text{halo}}, z) & = \left( \frac{M_{\text{halo}}}{M_{\text{knee}}(z)} \right)^{\Delta^{\text{knee}}(z)}, \end{cases} \quad (\text{B2})$$

where  $f_{\text{esc}}^{\text{max}}(z)$  is the maximum value below  $M_{\text{knee}}$ , it is  $f_{\text{esc}}^{0, \text{max}}$  at  $z = 6$ , and evolves as  $z^{\beta}$ .

$f_{\text{esc}}^{\text{knee}}(z)$  is where both power laws join at  $M_{\text{knee}}(z)$ , it is  $f_{\text{esc}}^{0,\text{knee}}$  at  $z = 6$ , and evolves as  $z^\gamma$ . It cannot be lower than  $f_{\text{min}}^{\text{knee}}$ .

$\Delta^{\text{max}}(z)$  is the slope of the power law above  $M_{\text{knee}}(z)$ , it is defined so as to reach  $f_{\text{esc}}^{\text{max}}(z)$  when  $M = 10^8 M_\odot$ .

$\Delta^{\text{knee}}(z)$  is the slope of the power law below  $M_{\text{knee}}(z)$ , it is defined so as to reach  $f_{\text{esc}}^{\text{knee}}(z)$  when  $M = M_{\text{knee}}(z)$ , and  $f_{\text{esc}}^{\text{join}}$  when  $M = M_{\text{join}}$ .

$M_{\text{knee}}(z)$  gives the mass where the power laws join, it is  $2.5 \times 10^9 M_\odot$  at  $z = 6$ , and evolves with  $z$  as  $z^\zeta$ .

$f_{\text{esc}}^{\text{max}}(M_{\text{halo}}, z)$  accounts for the mass slope for  $M_{\text{halo}} < M_{\text{knee}}(z)$  given by  $\Delta(z)$ .

$f_{\text{esc}}^{\text{knee}}(M_{\text{halo}}, z)$  accounts for the mass slope for  $M_{\text{halo}} \geq M_{\text{knee}}(z)$  given by  $\delta$ .

The adjusted values are summarized in equation (B3) next:

$$\begin{cases} f_{\text{esc}}^{0,\text{max}} = 0.98, \\ f_{\text{esc}}^{0,\text{knee}} = 0.8, \\ M_{\text{join}} = 10^{11} M_\odot, \\ f_{\text{esc}}^{\text{join}} = 0.1, \\ f_{\text{min}}^{\text{knee}} = 0.23, \\ \beta = -0.5, \\ \gamma = -3.5, \\ \zeta = -2.0. \end{cases} \quad (\text{B3})$$

The average curves corresponding to  $10 \leq z \leq 14.9$  behave somewhat differently. They appear to be simple power laws, offset in escape fraction by  $\sim 0.25$  dex. Hence, instead of using the previously discussed formulas for this redshift range, we tentatively provide the following fit (equation B4) for  $f_{\text{esc}}^{\text{ray}}$  for  $10 \leq z \leq 14.9$ .

$$f_{\text{esc}}^{\text{ray}, 10.1 \leq z \leq 14.9}(M_{\text{halo}}, z) = \min \left( f_{\text{esc},0}^{10 \leq z \leq 14.9} \times \left( \frac{M_{\text{halo}}}{10^8 M_\odot} \right)^{\Delta^{10 \leq z \leq 14.9}}, \right. \\ \left. \times \left( \frac{10.}{z} \right)^\gamma, 1 \right), \quad (\text{B4})$$

$$\text{with } \begin{cases} f_{\text{esc},0}^{10.1 \leq z \leq 14.9} = 0.77, \\ \Delta^{10.1 \leq z \leq 14.9} = \frac{\log_{10}(0.77) - \log_{10}(f_{\text{esc}}^{\text{join}})}{\log_{10}(10^8 M_\odot) - \log_{10}(M_{\text{join}})}, \\ \gamma = -1.64. \end{cases} \quad (\text{B5})$$

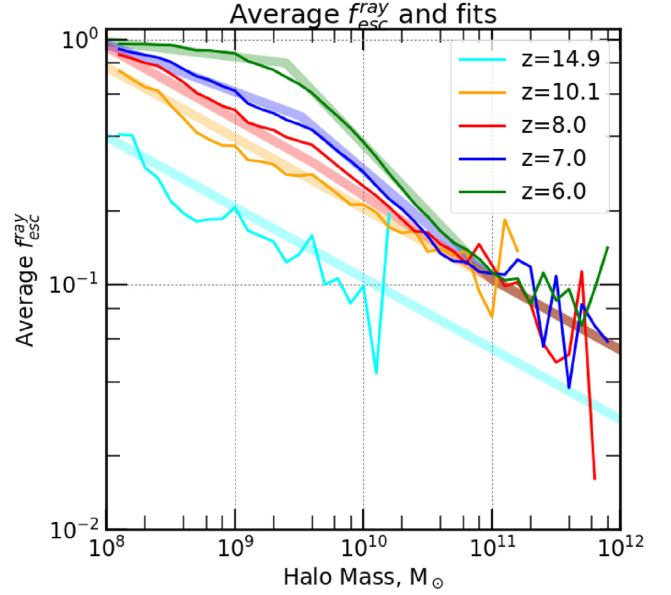
Once again, as we aim to represent the averages and not fit the data, we adjust these values by hand.

The aforementioned fits are presented in Fig. B1 (the full, thick lines), where they are also compared to the mass bin averages of  $f_{\text{esc}}^{\text{ray}}$  (the full, thin lines).

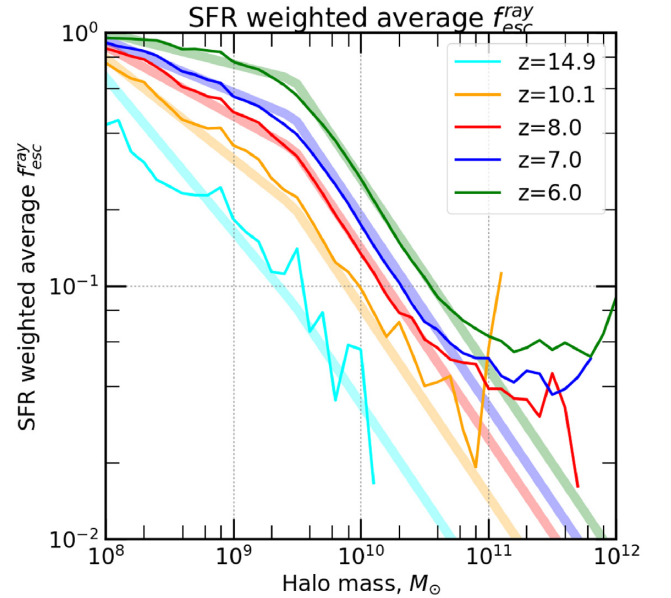
## B2 SFR-weighted average escape fraction

We also provide fits for the SFR-weighted average values of  $f_{\text{esc}}^{\text{ray}}$  as a function of mass and of redshift between  $z = 6$  and  $z = 14.9$ . As can be seen in Fig. B2, the SFR-weighted average of  $f_{\text{esc}}^{\text{ray}}$  has a slightly different behaviour. Indeed, in this case, the average curves do not join above a certain mass. To fit them, we opted for a slightly different double power law of halo mass, with a knee separating both laws. We propose the following functional form (equation B6):

$$f_{\text{esc}} = \begin{cases} f_{\text{esc}}^{\text{knee}}(z) f_{\text{esc}}^{\text{knee}}(M_{\text{halo}}, z); & \text{if } M_{\text{halo}} \geq M_{\text{knee}}(z), \\ f_{\text{esc}}^{\text{max}}(z) f_{\text{esc}}^{\text{max}}(M_{\text{halo}}, z); & \text{if } M_{\text{halo}} < M_{\text{knee}}(z), \end{cases} \quad (\text{B6})$$



**Figure B1.** Average escape fractions (thin lines) of star-forming haloes and corresponding fits (thick lines) at  $z = 6, 7, 8, 10.1, 14.9$ .



**Figure B2.** SFR-weighted average escape fractions (the thin lines) of star-forming haloes and corresponding fits (the thick lines) at  $z = 6, 7, 8, 10.1, 14.9$ .

$$\text{with } \begin{cases} z_6 = \left( \frac{1+z}{1+6} \right), \\ f_{\text{esc}}^{\text{max}}(z) = \min \left( f_{\text{esc}}^{0,\text{max}} z_6^\beta, 1.0 \right), \\ f_{\text{esc}}^{\text{knee}}(z) = \min \left( f_{\text{esc}}^{0,\text{knee}} z_6^\gamma, 1.0 \right), \\ \Delta(z) = \frac{\log_{10}(f_{\text{esc}}^{\text{knee}}(z)) - \log_{10}(f_{\text{esc}}^{\text{max}}(z))}{\log_{10}(M_{\text{knee}}(z)) - \log_{10}(10^8)}, \\ M_{\text{knee}}(z) = 3 \times 10^9 M_\odot, \\ f_{\text{esc}}^{\text{max}}(M_{\text{halo}}, z) = \left( \frac{M_{\text{halo}}}{M_{\text{knee}}(z)} \right)^{\Delta(z)}, \\ f_{\text{esc}}^{\text{knee}}(M_{\text{halo}}, z) = \left( \frac{M_{\text{halo}}}{M_{\text{knee}}(z)} \right)^\delta, \end{cases} \quad (\text{B7})$$



where  $f_{\text{esc}}^{\text{max}}(z)$  is the maximum value below  $M_{\text{knee}}(z)$ , it is  $f_{\text{esc}}^{0,\text{max}}$  at  $z = 6$ , and evolves as  $z^\beta$ .

$f_{\text{esc}}^{\text{knee}}(z)$  is where both power laws join at  $M_{\text{knee}}(z)$ , it is  $f_{\text{esc}}^{0,\text{knee}}$  at  $z = 6$ , and evolves as  $z^\gamma$ .

$\Delta(z)$  is the slope of the power law below  $M_{\text{knee}}(z)$ , it is defined so as to reach  $f_{\text{esc}}^{\text{max}}(z)$  when  $M_{\text{halo}} = 10^8 M_\odot$ .

$M_{\text{knee}}(z)$  gives the mass where the power laws join, it is  $3 \times 10^9 M_\odot$  at  $z = 6$ , and evolves with  $z$  as  $z^\zeta$ .

$f_{\text{esc}}^{\text{max}}(M_{\text{halo}}, z)$  accounts for the mass slope for  $M_{\text{halo}} < M_{\text{knee}}(z)$  given by  $\Delta(z)$ .

$f_{\text{esc}}^{\text{knee}}(M_{\text{halo}}, z)$  accounts for the mass slope for  $M_{\text{halo}} \geq M_{\text{knee}}(z)$  given by  $\delta$ .

Since we want to reproduce the behaviour with mass and with redshift of the SFR-weighted averages of  $f_{\text{esc}}^{\text{ray}}$ , and not model the full distribution of points, we again adopt the simple approach of adjusting the fit by hand (as opposed to computing the fit of the model to the data sample).

The adjusted values are summarized in equation (B8) next:

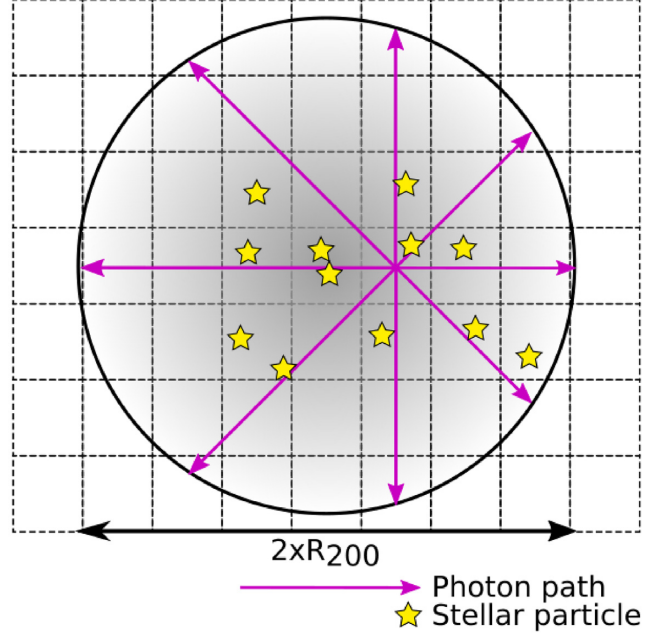
$$\begin{cases} f_{\text{esc}}^{0,\text{max}} = 1.0, \\ f_{\text{esc}}^{0,\text{knee}} = 0.65, \\ \beta = -0.5, \\ \gamma = -2.5, \\ \delta = -0.75, \end{cases} \quad (\text{B8})$$

The aforementioned fits are presented in Fig. B2 (the full, thick lines), where they are also compared to the SFR-weighted average  $f_{\text{esc}}^{\text{ray}}$  measure in CoDa II (the full, thin lines).

### APPENDIX C: COMPUTING ESCAPE FRACTION

Fig. C1 shows a simplified, explanatory drawing of the computation process for  $f_{\text{esc}}^{\text{ray}}$  for an individual star-forming halo cell.

### Example paths from a source cell



**Figure C1.** Explanatory drawing of the computation of  $f_{\text{esc}}^{\text{ray}}$ . For each cell containing an emitting particle, we trace 768 rays from its centre to a sphere of radius  $r_{200}$  centred on the halo centre given by the fof halo finder. By averaging this result for every emitting cell and weighting by SFR, we obtain the halo value  $f_{\text{esc}}^{\text{ray}}$  at  $r_{200}$ .

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