

The THESAN project: ionizing escape fractions of reionization-era galaxies

Jessica Y.-C. Yeh ^{1,★}, Aaron Smith ^{2,1,†}, Rahul Kannan ^{3,2}, Enrico Garaldi ⁴, Mark Vogelsberger ^{1,5},
Josh Borrow ¹, Rüdiger Pakmor ⁴, Volker Springel ⁴ and Lars Hernquist²

¹Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

³Department of Physics and Astronomy, York University, 4700 Keele Street, Toronto, Ontario, Canada, M3J 1P3

⁴Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany

⁵The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Massachusetts Institute of Technology, Cambridge MA 02139, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

A fundamental requirement for reionizing the Universe is that a sufficient fraction of the ionizing photons emitted by galaxies successfully escapes into the intergalactic medium. However, due to the scarcity of high-redshift observational data, the sources driving reionization remain uncertain. In this work, we calculate the ionizing escape fractions (f_{esc}) of reionization-era galaxies from the state-of-the-art THESAN simulations, which combine an accurate radiation-hydrodynamic solver (AREPO-RT) with the well-tested IllustrisTNG galaxy formation model to self-consistently simulate both small-scale galaxy physics and large-scale reionization throughout a large patch of the universe ($L_{\text{box}} = 95.5 \text{ cMpc}$). This allows the formation of numerous massive haloes ($M_{\text{halo}} \gtrsim 10^{10} M_{\odot}$), which are often statistically underrepresented in previous studies but are believed to be important to achieving rapid reionization. We find that low-mass galaxies ($M_{\text{stars}} \lesssim 10^7 M_{\odot}$) are the main drivers of reionization above $z \gtrsim 7$, while high-mass galaxies ($M_{\text{stars}} \gtrsim 10^8 M_{\odot}$) dominate the escaped ionizing photon budget at lower redshifts. We find a strong dependence of f_{esc} on the effective star formation rate (SFR) surface density defined as the SFR per gas mass per escape area, i.e. $\bar{\Sigma}_{\text{SFR}} = \text{SFR}/M_{\text{gas}}/R_{200}^2$. The variation in halo escape fractions decreases for higher mass haloes, which can be understood from the more settled galactic structure, SFR stability, and fraction of sightlines within each halo significantly contributing to the escaped flux. Dust is capable of reducing the escape fractions of massive galaxies, but the impact on the global f_{esc} depends on the dust model. Finally, active galactic nuclei are unimportant for reionization in THESAN and their escape fractions are lower than stellar ones due to being located near the centres of galaxy gravitational potential wells.

Key words: radiative transfer – methods: numerical – galaxies: high-redshift – cosmology: dark ages, reionization, first stars

1 INTRODUCTION

The formation of the first stars and galaxies marks the end of the cosmic dark ages and the beginning of the Epoch of Reionization (EoR; Bromm & Yoshida 2011; Loeb & Furlanetto 2013). Lyman continuum (LyC; photon energy $h\nu > 13.6 \text{ eV}$) radiation capable of ionizing atomic hydrogen was emitted by a growing number of sources, including young, hot massive stars. These photons primarily interacted with the surrounding interstellar medium (ISM), although a significant fraction successfully traversed through the circumgalactic medium (CGM) of their host galaxies to emerge into the vast intergalactic medium (IGM). The result was to transform the originally cold and neutral atoms into a hot plasma through photoionization and photoheating (Madau et al. 1999; Barkana & Loeb 2001). However, the complex nature of this last major phase transition of the Universe still presents many important unanswered questions. For instance, what are the main sources and sinks of ionizing photons in this period, what determines the morphology and coalescence

timeline of ionized bubbles, and how do these translate to observational signatures for current and upcoming facilities? To answer these questions, significant effort has been made to better understand the ionizing source emissivity, either through inferred measurements or theoretical modelling (e.g. Robertson et al. 2010; Wise 2019; Eide et al. 2020). It is widely believed that star-forming galaxies are the primary sources of reionization (Haardt & Madau 2012; Faucher-Giguère 2020). Throughout this paper we focus on the escape fraction f_{esc} , which we define as the ratio between the number of photons that escape galaxies to reach the IGM and the number of photons intrinsically emitted. As galaxy formation constraints improve, e.g. for stellar population synthesis modelling and stellar-to-halo mass relations, the escape fractions directly dictate the amount of radiation galaxies can contribute to reionization. However, the majority of the measured values of escape fractions to date are not high enough to fully reionize the Universe (Finkelstein et al. 2012; Robertson et al. 2013), and thus expose our lack of a complete understanding of reionization. In the following we briefly summarize observational and theoretical approaches that have been explored to rectify this inconsistency.

There are at least two major challenges when determining f_{esc} observationally. First, the high opacity of the intervening neutral IGM

★ E-mail: ycyeh@mit.edu

† E-mail: aaron.smith@cfa.harvard.edu; NHFP Einstein Fellow.

at high redshifts means that very limited observational data are available at present. In fact, certain properties such as the intrinsic LyC flux may never be directly observable, while other measurements such as Lyman- α (Ly α) absorption spectra become highly saturated. Secondly, it has been shown that f_{esc} has large sightline-to-sightline variability (Cen & Kimm 2015). In other words, individual observations may not be able to yield a representative picture and a large number of galaxies are needed to provide robust constraints. Despite these difficulties, several observations have been made by measuring the leakage of LyC photons, but due to detection biases most of which originate from starburst galaxies at $z \lesssim 3$, and almost all results indicate low f_{esc} . For example, f_{esc} of a few per cent at $z \approx 1$ were detected in star-forming galaxies by Bridge et al. (2010), Siana et al. (2010), and Rutkowski et al. (2016). More optimistically, averages of $f_{\text{esc}} \sim 10\%$ at $z \approx 3$ were found by Nestor et al. (2013), Grazian et al. (2017), Japelj et al. (2017), Steidel et al. (2018), and Pahl et al. (2021). Recently the Low-Redshift Lyman Continuum Survey (LzLCS) measured a median f_{esc} of 4% from galaxies at $z = 0.2\text{--}0.4$ (Flury et al. 2022; Saldana-Lopez et al. 2022). Only a few galaxies with uncharacteristically high f_{esc} have been reported (Vanzella et al. 2018; Izotov et al. 2018; Rivera-Thorsen et al. 2019). Looking forward, we expect more data to become available from upcoming facilities, including probing the faint and bright ends of the UV luminosity function with the *James Webb Space Telescope* (JWST) and *Nancy Grace Roman Space Telescope* (NGRST), as well as 21 cm cosmology measurements at high-redshift with the Low-Frequency Array (LOFAR), Hydrogen Epoch of Reionization Array (HERA), and Square Kilometer Array (SKA).

The uncertainty in determining escape fractions propagates to different predictions for the global reionization history. In particular, if the dominant photon sources are low-mass galaxies ($M_{\text{halo}} \lesssim 10^9 M_{\odot}$), then reionization will have started earlier (e.g. Finkelstein et al. 2019), reaching an IGM ionized volume filling fraction of 0.5 at $z = 9$. On the other hand, late reionization is favoured if massive galaxies ($M_{\text{halo}} \gtrsim 10^{10} M_{\odot}$) prevail, potentially delaying the mid-point of reionization to around $z = 6.8$ (e.g. Naidu et al. 2020). In this scenario reionization proceeds significantly more rapidly than early reionization models, especially at the tail end (Robertson et al. 2015). Recently, late reionization has gained popularity, increasing the importance of studying f_{esc} from massive galaxies. This is supported by the measurement of a much lower optical depth for electron scattering of cosmic microwave background (CMB) photons than previous studies (Planck Collaboration et al. 2020), a rapid decrease of the observed Ly α emission (Schenker et al. 2014; Mason et al. 2019), and the damping wing absorption on the spectrum of high-redshift quasars (Davies et al. 2018). Collectively, the observations suggest that reionization was nearly complete by $z \approx 6$ (McGreer et al. 2015) with large islands of neutral hydrogen remaining in underdense regions below $z \approx 5.5$ (Kulkarni et al. 2019).

The difficulty of obtaining statistically significant data at high redshift ($z \gtrsim 6$) makes theoretical modelling critical in the study of the EoR. However, determining f_{esc} with numerical simulations is also an extremely complicated problem since the details can depend on the models used for galaxy formation, feedback recipes, multiphase ISM, and dust physics. Some early studies based on cosmological simulations predict that escape fractions are correlated with the galaxy mass (Gnedin et al. 2008; Wise & Cen 2009) while some suggest the opposite trend (Yajima et al. 2011). On the contrary, a combination of studies from more recent radiation-hydrodynamic or post-processed radiative transfer simulations indicate a more complicated dependence of f_{esc} on halo mass: f_{esc} decreases with halo mass for haloes up to $10^9 M_{\odot}$ (Paardekooper et al. 2015; Xu et al. 2016)

with high f_{esc} in faint galaxies (Kimm et al. 2017). Ma et al. (2020) found the f_{esc} increases with halo mass for $M_{\text{halo}} = 10^8\text{--}10^{9.5} M_{\odot}$ and decreases for more massive haloes at $\gtrsim 10^{11} M_{\odot}$. Predicted f_{esc} from a wide range of hydrodynamical simulations have been found to be from a few percent to thirty percent (Ma et al. 2015; Paardekooper et al. 2015; Xu et al. 2016; Rosdahl et al. 2018; Trebitsch et al. 2018; Ma et al. 2020, 2021; Rosdahl et al. 2022). The potential boost in f_{esc} due to binary stars has also been explored (Ma et al. 2016, 2022). However, due to the high computational cost associated with large-volume simulations, most above-mentioned simulations either only have haloes with M_{halo} up to $10^9 M_{\odot}$ or do not have enough statistics in massive haloes ($M_{\text{halo}} \gtrsim 10^{11} M_{\odot}$) and therefore cannot provide a complete picture of the role of more massive galaxies in the reionization process. We note that the recent CoDa simulations (Ocvirk et al. 2020) feature large-volume (94.5 cMpc) boxes and also model haloes down to $\sim 10^8 M_{\odot}$. The associated ray-tracing analysis by Lewis et al. (2020) shows a decreasing f_{esc} with increasing halo mass and highlights the significance of intermediate-mass haloes ($6 \times 10^8 < M_{\text{halo}}/M_{\odot} < 3 \times 10^{10}$). However, the CoDa II (III) simulations have relatively low spatial resolution due to the fact that they use a uniform grid code with 23 (11.5) cMpc cells, which impacts both the galaxy formation physics as well as the sourcing and escape of ionizing photons. In fact, Lewis et al. (2020) found that cell escape fractions are bimodal depending on whether the host cell is ionized or neutral. It remains unclear whether coarse resolution within galaxies can accurately capture the physics of escaped emission, and thus it is important to compare different simulations to provide further insights.

In this work, we provide a detailed study of ionizing escape fraction statistics and the underlying properties of the galaxies that contribute to reionization by using the THESAN simulations (Kannan et al. 2022a; Garaldi et al. 2022; Smith et al. 2022a). The strength of THESAN is the significantly expanded range of physical processes, based on combining the IllustrisTNG galaxy formation model with a self-consistent on-the-fly treatment of radiation with AREPO-RT, while simulating a large volume of the Universe at state-of-the-art resolution. As a result, THESAN successfully produces realistic galaxies that match important physical quantities relevant to the EoR, and has demonstrated its strength with a breadth of topics including 21 cm cosmology (Kannan et al. 2022a; Qin et al. 2022), IGM–galaxy connections (Garaldi et al. 2022), Lyman- α emission and transmission (Smith et al. 2022a; Xu et al. 2022), and line intensity mapping (Kannan et al. 2022b).

The paper is organized as follows. In Section 2, we describe the THESAN simulations and radiative transfer analysis methodology. In Section 3, we present insights into the global production and escape statistics of LyC photons. In Section 4, we explore the dependence on halo properties and variation across galaxy populations. In Section 5, we discuss the halo-to-halo variations in escape fractions and characterize the anisotropic behaviour across sightlines from individual haloes. In Section 6, we discuss several important caveats, including the subresolution modelling, dust pre-absorption of ionizing photons, and contribution of active galactic nuclei (AGNs). Finally, in Section 7, we provide a summary of our main results and possible future investigations.

2 METHODS

We briefly describe the THESAN reionization simulations in Section 2.1, focusing on aspects of the model that are particularly relevant for our study of escape fractions in the EoR. In Section 2.3 we

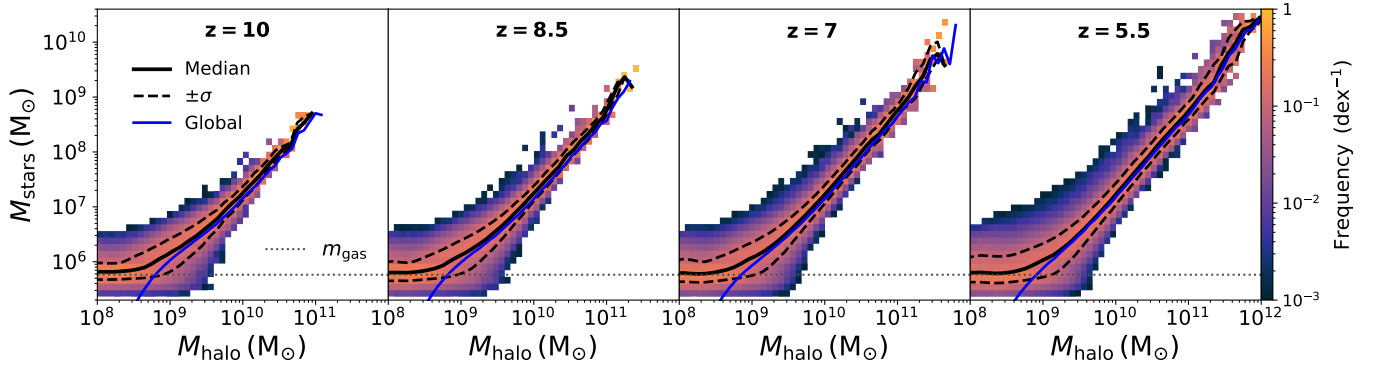


Figure 1. The stellar-to-halo mass relation for all stars within R_{200} for all selected haloes at $z = \{10, 8.5, 7, 5.5\}$. The black solid curves show the median while the black dashed curves show the 16th and 84th percentiles of the selected data. The blue curves show the average stellar mass from the full halo catalogue. The mismatch between the selected haloes and the full catalogue comes from the fact that only haloes with at least one star particle are selected. The flattening at the low-mass end is due to the baryonic mass resolution of $m_{\text{gas}} \approx 5.8 \times 10^5 M_{\odot}$ from the THESAN simulation, as indicated by the gray dotted lines. The coloured histogram illustrates the full distribution with each bin normalized by the number of haloes in the given halo mass range, i.e. the following equation holds: $\int p(M_{\text{star}}|M_{\text{halo}}) d\log M_{\text{star}} = 1$.

discuss the halo selection and completeness of our escape fraction catalogues. Finally, in Section 2.4 we summarize the outcomes of the Monte Carlo radiative transfer (MCRT) calculations employed to obtain f_{esc} from galaxies extracted from the THESAN simulation.

2.1 THESAN simulations

The THESAN project is a suite of large-volume radiation-magneto-hydrodynamic simulations that self-consistently model the hydrogen reionization process and the resolved properties of the sources (galaxies and AGN) responsible for it. The simulations were performed with AREPO-RT (Kannan et al. 2019), a radiation-hydrodynamic extension of the moving mesh code AREPO (Springel 2010; Weinberger et al. 2020), which solves the fluid equations on an unstructured Voronoi mesh that is allowed to move along with the fluid flow for an accurate quasi-Lagrangian treatment of cosmological gas flows. Gravitational forces are calculated using a hybrid Tree-PM approach, which splits the force into a short-range force that is computed using an oct-tree algorithm (Barnes & Hut 1986) and a long-range force, estimated using a particle mesh approach.

Radiation fields are modelled using a moment based approach that solves the zeroth and first moments of the radiative transfer equation (Rybicki & Lightman 1986), coupled with the M1 closure relation, that approximates the Eddington tensor based on the local properties of each cell (Levermore 1984). They are coupled to the gas via a non-equilibrium thermochemistry module, which self-consistently calculates the ionization states and cooling rates from hydrogen and helium, while also including equilibrium metal cooling and Compton cooling of the CMB (see Section 3.2.1 of Kannan et al. 2019, for more details). Both stars and AGN act as sources of radiation, with the spectral energy distribution (SED) of stars taken from the Binary Population and Spectral Synthesis models (BPASS; Eldridge et al. 2017). The AGN radiation output is scaled linearly with the mass accretion rate with a radiation conversion efficiency of 0.2 (Weinberger et al. 2018) and a Lusso et al. (2015) parametrization for the shape of its spectrum.

The prescriptions for processes happening below the resolution limit of the simulations, such as star and black hole formation and feedback and metal production and enrichment, are taken from the IllustrisTNG model (Weinberger et al. 2017; Springel et al. 2018;

Pillepich et al. 2018a,b; Naiman et al. 2018; Marinacci et al. 2018; Nelson et al. 2018, 2019; Pillepich et al. 2019). The model is augmented with a scalar dust model that tracks the production, growth, and destruction of dust using the formalism outlined in McKinnon et al. (2017). An additional birth cloud escape fraction parameter, $f_{\text{esc}}^{\text{cloud}} = 0.37$, is added to mimic the absorption of LyC photons happening below the resolution scale of the simulation. The parameter is tuned such that the simulation reproduces a realistic late-reionization history (Kannan et al. 2022a), which matches the observed neutral fraction evolution in the Universe (Greig et al. 2017).

All THESAN simulations follow the evolution of a cubic patch of the universe with linear comoving size $L_{\text{box}} = 95.5 \text{ cMpc}$, and utilize variance-suppressed initial conditions (Angulo & Pontzen 2016). We employ a Planck Collaboration et al. (2016) cosmology (more precisely, the one obtained from their TT, TE, EE+lowP+lensing+BAO+JLA+H₀ dataset), i.e. $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$ with $h = 0.6774$, $\Omega_m = 0.3089$, $\Omega_{\Lambda} = 0.6911$, $\Omega_b = 0.0486$, $\sigma_8 = 0.8159$, and $n_s = 0.9667$, where all symbols have their usual meanings. In this paper we focus on the highest resolution THESAN-1 simulation, which has a dark matter mass resolution of $3.12 \times 10^6 M_{\odot}$ and a baryonic mass resolution of $5.82 \times 10^5 M_{\odot}$. The gravitational forces are softened on scales of 2.2 ckpc with the smallest cell sizes reaching 10 pc. This allows us to model atomic cooling haloes throughout the entire simulation volume.

2.2 Caveats about feedback and ISM physics

We now briefly discuss some relevant aspects of the galaxy formation model, which uses a subresolution treatment of the ISM as a two-phase gas where cold clumps are embedded in a smooth, hot phase produced by supernova explosions (Springel & Hernquist 2003). Such subgrid modelling comes with the territory of large-volume simulations with demonstrated agreement with observations down to $z = 0$ (Vogelsberger et al. 2020a), but certainly has implications for ionizing escape fractions. In particular, the temporary decoupling of wind particles from the hydrodynamics reduces the ability to create low-density channels through which ionizing radiation can escape (Trebitsch et al. 2017). Additionally, the effective equation of state (EoS) parametrization is known to result in too smooth of an ISM. The reduced clumpiness compared to simulations that allow

the formation of a cold phase leads to high volume-weighted densities in the ISM. This in turn artificially lowers the median escape fractions along with variations across sightlines and haloes. In the context of the subgrid ISM model used in THESAN, it is also difficult to self-consistently decouple escape fractions from the birth cloud and EoS cells, which correspond to cells with hydrogen number densities larger than $n_{\text{H}} > 0.13 \text{ cm}^{-3}$. The thermochemistry and opacity equations are treated the same for all cells during RT subcycles between hydrodynamical time-steps, but afterwards cells below the EoS pressure floor are reset to be ionized. Therefore, the birth cloud escape fraction calibration systematically accounts for both unresolved clouds and semitransparent EoS cells. This approach works on larger scales because these uncertainties are bundled into the escape fraction parameter, which is tuned to produce the correct reionization history. We are not aware of any large-volume reionization simulations that claim to fully resolve LyC escape down to the level of individual stars, as this may require subparsec cell sizes. Even simulations like SPHINX with a stellar LyC luminosity factor of unity should not be interpreted as properly resolving cloud absorption or escape channels, but rather as a favourable by-product of the combination of resolution, subgrid models, and assumed stellar evolution models, given that the overall escape fractions are not converged with resolution (see Appendix C of Rosdahl et al. 2022). To help interpret THESAN results more broadly, we explore the quantitative impact of altering the EoS ionization state in post-processing in Appendix B.

There may be several different pathways forward to connect sub-resolution physics with large-volume galaxy formation simulations. For example, compressible multifluid hydrodynamic schemes can help incorporate multiscale unresolved phase mixing (Weinberger & Hernquist 2023) while analytic and tuned prescriptions for cloud-wind interactions can help integrate turbulent mixing and radiative cooling (Fielding & Bryan 2022). Such effective clumping models may also inform the radiation transport to further our understanding of effective escape fractions through low resolution simulation cells (Buck et al. 2022) or via subgrid clumping factors in coarse-grained reionization simulations (Bianco et al. 2021). However, these may still produce too smooth of an ISM for self-consistent escape fractions from galaxy formation simulations. In fact, approximately converged first principles multiphase ISM simulations are only feasible for local stellar neighbourhoods and patches of Milky way disks (Kim & Ostriker 2017; Rathjen et al. 2021) or low-mass dwarf galaxies (Lahén et al. 2020; Steinwandel et al. 2020; Gutcke et al. 2021). We remain optimistic that high-resolution multiphase zoom-in and small-volume radiation-hydrodynamic simulations can be sufficiently realistic that comparisons to observations also inform us about feedback and ISM physics (e.g. Hopkins et al. 2018; Marinacci et al. 2019; Kannan et al. 2020).

2.3 Halo selection

The THESAN simulation uses the friends-of-friends (FoF; Springel 2005) algorithm as implemented on-the-fly in AREPO to group dark matter particles into resolved structures along with their associated gas and star particles. The SUBFIND algorithm as first described in Springel et al. (2001) is then used to identify gravitationally bound substructures, which form the basis of the haloes hosting the galaxies referred to throughout this paper. However, for clarity we define M_{halo} to be the halo mass within the radius at which the average density reaches 200 times the density of the universe, M_{200} . Similarly, we adopt this same definition for M_{stars} as the stellar mass within R_{200} , the star formation rate (SFR) as the sum of the instantaneous SFRs of gas particles within R_{200} , and the virial radius, i.e. $R_{\text{vir}} \equiv R_{200}$.

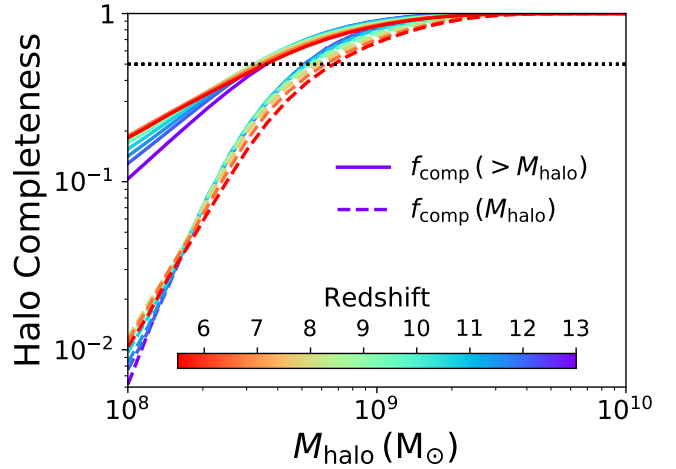


Figure 2. The completeness is defined as the fraction of selected haloes and haloes in the catalogue. The solid curves show the cumulative completeness for all haloes above a given mass and the dashed curves show the completeness at a specific mass at several different redshifts. The cumulative completeness exceeds 90% when $M_{\text{halo}} \gtrsim 10^9 M_{\odot}$ for nearly all redshifts, and reach 50% when $M_{\text{halo}} \approx 3 \times 10^8 M_{\odot}$ as shown by the black dotted line.

These definitions are chosen to ensure a consistent accounting of properties contributing to the escape fraction calculations.

At this point we select a subsample of haloes from the full catalogues that contain at least one star and four gas cells within the virial radius. On a practical level, these criteria are chosen to ensure Voronoi mesh reconstructions in ray-tracing are robust and therefore reasonable escape fractions can be defined. Although we do not claim all selected haloes are well resolved, our criteria maximize the completeness in halo selection and thus provide a less biased transition from selected haloes to non-selected haloes in the catalogue. We emphasize that any requirement on the number of gas or star particles induces selection biases, which in the context of the IllustrisTNG galaxy formation model arise from stochastic star formation leading to characteristically bursty star formation histories increasing the variation at the low-mass end of the stellar-to-halo mass relation (e.g. Genel et al. 2019; Iyer et al. 2020).

In Fig. 1 we show the stellar-to-halo mass relation for our selected sample of haloes at $z = \{10, 8.5, 7, 5.5\}$. The above selection criteria result in a wide range of haloes from $M_{\text{halo}} = 10^8 - 10^{12} M_{\odot}$. Haloes below $10^8 M_{\odot}$ are not well-resolved in the THESAN simulation, but these are only expected to be important during the earliest stages of reionization (e.g. Paardekooper et al. 2013; Wise et al. 2014). The stellar-to-halo mass relation in THESAN is in agreement with current observational estimates as shown in Fig. 9 of Kannan et al. (2022a). In Fig. 2 we show that our selection criteria result in a very high completeness for haloes more massive than $10^9 M_{\odot}$. That is, our selection covers most of the haloes with $M_{\text{halo}} > 10^9 M_{\odot}$ in the catalogue. Specifically, near the midpoint of reionization at $z \approx 7.67$ we achieve greater than {10, 50, 90}% cumulative completeness at halo masses of $M_{\text{halo}} \approx \{1.0 \times 10^8 M_{\odot}, 3.2 \times 10^8 M_{\odot}, 1.2 \times 10^9 M_{\odot}\}$.

2.4 Post-processing radiative transfer

As alluded to above, we define the escape fraction of a halo as the ratio between the number of ionizing photons that escape and are

intrinsically emitted from all sources within the virial radius, i.e.

$$f_{\text{esc}} \equiv \frac{\dot{N}_{\text{esc}}}{\dot{N}_{\text{int}}}. \quad (1)$$

The escape fractions for star particles outside the virial radius of any halo are assumed to be maximal (recall that $f_{\text{esc}}^{\text{cloud}} = 0.37$ from the birth cloud) as these directly contribute to reionizing the IGM. The same applies for stars in haloes with fewer than four gas particles (non-selected haloes), which do not possess enough opacity to lead to a significant amount of photon absorption in agreement with calculations of f_{esc} in low-mass haloes. The intrinsic photon rate for each subhalo (including stars outside R_{200}) is provided in the Ly α emission catalogue described in Smith et al. (2022a), which is calculated from the Binary Population and Spectral Synthesis models (BPASS version 2.2.1; Eldridge et al. 2017) with a Chabrier IMF (Chabrier 2003). The photon rates within R_{200} are calculated from the same exact age-metallicity tables as part of the radiative transfer calculations within the Cosmic Ly α Transfer (COLT) code (Smith et al. 2015, 2019). This ensures consistent results for global properties when accounting for all ionizing sources. The details of the photoionization physics and MCRT implementation are described in Smith et al. (2022b), and we only summarize the main features and differences here. Most importantly, we perform passive RT as our goal is to obtain escape fraction statistics based on the on-the-fly ionization states in the THESAN simulation rather than recompute new ones based on the assumption of photoionization equilibrium. Our escape fraction calculations employ 10^5 photon packets for haloes with fewer than 10 star particles, while the photon count increases linearly with the number of star particles up to a maximum of 10^7 packets for well-converged statistics. Similar to THESAN, we model the multi-frequency ionizing radiation field in three bands, H I, He I, and He II with energy edges of [13.6, 24.59, 54.42, ∞] eV. We employ continuous absorption algorithms for the treatment of photoionization, with frequency-dependent cross-sections taken from Verner et al. (1996). The emission direction is isotropic and the initial position corresponds to the location of the star. We employ native ray-tracing through the Voronoi tessellation.

Dust absorption has a non-negligible effect on the escape of LyC photons and has been discussed in previous studies (see e.g. Puglisi et al. 2016; Tacchella et al. 2018; Ma et al. 2020). THESAN integrates the production and destruction of dust based on the empirical model of McKinnon et al. (2017), and treats dust as a property of the gas resolution elements. As in Smith et al. (2022b), we can isolate the pre-absorption of ionizing photons by dust, according to

$$f_{\text{abs}} \equiv \int_0^\ell k_{\text{a,d}} e^{-k_{\text{a}} \ell'} d\ell', \quad (2)$$

where $k_{\text{a,d}}$ and k_{a} denote the dust and total absorption coefficients along the path integration of length ℓ . The absorption by hydrogen and helium can also be calculated in a similar way by replacing $k_{\text{a,d}}$ with $k_i = n_i \sigma_i$, $i \in \{\text{H I}, \text{He I}, \text{He II}\}$. We note that it is challenging to resolve dust in ionized H II regions with the resolution of the THESAN simulation because of the ISM and feedback models. Hence, the fiducial MCRT calculation in this paper assumes a no-dust scenario, and we postpone a more detailed estimation of the effect of dust absorption to Sec. 6. Finally, the ray-tracing procedure also allows us to conveniently track the average distance to absorption, defined as:

$$\langle \ell \rangle \equiv \ell^{-1} \int_0^\ell \ell' e^{-k_{\text{a}} \ell'} d\ell', \quad (3)$$

which provides complementary information about the ionization and dust absorption of photons that we explore in later sections.

3 GLOBAL STATISTICS

3.1 Large-scale structure

In Fig. 3, we present a visualization of the spatial distribution of the volume-weighted H II fraction $\langle x_{\text{H II}} \rangle$, gas overdensity $\delta \equiv \rho/\bar{\rho} - 1$, escaped ionizing photon rate densities ρ_{esc} , and LyC escape fractions f_{esc} across a $95.5 \times 95.5 \times 9.55 \text{ cMpc}^3$ subvolume at redshifts $z = \{10, 8.5, 7, 5.5\}$. ρ_{esc} and f_{esc} are calculated by dividing the subvolume into a grid and summing up the contributing haloes in each pixel region. We also apply a Gaussian filter over the grid to smooth out the halo distribution. The similarity in morphology among the four quantities can be readily seen. At $z \gtrsim 8$, the reionization of hydrogen corresponds with the highest overdensity regions, which are the earliest sites of galaxy formation and are thus capable of providing the most intrinsic and escaped ionizing photons at these epochs. The H II regions then expand and fill the entire simulation volume at the end of the simulation at $z = 5.5$ (bottom panel). Regions with high ρ_{esc} and f_{esc} also generally align with the high overdensity regions at higher redshifts, and they expand into the nearby regions as time increases. At the end, even regions with overdensity less than zero (underdensities) can have relatively high ρ_{esc} and f_{esc} . This emphasizes the important role environment plays during reionization, and neighborhoods without significant star formation can be externally reionized via the ionizing sources nearby.

3.2 Redshift evolution

We next present quantitative results for the global statistics, which are critical for understanding the evolution of the IGM ionized volume fraction and therefore the reionization history (Madau et al. 1999). In the left-hand panels of Fig. 4 we show the global escape fraction (thick black curves) defined as $\langle f_{\text{esc}} \rangle \equiv \sum \dot{N}_{\text{esc}} / \sum \dot{N}_{\text{int}}$, where the summations are over all photon sources. In particular, this includes all resolved subhaloes with MCRT calculations but also all stars outside the virial radius of any halo assuming maximal escape fractions. In fact, the plotted range corresponds to $f_{\text{esc}}^{\text{cloud}} = 0.37$ for consistency with the birth cloud value used in THESAN to match available constraints on the observed reionization history. To reduce distracting numerical fluctuations, we have applied Savitzky-Golay smoothing filter to all values. To show the dependence on the halo mass, in the upper panels we also calculate the global contributions from sources split into five different mass ranges defined by the following thresholds: $M_{\text{halo}} \in (0, 10^8, 10^9, 10^{10}, 10^{11}, \infty) M_{\odot}$. Likewise, in the lower panels we separate the dependence on the stellar mass in the following ranges: $M_{\text{stars}} \in (0, 10^6, 10^7, 10^8, 10^9, \infty) M_{\odot}$. In comparison to previous similar studies (e.g. Paardekooper et al. 2015; Rosdahl et al. 2018), our larger simulation volume provides us with significantly more data for massive galaxies. For reference, we also include the result from Rosdahl et al. (2018) as shown with the blue-dashed curve (the luminosity-weighted mean over the last 100 Myr $f_{\text{esc},100}$ from the binary SED model in their Fig. 13), and Rosdahl et al. (2022) with the blue-dotted curve ($f_{\text{esc},100}$ in their Fig. 16). Last, we compare our post-processing escape fractions to the effective escape fractions $f_{\text{esc}}^{\text{eff}}$ (black-dashed lines) calculated in Sec. 3.3 of Kannan et al. (2022a) using the equations outlined in Madau (2017). Our results show that the effective escape fractions $f_{\text{esc}}^{\text{eff}}$ overestimate the escape fractions before $z \approx 6$. This is due to the oversimplification of the equation used to obtain $f_{\text{esc}}^{\text{eff}}$, which only treats the LyC production and absorption within a clumpy medium and overlooks smaller structures. Thus, a more careful approach is necessary to determine escape fractions from simulations. In addition, our ray-tracing RT ap-

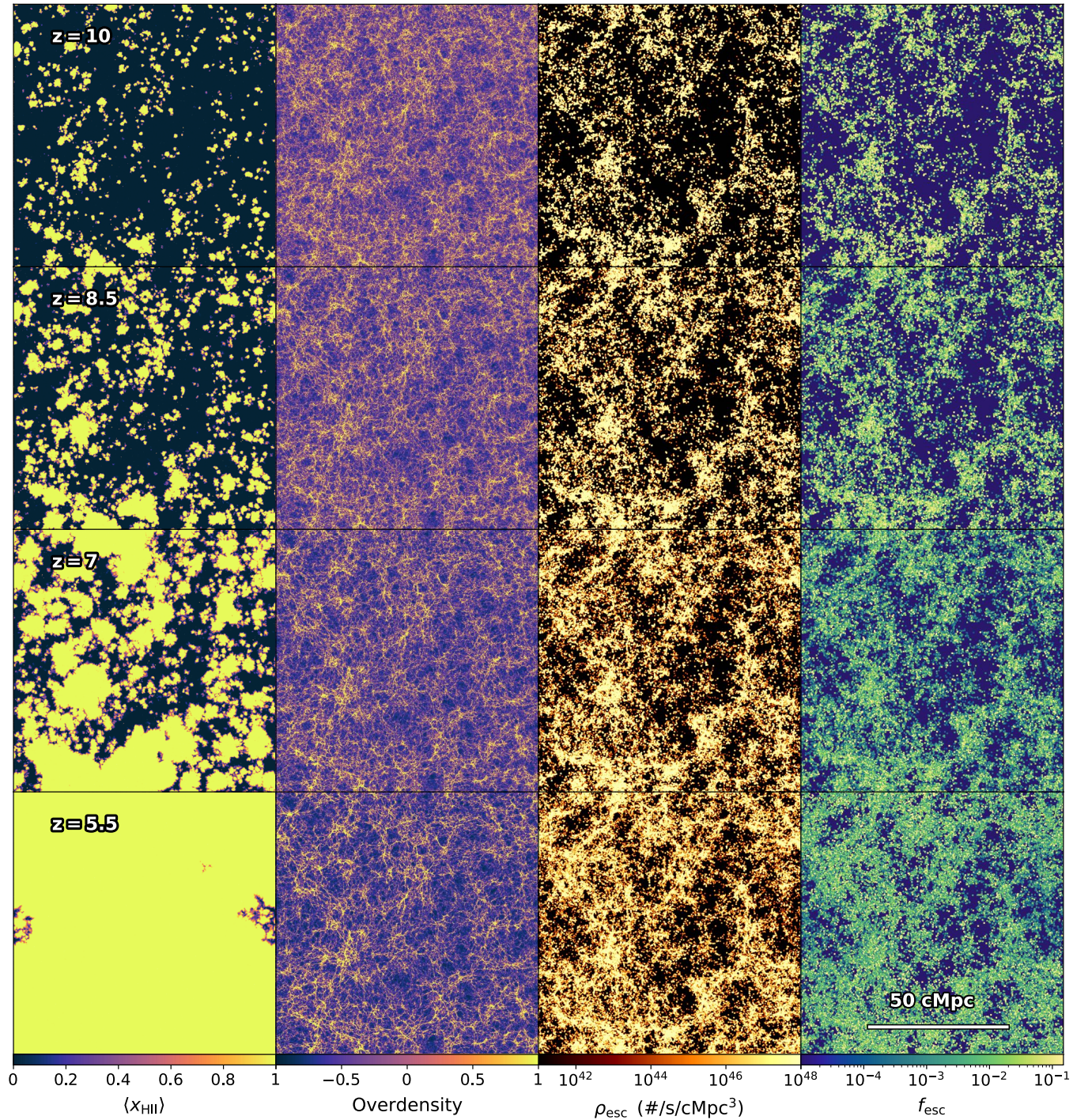


Figure 3. Left to right: Visualization of the spatial distribution of the H II fraction $\langle x_{\text{H II}} \rangle$, gas overdensity $\delta \equiv \rho/\bar{\rho} - 1$, escaped ionizing photon rate densities ρ_{esc} , and LyC escape fractions f_{esc} , covering the same $95.5 \times 95.5 \times 9.55 \text{ cMpc}^3$ subvolume at redshifts $z = \{10, 8.5, 7, 5.5\}$, from top to bottom. The similarity in morphology among these quantities is striking, and indicates that environment also influences the production and escape of ionizing photons in the EoR.

proach results in a redshift-dependent, galaxy-mass dependent f_{esc} , which is not seen in the zeroth-order estimation of $f_{\text{esc}}^{\text{eff}}$.

In the right-hand panels, we show the total intrinsic (dashed curves) and escaped (solid curves) cosmic photon rate densities, i.e. normalized by the volume of the THESAN simulation. The intrinsic photon rate density rises from $\sim 10^{51} \text{ photons s}^{-1} \text{ cMpc}^{-3}$ at $z = 13$ to $\sim 10^{52} \text{ photons s}^{-1} \text{ cMpc}^{-3}$ at $z = 5.5$, while the escaped photon

rate density has a slightly flatter evolution over the same period of time. In fact, there is a clear decrease in the global effective escape fraction from $\approx 15\%$ to $\approx 5\%$, with varying behaviour across the individual halo and stellar mass ranges. Notably, in our model the escape fractions of the lowest mass unresolved haloes ($< 10^8 M_{\odot}$) are close to the maximal values but have a small contribution to reionization due to their low intrinsic photon rates. As reionization

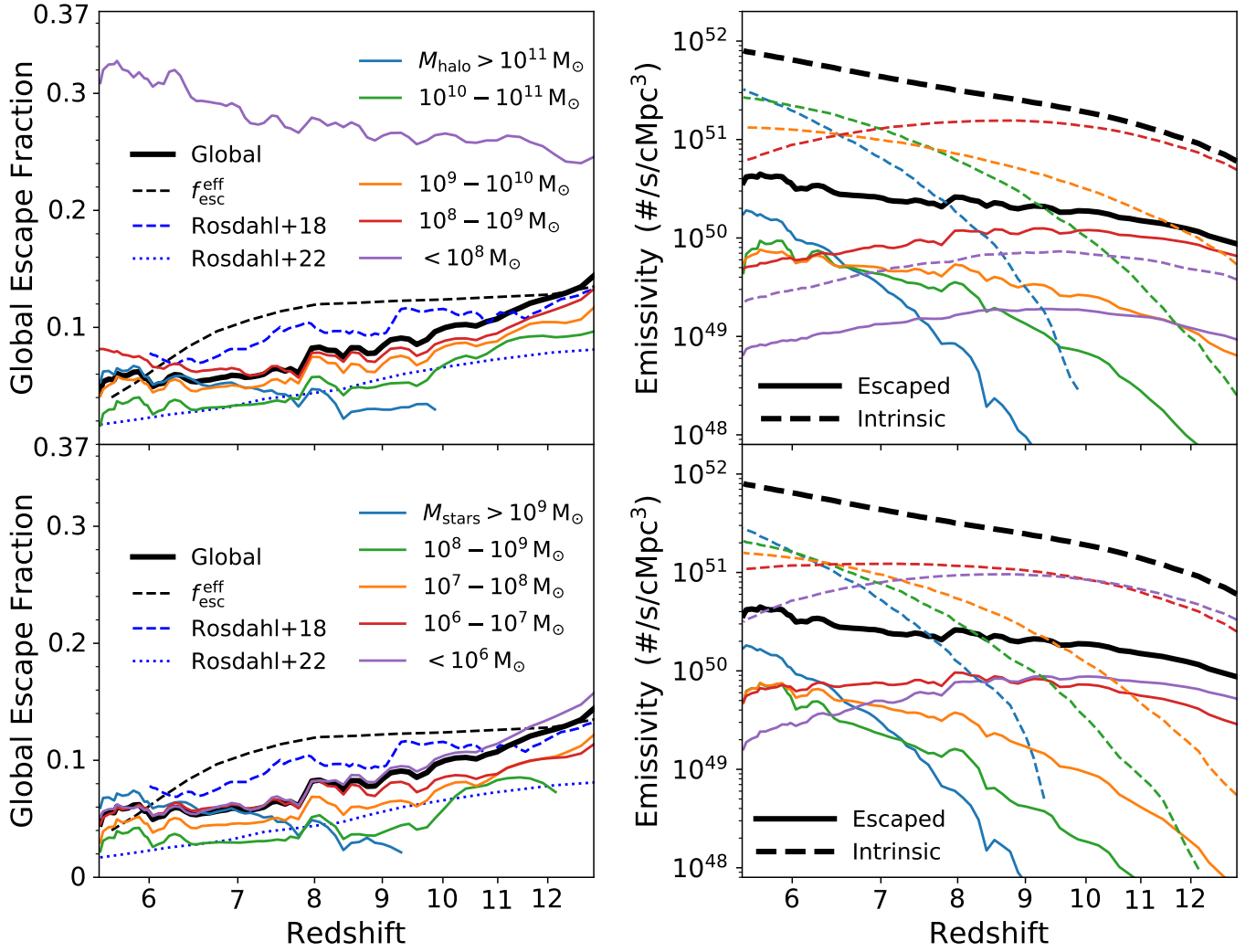


Figure 4. *Left:* Global photon rate weighted escape fractions $\langle f_{\text{esc}} \rangle \equiv \sum \dot{N}_{\text{esc}} / \sum \dot{N}_{\text{int}}$ as a function of redshift (thick black curves). This contains all photon sources in the simulated volume, including all resolved haloes and a small fraction of unassigned star particles. The overall escaped photon rates are obtained using the f_{esc} calculated by COLT for all selected haloes and assuming maximal escape $f_{\text{esc}}^{\text{cloud}} = 0.37$ for the remaining sources. We also show global escape fractions for haloes in five different halo (upper panels) and stellar (lower panels) mass ranges (coloured curves), which reflect the nontrivial mass dependence with redshift. The lowest mass haloes have lower completeness for stars within the virial radius and thus have maximal escape properties. Overall, the weighted- f_{esc} decreases with increasing halo mass except for the most massive haloes (either $M_{\text{halo}} > 10^{11} M_{\odot}$ or $M_{\text{stars}} > 10^9 M_{\odot}$). The black dashed curve shows the effective escape fraction $f_{\text{esc}}^{\text{eff}}$ calculated as described in Kannan et al. (2022a). For reference, we also include the result from Rosdahl et al. (2018, blue dashed curves) and Rosdahl et al. (2022, blue dotted curves). *Right:* Global intrinsic (solid) and escaped (dashed) photon rate densities as functions of redshift (thick black curves), also plotted with the same mass bin ranges (coloured curves). The emissivity evolution is mainly due to the formation of massive galaxies at lower redshifts, with a slightly flattened slope when going from intrinsic to escaped values. The halo and stellar mass dependence clearly show that massive haloes become the dominant photon sources by $z \approx 7$.

progresses the emissivity transitions from being dominated by atomic cooling haloes ($M_{\text{halo}} \approx 10^8 - 10^9 M_{\odot}$) to more massive haloes. The suppressed star formation in this mass range is likely due to photo-heating feedback by the reionization process itself, which reduces the accretion of gas onto smaller haloes (e.g. Gnedin 2014; Dawoodbhoj et al. 2018; Wu et al. 2019).

3.3 Importance of halo mass

The increasing photon rate densities can also be understood by considering the redshift evolution of the halo and stellar mass distributions towards more massive objects. Conceptually, the intrinsic

luminosity of a galaxy is roughly proportional to its SFR, which is correlated with the halo and stellar mass. In Fig. 5 we show the global intrinsic (top) and escaped (middle) photon rate density distributions along with the escape fraction (bottom) as functions of halo mass ($10^8 - 10^{12} M_{\odot}$) for several different redshift bins across the range $z = 5.5 - 13$ as indicated in the colour bar. Data in the same redshift bins are combined to improve the statistics. The solid curves are the underlying mass distributions while the dashed curves represent cumulative values from the massive end. For reference, we also include a grayscale background to show the cumulative completeness of the halo selection. The solid black line indicates the point of 90% completeness, which serves as a reasonable signal for the resolution limit of our escape fraction analysis. At nearly all redshifts

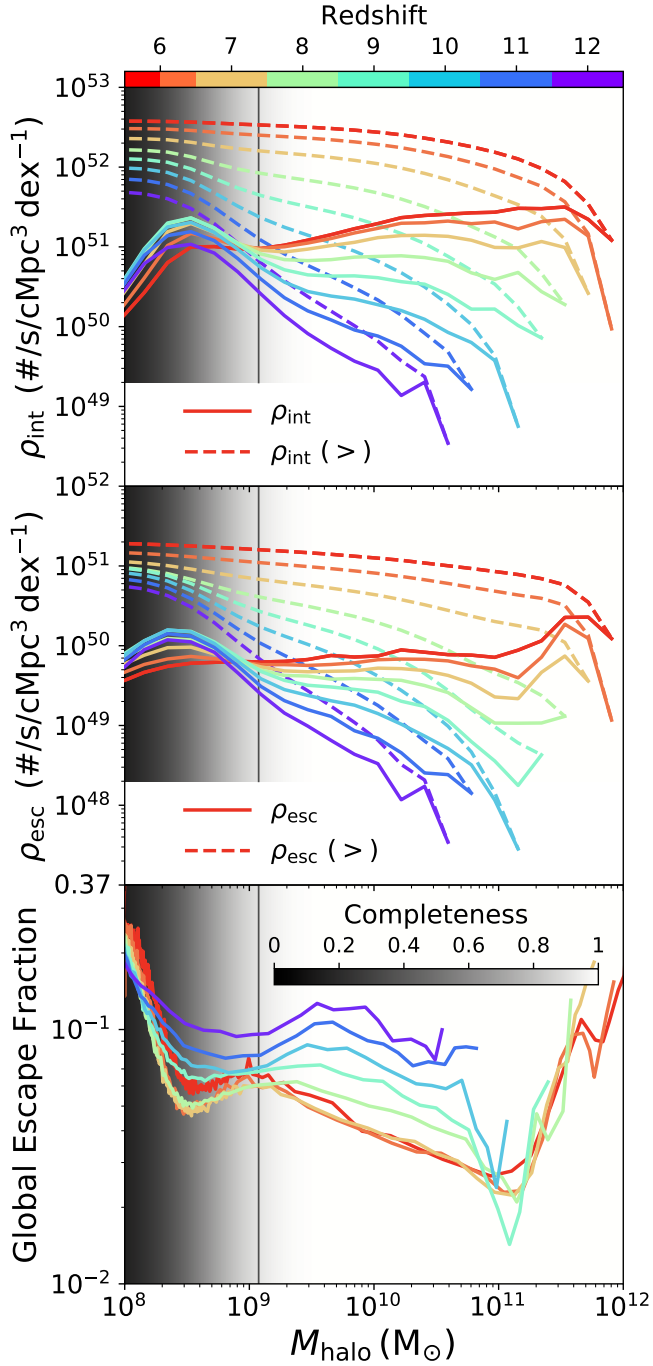


Figure 5. The distribution of global intrinsic (top) and escaped (middle) photon rate densities along with the escape fraction (bottom) as functions of halo mass for several redshift ranges covering $z = 5.5\text{--}13$ as indicated in the colour bar. The solid curves are mass distributions while the dashed curves are cumulative values from the massive end. The grayscale background shows the cumulative completeness of the halo selection, and the solid black line indicates the point of 90% completeness. Haloes at the low-mass end ($M_{\text{halo}} \lesssim 10^9 M_{\odot}$) are less resolved while more massive ones ($M_{\text{halo}} \gtrsim 10^{11} M_{\odot}$) only form at lower redshifts given the moderate volume of the THESAN simulation. Lower mass haloes are the main intrinsic and escaped photon sources at $z \gtrsim 8$ while higher mass haloes become dominant at lower redshifts. The most massive haloes are especially strong sources of escaped photons at $z \lesssim 7$ as indicated by the noticeable peak in the middle panel and upturn in the escape fraction shown in the bottom panel.

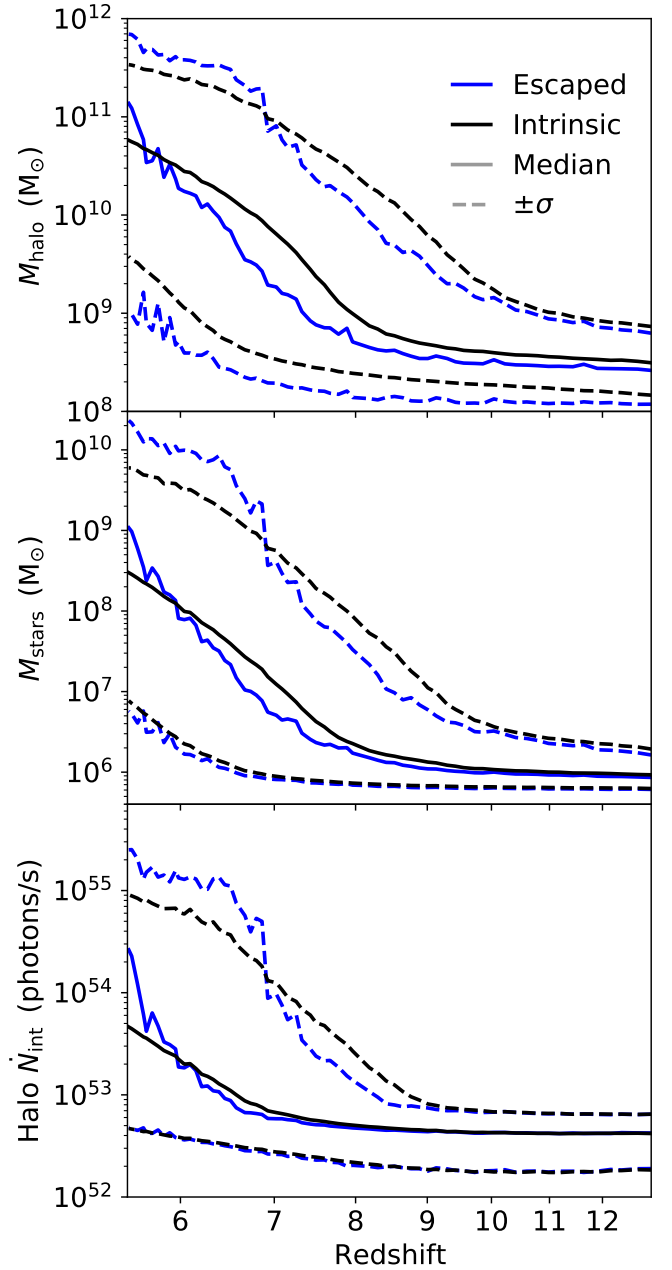


Figure 6. Halo masses, stellar masses, and halo intrinsic photon rates (top, middle, and bottom panels, respectively) that contribute to {16, 50, 84}% of the global escaped and intrinsic photons (blue and black curves, respectively) when sorted by these quantities across redshifts $z = 5.5\text{--}13$. The increasing median M_{halo} , M_{stars} , and $\dot{N}_{\text{int}}$ as redshift decreases clearly illustrates the growing importance of massive and bright haloes towards lower redshifts.

the low-mass peak coincides with the range of marginally resolved haloes in the simulation $M_{\text{halo}} \approx 10^8\text{--}10^9 M_{\odot}$, i.e. haloes below this have low stellar occupation fractions so do not emit significant ionizing radiation. The low-mass bump gradually becomes less dominant with time as more massive haloes start to dominate the production budget. Furthermore, the increasing intrinsic photon rate densities from the formation of more massive galaxies changes the negative slope at $z \approx 7.5\text{--}12.5$ to the positive slope at $z < 7.5$, with the exception at $M_{\text{halo}} \gtrsim 3 \times 10^{11} M_{\odot}$ where the slope drives downward due to the rapidly decreasing number of haloes in this range given

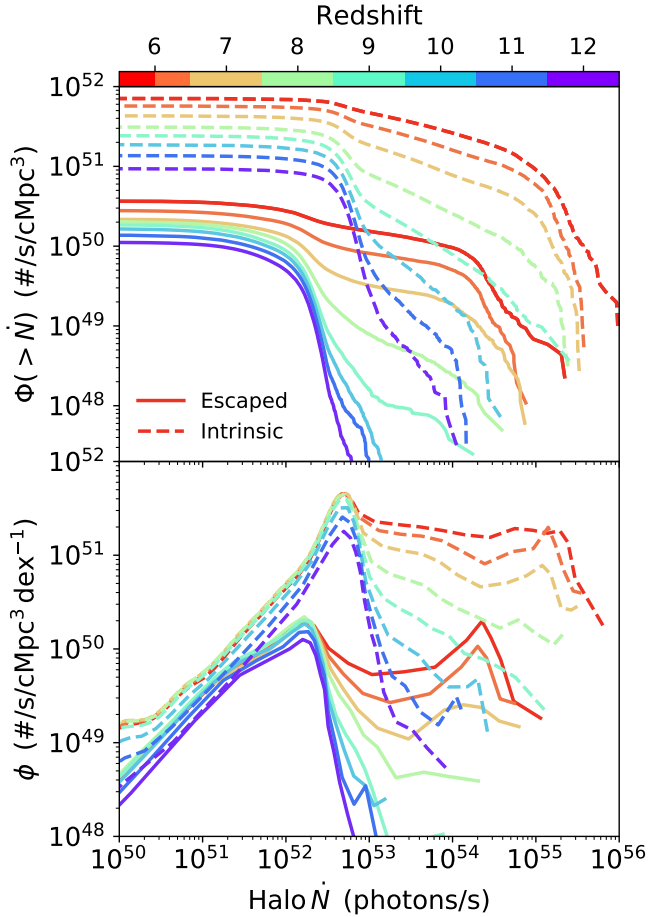


Figure 7. *Top:* Cumulative photon rate density functions for the intrinsic (dashed) and escaped (solid) photon rates $\dot{N}$ as a function of halo luminosity for all haloes in redshift bins covering redshifts from $z = 5.5$ – 13 as ranges indicated in the colour bar. There is a clear evolution revealing an increase of bright haloes starting to dominate the total photon rate density as redshift decreases. The rightmost points show the largest halo photon rate in the redshift bin, which also increases with the formation of more massive haloes. The curves flatten off when the cumulative contribution of fainter haloes is no longer significant. *Bottom:* Distributions of photon rate densities for both intrinsic and escaped halo photon rates. The combined contribution of massive haloes gains importance as redshift decreases in both escaped and intrinsic emission. The peaks at the low-mass end again indicate resolution limits and the point below which fainter haloes do not play important roles.

the limited simulation volume. This shows that the combined photon rate densities from massive haloes eventually exceed the contribution from lower mass haloes. We emphasize that the intrinsic LyC production is a result of the galaxy formation model, and smaller simulation boxes may significantly underrepresent the role of bright galaxies. A similar trend is also identified in the distribution of escaped photon rate densities (middle panel) with a notable difference at lower redshifts, where the small bumps at $M_{\text{halo}} > 10^{11} M_{\odot}$ are caused by the sharply increasing escape fractions at the massive end in our model. In bottom panel, we observe decreasing global escape fractions as the halo mass increases up until $M_{\text{halo}} \sim 10^{11} M_{\odot}$. Lewis et al. (2020) post-processed Coda II simulations and also reported a decreasing escape fraction with halo mass. We note that the uptick at $M_{\text{halo}} > 10^{11} M_{\odot}$, which is a unique feature from our post-processing results, can likely be attributed to the different feedback

mechanisms in the simulations. For example, AGN sources are ignored in the CoDa simulations even though they are known to impact star formation in high-mass galaxies. In addition, even if the direct contribution of AGN to escaped emission is negligible, they can still alter the thermal and ionization states within their host galaxies.

We directly demonstrate the importance of massive galaxies in Fig. 6, where we plot the corresponding M_{halo} (top panel), M_{stars} (middle panel), and halo $\dot{N}_{\text{int}}$ (bottom panel) that contribute to {16, 50, 84}% of the global escaped (blue) and intrinsic (black) photons when sorted by these quantities, all as functions of redshift. The median M_{halo} , M_{stars} , and $\dot{N}_{\text{int}}$ increase over time in terms of both intrinsic and escaped photons. Moreover, a particularly sharp increase in the escaped distribution can be seen below $z \lesssim 7$, emphasizing the role of the brightest and most massive galaxies in accelerating reionization. Their importance mainly stems from their exceedingly high intrinsic photon rates in combination with moderately higher escape fractions ($f_{\text{esc}} \sim 0.1$). Similarly, although there are larger numbers of faint galaxies with high f_{esc} , their combined impact on reionization is minor. These conclusions are consistent with the late-reionization model supported by recent observations and simulations (e.g. Naidu et al. 2020; Ma et al. 2020; Yung et al. 2020).

4 DEPENDENCE ON HALO PROPERTIES

4.1 Photon rate density functions

We now investigate the key physical quantities that determine the escape fractions of individual haloes. The dependence on halo mass has been studied in previous cosmological simulations where the results suggest an anticorrelation between halo mass and escape fraction (Paardekooper et al. 2015; Kimm et al. 2017). However, these specific studies were limited to haloes with $M_{\text{halo}} \lesssim 10^9 M_{\odot}$ due to the high computational cost associated with simulating larger volumes at comparable resolutions. Here we take advantage of the significantly improved statistics provided by the THESAN simulation to explore the halo mass dependence of f_{esc} extended to higher mass galaxies. Specifically, we examine trends for individual haloes covering a mass range of $M_{\text{halo}} \in 10^8$ – $10^{12} M_{\odot}$. Haloes with masses below this range are generally not resolved in the simulation. On the other hand, haloes with masses above this range are not found in the THESAN simulation due to the limited volume of the simulation.

To characterize the halo source population, in the top panel of Fig. 7 we show cumulative photon rate density against the halo photon rate for the intrinsic (dashed) and escaped (solid) photon rates $\dot{N}$ for all haloes in the same redshift bin ranges as before covering $z = 5.5$ – 13 . The photon rate density in this section includes only stars in selected haloes within their virial radii. This is different from Fig. 5, where all LyC sources are included. There is a clear evolution with redshift due to the increase of bright haloes as they start to dominate the total photon rate density. The rightmost ends of the curves show the largest photon rate from a single halo at a given redshift, which also increases with time as more massive haloes form. The curves become increasingly flat as fainter halo sources no longer contribute significantly compared to brighter ones. The bottom panel of Fig. 7 shows the distribution of photon rate densities, again with intrinsic (dashed) and escaped (solid) photon rates in the same redshift range $z = 5.5$ – 13 . The shape of the escaped photon rate density distribution resembles the intrinsic one with overall lower values and more evident peaks at the lower redshifts. This is due to the contribution from massive haloes with moderate f_{esc} that are formed at low redshifts.

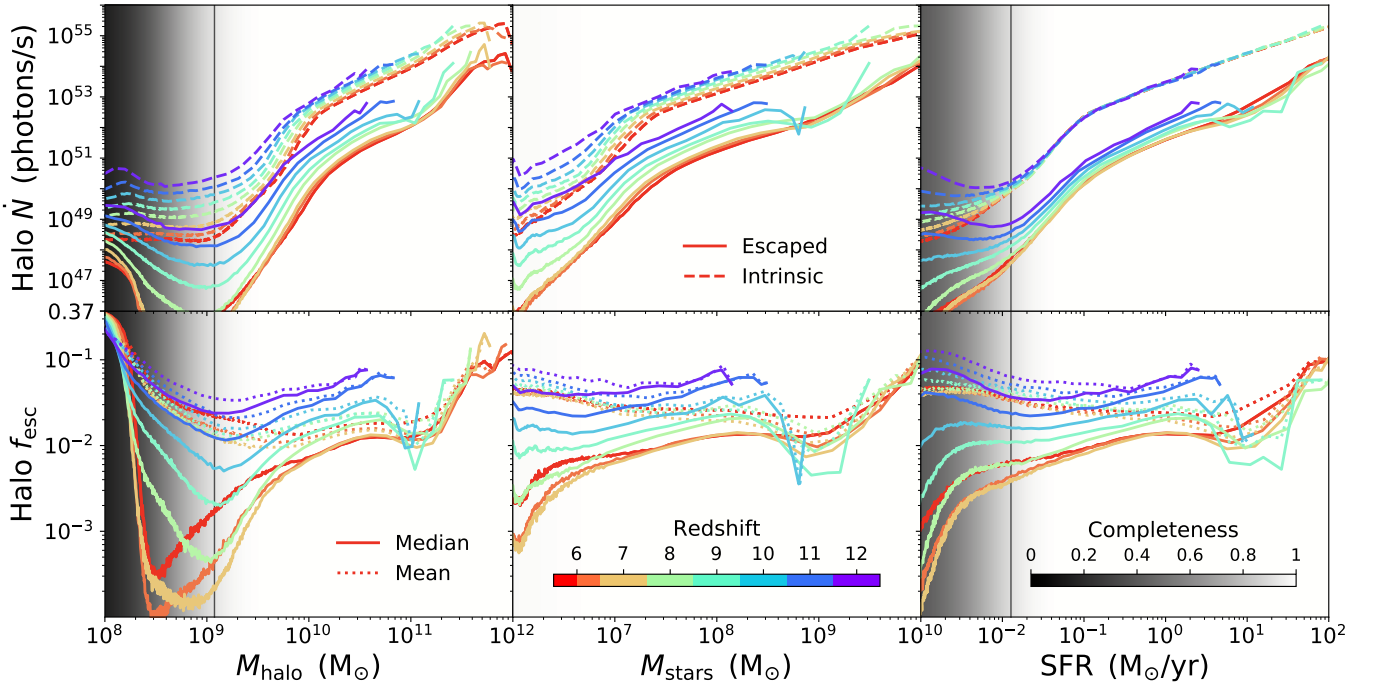


Figure 8. *Top:* The median values of the intrinsic (dashed) and escaped (solid) photon rates as a function of halo mass (left), stellar mass (middle), and SFR (right) for all selected haloes across several redshift ranges covering $z = 5.5\text{--}13$ as indicated in the colour bar. Haloes falling within the same redshift ranges are combined. The binning is adapted based on the halo number distribution with more bins towards the high-mass end to account for the fewer amount of haloes available in these ranges. The figure includes all the selected galaxies with criteria defined in 2.3, ranging from $10^8\text{--}10^{12} M_\odot$. The grayscale background shows the cumulative completeness of the halo selection, and the solid black line indicates the point of 90% completeness. The intrinsic photon rate increases as halo mass increases, and the flattening at the low-mass end is again due to the baryonic mass resolution of the THESAN simulation. The escaped photon rate drops to the minimal at $M_{\text{halo}} \approx 10^9 M_\odot$ and increases beyond this point. *Bottom:* The median (solid) and mean (dashed) of f_{esc} as a function of M_{halo} with the same binning method and halo selection criteria as above. The escape fractions reach the maximal value of 0.37 for lowest mass haloes, and decrease as a function of halo mass afterwards. f_{esc} reach the minimal values at $M_{\text{halo}} \approx 10^9 M_\odot$. For haloes more massive than this value, the f_{esc} gradually increases as halo mass increases. Some very massive haloes at low redshifts ($M_{\text{halo}} > 10^{11} M_\odot$) can reach $f_{\text{esc}} = 0.1$. A clear redshift dependence can be seen from the figure with higher f_{esc} at higher redshifts.

4.2 Median escape fractions

To better connect the halo properties with these sources, in Fig. 8 we show the dependence of the median halo intrinsic (dashed) and escaped (solid) photon rate ($\dot{N}$; top) and the median (solid) and mean (dashed) escape fraction (f_{esc} ; bottom) on halo mass ($M_{\text{halo}} \in 10^8\text{--}10^{12} M_\odot$; left), stellar mass ($M_{\text{stars}} \in 10^6\text{--}10^{10} M_\odot$; middle), and star formation rate (SFR $\in 10^{-3}\text{--}10^2 M_\odot \text{ yr}^{-1}$; right) for all selected haloes in different redshift ranges from $z = 5.5\text{--}13$. The grayscale backgrounds show the cumulative completeness of the halo selection, and the solid black lines indicate the mass of 90% completeness. The absent grayscale background in the middle panels is due to the high completeness for $M_{\text{stars}} > 10^6 M_\odot$ and reflects our selection criterion of at least one star in haloes. We note that the mean escape fractions here are different from the global averaged escape fractions presented in Fig. 5, where the intrinsic luminosity of individual haloes are taken into account instead of a naive mean of escape fractions from haloes in the same mass range. These two results converge when haloes in the same mass ranges have similar intrinsic photon rates. Since this is not generally true, the median/mean escape fractions here behave differently from the ones in Fig. 5. We postpone a closer analysis to Sec. 5.1, where we discuss the variation across haloes. However, it is important to note that mean statistics are biased to represent the presence of high values, while medians are sensitive to resolution and completeness. In addition, the global averages

include both bright outliers and stars outside the virial radius, and thus are slightly different than the f_{esc} we defined in Eqn. (1). The haloes at the low-mass end have escape fractions close to the birth cloud value of $f_{\text{esc}}^{\text{cloud}} = 0.37$, indicating that almost all photons reach the IGM once they escape the birth cloud. This maximal escape behaviour of low-mass haloes also justifies our assumption of adopting this value for non-selected haloes in Section 3. As the halo mass increases, the median f_{esc} drops significantly with a minimum around $M_{\text{halo}} \sim 10^9 M_\odot$. The decrease of f_{esc} at the low-mass end agrees with the results from the above-mentioned studies, but in our case this is clearly also driven by resolution limitations and a selection bias towards haloes with higher stellar-to-halo mass ratios, which might lead to an overestimation of f_{esc} at the low-mass end. The median f_{esc} remains low with $f_{\text{esc}} \sim 0.01$ at $z \sim 6$ in intermediate-mass haloes with an observed gradual increase in f_{esc} as halo-mass increases up to about $M_{\text{halo}} \sim 10^{11} M_\odot$ at all redshifts. Finally, f_{esc} has a steep increase for haloes above this mass, as can be seen at lower redshifts. This may result from ionizing channels created by other mechanisms, e.g. powerful feedback associated with massive galaxies. We will discuss the impact of dust at the high-mass end in Section 6.1. Comparing the curves in different redshifts, we see f_{esc} increases with redshift in all mass ranges. Similar trends are also reported in several previous studies (Eide et al. 2020; Faucher-Giguère 2020; Ma et al. 2020). Finally, comparing median and mean values of f_{esc} , we see an overall similar pattern but a much less extreme change

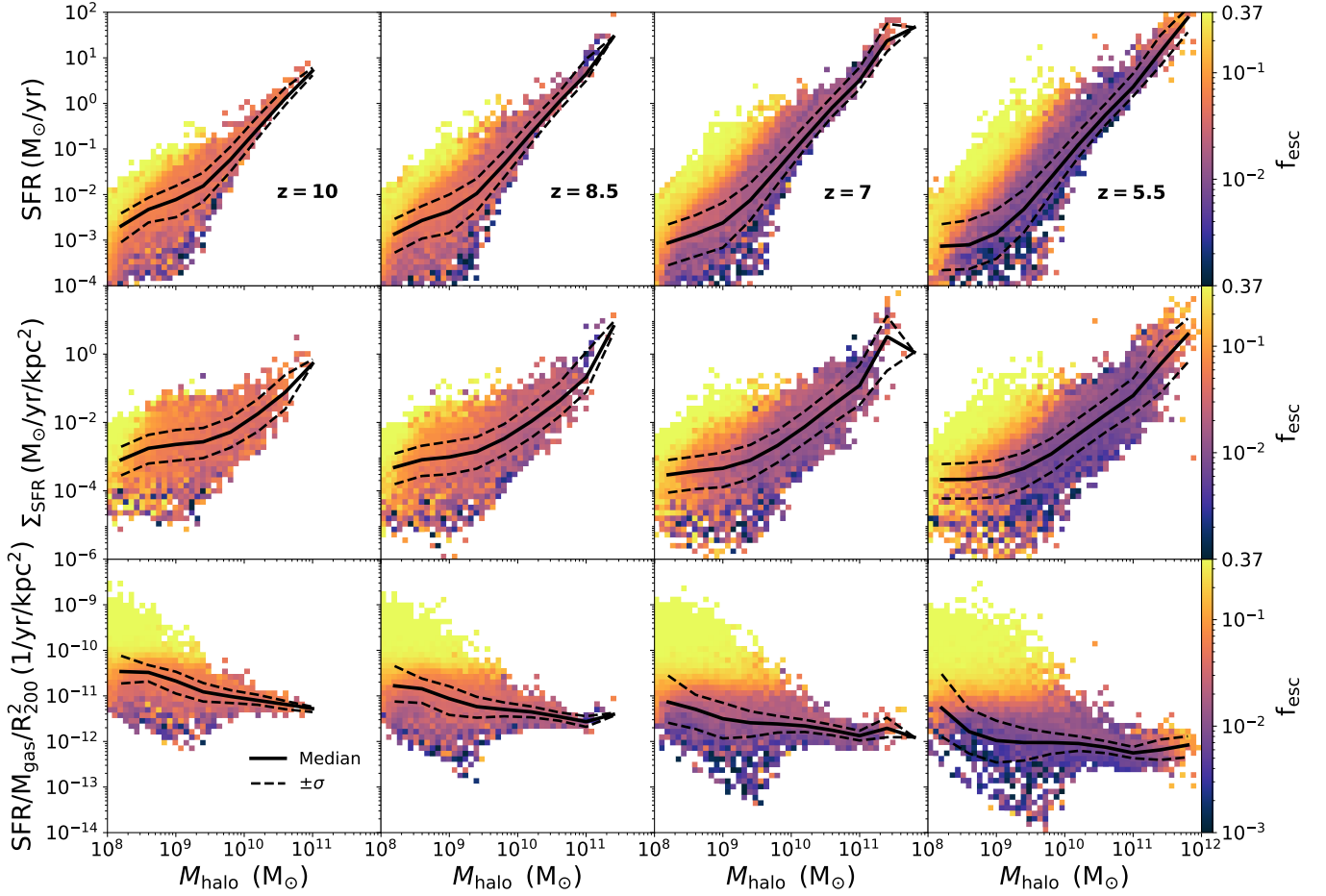


Figure 9. Distribution of haloes at $z = \{10, 8.5, 7, 5.5\}$, from left to right, with various masses and SFRs (top), surface SFR (Σ_{SFR} as defined in Pahl et al. (2022) with r_e = half stellar mass radius, middle), and modified surface SFR per gas mass ($\bar{\Sigma}_{\text{SFR}} = \text{SFR}/M_{\text{gas}}/R_{200}^2$, bottom). The 2D histograms are coloured by mean f_{esc} with 50 bins in both axes, and each bin is normalized by the number of haloes as in Fig. 1. The median values are shown in the black solid curves and the 16th and 84th percentiles are shown in the black dashed curves. At each redshift, haloes with relatively low masses and high SFRs have the highest escape fractions. For massive haloes ($M_{\text{halo}} > 10^{10} M_{\odot}$), $f_{\text{esc}} < 10\%$ even with high SFRs. In addition, in the low-mass haloes, the distribution of SFRs is wider and thus allows haloes with higher than usual SFR to have high f_{esc} . On the other hand, the distribution of SFR in high-mass regions is relatively narrow and a typical massive halo does not have sufficient SFR to produce high f_{esc} in our simulation. In the middle panel, Σ_{SFR} shows a similar trend as SFR. In the bottom panel, the histograms are seen to be divided into two regions by $\bar{\Sigma}_{\text{SFR}}$, suggesting that this quantity is a potential discriminator for f_{esc} .

in mean statistics below the mass of 90% completeness. This implies the medians are more sensitive to the completeness compared to the means, although the two measures have reasonable agreement (or at least interpretations) in regions with full completeness.

Furthermore, since f_{esc} is defined to be the ratio of the escaped and intrinsic photon rates, it is interesting to also examine the change in both quantities as a function of halo mass. At all redshifts, the medians of intrinsic and escaped photon rates increase with the halo mass, stellar mass, or star formation rate, although the curves become relatively flat at the low-mass end (Fig. 8). The MCRT procedure requires at least one star within the virial radius as a photon source in the calculation, and therefore the intrinsic photon rates do not further decrease as halo mass decreases at the low-mass end due to the limited mass resolution in the simulation. The escaped photon rates reach the lowest point at $\approx 10^9 M_{\odot}$, and reach the highest point at the high-mass end. In all redshifts, we notice higher photon rates at higher redshifts for a given halo mass range. This can be attributed to changes in the stellar population as galaxies assemble and evolve. In fact, the specific SFR in the halo decreases with time

(see Figure 1 of Kannan et al. 2022b), which explains why this evolution is not apparent in the rightmost panel of Fig. 8 showing the intrinsic photon rate as a function of the SFR. In fact, intrinsic photon rates are expected to be proportional to SFRs if galaxies have continuous star formation histories (Kennicutt 1998; Dijkstra 2019). The constant of proportionality depends on the IMF, and in our case we find $\dot{N}_{\text{int}} = 2.34 \times 10^{53} \text{ s}^{-1} [\text{SFR}/(M_{\odot} \text{ yr}^{-1})]$ by fitting the values above the 90% completeness with a linear function with a goodness of fit of $R^2 > 0.99$. The dip below $\text{SFR} \sim 0.1 M_{\odot} \text{ yr}^{-1}$ is a result of poorly resolved stochastic star formation histories in low-mass galaxies, which is exaggerated by plotting the median rather than the mean and in showing the gas-based instantaneous SFR against the star-based LyC production rate that is sensitive to the sampled ages.

The non-monotonic dependence of f_{esc} on the halo mass strongly suggests that there are additional factors that can affect the escape of ionizing photons. In particular, the stellar mass and star formation rate are two of the more intuitive quantities since star-forming galaxies are believed to play a crucial role in reionization and many reported measurements of f_{esc} come from starburst galaxies. Overall, the

median curves in Fig. 8 reveal a flatter or more moderate dependence on M_{stars} and SFR, except at the highest ranges, while the redshift evolution is strong before the midpoint of reionization.

4.3 Effective SFR surface density

Beyond the cumulative and instantaneous star formation history, we expect the relative amount and density of absorbing hydrogen atoms in the surrounding gas reservoir to affect halo escape fractions. We therefore explore several of these features in Fig. 9. In the top panel we show the distribution of mean f_{esc} in a 2D plane of SFR and M_{halo} for galaxies at redshifts of $z = \{10, 8.5, 7, 5.5\}$. The median and 1σ variation (16th and 84th percentiles) of f_{esc} at each mass bin is shown in the black solid and dashed curves, respectively. It is clear that the SFR increases as the halo mass increases in all mass ranges as expected. Galaxies with the highest f_{esc} have a combination of lower mass and higher SFR as highlighted in the yellow regions. For slightly more massive haloes ($M_{\text{halo}} > 10^{10} M_{\odot}$), $f_{\text{esc}} < 10\%$ even with high SFRs. In the low-mass region, the distribution of SFR at a given halo mass is wide and thus some haloes with SFRs 1–2 orders of magnitude higher than the median can have high f_{esc} , which is expected due to starburst duty cycles. On the other hand, the distribution of SFR in the high-mass region is relatively narrow and most massive haloes in our simulation do not have SFRs far from the median to reach high f_{esc} . In addition, we note that since maximal escape fractions are achieved at $M_{\text{halo}} \sim 10^8 M_{\odot}$ due to the resolution limit and the mean f_{esc} is higher than the median f_{esc} at $M_{\text{halo}} \lesssim 10^{10} M_{\odot}$, it is expected that most of the high escape fraction regions (yellow to orange) are located in this low-to-intermediate mass range.

In the middle panel of Fig. 9, we show the SFR surface density $\Sigma_{\text{SFR}} \equiv \text{SFR}_{r_e} / (2\pi r_e^2)$, with r_e denoting the half stellar mass radius as this is observationally accessible (although see Trujillo et al. 2020), and SFR_{r_e} to be the SFR within the half stellar mass radius. This is additionally motivated by the fact that many of the LyC leakers have higher than average Σ_{SFR} (Naidu et al. 2020). Recent observational efforts have also been focused on the dependence of f_{esc} on Σ_{SFR} . In particular, Pahl et al. (2022) points out that a large sample is required to identify the relation between f_{esc} and Σ_{SFR} . Our figure shows that at a given halo mass, galaxies with higher Σ_{SFR} have higher f_{esc} , but it does not solely determine f_{esc} either as there is a significant degree of mixing between higher and lower f_{esc} values.

In an attempt to find a discriminating quantity that correlates with f_{esc} in a monotonic manner, we modify the concept of Σ_{SFR} by dividing SFR by the gas mass M_{gas} and replacing a stellar-based radius with the virial radius R_{200} . We denote this new heuristic but theoretically motivated quantity as $\bar{\Sigma}_{\text{SFR}} = \text{SFR} / M_{\text{gas}} / R_{200}^2$. This particular combination is chosen since the gas mass of a galaxy roughly correlates with the amount of neutral hydrogen needed to be ionized before LyC photons can escape to the IGM. The bottom panel in Fig. 9 shows the distribution of f_{esc} with respect to $\bar{\Sigma}_{\text{SFR}}$ and the halo mass. The median values $\bar{\Sigma}_{\text{SFR}}$ drop slightly as halo mass increase from 10^8 to $10^9 M_{\odot}$, but do not change significantly as the halo mass increases further. Moreover, there is a strong colour gradient in the vertical direction and relatively little variation along the horizontal direction indicating that $\bar{\Sigma}_{\text{SFR}}$ correlates with f_{esc} in a simpler way than using M_{halo} or SFR alone. However, we note that the escape of ionizing photons is a complex process, and the construction of $\bar{\Sigma}_{\text{SFR}}$ does not include other potentially important mechanisms such as AGN feedback in massive haloes.

5 VARIATIONS IN THE ESCAPE FRACTIONS

In the previous section, we discussed relationships between various physical quantities and the *median/mean* escape fractions. While this provides an intuitive picture for the sources of reionization, it does not fully represent the rich statistics across the entire ensemble of galaxies. In particular, bright outliers with $\dot{N}_{\text{esc}}$ above the median can play more important roles in the reionization process. In this section, we further discuss the distribution of f_{esc} and photon rates to bring a more complete picture of halo escape fractions.

5.1 Halo-to-halo variations

In Fig. 10 we show the normalized distribution of halo escape fractions f_{esc} (top panels) along with intrinsic and escaped photon rates $\dot{N}$ (bottom panels) as functions of halo mass M_{halo} at the specific redshifts $z = \{10, 8.5, 7, 5.5\}$. The figure includes all selected haloes from the individual snapshots and each vertical histogram is normalized by the number of haloes in the given mass bin. The red solid lines in the top panels show the median halo f_{esc} in each mass bin, and the dashed lines indicate the 16th and 84th percentiles. Likewise, the black (blue) solid lines show the medians of the intrinsic (escaped) photon rates, and the associated dashed lines correspond to the 16th and 84th percentiles. At low halo mass, f_{esc} exhibits a wide distribution from 0.1–37% with the medians between 1–10% at $z = 10$ but extending down to 0.1% at later redshifts. This is consistent with the picture in Fig. 9 where f_{esc} can vary for galaxies with similar masses because of different SFRs in different galaxies. In addition, a significant amount of medium-sized galaxies with masses below $\lesssim 10^{10} M_{\odot}$ have $f_{\text{esc}} \lesssim 10^{-3}$ at $z = 7$. As M_{halo} further increases, the distribution becomes narrower and most haloes have percent level escape fractions. The largest haloes with $M_{\text{halo}} \gtrsim 10^{11} M_{\odot}$ generally have a higher f_{esc} , but are not commonly found in our simulated volume above redshift $z = 8.5$. In the bottom panel, both intrinsic and escaped photon rates again have wide distributions at the low-mass end and narrower distributions as the halo mass increases. This again agrees with the observation of wide SFR distributions at the low-mass end in Fig. 9. At the high-mass end, the intrinsic photon rates align well with M_{halo} , and the main contribution of the f_{esc} comes from the variation in the escaped photon rates.

In Fig. 11 we show the same distributions for f_{esc} and $\dot{N}$ as functions of SFR ranging from 0.1–100 $M_{\odot}/\text{yr}$ at redshifts $z = \{10, 8.5, 7, 5.5\}$. Haloes with SFRs higher than 100 $M_{\odot}/\text{yr}$ are rarely seen in the simulation, and the region below 0.1 $M_{\odot}/\text{yr}$ is clipped due to the lower resolution below this range. In comparison to Fig. 10, SFRs better characterize the intrinsic photon rates of a halo as the distributions are generally narrower for SFRs than M_{halo} . At all redshifts, the variation in the intrinsic photon rates as a function of SFR is relatively small compared to that of halo mass, especially for haloes with SFRs higher than 1 $M_{\odot}/\text{yr}$. This further implies that the wider distributions of the escaped photon rates are mostly from the complex escape fraction physics within the virial radius. That is, the strong sensitivity of f_{esc} to ISM and CGM scale processes is the main factor in setting the variation in escaped sources into the IGM throughout the EoR.

Lastly, we argue that the wide distributions of intrinsic photon rates at the low-mass end leads to the difference between the global averaged escape fractions in Fig. 5 and the mean f_{esc} in Fig. 8. Since the brighter haloes usually have higher escape fractions as seen in Fig. 9, the global averaged escape fractions are higher than a naive mean. This is especially evident at $M_{\text{halo}} \lesssim 10^{10} M_{\odot}$. At the high-

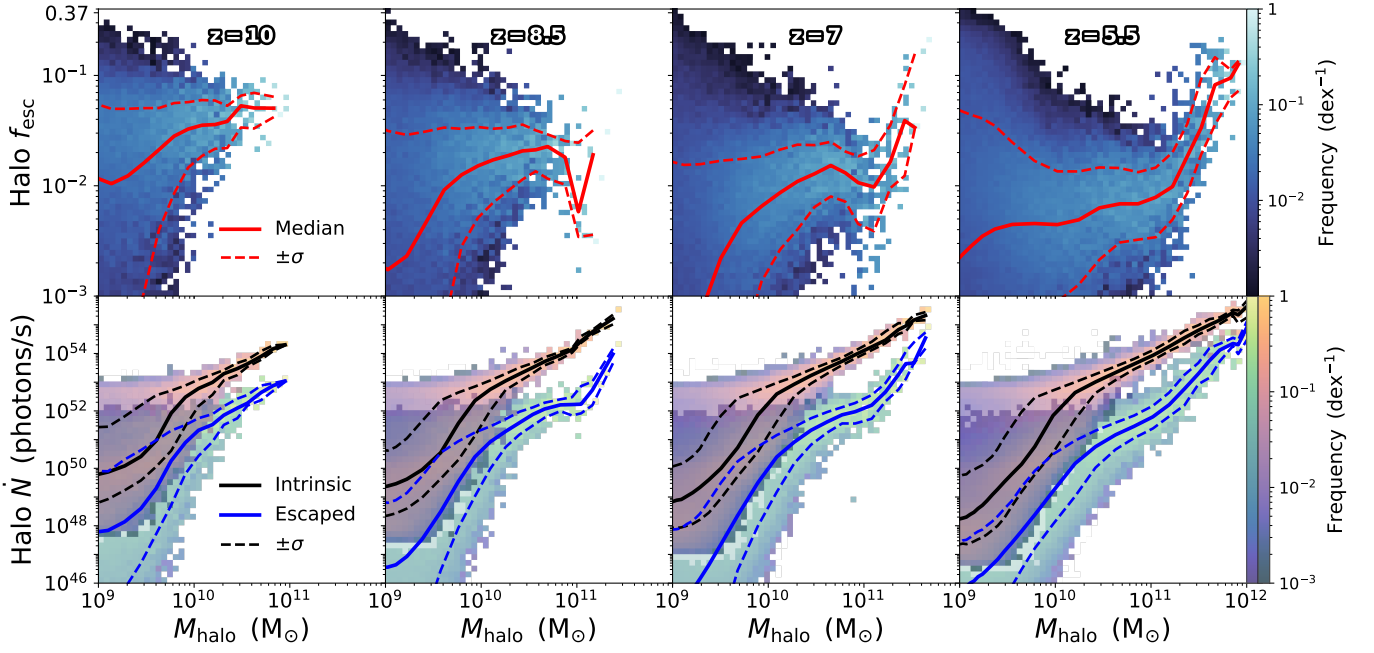


Figure 10. *Top:* Distribution of halo escape fractions f_{esc} as a function of halo mass M_{halo} at redshifts of $z = \{10, 8.5, 7, 5.5\}$, from left to right. Data below $10^9 M_{\odot}$ are not shown due to the lower completeness. The distribution is normalized by the number of haloes in each mass bin similarly to Fig. 1. The red solid curves show the median f_{esc} at a given mass bin, and the red dashed curves indicate the 16th and 84th percentiles. f_{esc} starts with a wide distribution at the low-mass end and becomes narrower for more massive haloes ($M_{\text{halo}} \gtrsim 10^{10} M_{\odot}$). At the massive end ($M_{\text{halo}} \gtrsim 10^{11} M_{\odot}$), almost all haloes have $f_{\text{esc}} \sim 10\%$ although they only occur at $z \lesssim 7$. *Bottom:* The distributions of intrinsic and escaped photon rates $\dot{N}$ as functions of halo mass. In both cases, the scatter is large at the low-mass end and becomes smaller above $M_{\text{halo}} \gtrsim 10^{10} M_{\odot}$, with the escaped photon rate distributions being wider than the intrinsic ones.

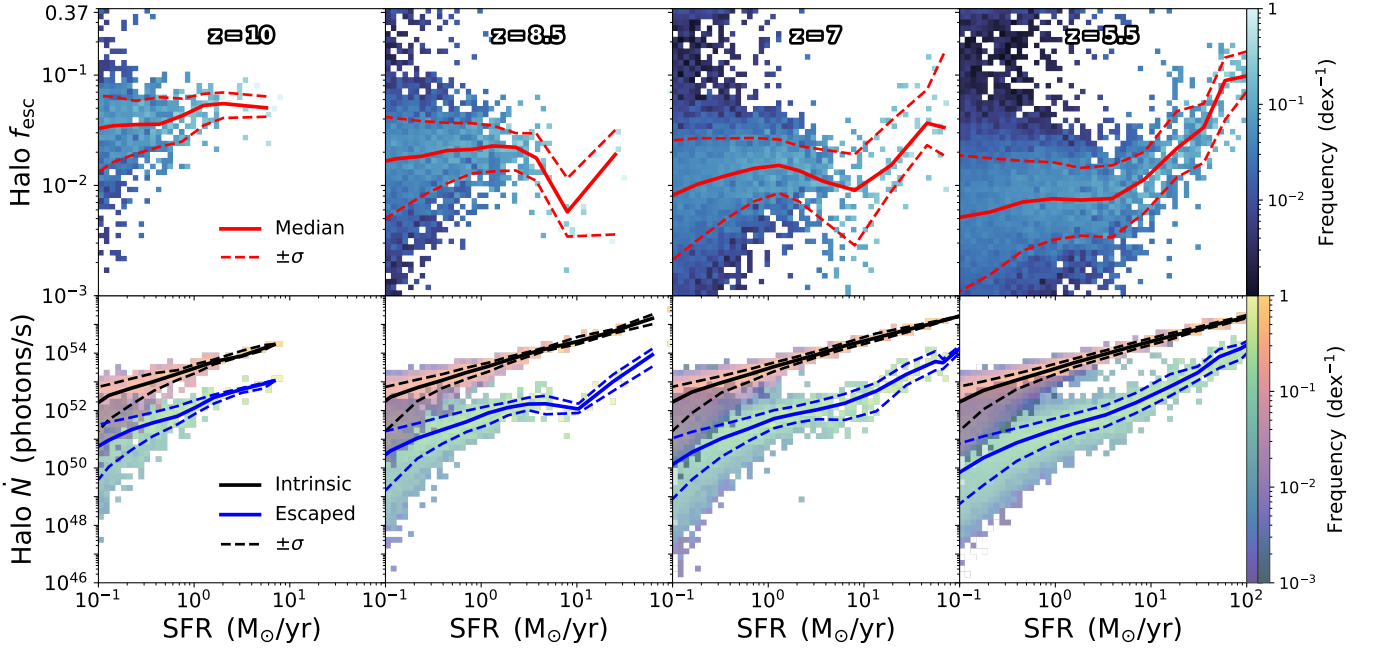


Figure 11. *Top:* Distribution of halo escape fractions f_{esc} as a function of the star-formation rate at redshifts of $z = \{10, 8.5, 7, 5.5\}$, from left to right. All plotting styles and data representations are identical to Fig. 10, specifically the solid and dashed coloured curves indicate the median and 1σ percentiles. As before, f_{esc} have a large scatter at the low-SFR end and a narrower distribution as the SFR increases. *Bottom:* The distributions of intrinsic and escaped photon rates $\dot{N}$ as functions of the SFR. Intrinsic photon rates correlate extremely well with SFR, such that the variation in escape fractions drives the IGM source rate distribution at a given SFR.

mass end, the distributions of intrinsic photon rates are relatively narrow and the two escape fractions are very similar.

5.2 Sightline statistics

In the previous sections, we have presented results from the angular-averaged f_{esc} from all stellar sources escaping the surface of a sphere at R_{200} . However, it should not be surprising that the escape fractions also have a large sightline-to-sightline variability due to the anisotropic distribution of the matter within galaxies. In this subsection, we aim to capture the anisotropic structure using the ray-tracing techniques implemented in COLT to calculate f_{esc} for observers oriented along each of $N = 3072$ healpix directions of equal solid angle.

Analyzing the data from sightline statistics allows for deeper understandings into the underlying escape fraction physics. For example, we can answer the following questions: What is the variability of f_{esc} from different sightlines? Do most of the photons escape through a handful of ionizing channels consistent with the concept of a covering fraction, or is a more isotropic picture favoured? What is the percentage of sightlines with f_{esc} higher than the mean escape fraction? How do these answers also depend on halo properties?

To understand the anisotropic features of sightline dependent escape fractions in a more quantitative way, we define the following five quantities for each selected halo: (1) normalized covering fraction $P(f_{\text{esc}} > \langle f_{\text{esc}} \rangle)$ as the fraction of sightlines that have escape fractions higher than the mean escape fraction $\langle f_{\text{esc}} \rangle \equiv \frac{1}{N} \sum_{i=1}^N f_{\text{esc},i}$, (2) standard deviation of the sightline escape fractions $\sigma_{\text{esc}}^2 \equiv \frac{1}{N} \sum_{i=1}^N (f_{\text{esc},i} - \langle f_{\text{esc}} \rangle)^2$, (3) normalized standard deviation of the sightline escape fractions as $\sigma_{\text{esc}} / \langle f_{\text{esc}} \rangle$, (4) half flux area, which is defined as the fractional area of the sphere contributing half of the escaped photons when sorted from brightest to faintest, and (5) the Gini coefficient G measuring the statistical dispersion in terms of equity of sightline distributions. Specifically, the Gini coefficient

$$G = \frac{2 \sum_{i=1}^N i f_{\text{esc},i}}{N \sum_{i=1}^N f_{\text{esc},i}} - \frac{N+1}{N}, \quad (4)$$

always lies in the range $[0, 1]$ with values near unity representing significant anisotropy (escape dominated by a few sightlines) and small values representing significant isotropy (all sightlines contributing equally). As we will also make our escape fraction catalogues publicly available we defer more specialized viewing dependence analyses to future studies.

In Fig. 12, we present the distributions of these five quantities as functions of halo mass at redshifts of $z = \{10, 8.5, 7, 5.5\}$. The median and 1σ deviation (16th and 84th percentiles) are also shown in black solid and dashed lines, respectively. In the top panel, P represents the fraction of sightlines with higher-than-average escape fractions for each selected halo. At all redshifts, these normalized covering fractions have a wide distribution for less massive haloes $M_{\text{halo}} \lesssim 10^{10} M_{\odot}$, ranging from 0.01 to almost 1 with 68% of haloes having $P \in [0.05, 0.4]$. The black solid curves suggest that in the median case, only 10–20% of sightlines have f_{esc} higher than their mean at all redshifts, revealing a substantial degree of anisotropy as $\langle f_{\text{esc}} \rangle$ is influenced by a small minority of sightlines. On the other hand, the distribution of P is much narrower among high-mass haloes. For example, most of the haloes with $M_{\text{halo}} > 10^{11} M_{\odot}$ have normalized covering fractions increasing to 30–50% by $z = 5.5$. This indicates that the sightline variability is more drastically different from halo to halo in low-mass haloes than in high-mass ones.

In the second and third panels, the standard deviation and normalized standard deviation provide a direct intuitive measure of the

variability among sightlines. In particular, $\sigma_{\text{esc}} / \langle f_{\text{esc}} \rangle$ evaluate the standard deviations in the unit of the angle-averaged escape fractions. The third panel shows that at all redshifts the majority of haloes have sightline standard deviations larger than $\langle f_{\text{esc}} \rangle$. Similar to the distribution of the covering fractions P , standard deviations also have a wider distribution among lower mass haloes while transitioning to a smaller scatter for the highest mass haloes. Given these trends of increasing standard deviation and decreasing normalized standard deviation with halo mass, we interpret this as a transition from a low-escape high-anisotropic mode at $M_{\text{halo}} \lesssim 10^{10} M_{\odot}$ to a more structured, large opening angle, and steady mode of escape at the highest masses.

The second to last panel in Fig. 12 shows the half flux area as defined above. The area gradually increases with halo mass, showing a notable increase above $M_{\text{halo}} \gtrsim 10^{10} M_{\odot}$ at $z \lesssim 7$. The low median values for low-mass haloes indicate that the halo f_{esc} are largely dominated by few sightlines with high escape fractions. In particular, half of the escaped photons are decided by the highest 1–10% of sightlines for low-mass haloes, but the number rises to $\approx 25\%$ with less variation for the most massive haloes. This shows that the escape is not limited to a few high- f_{esc} sightlines, and more ionizing channels are present in massive haloes. This picture provides a potential explanation of the increase in halo f_{esc} with M_{halo} as shown in Fig. 8. The more extreme intrinsic luminosities and structured environments of massive haloes foster the creation of either more ionized channels or a few dominant ones with larger opening angles, and thus leads to overall higher escape fractions.

Finally, the Gini coefficient G in the bottom panel compares the contribution from high- f_{esc} and low- f_{esc} sightlines. At all redshifts, most low-mass haloes have G close to unity, meaning the sightline distributions are highly skewed. G deviates from unity as halo mass increases, and reaches $G \approx 0.5$ for the most massive haloes at $z = 5.5$. This shows a more equally distributed contribution from all sightlines. Therefore, the measure from the Gini coefficient G is consistent with the result from the half flux area.

5.3 Angular structure correlations

The anisotropic behaviour has also been discussed in other studies (e.g. Paardekooper et al. 2015; Trebitsch et al. 2017), showing that most of the ionizing photons escape from galaxies through low column density channels. In Fig. 13 we provide visualizations of the angular distribution of the escape fraction f_{esc} , hydrogen absorption $f_{\text{H I}}$, helium absorption $f_{\text{He I}} + f_{\text{He II}}$, hydrogen column density N_{H} (normalized by the mean), and mean absorption distance $\langle \ell \rangle$ (normalized by the virial radius) in the most massive haloes at redshifts of $z = \{10, 8.5, 7, 5.5\}$, respectively. Below each image, we show the mean (left) and median $\pm 1\sigma$ (right) statistics based on 3072 healpix directions, with the exception of the hydrogen column density and absorption distance where for convenience we report the halo-dependent mean logarithmic column density (in $\log \text{cm}^{-2}$) and virial radius (in kpc) at the left. The absorption from hydrogen and helium is calculated using Eqn. (2), with f_{He} representing the combined absorption of He I and He II. The dust density has been set to zero and thus $f_{\text{abs}} = 0$ in this calculation. In general, by construction the sum of f_{esc} and the absorption from different species equals the birth cloud escape fraction as can be calculated from the following equation: $f_{\text{esc}}^{\text{cloud}} = f_{\text{esc}} + f_{\text{H I}} + f_{\text{He}} + f_{\text{abs}}$.

In greater detail, to more fairly compare the four different haloes, the hydrogen column density has been normalized by the angular-averaged value for each halo. Each sightline is the intrinsic photon rate weighted value from each star, i.e. $N_{\text{H}} \equiv \sum \dot{N}_{\text{int},i} N_{\text{H},i} / \sum \dot{N}_{\text{int},i}$.

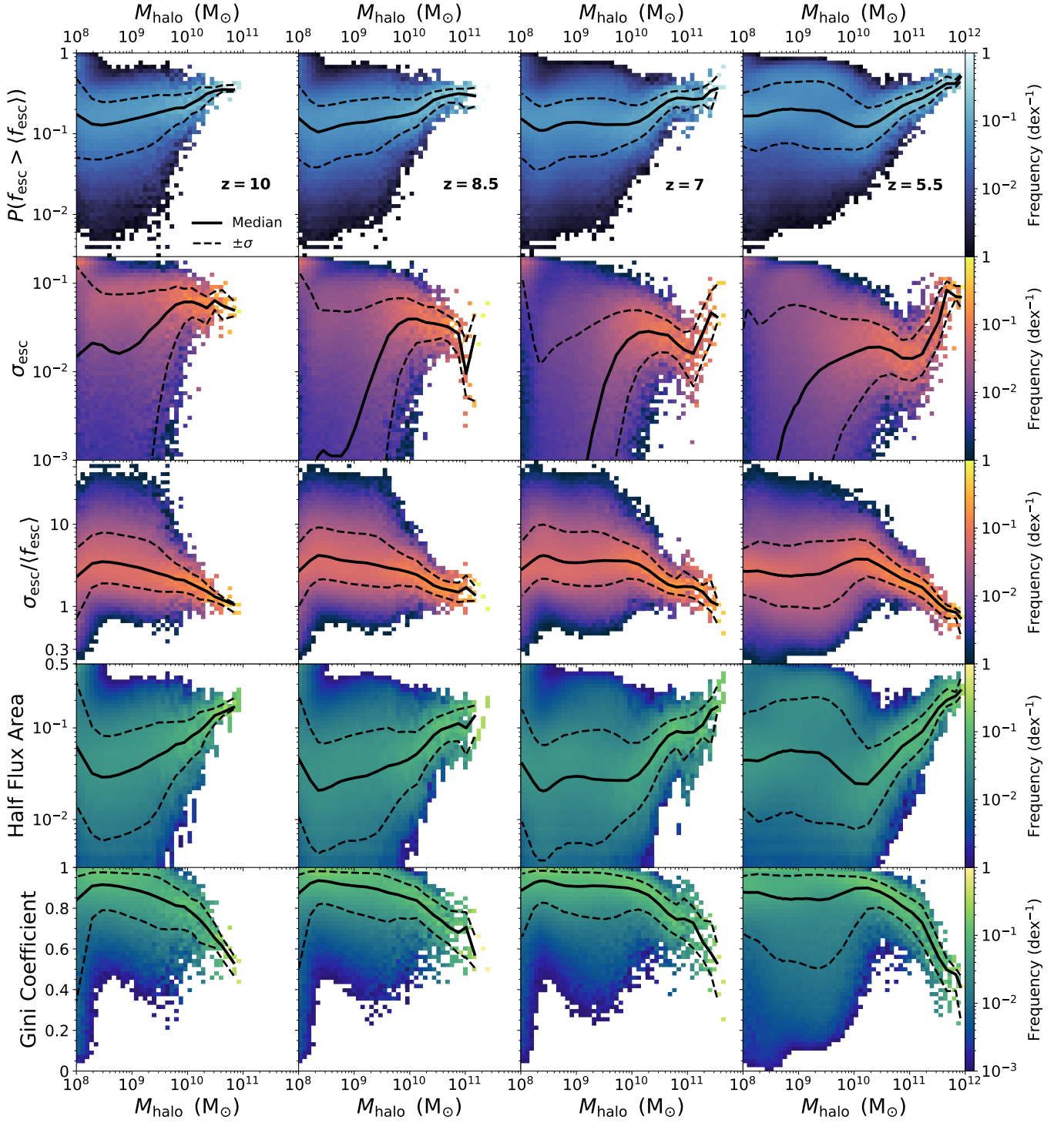


Figure 12. Distribution of covering fractions relative to the mean $P(f_{\text{esc}} > \langle f_{\text{esc}} \rangle)$ (top), standard deviations of escape fractions σ_{esc} (second), and normalized standard deviations $\sigma_{\text{esc}}/\langle f_{\text{esc}} \rangle$ (third), half flux area (fourth), and the Gini coefficient G (bottom; defined in Eqn. 4) of all selected haloes as a function of M_{halo} at redshifts of $z = \{10, 8.5, 7, 5.5\}$. The distribution is normalized by the number of haloes in each mass bin similarly to Fig. 1. The median and 16th and 84th percentiles are also shown in black solid and dashed lines, respectively. Most haloes have covering fractions P falling between 0.05 and 0.4, i.e. a minority of sightlines influences the mean. There is no significant dependence on redshift but a qualitative increase for the highest mass haloes. The third panel shows that the standard deviations of sightline escape fractions in most haloes are above $\langle f_{\text{esc}} \rangle$, and the median of the normalized standard deviation decreases as halo mass and mean escape fraction increase. The fourth panel shows an increasing half flux area as halo mass increases, emphasizing the importance of sightlines that are not among the highest ones. The Gini coefficient in the bottom panel similarly shows a more isotropic and equitably distributed picture at the high-mass end.

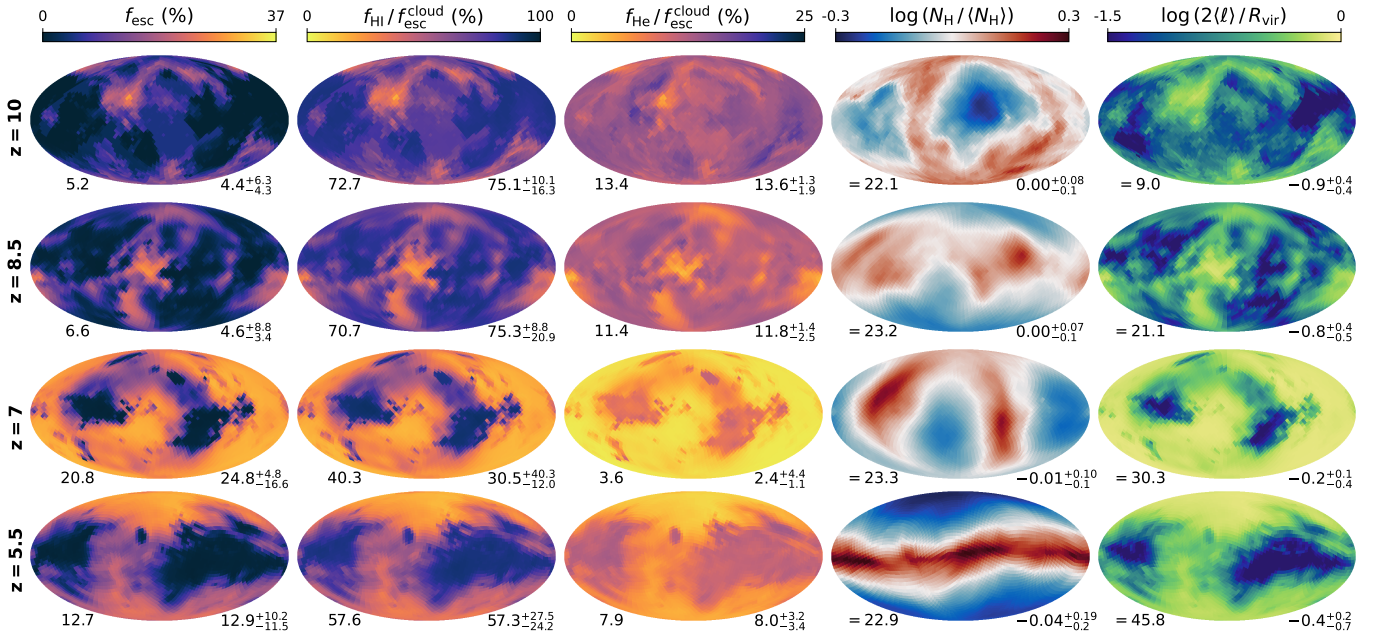


Figure 13. Healpix plots of various physical quantities in the most massive galaxies at redshifts of $z = \{10, 8.5, 7, 5.5\}$, respectively. From left to right these are: LyC escape fraction f_{esc} , hydrogen absorption fraction f_{HI} (normalized by $f_{\text{esc}}^{\text{cloud}}$), helium absorption fraction $f_{\text{He I}} + f_{\text{He II}}$ (also normalized), hydrogen column density N_{H} (normalized by the mean hydrogen column density of 3072 directions), and mean absorption distance $\langle \ell \rangle$. The dust density has been set to zero in this calculation. In the lower left (right) of each map we provide the average (median and 1σ variation) of all 3072 sightlines, with the exception of the hydrogen column density and absorption distance where for convenience we report the halo-dependent mean logarithmic column density (in cm^{-2}) and virial radius (in kpc), respectively. Formulae used in the calculation of absorption and mean absorption distance are discussed in Sec. 2.

Similarly, the mean absorption distance is normalized by the virial radius of the halo. The mean absorption distance corresponds to the typical distance photons travel before being absorbed, as defined in Eqn. (3). Note that the maximum mean optically thin absorption distance equals half the virial radius as can be seen by setting $k_a = 0$ in Eqn. (3). The median values in this example give $\langle \ell \rangle = \{0.57, 1.56, 10.34, 8.32\}$ kpc with $R_{\text{vir}} = \{9.05, 21.1, 30.3, 45.8\}$ kpc at $z = \{10, 8.5, 7, 5.5\}$. Sightlines with particularly low absorption distances are expected to exist and are due to the geometry (e.g. covering) of neutral gas and star particle migration from dense environments.

Our results clearly demonstrate the anisotropic behaviour of f_{esc} . Moreover, the morphologies of the direct radiative transfer quantities (i.e. with some exceptions for N_{H}) are strikingly similar, indicating correlations among them at the sightline level. In particular, f_{esc} strongly correlates with the mean absorption distance but anti-correlates with the hydrogen and helium absorption fractions. In other words, the photons can travel longer distances and escape the halo along low column density sightlines, resulting in lower f_{HI} , f_{He} and thus higher f_{esc} . Finally, disk-like structures from the forming galaxy manifest themselves in the bipolar distribution of the sightline statistics, with higher f_{esc} in the direction perpendicular to the plane of the disk. The high anisotropy found in the values of f_{esc} from various sightlines also implies the measured f_{esc} can be highly dependent on the direction being measured (edge-on or face-on). It therefore presents a potential challenge for comparing simulations to observations, which are more or less randomly oriented. Of course, lower mass galaxies will have different morphologies and characterizing the anisotropic sourcing of LyC radiation during the EoR is worth a more detailed investigation, especially if small-to-large scale

(ISM, CGM, and IGM) connections cascade to affect the bubble topology of reionization.

6 MODEL UNCERTAINTIES

We now discuss the main caveats of our analysis and when possible quantify the level of uncertainties, starting with dust absorption, AGN emission, and galaxy formation modelling. A discussion of resolution and convergence is presented in Appendix A.

6.1 Dust absorption

As briefly noted in Sec. 2, THESAN does not include on-the-fly dust absorption for ionizing photons, mostly because dust modelling in the high-redshift Universe remains relatively unconstrained. Including dust would mainly translate to a slightly different behaviour for the radiation field in high-mass galaxies. While it is worth exploring the impact of on-the-fly dust absorption in THESAN, also in the context of subresolution ISM and feedback models, for consistency with the simulated ionization states we also do not include dust in our fiducial MCRT calculations. In this subsection, we focus on estimating the impact of dust on post-processed MCRT escape fractions.

Although THESAN incorporates the creation and destruction of dust, many physical processes relevant to dust occur at subparsec scales and thus the precise amount of dust in each cell might not be accurately determined. Thus, in addition to using the dust abundance modelled from the simulation, we also insert dust into the post-processing MCRT calculation using the redshift-dependent dust-to-metal ratio suggested by Vogelsberger et al. (2020b) (model C), which has been calibrated to match observed UV luminosity functions. In the latter

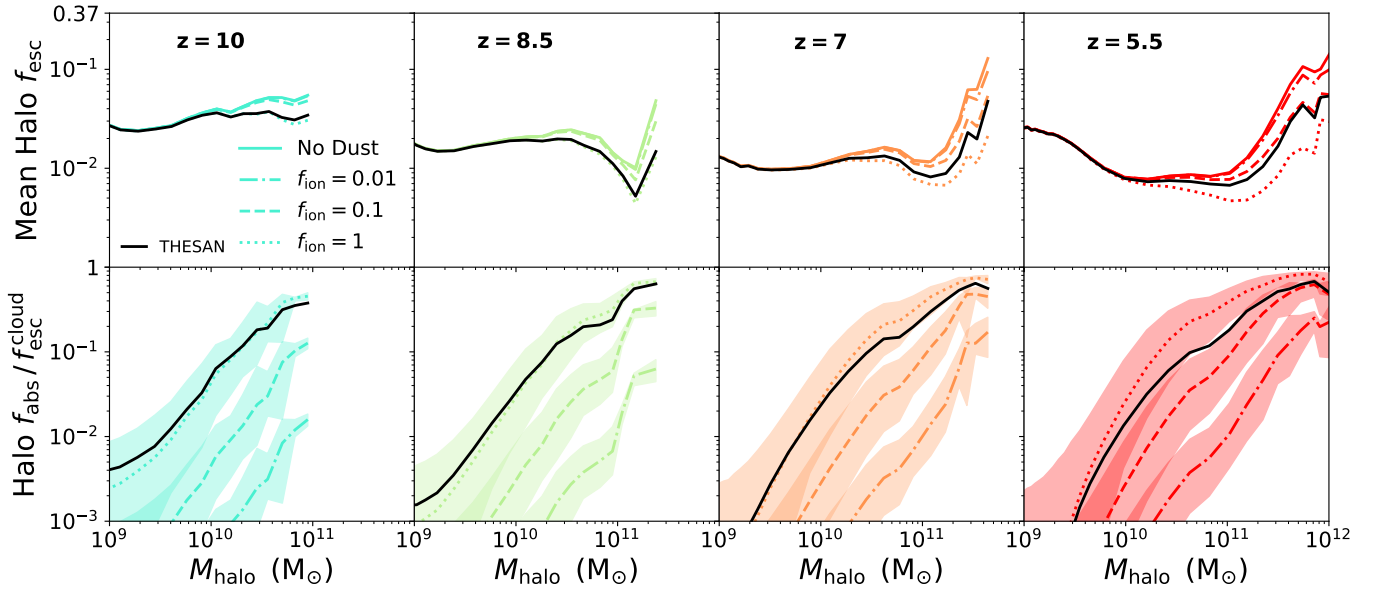


Figure 14. *Top:* Mean values of halo escape fraction f_{esc} as a function of M_{halo} at $z = \{10, 8.5, 7, 5.5\}$ for five different dust scenarios: no-dust (fiducial; solid), redshift-dependent dust-to-metal ratio from Vogelsberger et al. (2020b) together with three different dust survival fractions in H II regions $f_{\text{ion}} = \{0.01, 0.1, 1\}$ (dash-dotted, dashed, dotted), and on-the-fly dust modelling from THESAN (black solid). The inserted dust in all cases does not affect the overall dependence of f_{esc} on the halo mass except at the high-mass end where a lower f_{esc} results from the higher dust content. The on-the-fly simulated dust model lies between $f_{\text{ion}} = 0.1$ – 1 . *Bottom:* Median values of f_{abs} as a function of halo mass shown for the same redshifts and same five dust scenarios as above. The shaded regions indicate the 16th and 84th percentiles across different haloes. f_{abs} is calculated by Eqn. (2) and increases as dust mass increases as expected.

case, we also adopt three different dust survival fractions in the H II regions of $f_{\text{ion}} = \{0.01, 0.1, 1\}$ to empirically explore scenarios with varying degrees of reduced dust abundances in ionized regions. This is a common parametrization used within the Lyman-alpha radiative transfer community to increase escape fractions with the justification of being a proxy for generally harsh environments (Laursen et al. 2009). Specifically, the destruction processes correlate with local feedback and unresolved turbulence or shocks, as well as other mechanisms such as radiation pressure (Draine 2011) or rotational disruption of dust grains by radiative torques in strong radiation fields (Hoang et al. 2019). We warn that values below $f_{\text{ion}} \lesssim 0.1$ may not be realistic (see e.g. Hu et al. 2019; Kannan et al. 2020), however for our purposes the empirical model nonetheless provides a more continuous transition between the no dust and full dust scenarios. All the dust absorption opacities and anisotropic scattering are included based on the scattering albedo and Henyey–Greenstein phase function asymmetry parameter assuming the fiducial Milky Way dust model from Weingartner & Draine (2001).

In the top panel of Fig. 14 we show the mean halo f_{esc} at redshifts of $z = \{10, 8.5, 7, 5.5\}$ in the presence of dust with the models described above. Results using the simulation-determined dust distributions and constant dust-to-metal scaling are shown as black and coloured curves, respectively. For reference, the coloured solid curves show the values from the fiducial no-dust scenario, while the various dashed and dotted line styles indicate partial or full survival in H II regions. The f_{esc} dependence on M_{halo} has a similar trend in all cases, with a noticeable difference at the high-mass end where f_{esc} decreases as f_{ion} increases. This is due to a higher amount of dust in high-mass haloes. In fact, the dust attenuation seems to compensate for the increased f_{esc} to flatten out the high-mass end. The impact of dust attenuation is also higher at the lowest redshift ($z = 5.5$) due to the production of metals and dust throughout the stellar evolution process along with the increased presence of massive galaxies. We

note that the abundance of simulation-modelled dust lies between $f_{\text{ion}} = 0.1$ and $f_{\text{ion}} = 1$ at all redshifts examined.

The bottom panel of Fig. 14 presents the median dust absorption f_{abs} as calculated by Eqn. (2) under the same five dust scenarios. The shaded regions illustrate the 1σ variation among the halo population. There is a clear increase in f_{abs} with halo mass, again due to a larger amount of dust present in massive haloes, but it is also interesting to see the order-of-magnitude offsets with f_{ion} . The lowest redshift ($z = 5.5$) also reaches the most extreme values with $f_{\text{abs}} \approx \{0.25, 0.62, 0.83\}$ when $f_{\text{ion}} = \{0.01, 0.1, 1\}$. We emphasize that f_{abs} represents the degree of dust pre-absorption and is therefore indicative of additional dust within ionized regions, which could shrink if on-the-fly dust absorption is included self-consistently. The impact on halo escape fractions is not as large as one might expect based on f_{abs} alone because sightlines with low escape fractions will remain saturated. The global escaped photon rates are most affected if high f_{esc} sightlines also undergo significant dust absorption. Keeping this in mind, our estimates are roughly consistent with other post-processing results from simulations with a multiphase ISM (e.g. Kado-Fong et al. 2020; Smith et al. 2022b; Tacchella et al. 2022).

In addition to examining the effect of dust on the individual level, we also calculate the change in global escape fractions for different dust survival fractions as summarized in Table 1. The dust effect is almost negligible at pre-reionization redshifts ($z \gtrsim 8$). In the most extreme case of full dust ($f_{\text{ion}} = 1$), at the end of the simulation ($z = 5.5$) the global escape fraction is reduced to $\approx 57\%$ of the fiducial value from the no-dust scenario. We conclude that dust can be important during the latter half of the reionization process and incorporating on-the-fly prescriptions should be explored further in future simulation models. It remains unclear whether dust mainly adds degeneracy with other parameters such as the birth cloud escape fraction or if there are more fundamental differences.

Global ionizing escape fraction f_{esc} [%]						
Redshift	No dust	$f_{\text{ion}} = 0.01$	$f_{\text{ion}} = 0.1$	$f_{\text{ion}} = 1$	THESAN	$f_{\text{ion}}=1$ No dust
13	14.3	14.3	14.3	14.3	14.3	99.8
11	10.7	10.7	10.7	10.7	10.7	99.5
10	10.9	10.9	10.9	10.8	10.8	99.2
9	7.91	7.91	7.89	7.83	7.82	99.0
8	6.37	6.35	6.30	6.16	6.18	96.8
7	5.82	5.69	5.45	5.12	5.27	88.0
6	5.92	5.63	4.96	4.22	4.67	71.3
5.5	4.89	4.40	3.51	2.77	3.31	56.7

Table 1. The impact on the global escape fraction assuming the same dust scenarios as in Fig. 14, including the fiducial no-dust case, redshift-dependent dust-to-metal ratio model with various empirical dust survival fractions of $f_{\text{ion}} = \{0.01, 0.1, 1\}$, and the THESAN-modelled dust. Note that here we only present the instantaneous global escape fractions at $z = \{13, 11, 10, 9, 8, 7, 6, 5.5\}$, unlike in Fig. 4 where the values have been smoothed for presentation purposes. Including dust into the post-processing MCRT calculations has little influence at higher redshifts due to the relatively low dust abundances and halo masses at pre-reionization epochs. The impact increases as redshift decreases with a factor of ≈ 2 difference in the most extreme case by the end of the simulation.

6.2 Active galactic nuclei

The evident rise in f_{esc} at the high-mass end as seen in Fig. 8 suggests the potential presence of efficient ionizing mechanisms beyond the modest increase in stellar mass seen in Fig. 1. In particular, AGN formed in massive haloes can create highly energetic jets and significant UV to X-ray radiation. In the context of THESAN, AGN also provide important thermal and kinetic feedback to the surrounding gas (Weinberger et al. 2017). These collectively help generate additional escape channels for ionizing photons. The effects of AGN on reionization have been discussed in some previous studies (e.g. Hassan et al. 2018; Trebitsch et al. 2018; Finkelstein et al. 2019; Dayal et al. 2020), however THESAN offers a unique perspective because AGN sources are included on-the-fly throughout a large simulated volume. Thus, the effects of AGN on galaxies are included in our post-processing MCRT calculations in previous sections although only stars were considered as photon sources. In this subsection, we briefly present the results when treating AGN as additional photon sources.

In determining the AGN luminosity, we adopt the AGN spectral energy distribution (SED) from Lusso et al. (2015) with 35.5% of bolometric luminosity (as given in the IllustrisTNG model) contributing to the LyC photons with a subgrid escape fraction of 100%. We then perform the same MCRT calculations with AGN as the only photon source, assuming isotropic emission and no dust absorption. In the top panel of Fig. 15 we show a comparison of the global intrinsic (escaped) photon rate density as dashed (solid) curves between AGN (red) and stars (black) as functions of redshift. The total intrinsic photon rate density ρ_{int} from AGN is negligible at high redshifts, but rapidly rises as z decreases due to the formation of more AGN within massive galaxies. By the end of the simulation at $z = 5.5$, ρ_{int} remains below 2×10^{50} photons $\text{s}^{-1} \text{cMpc}^3$, which is approximately 100 times lower than ρ_{int} from stars. The low contribution from AGN is in agreement with Eide et al. (2020) and is also demonstrated in the original THESAN paper (Sec. 3.2.2 in Kannan et al. (2022a)). Nevertheless, AGN could potentially increase or even dominate locally

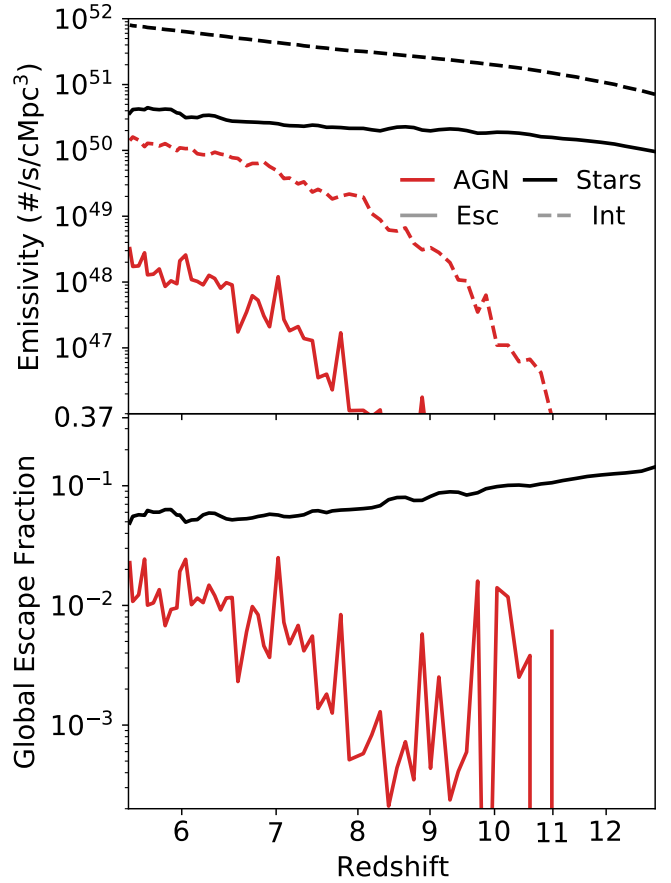


Figure 15. *Top:* Global intrinsic (dashed) and escaped (solid) photon rate densities from AGN (red) and stars (black) as functions of redshift. The AGN curves are analogous to Fig. 4 but with AGN as the only photon sources. Even at the lowest redshift ($z = 5.5$), AGN only contribute 1% and 0.5% of the total intrinsic and escaped photon rates, respectively. *Bottom:* Comparison of the corresponding global escape fractions for AGN and stars as functions of redshift. AGN f_{esc} are generally much lower than that of stars as a result of being located within galaxy centres. High redshift AGN statistics have large fluctuations due to the lack of massive galaxies in that regime given the limited simulation volume.

in simulations with substantially larger cosmic volumes or different physics modelling.

Similarly, the escaped photon rate density ρ_{esc} from AGN is also negligible, and at $z = 5.5$ reaches global levels that are only about 0.5% of the value from stars. In the bottom panel of Fig. 15 we show the corresponding global escape fractions showing that for AGN $f_{\text{esc}} \sim 1\%$, which is significantly below that of stars at all times. Contrary to the trend in the stellar f_{esc} , which drops slightly as redshift decreases, the AGN f_{esc} rises as redshift decreases. The oscillating AGN f_{esc} at high redshifts is due to the lack of statistics. We attribute the relatively low AGN f_{esc} to the fact that AGN are positioned at the centres of galaxy gravitational potential wells, where the abundance of high-density gas results in higher self-shielding covering fractions. Thus, escaping from an AGN source requires substantially more ionizing photons to compensate for the systematically larger gas column densities. However, although AGN do not have a significant contribution to the global escaped photons, they potentially play important roles in creating H II regions and ionizing channels for the more diffuse stellar sources.

7 CONCLUSIONS

In this paper, we have performed Monte Carlo radiative transfer calculations on all haloes from the THESAN simulation to study the production and escape of ionizing photons during the EoR. Our sample consists of haloes in the range $M_{\text{halo}} \in [10^8, 10^{12}] M_{\odot}$, and covers 90% of haloes with $M_{\text{halo}} > 1.2 \times 10^9 M_{\odot}$ at the midpoint of reionization at $z \approx 7.67$. The THESAN simulations have been shown to produce realistic galaxy and IGM properties that match the current observations and thus provide a self-consistent framework for studying the escape fractions of reionization-era galaxies (Kannan et al. 2022a; Garaldi et al. 2022; Smith et al. 2022a). The large volume of the simulation ($L_{\text{box}} = 95.5 \text{ cMpc}$) provides rich statistics for massive haloes ($M_{\text{halo}} > 10^{10} M_{\odot}$) that are important in driving the reionization process but are often statistically underrepresented in previous studies. We summarize our main conclusions in the following points:

(i) Lower mass haloes ($M_{\text{halo}} \lesssim 10^9 M_{\odot}$) provide the majority of both intrinsic and escaped photons above $z \gtrsim 7$, while higher mass haloes ($M_{\text{halo}} \gtrsim 10^{10} M_{\odot}$) dominate the photon budget thereafter. The most massive galaxies ($M_{\text{stars}} \gtrsim 10^9 M_{\odot}$) are particularly important at the later stages of reionization due to their proportionately high intrinsic luminosities. In fact, we find the decrease in the global escape fraction with decreasing redshift can be attributed to this transition from low-to-high mass dominance.

(ii) Halo escape fractions have a non-trivial dependence on M_{halo} : f_{esc} reaches a maximal value of $f_{\text{esc}}^{\text{cloud}} = 0.37$ for unresolved haloes ($M_{\text{halo}} \lesssim 10^8 M_{\odot}$), meaning that almost all ionizing photons escape after emerging from their local birth clouds. While this agrees with our physical intuition, it is essentially an artefact of numerical resolution in our model and we warn that previous studies reporting high f_{esc} close to the resolution limit of their simulations also likely have this issue. Specifically, the median f_{esc} decreases as halo mass increases, reaching a minimum once haloes are more robustly resolved around $3 \times 10^8 M_{\odot}$ at $z \lesssim 6$ and $10^9 M_{\odot}$ at $z \gtrsim 8$. After reaching the minimum value, f_{esc} increases with halo mass, stellar mass, and SFR at all redshifts. This trend still holds even with the presence of dust attenuation, which has a larger effect on massive haloes. We note that while the halo mass corresponding to the minimum f_{esc} shifts as the resolution changes, we find reasonable agreement in the global and resolved halo escape fractions with lower resolution simulations (see Appendix A).

(iii) Halo f_{esc} are not a simple function of either M_{halo} , M_{stars} , SFR, or Σ_{SFR} . However, given a certain halo mass range, galaxies with higher SFR or Σ_{SFR} tend to have higher f_{esc} . Furthermore, we found by considering SFR, gas mass, and the size of a halo, high f_{esc} can be fairly well distinguished by a theoretical effective SFR surface density $\tilde{\Sigma}_{\text{SFR}} = \text{SFR}/M_{\text{gas}}/R_{200}^2$.

(iv) We find a wide range of values for f_{esc} and $\dot{N}_{\text{esc}}$ at all redshifts, with lower mass haloes exhibiting broader distributions than high-mass ones. On the other hand, a tight relation can be observed between intrinsic photon rates and SFR, implying that the large variation of f_{esc} is indeed due to the specific star formation histories and complex environmental dependence.

(v) Channels with lower escape fractions (f_{esc}) have higher hydrogen and helium absorption ($f_{\text{H I}}$, f_{He}) and correspondingly lower absorption distances ($\langle \ell \rangle$). This implies that the break out of ionizing photons tends to be highly anisotropic. In the case of disk-like galaxies, escape fractions are shown to be higher in face-on directions and essentially zero for edge-on viewing angles. Lower mass galaxies tend to be more anisotropic in the sense that half of the escaping flux emerges from a smaller fraction of sightlines and the

Gini coefficients are close to unity. The most massive galaxies exhibit a significant trend towards more isotropic and equitable escape.

(vi) Including dust in the MCRT calculations has observable consequences for high-mass haloes, especially at low redshifts. By the end of the simulation at $z = 5.5$, the global escape fraction decreases from roughly 5% in the no-dust scenario to nearly half that value with the added presence of dust for either the on-the-fly or constant dust-to-metal ratio models.

(vii) We find that AGN intrinsic LyC production and escape fractions are generally lower than stellar sources. Therefore, while it is important to include AGN in large-volume simulations, they do not contribute significantly to hydrogen reionization.

Our work analyzes millions of resolved galaxies in the large-volume THESAN simulation for haloes with mass between 10^8 – $10^{12} M_{\odot}$. We demonstrate the dominant role of massive haloes in driving cosmic reionization by considering the escaped photons from the ensemble. We plan to include our escape fraction catalogues as part of the upcoming THESAN public data release, which can be utilized in follow-up analyses. Our results can also be used to inform and calibrate contemporary reionization models. In fact, we argue that including on-the-fly dust absorption is important for simulation volumes with significant numbers of massive haloes ($M_{\text{halo}} \gtrsim 10^{10} M_{\odot}$). In the future, we plan to study the time-evolution of f_{esc} from individual haloes by utilizing merger tree catalogues to correlate the production and escape of ionizing photons with properties of star formation and local environment in greater detail. Furthermore, the THESAN project will soon include high-resolution zoom-in resimulations of a wide range of galaxies including a multiphase ISM framework (Marinacci et al. 2019; Kannan et al. 2020) and self-consistent meso-scale reionization environment inherited directly from the flagship simulation. These efforts will provide additional insights into the physics of escape from reionization-era galaxies and their evolved lower redshift counterparts for direct comparison to current and upcoming observational data.

ACKNOWLEDGEMENTS

We thank the referee for comments and suggestions which have improved the quality of this work. JY acknowledges support from the MIT Undergraduate Research Opportunities Program. AS acknowledges support for Program number *HST*-HF2-51421.001-A provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, incorporated, under NASA contract NAS5-26555. MV acknowledges support through NASA ATP grants 16-ATP16-0167, 19-ATP19-0019, 19-ATP19-0020, 19-ATP19-0167, and NSF grants AST-1814053, AST-1814259, AST-1909831 and AST-2007355. 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). Additional computing resources were provided by the MIT Engaging cluster. We are thankful to the community developing and maintaining software packages extensively used in our work, namely: *matplotlib* (Hunter 2007), *numpy* (Walt et al. 2011) and *scipy* (Jones et al. 2001).

DATA AVAILABILITY

All THESAN simulation data will be made publicly available in the near future, including escape fraction catalogues. Data will be distributed

via www.thesan-project.com. Before the public data release, data underlying this article will be shared on reasonable request to the corresponding author(s).

REFERENCES

- Angulo R. E., Pontzen A., 2016, *MNRAS*, **462**, L1
- Barkana R., Loeb A., 2001, *Phys. Rep.*, **349**, 125
- Barnes J., Hut P., 1986, *Nature*, **324**, 446
- Bianco M., Iliev I. T., Ahn K., Giri S. K., Mao Y., Park H., Shapiro P. R., 2021, *MNRAS*, **504**, 2443
- Borrow J., Kannan R., Garaldi E., Smith A., Vogelsberger M., Pakmor R., Springel V., Hernquist L., 2022, arXiv e-prints, [p. arXiv:2212.03255](https://arxiv.org/abs/2212.03255)
- Bridge C. R., et al., 2010, *ApJ*, **720**, 465
- Bromm V., Yoshida N., 2011, *ARA&A*, **49**, 373
- Buck T., Pfrommer C., Girichidis P., Corobean B., 2022, *MNRAS*,
- Cen R., Kimm T., 2015, *ApJ*, **801**, L25
- Chabrier G., 2003, *ApJ*, **586**, L133
- Davies F. B., et al., 2018, *ApJ*, **864**, 142
- Dawoodbhoy T., et al., 2018, *MNRAS*, **480**, 1740
- Dayal P., et al., 2020, *MNRAS*, **495**, 3065
- Dijkstra M., 2019, *Saas-Fee Advanced Course*, **46**, 1
- Draine B. T., 2011, *ApJ*, **732**, 100
- Eide M. B., Ciardi B., Graziani L., Busch P., Feng Y., Di Matteo T., 2020, *MNRAS*, **498**, 6083
- 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. Australia*, **34**, e058
- Faucher-Giguère C.-A., 2020, *MNRAS*, **493**, 1614
- Fielding D. B., Bryan G. L., 2022, *ApJ*, **924**, 82
- Finkelstein S. L., et al., 2012, *ApJ*, **758**, 93
- Finkelstein S. L., et al., 2019, *ApJ*, **879**, 36
- Flury S. R., et al., 2022, *ApJS*, **260**, 1
- Garaldi E., Kannan R., Smith A., Springel V., Pakmor R., Vogelsberger M., Hernquist L., 2022, *MNRAS*, **512**, 4909
- Genel S., et al., 2019, *ApJ*, **871**, 21
- Gnedin N. Y., 2014, *ApJ*, **793**, 29
- Gnedin N. Y., Kravtsov A. V., Chen H.-W., 2008, *ApJ*, **672**, 765
- Grazian A., et al., 2017, *A&A*, **602**, A18
- Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, *MNRAS*, **466**, 4239
- Gutcke T. A., Pakmor R., Naab T., Springel V., 2021, *MNRAS*, **501**, 5597
- Haardt F., Madau P., 2012, *ApJ*, **746**, 125
- Hassan S., Davé R., Mitra S., Finlator K., Ciardi B., Santos M. G., 2018, *MNRAS*, **473**, 227
- Hoang T., Tram L. N., Lee H., Ahn S.-H., 2019, *Nature Astronomy*, **3**, 766
- Hopkins P. F., et al., 2018, *MNRAS*, **480**, 800
- Hu C.-Y., Zhukovska S., Somerville R. S., Naab T., 2019, *MNRAS*, **487**, 3252
- Hunter J. D., 2007, *Computing In Science & Engineering*, **9**, 90
- Iyer K. G., et al., 2020, *MNRAS*, **498**, 430
- Izotov Y. I., Schaerer D., Worseck G., Guseva N. G., Thuan T. X., Verhamme A., Orlitová I., Fricke K. J., 2018, *MNRAS*, **474**, 4514
- Japelj J., et al., 2017, *MNRAS*, **468**, 389
- Jones E., Oliphant T., Peterson P., et al., 2001, *SciPy: Open source scientific tools for Python*, <http://www.scipy.org/>
- Kado-Fong E., Kim J.-G., Ostriker E. C., Kim C.-G., 2020, *ApJ*, **897**, 143
- Kannan R., Vogelsberger M., Marinacci F., McKinnon R., Pakmor R., Springel V., 2019, *MNRAS*, **485**, 117
- Kannan R., Marinacci F., Vogelsberger M., Sales L. V., Torrey P., Springel V., Hernquist L., 2020, *MNRAS*, **499**, 5732
- Kannan R., Garaldi E., Smith A., Pakmor R., Springel V., Vogelsberger M., Hernquist L., 2022a, *MNRAS*, **511**, 4005
- Kannan R., Smith A., Garaldi E., Shen X., Vogelsberger M., Pakmor R., Springel V., Hernquist L., 2022b, *MNRAS*, **514**, 3857
- Kennicutt Robert C. J., 1998, *ARA&A*, **36**, 189
- Kim C.-G., Ostriker E. C., 2017, *ApJ*, **846**, 133
- Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, *MNRAS*, **466**, 4826
- Kostyuk I., Nelson D., Ciardi B., Glatzle M., Pillepich A., 2022, arXiv e-prints, [p. arXiv:2207.11278](https://arxiv.org/abs/2207.11278)
- Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, *MNRAS*, **485**, L24
- Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., Steinwandel U. P., Moster B. P., 2020, *ApJ*, **891**, 2
- Laursen P., Sommer-Larsen J., Andersen A. C., 2009, *ApJ*, **704**, 1640
- Levermore C. D., 1984, *J. Quant. Spectrosc. Radiative Transfer*, **31**, 149
- Lewis J. S. W., et al., 2020, *MNRAS*, **496**, 4342
- Loeb A., Furlanetto S. R., 2013, *The First Galaxies in the Universe*. Princeton Univ. Press, Princeton, NJ
- Lusso E., Worseck G., Hennawi J. F., Prochaska J. X., Vignali C., Stern J., O’Meara J. M., 2015, *MNRAS*, **449**, 4204
- Ma X., Kasen D., Hopkins P. F., Faucher-Giguère C.-A., Quataert E., Kereš D., Murray N., 2015, *MNRAS*, **453**, 960
- Ma X., Hopkins P. F., Kasen D., Quataert E., Faucher-Giguère C.-A., Kereš D., Murray N., Strom A., 2016, *MNRAS*, **459**, 3614
- Ma X., Quataert E., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2020, *MNRAS*, **498**, 2001
- Ma X., Quataert E., Wetzel A., Faucher-Giguère C.-A., Boylan-Kolchin M., 2021, *MNRAS*, **504**, 4062
- Ma Q.-B., Fiaschi S., Ciardi B., Busch P., Eide M. B., 2022, *MNRAS*,
- Madau P., 2017, *ApJ*, **851**, 50
- Madau P., Haardt F., Rees M. J., 1999, *ApJ*, **514**, 648
- Marinacci F., et al., 2018, *MNRAS*, **480**, 5113
- Marinacci F., Sales L. V., Vogelsberger M., Torrey P., Springel V., 2019, *MNRAS*, **489**, 4233
- Mason C. A., et al., 2019, *MNRAS*, **485**, 3947
- McGreer I. D., Mesinger A., D’Odorico V., 2015, *MNRAS*, **447**, 499
- McKinnon R., Torrey P., Vogelsberger M., Hayward C. C., Marinacci F., 2017, *MNRAS*, **468**, 1505
- Naidu R. P., Tacchella S., Mason C. A., Bose S., Oesch P. A., Conroy C., 2020, *ApJ*, **892**, 109
- Naiman J. P., et al., 2018, *MNRAS*, **477**, 1206
- Nelson D., et al., 2018, *MNRAS*, **475**, 624
- Nelson D., et al., 2019, *MNRAS*, **490**, 3234
- Nestor D. B., Shapley A. E., Kornei K. A., Steidel C. C., Siana B., 2013, *ApJ*, **765**, 47
- Ocvirk P., et al., 2020, *MNRAS*, **496**, 4087
- Paardekooper J. P., Khochfar S., Dalla C. V., 2013, *MNRAS*, **429**, L94
- Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, *MNRAS*, **451**, 2544
- Pahl A. J., Shapley A., Steidel C. C., Chen Y., Reddy N. A., 2021, *MNRAS*, **505**, 2447
- Pahl A. J., Shapley A., Steidel C. C., Reddy N. A., Chen Y., 2022, *MNRAS*, **516**, 2062
- Pillepich A., et al., 2018a, *MNRAS*, **473**, 4077
- Pillepich A., et al., 2018b, *MNRAS*, **475**, 648
- Pillepich A., et al., 2019, *MNRAS*, **490**, 3196
- Planck Collaboration et al., 2016, *A&A*, **594**, A13
- Planck Collaboration et al., 2020, *A&A*, **641**, A6
- Puglisi A., et al., 2016, *A&A*, **586**, A83
- Qin W., Schutz K., Smith A., Garaldi E., Kannan R., Slatyer T. R., Vogelsberger M., 2022, *Phys. Rev. D*, **106**, 123506
- Rathjen T.-E., et al., 2021, *MNRAS*, **504**, 1039
- Rivera-Thorsen T. E., et al., 2019, *Science*, **366**, 738
- Robertson B. E., Ellis R. S., Dunlop J. S., McLure R. J., Stark D. P., 2010, *Nature*, **468**, 49
- Robertson B. E., et al., 2013, *ApJ*, **768**, 71
- Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, *ApJ*, **802**, L19
- Rosdahl J., et al., 2018, *MNRAS*, **479**, 994
- Rosdahl J., et al., 2022, *MNRAS*, **515**, 2386
- Rutkowski M. J., et al., 2016, *ApJ*, **819**, 81
- Rybicki G. B., Lightman A. P., 1986, *Radiative Processes in Astrophysics*
- Saldana-Lopez A., et al., 2022, *A&A*, **663**, A59
- Schenker M. A., Ellis R. S., Konidaris N. P., Stark D. P., 2014, *ApJ*, **795**, 20
- Siana B., et al., 2010, *ApJ*, **723**, 241

Smith A., Safranek-Shrader C., Bromm V., Milosavljević M., 2015, *MNRAS*, **449**, 4336

Smith A., Ma X., Bromm V., Finkelstein S. L., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2019, *MNRAS*, **484**, 39

Smith A., Kannan R., Garaldi E., Vogelsberger M., Pakmor R., Springel V., Hernquist L., 2022a, *MNRAS*, **512**, 3243

Smith A., et al., 2022b, *MNRAS*, **517**, 1

Springel V., 2005, *MNRAS*, **364**, 1105

Springel V., 2010, *MNRAS*, **401**, 791

Springel V., Hernquist L., 2003, *MNRAS*, **339**, 289

Springel V., Yoshida N., White S. D., 2001, *New Astronomy*, **6**, 79

Springel V., et al., 2018, *MNRAS*, **475**, 676

Steidel C. C., Bogosavljević M., Shapley A. E., Reddy N. A., Rudie G. C., Pettini M., Trainor R. F., Strom A. L., 2018, *ApJ*, **869**, 123

Steinwandel U. P., Moster B. P., Naab T., Hu C.-Y., Walch S., 2020, *MNRAS*, **495**, 1035

Tacchella S., et al., 2018, *ApJ*, **859**, 56

Tacchella S., et al., 2022, *MNRAS*, **513**, 2904

Trebtsch M., Blaizot J., Rosdahl J., Devriendt J., Slyz A., 2017, *MNRAS*, **470**, 224

Trebtsch M., Volonteri M., Dubois Y., Madau P., 2018, *MNRAS*, **478**, 5607

Trujillo I., Chamba N., Knapen J. H., 2020, *MNRAS*, **493**, 87

Vanzella E., et al., 2018, *MNRAS*, **476**, L15

Verner D. A., Ferland G. J., Korista K. T., Yakovlev D. G., 1996, *ApJ*, **465**, 487

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020a, *Nature Reviews Physics*, **2**, 42

Vogelsberger M., et al., 2020b, *MNRAS*, **492**, 5167

Walt S. v. d., Colbert S. C., Varoquaux G., 2011, *Computing in Science & Engineering*, **13**, 22

Weinberger R., Hernquist L., 2023, *MNRAS*, **519**, 3011

Weinberger R., et al., 2017, *MNRAS*, **465**, 3291

Weinberger R., et al., 2018, *MNRAS*, **479**, 4056

Weinberger R., Springel V., Pakmor R., 2020, *ApJS*, **248**, 32

Weingartner J. C., Draine B. T., 2001, *ApJ*, **548**, 296

Wise J. H., 2019, *Contemporary Physics*, **60**, 145

Wise J. H., Cen R., 2009, *ApJ*, **693**, 984

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

Xu H., Wise J. H., Norman M. L., Ahn K., O'Shea B. W., 2016, *ApJ*, **833**, 84

Xu C., et al., 2022, arXiv e-prints, p. arXiv:2210.16275

Yajima H., Choi J.-H., Nagamine K., 2011, *MNRAS*, **412**, 411

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., Venkatesan A., Behroozi P., Ferguson H. C., 2020, *MNRAS*, **496**, 4574

APPENDIX A: RESOLUTION DEPENDENCE

THESAN also includes a suite of medium resolution simulations with different features aiming to investigate the impacts on reionization caused by different physics as described in Kannan et al. (2022a). In this appendix, we analyze the results from several of the medium resolution simulations to compare with the main results from the flagship THESAN-1 simulation explored throughout the body of this paper. We mainly focus on the THESAN-2 simulation, which has the same initial conditions as THESAN-1 but with two (eight) times lower spatial (mass) resolution. When relevant we also include the THESAN-WC-2 simulation, where the birth cloud escape fraction is slightly higher to compensate for lower star formation in the medium resolution runs. In addition, THESAN-SDAO-2 assumes an alternative dark matter model that includes couplings to relativistic particles giving rise to strong Dark Acoustic Oscillations (sDAOs) cutting off the linear matter power spectrum at small scales, and THESAN-HIGH-2 and THESAN-LOW-2 are designed to better understand

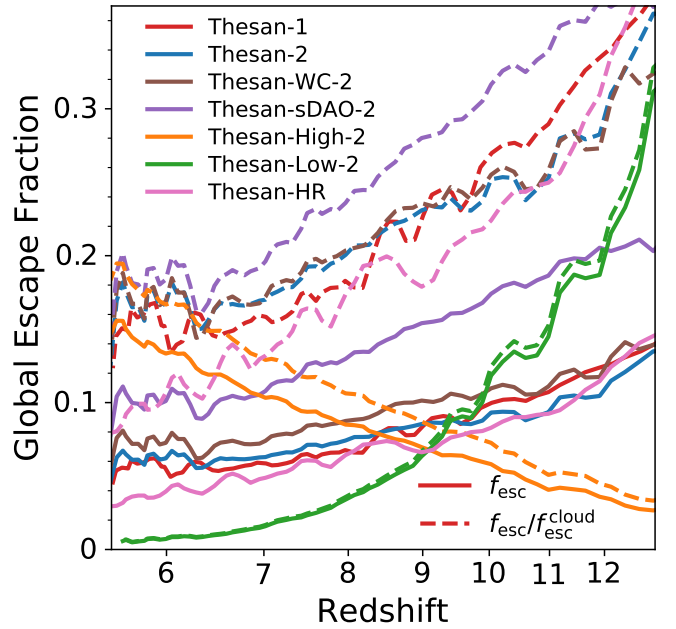


Figure A1. Comparison of the global photon rate weighted escape fractions from different THESAN simulations (see Fig. 4): THESAN-1 (red), THESAN-2 (blue), THESAN-WC-2 (brown), THESAN-SDAO-2 (purple), THESAN-HIGH-2 (orange), THESAN-LOW-2 (green), and THESAN-HR (pink). The original escape fractions are shown in solid lines, and the values after normalizing by the birth cloud escape fractions are in dashed lines. Results from THESAN-1, THESAN-2, and THESAN-WC-2 agree reasonably well after the normalization, demonstrating the robustness of the global escape fractions. Discussions on the comparisons with other simulations can be found in the associated texts.

whether high or low-mass galaxies dictate the reionization process with only haloes above/below $10^{10} M_{\odot}$ contributing to the escaped emission in THESAN-HIGH-2/THESAN-LOW-2. We note that in order to match the observed neutral hydrogen fraction, the different THESAN runs adopt different birth cloud escape fractions $f_{\text{esc}}^{\text{cloud}}$: THESAN-1 and THESAN-2 have $f_{\text{esc}}^{\text{cloud}} = 0.37$, THESAN-WC-2 has $f_{\text{esc}}^{\text{cloud}} = 0.43$, THESAN-SDAO-2 has $f_{\text{esc}}^{\text{cloud}} = 0.55$, THESAN-HIGH-2 has $f_{\text{esc}}^{\text{cloud}} = 0.8$, and THESAN-LOW-2 has $f_{\text{esc}}^{\text{cloud}} = 0.95$.

In addition to the medium resolution simulations, we also compare the results with a higher-resolution small box ($L_{\text{box}} = 12.5 \text{ cMpc}$, $N_{\text{particles}} = 2 \times 512^3$). This comes from a suite of simulations designed to investigate the impact of various reionization models on the properties of low-mass galaxies, and as such were simulated at much higher resolution than the original THESAN volumes. Our box, L8_N512 denoted here as THESAN-HR, has a baryonic mass resolution of $9.0 \times 10^4 M_{\odot}$, with a corresponding dark matter mass of $4.8 \times 10^5 M_{\odot}$, roughly a factor of 6.5 lower than THESAN-1. The gravitational softening is scaled appropriately, with the gravitational forces softened on a scale of 0.85 ckpc , but all other parameters remain fixed with respect to THESAN-1. This provides a unique opportunity to investigate the resolution dependence of the THESAN model, with these volumes discussed in detail in Borrow et al. (2022).

A1 Change in global statistics

In Fig. A1 we show the global escape fractions (solid) as a function of redshift for each of the above-mentioned THESAN variations. The THESAN-1 result is shown in the red curve for comparison (the same as in Fig. 4). The dashed lines show the values after normalizing the

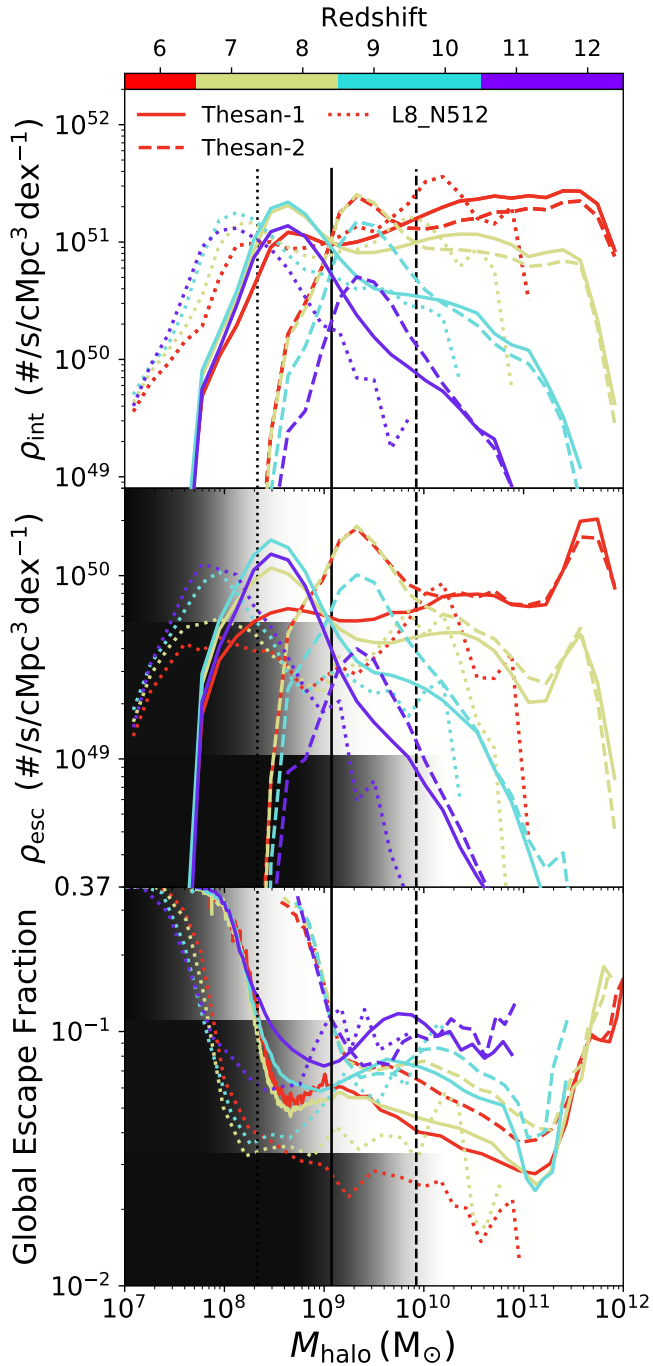


Figure A2. Resolution comparison between THESAN-1 (solid), THESAN-2 (dashed), and THESAN-HR (dotted) for the global intrinsic ρ_{int} (top) and escaped ρ_{esc} (middle) photon rate densities along with their escape fraction ratio (bottom) as functions of halo mass for several redshift ranges covering $z = 5.5\text{--}13$ (see Fig. 5). The grayscale backgrounds indicate the cumulative completeness, with vertical lines indicating the 90% level. The intrinsic and escaped values in THESAN-1 and THESAN-2 simulations agree at the high-mass end but differ at the low-mass end as halo structures and star formation histories become poorly resolved. THESAN-HR shows a different structure at the high-mass end due to its limited volume. The peaks at the low-mass end shift to even lower masses in THESAN-1 and THESAN-HR as the higher resolution allows the formation of lower mass galaxies. In the bottom panel, we see an overall reduction in global escape fractions as the resolution increases across most mass ranges, which is conveniently offset by the higher total intrinsic luminosities, such that there are similar amounts of escaped photons in all scenarios.

escape fractions by their birth cloud values to focus on differences in the MCRT calculations. The results from THESAN-1, THESAN-2, and THESAN-WC-2 simulations are in reasonable agreement after the normalization, implying that the change in mass resolution has relatively little impact on the global result. Thus, the birth cloud escape fraction acts as a viable way to account for the reduced SFR density of the medium resolution runs. The trends in THESAN-HIGH-2 and THESAN-LOW-2 are also expected since the growing number of massive haloes that have a significant contribution to escaped emission only affect THESAN-HIGH-2, and are completely turned off in THESAN-LOW-2. The escape fractions from THESAN-HR are seen to be lower at low-redshifts, which is likely due to the lack of high-mass haloes in the smaller box and consequently later reionization history rather than the improved resolution. Lastly, the ordering at $z \lesssim 7$ is as expected, i.e. if massive haloes start to dominate and are more or less converged with resolution then the curves should roughly line up near the end of the simulation.

Although the global results remain stable, adjusting the resolution creates a noticeable difference at the low-mass end as lower resolution limits the range of the least massive galaxies modelled. On the other hand, a simulation with a smaller box size limits the range of the most massive galaxies modelled and leads to differences at the high-mass end as well. Similar to Fig. 5, Fig. A2 shows the global intrinsic photon rate densities (top), escaped photon rate densities (middle), and escape fractions (bottom) from THESAN-1 (solid), THESAN-2 (dashed), and THESAN-HR (dotted). The three grayscale backgrounds indicate the cumulative completeness of THESAN-HR, THESAN-1, and THESAN-2 from top to bottom with three vertical lines to indicate 90% completeness. At the high-mass end, both intrinsic and escaped rates have similar values between THESAN-1 and in THESAN-2. Specifically, we note that they are reasonably converged for the dominant drivers of reionization ($M_{\text{halo}} \gtrsim 10^{10} M_{\odot}$). The discrepancy between two simulations increases as the mass decreases and resolution artifacts become apparent. The positions of the peaks at the low-mass end shift to the left as the resolution increases and more low-mass galaxies are being resolved in THESAN-1. The low-mass peak in THESAN-HR shifts to further left due to its higher resolution. Furthermore, it shows a different structure at the high-mass end due to its limited box size ($L_{\text{box}} = 12.5 \text{ cMpc}$) compared to THESAN-1 and THESAN-2 (both with $L_{\text{box}} = 95.5 \text{ cMpc}$). The lack of massive haloes above $10^{11} M_{\odot}$ is the main reason for the observed discrepancy. In terms of global escape fractions (bottom panel), the positions of the sharp upturn to maximal escape fractions are similarly set by the resolution. Quantitatively speaking, we found that increasing resolution leads to overall lowered values, which can be understood by requiring the total amount of photons to be the same for all simulations to have a consistent description of the reionization history. For example, the higher total luminosity in THESAN-1 due to a larger number of resolved stars is compensated by lower global escape fractions compared to THESAN-2.

Fig. A2 shows that although the global emission is roughly converged at the high-mass end for simulations with the same box size, they generally do not agree at the low-mass end. However, in this case it is misleading to judge the convergence of a higher resolution simulation by comparing to a lower resolution one. It is better to compare to even higher resolution simulations, but to control computational cost it is inevitable to also reduce the box size when increasing resolution, which leads to different halo mass functions and other discrepancies. This being said, we are not aware of any reionization simulations claiming to have converged escape fractions. Systematic factors of a few are common, especially when considering individual haloes. For example, for the zoom-in simulations in Ma et al. (2020), the absolute f_{esc} does not converge in all mass ranges. This is also

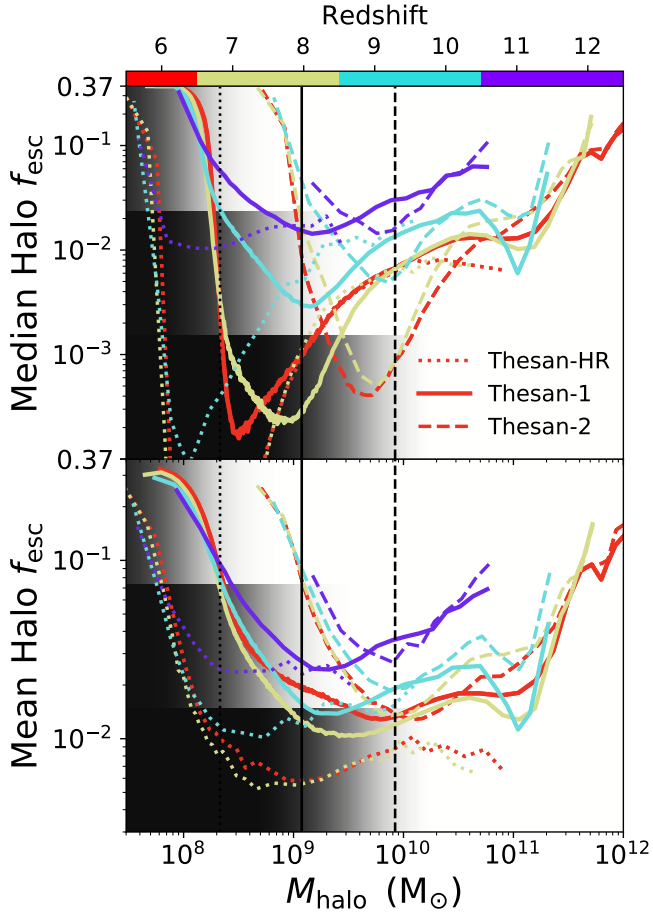


Figure A3. Resolution comparison between THESAN-1 and THESAN-2, and THESAN-HR for the median (top) and mean (bottom) halo values of f_{esc} as a function of M_{halo} across several redshifts in the range $z = 5.5\text{--}13$ (see Fig. 8). The vertical lines correspond to 90% completeness and the grayscale backgrounds indicate the cumulative completeness of THESAN-HR, THESAN-1, and THESAN-2 from top to bottom. Three simulations admit the same pattern of the f_{esc} dependence on halo mass as discussed in the text with a dip and gradual rise towards more massive haloes. The halo mass at which the minimal f_{esc} occurs increases from $\approx 2 \times 10^8 M_{\odot}$ in THESAN-HR to $\approx 10^9 M_{\odot}$ in THESAN-1 to $\approx 8 \times 10^9 M_{\odot}$ in THESAN-2 as the resolution is lowered.

seen in other large-volume simulations such as Lewis et al. (2020) and Rosdahl et al. (2022). We list several reasons that simulations might not demonstrate convergence: (i) changing the box size impacts the halo mass function especially at the high-mass end, and this leads to missing impacts on neighbouring haloes, (ii) some lower resolution simulations might not be able to capture key physical processes as escape fractions ultimately need to be resolved down to cloud scales, and (iii) some resolution dependent effects like feedback are difficult to account for in post-processing (Trebitsch et al. 2018; Kimm et al. 2017). To the best of our knowledge, there are no reionization simulations that can fully address the resolution convergence problem. However, we emphasize that in the above cases and ours that global results are sufficiently robust to maintain the same mechanisms for reionization and qualitative halo trends.

A2 Change in individual halo dependence

We also argue that the change in resolution does not significantly alter the pattern of the halo median/mean f_{esc} dependence on halo mass. As shown in Fig. A3, both the median (top) and mean (bottom) f_{esc} dependence on M_{halo} in THESAN-2 and THESAN-HR resembles the result from THESAN-1 except that the dip marking the transition from low completeness of low-mass haloes and sufficient resolution for richer substructures is shifted to higher masses for lower resolution simulations. This dip and gradual rise of f_{esc} towards more massive haloes roughly corresponds to the point of 90% completeness as indicated in the vertical lines, which is $\approx 2 \times 10^8 M_{\odot}$ in THESAN-HR, $\approx 10^9 M_{\odot}$ in THESAN-1, and $\approx 8 \times 10^9 M_{\odot}$ in THESAN-2. This is again due to resolution effects and this correspondence also implies that completeness can act as a reasonable proxy for the resolution. Overall, it is clear that the resolution requirements for obtaining robust halo escape fractions are likely more demanding than for other galaxy formation properties such as SFRs or gas fractions.

APPENDIX B: EQUATION OF STATE CELLS

In Sec. 2.2, we have discussed the model uncertainty caused by EoS cells. To provide a more quantitative view, we briefly explore the impact of the on-the-fly coupling to the equation of state (EoS) for escape fractions by modifying the ionization states of EoS cells in post-processing. Fig. B1 shows median (top) and mean (bottom) halo f_{esc} as a function of halo mass at $z = 5.5$ (red) and $z = 9$ (blue) in the following four scenarios: (i) directly adopting the original ionization states from the THESAN-1 simulation as is done throughout paper (THESAN-1, solid), (ii) performing MCRT photoionization equilibrium calculations (as in Smith et al. 2022b) to update the EoS ionization states while retaining the states of non-EoS cells (EoS-eq, dashed), (iii) same as (ii) but in addition also requiring the temperature of EoS cells to be $T = 10^4$ K (EoS-eq-1e4, dash-dotted), and (iv) setting all EoS cells to be fully neutral (Neutral, dotted). The grayscale background is again shown to represent the cumulative completeness of the halo selection and the vertical lines show the 90% level. We find the Neutral scenario shows a much lower f_{esc} compared to the fiducial THESAN-1 results and establishes a lower limit on the impact of the EoS on f_{esc} . This comparison reflects that EoS cells are mostly ionized in simulations. To assess the validity of the effective ionization states in EoS cells, we compare THESAN-1 with EoS-eq, where photoionization equilibrium ionization states are used. From Fig. B1, we see that these two results agree well at the high-mass end in both median and mean. At the low-mass end, they show reasonable agreement in mean statistics but differ in median upon entering the range with low completeness. Overall, the escape of ionizing photons using ionization states from the simulation are comparable to those obtained by the photoionization equilibrium. Moreover, lowered values from EoS-eq-1e4 compared to EoS-eq imply that some ionizing photons escape as a result of collisional ionization from artificially high temperatures. Quantitatively, we find the global escape fraction at $z = 5.5$ (9) decreases from 4.89% (7.91%) in the fiducial model to 4.30% (7.42%) in EoS-eq, 2.43% (7.08%) in EoS-eq-1e4, and 1.67% (4.25%) in Neutral, i.e. a modest factor $\lesssim 2$ uncertainty if the birth cloud escape fraction accounts for local cold phase absorption while most transport is through the hot ionized phase (e.g. Buck et al. 2022). We conclude that on-the-fly EoS coupling should be studied in greater detail as the low $f_{\text{esc}}^{\text{cloud}}$ calibration is a direct result of the adopted EoS absorption model (see also Kostyuk et al. 2022). The fiducial passive RT method used

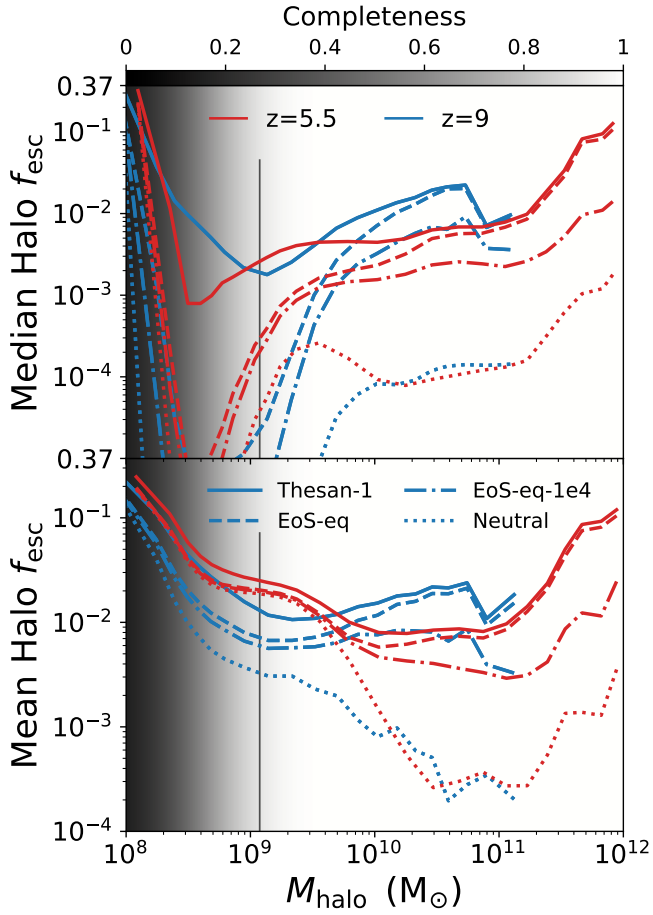


Figure B1. The change in median (top) and mean (bottom) f_{esc} as a function of M_{halo} while altering the ionization states in the equation of states (EoS) cells. We show the following four cases: (i) *THESAN-1* (solid): using the ionization states from the simulation, (ii) *EoS-eq* (dashed): updating the EoS cells from converged values in photoionization equilibrium, (iii) *EoS-eq-1e4* (dash-dotted): same as (ii) but first setting $T = 10^4$ K in EoS cells, and (iv) *Neutral* (dotted): setting all EoS cells to be completely neutral.

throughout the body of this paper is the correct approach for interpreting *THESAN* escape fractions because it most closely reflects the on-the-fly results for this particular model.

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