

Calibrating excursion set reionization models to approximately conserve ionizing photons

Jaehong Park ¹★, Bradley Greig ^{2,3} and Andrei Mesinger ⁴

¹*School of Physics, Korea Institute for Advanced Study (KIAS), 85 Hoegiro, Dongdaemun-gu, Seoul 02455, Republic of Korea*

²*ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D), University of Melbourne, VIC 3010, Australia*

³*School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia*

⁴*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*

Accepted 2022 September 23. Received 2022 September 23; in original form 2021 December 10

ABSTRACT

The excursion set reionization framework is widely used due to its speed and accuracy in reproducing the 3D topology of reionization. However, it is known that it does not conserve photon number. Here, we introduce an efficient, on-the-fly recipe to approximately account for photon conservation. Using a flexible galaxy model shown to reproduce current high- z observables, we quantify the bias in the inferred reionization history and galaxy properties resulting from the non-conservation of ionizing photons. Using a mock 21-cm observation, we perform inference with and without correcting for ionizing photon conservation. We find that ignoring photon conservation results in very modest biases in the inferred galaxy properties for our fiducial model. The notable exception is in the power-law scaling of the ionizing escape fraction with halo mass, which can be biased from the true value by $\sim 2.4\sigma$ (corresponding to ~ -0.2 in the power-law index). Our scheme is implemented in the public code 21cmFAST.

Key words: galaxies: high-redshift – intergalactic medium – dark ages, reionization, first stars – diffuse radiation – early Universe – cosmology: theory.

1 INTRODUCTION

Due to the short mean free paths of Lyman limit photons in the neutral intergalactic medium (IGM), cosmic reionization is inherently inhomogeneous. Ionized regions surrounding the first galaxies expand into the neutral IGM, growing and percolating as new galaxies are formed (e.g. Friedrich et al. 2011; Furlanetto & Oh 2016). This bimodality of the ionization state of the IGM motivated the development of excursion set models of reionization (Furlanetto, Zaldarriaga & Hernquist 2004; Paranjape, Choudhury & Padmanabhan 2016). In the excursion set framework, the cumulative number of ionizing photons is compared to the number of baryons plus recombinations in spheres of decreasing radii around a given patch of the IGM. If there are more ionizing photons, the patch is determined to be ionized.

These models can be easily implemented on 3D grids using a series of fast Fourier transforms (FFTs), allowing for fast realizations of the reionization topology on large scales. This has made excursion set models very popular in seminumerical Epoch of Reionization (EoR) simulations, allowing us to efficiently explore the large-parameter space of astrophysical and cosmological uncertainties. Excursion set models have been tested numerous times against more accurate radiative transfer (RT) simulations (Zahn et al. 2007, 2011; Mesinger, Furlanetto & Cen 2011; Majumdar et al. 2014; Hutter 2018; Thélie et al. 2021). When compared at a fixed stage of the EoR (i.e. a given filling factor of H II regions, Q_{HII}), excursion set and more accurate RT approaches result in ionization maps that agree well on moderate to large scales ($\gtrsim 1$ Mpc).

However, it is well known that excursion set approaches do not conserve the ionizing photon number (Zahn et al. 2007, 2011; Paranjape et al. 2016; Choudhury & Paranjape 2018; Hutter 2018). Fundamentally, this problem occurs when neighbouring H II regions overlap. Photons in excess of the criteria for ionization are not correctly re-distributed into neighbouring cells (Zahn et al. 2007). Therefore, without accounting for this non-conservation when using seminumerical EoR simulations to infer galaxy properties from observations, we expect a bias in the recovered ionizing emissivity (e.g. an overestimate of the ionizing photon escape fraction; Hutter 2018; Molaro et al. 2019).

Several approaches have been developed to conserve ionizing photons in approximate radiative transfer schemes (e.g. Kim et al. 2016; Paranjape et al. 2016; Hassan et al. 2017; Choudhury & Paranjape 2018; Molaro et al. 2019). However, these can come at the cost of significantly increased computation time and/or stepping away from the well-tested excursion set framework entirely.

Motivated by the need for rapid parameter space exploration, here we develop a computationally efficient, approximate correction for ionizing photon conservation in 3D excursion set models. Rather than explicitly tracking the number of ionizing photons, we calibrate the simulated EoR history to match an analytically derived EoR history. We effectively re-scale (increase) the ionizing emissivity as a function of redshift to compensate for the loss of ionizing photons. This calibration therefore preserves the well-tested reionization topology at a fixed stage of the EoR, resulting from excursion set models (Zahn et al. 2007, 2011; Mesinger et al. 2011; Hutter 2018; Thélie et al. 2021).

We also quantify the biases in the recovered astrophysical parameters when not accounting for the photon conservation using a mock

* E-mail: jaehongpark@kias.re.kr

observation of the 21-cm signal. We infer the astrophysical parameters within a fully Bayesian framework using the public MCMC sampler of 3D EoR simulations 21cmMC¹ (Greig & Mesinger 2015, 2017b; Greig & Mesinger 2018). Our correction has been implemented in the (v3.0²; Murray et al. 2020) release of 21cmFAST (Mesinger & Furlanetto 2007; Mesinger et al. 2011).

This paper is organized as follows. In Section 2, we summarize how 21cmFAST computes ionization fields before we introduce the methodology for correcting the photon non-conservation in Section 3. In Section 4, we outline the set-up for Bayesian inference of astrophysical parameters, and quantify biases from photon non-conservation in Section 5. Finally, we summarize our results in Section 6. We assume a standard Λ CDM cosmology based on *Planck* 2016 (Planck Collaboration et al. 2016): $(h, \Omega_m, \Omega_b, \Omega_\Lambda, \sigma_8, n_s) = (0.678, 0.308, 0.0484, 0.692, 0.815, 0.968)$. Unless stated otherwise, we quote all quantities in comoving units, and UV magnitudes refer to the rest-frame 1500 Å AB magnitude.

2 IONIZATION FIELDS FROM THE EXCURSION SET

Here we briefly summarize the excursion set EoR model. The general analytic framework was first developed by Furlanetto et al. (2004), and subsequently applied to 3D realizations in Zahn et al. (2005), Mesinger & Furlanetto (2007). Since then various seminumerical simulations have adopted the excursion set framework when generating reionization fields (e.g. Zahn et al. 2005; Mesinger & Furlanetto 2007; Geil & Wyithe 2008; Alvarez et al. 2009; Choudhury, Haehnelt & Regan 2009; Santos et al. 2010; Mesinger et al. 2011; Visbal et al. 2012; Mutch et al. 2016; Hutter 2018). Below we focus our discussion on its implementation in the public 21cmFAST code.

As mentioned above, the excursion set model divides the IGM during the EoR into (mostly) ionized and (mostly) neutral regions.³ To determine whether or not an IGM patch is ionized, the cumulative number of ionizing photons is compared to the number of atoms plus recombinations. Specifically, a voxel at spatial location and redshift (\mathbf{x}, z) is flagged as ionized if the following criterion is satisfied:

$$n_{\text{ion}}(\mathbf{x}, z|R, \delta_R) \geq (1 + n_{\text{rec}})(1 - x_e). \quad (1)$$

Here n_{rec} is the cumulative number of recombinations per baryon, x_e is the fraction of ionizations by X-rays. The criterion in equation (1) is checked in spheres of decreasing radii starting from the maximum radius (set by the mean free path through the ionized IGM) to a radius equivalent in size to the cell resolution. In 21cmFAST, the cumulative number of ionizing photons per baryon inside a spherical region of scale R with overdensity δ_R is given by

$$n_{\text{ion}} = \bar{\rho}_b^{-1} \int_0^\infty dM_h \frac{dn(M_h, z|R, \delta_R)}{dM_h} f_{\text{duty}} M_* f_{\text{esc}} N_{\gamma/b}, \quad (2)$$

where $\bar{\rho}_b$ is the mean baryon density, $dn(M_h, z|R, \delta_R)/dM_h$ is the conditional Press–Schechter halo mass function (HMF) normalized to the mean of the Sheth–Tormen HMF (Sheth & Tormen 1999),

and $N_{\gamma/b}$ is the number of ionizing photons per stellar baryon.⁴ M_* , f_{esc} , f_{duty} are the stellar mass of a galaxy, the ionizing photon escape fraction, and the duty cycle, respectively, which will be described in Section 4.1.

21cmFAST calculates inhomogeneous recombinations according to the subgrid prescription of Sobacchi & Mesinger (2014). The hydrogen recombination rate at a spatial position \mathbf{x} and redshift z is:

$$\frac{dn_{\text{rec}}}{dt}(\mathbf{x}, z) = \bar{n}_H \alpha_B \Delta_{\text{cell}}^{-1} \int_0^{180} [1 - x_{\text{HII}}]^2 P_V \Delta^2 d\Delta, \quad (3)$$

where $\Delta_{\text{cell}} \equiv 1 + \delta_{\text{nl}}$ is density contrast of the cell, P_V is the volume-averaged probability distribution function of subgrid overdensity, Δ , taken from Miralda-Escudé, Haehnelt & Rees (2000) and adjusted for the mean contrast Δ_{cell} , α_B is the case-B recombination coefficient, and x_{HII} is the neutral fraction at the overdensity Δ with the attenuation of the local, inhomogeneous ionizing background, Γ , computed using the fitting formula from Rahmati et al. (2013). For further details, interested readers are referred to Sobacchi & Mesinger 2014 (see also Hutter 2018).

3 IONIZING PHOTON NON-CONSERVATION CORRECTION

The excursion set approach, outlined in equation (1) does not explicitly follow the propagation of ionizing photons. Because the diffusion of ionizing photons when H II regions overlap is not properly accounted for, the ionizing photon number is not conserved. This effect has been studied in detail, and depends on the specific implementation of the excursion set algorithm (e.g. Zahn et al. 2007, 2011; Paranjape et al. 2016; Choudhury & Paranjape 2018; Hutter 2018). In the most common implementation described in the previous section, ionizing photons are effectively ‘lost’, resulting in a slower EoR history. This might not be a problem if the ionizing efficiency or escape fraction are treated as effective (i.e. ‘tuning knob’) parameters. However, physically interpreting the constraints on these parameters from observations requires understanding and correcting for the biases caused by ionizing photon non-conservation.

Here, we introduce an efficient method for calibrating excursion set algorithms on-the-fly using an exact analytic solution for a related EoR history. We estimate the (model-dependent) redshift delay resulting from the photon loss, shifting the excursion set EoR history according to this calibration curve. Our approach is motivated by the fact that the EoR topology from the excursion set has been extensively tested and shown to reproduce RT simulations, at a given Q_{HII} (e.g. Zahn et al. 2007, 2011; Mesinger et al. 2011; Hutter 2018). Our approach therefore preserves the EoR topology, shifting only the redshift associated with a given Q_{HII} . Below we describe the algorithm in more detail.

For a given choice of galaxy parameters (c.f. equation 2), we first simulate the reionization history using 21cmFAST including only ionizations by UV sources (i.e. we do not consider recombinations, sinks of ionizing photons referred to as Lyman limit systems, or X-ray preionization at this stage). This EoR history is designated as the calibration curve. We then solve the corresponding evolution of the H II filling factor analytically:

$$\frac{dQ_{\text{HII}}}{dt} = \frac{dn_{\text{ion}}}{dt}. \quad (4)$$

⁴We set $N_{\gamma/b} = 5000$, motivated by a Salpeter initial mass function (Salpeter 1955). Note that this is degenerate with $f_{\text{esc}}(M_h)$, which is treated as a free parameter when performing inference.

¹<https://github.com/21cmfast/21cMMC>

²<https://github.com/21cmfast/21cmFAST>

³The neutral IGM component could be partially ionized by X-ray photons that have long mean free paths. These partial ionizations are followed for each cell by integrating the emissivity back along the light-cone (Mesinger et al. 2011). Furthermore, the ionized IGM component has some residual neutral gas that depends on photo-ionization equilibrium with the local UV background (UVB), accounting also for self-shielding (Sobacchi & Mesinger 2014).

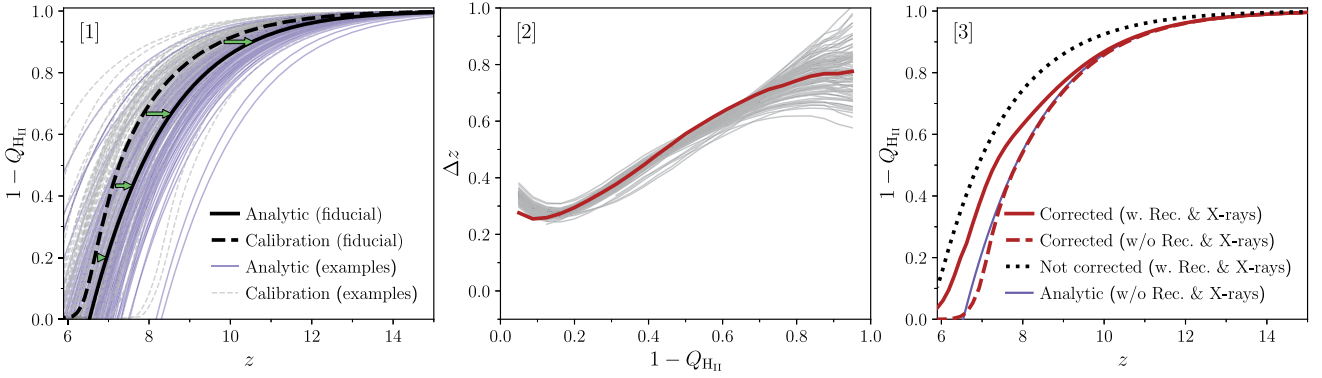


Figure 1. Each panel corresponds to a step of our calibration procedure described in the text. [1]: the EoR history from the analytic expression and the calibration curve (from 21cmFAST); both include only ionizations by UV sources. Thick lines correspond to the fiducial astrophysical parameters (listed in Table 1). Thin solid (violet) and thin dashed (grey) lines represent analytic EoR histories and calibration curves using 100 different combinations of parameters, which are randomly selected from the MCMC chain (see Section 5). Arrows (lime green) illustrate the calibration vector for the fiducial parameter set. [2]: the calibration vector, $\Delta z \equiv z_{\text{analytic}} - z_{\text{calibration}}$ (i.e. Δz corresponds to the size of arrows in panel [1]). Thick solid (red) and thin (grey) lines represent Δz for the fiducial parameter set and the 100 different parameter combinations, respectively. [3]: final EoR histories. The thick solid (red) line represent the corrected EoR history when considering inhomogeneous recombinations and X-rays. The thick dotted line shows the same EoR history, but without photon conservation (i.e. not corrected). We also show the corrected EoR histories when not considering X-rays and inhomogeneous recombinations (thick red dashed line), which is almost identical with the analytic EoR history (thin violet solid line).

Table 1. The medians and 68 per cent C.I. of the eight parameter astrophysical model (see §. 4.1) inferred from a mock 1000h HERA observation, computed with the fiducial parameters and a different initial seed. The likelihood also includes existing observational constraints from UV LFs, the effective optical depth to the CMB and the dark fraction of pixels in QSO spectra.

	Parameters							
	$\log_{10}(f_{*,10})$	α_*	$\log_{10}(f_{\text{esc},10})$	α_{esc}	$\log_{10}(M_{\text{turn}})$ [M_{\odot}]	t_*	$\log_{10}(\frac{L_{X<2\text{keV}}}{\text{SFR}})$ [$\text{erg s}^{-1} M_{\odot}^{-1} \text{yr}$]	E_0 [keV]
Fiducial values	-1.30	0.50	-1.10	-0.40	8.70	0.5	40.50	0.50
Photon conservation	$-1.23^{+0.16}_{-0.20}$	$0.54^{+0.06}_{-0.07}$	$-1.17^{+0.21}_{-0.16}$	$-0.44^{+0.06}_{-0.06}$	$8.74^{+0.17}_{-0.17}$	$0.55^{+0.21}_{-0.19}$	$40.50^{+0.03}_{-0.03}$	$0.48^{+0.02}_{-0.02}$
Without photon conservation	$-1.22^{+0.15}_{-0.18}$	$0.51^{+0.06}_{-0.06}$	$-1.10^{+0.19}_{-0.17}$	$-0.62^{+0.08}_{-0.09}$	$8.87^{+0.13}_{-0.15}$	$0.53^{+0.21}_{-0.17}$	$40.52^{+0.03}_{-0.03}$	$0.48^{+0.01}_{-0.01}$

We assume this analytic expression is the *true* EoR history, given our assumptions.⁵ In Fig. 1 (panel [1]), we show the analytic EoR history (thick black solid) and the calibration curve from 21cmFAST (thick black dashed) for our fiducial parameter set listed in Table 1. We also show the analytic EoR histories and calibration curves (thin grey and violet curves, respectively) for 100 samples of different astrophysical parameters corresponding to the posterior discussed in Section 5.

Second, after obtaining the two EoR histories, we calculate the corresponding redshift calibration vector, $\Delta z(Q_{\text{HII}}) \equiv z_{\text{analytic}}(Q_{\text{HII}}) - z_{\text{calibration}}(Q_{\text{HII}})$, where z_{analytic} and $z_{\text{calibration}}$ represent the redshift of the analytic EoR history from equation (4) and that of the calibration curve from 21cmFAST, respectively.⁶ We show these calibration

vectors in panel [2] of Fig. 1 for the fiducial parameter sets (listed in Table 1, thick red curve) together with the 100 posterior samples (thin grey curves) as in [1]. We see Δz is always positive, confirming that our implementation of the excursion set loses ionizing photons (i.e. has a delayed EoR history relative to the exact analytic solution; c.f. Paranjape & Choudhury 2014). From the figure, we see that even without accounting for photon conservation, the impact on the EoR history is relatively modest: Δz ranges approximately between 0.2 and 1.0. However, even such a relatively modest error can bias our interpretation of the ionizing escape fraction, as we quantify in Section 5. It is also important to note that, despite the wide range of EoR histories shown in panel [1], the calibration curves all cluster together in a narrow region in this $\Delta z - Q_{\text{HII}}$ plane. This shows that the photon non-conservation is a self-similar process when defined as $\Delta z(Q_{\text{HII}})$. The fact that the calibration is mostly dependent on Q_{HII} and only weakly depends on the details of the astrophysics is strongly suggestive that it can be applied more generally to models that include inhomogeneous recombinations for which we do not have an exact analytic solution (as we do in the next step).

Finally, we run the full 21cmFAST simulation, i.e. turning on all the desired physics such as inhomogeneous recombinations and X-ray ionizations. At any redshift, z , 21cmFAST computes all the

⁵Unfortunately, there is no straightforward analytic way of including inhomogeneous recombinations in the exact analytic solution as we cannot a priori determine effective recombination times or subgrid clumping factors. Commonly-used approximations such as constant clumping factors (e.g. Maity & Choudhury 2021) do not capture the interplay of sources and sinks in our subgrid recombination model (c.f. Figure 11 in Sobacchi & Mesinger 2014). Since we have to compare the excursion set and the exact analytic solution using the same assumptions, and iterating on the solution would be prohibitively expensive, we ignore recombinations in this first step.

⁶To minimize numerical effects associated with the end stages of the EoR, where recombinations can be important, we linearly extrapolate the calibration vector Δz above $Q_{\text{HII}} > 0.7$. We also note that the photon non-conservation correction is only applied once individual H II regions start

overlapping (e.g. Friedrich et al. 2011; Furlanetto & Oh 2016), specifically at $Q_{\text{HII}} > 0.05$.

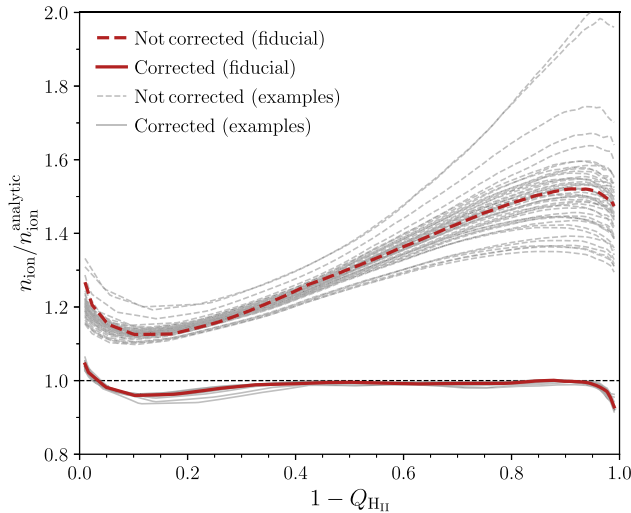


Figure 2. The ratio of the cumulative ionization emissivity from 21cmFAST to the cumulative ionization emissivity from the analytic expression as a function of the mass-weighted ionized filling factor (Q_{HII}). Thick dashed (red) and thick solid (red) lines represent the ratio before and after the correction for our fiducial model, respectively. Thin dashed (grey) and solid (grey) lines represent 50 different parameter combinations. We note that most of thin dashed (grey) lines are overlapping with the thick solid (red) line.

relevant fields (e.g. density, velocity, spin temperature, recombination, etc.). We then apply the redshift calibration vector (c.f. panel [2]) to adjust the excursion set ionization field. Specifically, we calculate the ionization field at an adjusted redshift, $z_{\text{adj}} = z - \Delta z$, linearly evolving the density field to z_{adj} , using the growth factor $D(z)$, i.e. $\delta(x, z_{\text{adj}}) = \delta(x, z) \times D(z_{\text{adj}})/D(z)$. Effectively, this correction amounts to boosting the number of ionizing photons produced by increasing the effective number density of sources. The corrected EoR history is shown in [3] (thick red solid line) of Fig. 1, together with the EoR history without the correction (thick dotted line) and that of analytic EoR history (thin violet solid line).

As we expect, the EoR history with photon conservation (thick red solid line in [3] of Fig. 1) is consistent with that of analytic EoR history at the early stage of reionization, because there are fewer overlapping regions at this stage. As reionization progresses, the corrected EoR history (i.e. with photon conservation) from 21cmFAST shows an offset from that of the analytic EoR history, because the former includes recombinations.⁷ On the other hand, when recombinations are not included, the corrected (thick red dashed line in [3] of Fig. 1) and analytic EoR histories (thin violet solid line) are almost identical, by construction.

Thus so far, we have shown the impact of photon conservation on the evolution of the EoR history, $Q_{\text{HII}}(z)$. Because our calibration is not perfect, the resulting misestimate of Q_{HII} post-calibration will still result in a misestimate of the ionizing photon production

⁷We note that there is a mild ‘knee’ in the EoR history with photon conservation, due to the aforementioned extrapolation of the calibration curve at the late stages of reionization. As can be seen in [1] of Fig. 1, the calibration curve is non-monotonic for the final EoR stages, when multiple bubble overlap starts being significant. Applying this non-monotonic calibration curve to the model that includes recombinations can introduce unphysical features in the EoR history. To minimize this effect, we force the calibration curve to be monotonic, linearly extrapolating below $1 - Q_{\text{HII}} < 0.3$. While there is still a slight knee in some EoR histories, we find this implementation is robust to a wide range of our astrophysical parameter space.

rate of galaxies. In Fig. 2, we quantify this error in the limit of no recombinations from equation (4), showing the ratio of n_{ion} inferred from the EoR history produced by 21cmFAST to the true value $n_{\text{ion}}^{\text{analytic}}$. If the calibration were perfect, this ratio would be unity. As expected, the ratio deviates from unity without photon conservation (dashed lines). On the other hand, with photon conservation (solid lines), the ratio is close to unity. Over the middle stages of the EoR (which are the easiest to detect and thus provide the most constraining power; e.g. Greig & Mesinger 2017a), the ratio is indeed unity with errors of up to a few per cent at the early and late stages. This is a significant improvement compared to the tens – hundred per cent error pre-calibration, as well as the current factor of ~ 10 uncertainty on this quantity (e.g. Bouwens et al. 2015b).

Our new recipe to, approximately, correct ionizing photon non-conservation increases the computing time by only ~ 20 per cent for a full run, including recombinations and spin temperatures. This modest additional overhead means that it can be applied *on-the-fly* for the Bayesian inference. And indeed it was already included in the analysis of recent 21-cm upper limits (Greig et al. 2021b, a; The HERA Collaboration 2021).

4 SETTING UP BAYESIAN INFERENCE

In order to quantify the impact of ignoring ionizing photon non-conservation, we compare the results of Bayesian parameter inference with and without applying the correction discussed in the previous section. Here we summarize the model and observations we will use in this inference.

4.1 The galaxy model

We use the galaxy model of Park et al. (2019), which assumes flexible, empirical scaling relations connecting high-redshift galaxy properties to host dark matter halo masses (see also, e.g. Kuhlen & Faucher-Giguère 2012; Dayal et al. 2014; Behroozi & Silk 2015; Mitra, Roy Choudhury & Ferrara 2015; Mutch et al. 2016; Ocvirk et al. 2016; Sun & Furlanetto 2016; Yue, Ferrara & Xu 2016; Salcido, Bower & Theuns 2019; Hutter et al. 2021). This model can reproduce all current high- z observables, including the rest-frame UV luminosity functions (e.g. Bouwens et al. 2016, 2015a; Oesch et al. 2017), Lyman alpha forest statistics (e.g. Bosman et al. 2018, 2021), and various probes of the EoR history (e.g. McGreer, Mesinger & D’Odorico 2015; Planck Collaboration 2016; Pagano et al. 2020). Furthermore, the galaxy scaling relations are consistent with results from hydrodynamic simulations (e.g. Xu et al. 2016; Ma et al. 2020; Park et al. 2020) and semi-analytical galaxy formation models (e.g. Mutch et al. 2016; Sun & Furlanetto 2016; Tacchella et al. 2018; Behroozi et al. 2019; Yung et al. 2019). We briefly summarize the model here and its free parameters; interested readers can find more details in Park et al. (2019).

We are only concerned with the faint end of the high- z galaxy luminosity functions as these galaxies dominate the photon budget (e.g. Qin et al. 2021). We thus characterize the specific stellar mass with a single power law:

$$\frac{M_*}{M_h} = f_{*,10} \left(\frac{M_h}{10^{10} M_\odot} \right)^{\alpha_*} \left(\frac{\Omega_b}{\Omega_m} \right), \quad (5)$$

where (Ω_b/Ω_m) represents the mean baryon fraction and $f_*(M_h) = f_{*,10}(M_h/10^{10} M_\odot)^{\alpha_*}$ is the fraction of the galactic gas in stars. Here, $f_{*,10}$ and α_* are free parameters and $f_*(M_h)$ is allowed to vary between 0 and 1. Consistent with current observations (e.g. Oesch et al. 2017; Harikane et al. 2018; Bouwens et al. 2021), the star formation

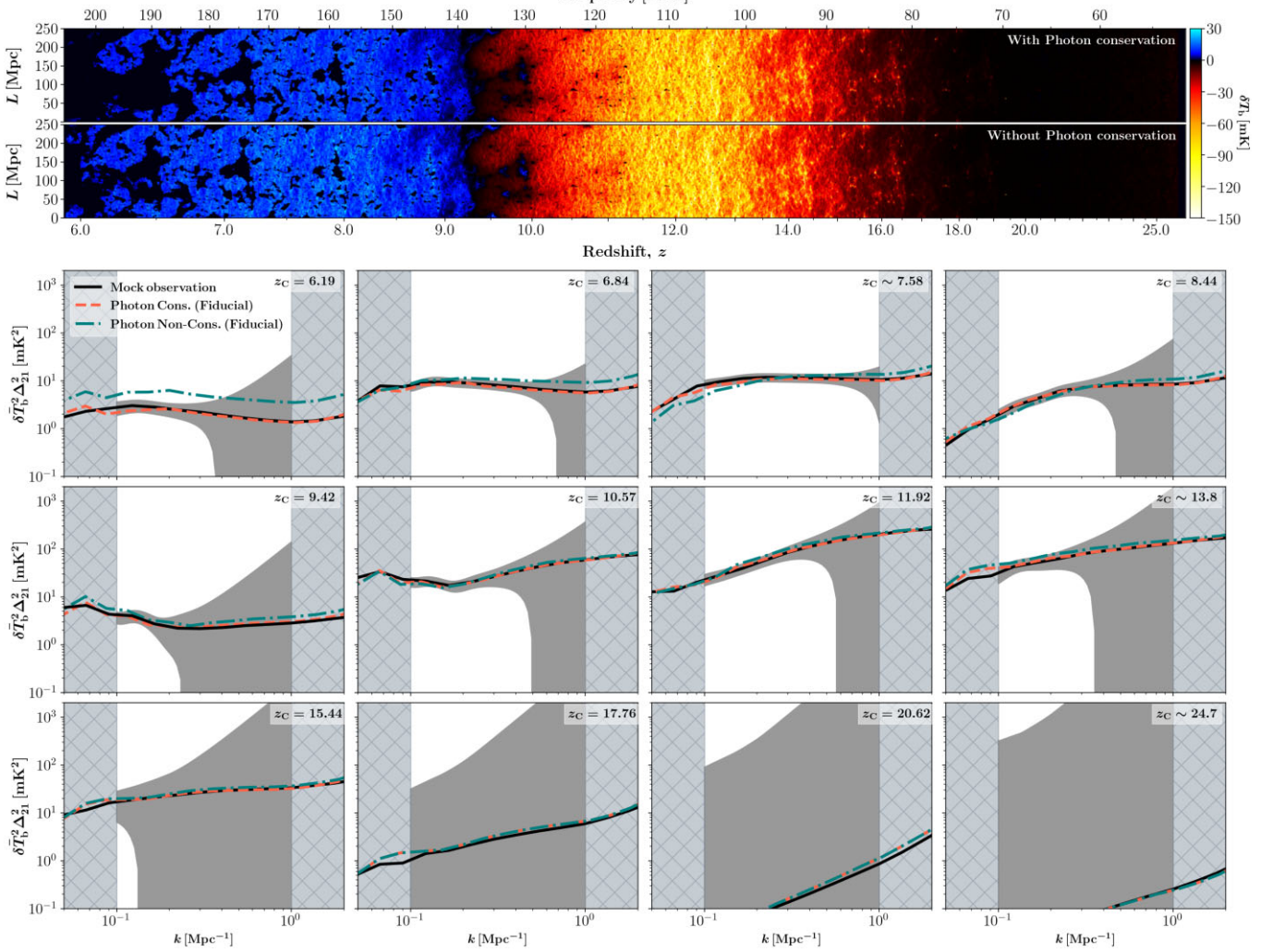


Figure 3. Top panels: 2D light-cone slices with and without photon conservation for comparison. Both light-cones are generated from a box size of 250^3 Mpc^3 using the fiducial parameter set. Bottom panels: the mock 21-cm PS (black solid) calculated from twelve independent ‘chunks’ with 1σ noise assuming a HERA 1000h observation (shaded grey region). In each panel, hatched region outside of $k = 0.1$ and 1 Mpc^{-1} represent excluded regions in our likelihood calculation. Dashed (red) and dot–dashed (green) lines represent PS generated from the box size of 250^3 Mpc^3 with and without photon conservation, respectively. Neglecting ionizing photon conservation delays the EoR and the corresponding 21-cm PS evolution; as we see in the following section, this effect is mimicked by a shift in the α_{esc} parameter.

efficiency is independent of redshift. The star formation rate (SFR) is characterized by a time-scale that scales with the halo dynamical time (or analogously with the Hubble time, $H(z)^{-1}$):

$$\dot{M}_*(M_h, z) = \frac{M_*}{t_* H(z)^{-1}}, \quad (6)$$

where t_* is a free parameter. We account for inefficient cooling and/or feedback processes in low-mass haloes (e.g. Giroux, Sutherland & Shull 1994; Shapiro, Giroux & Babul 1994; Hui & Gnedin 1997; Barkana & Loeb 2001; Springel & Hernquist 2003; Mesinger & Dijkstra 2008; Okamoto, Gao & Theuns 2008; Sobacchi & Mesinger 2013a, b) by including a duty cycle (i.e. fraction of haloes hosting star forming galaxies; c.f. equation 3): $f_{\text{duty}}(M_h) = \exp(-M_{\text{turn}}/M_h)$.

We also characterize the UV ionizing escape fraction, f_{esc} with a power law:

$$f_{\text{esc}}(M_h) = f_{\text{esc},10} \left(\frac{M_h}{10^{10} M_\odot} \right)^{\alpha_{\text{esc}}}, \quad (7)$$

where $f_{\text{esc},10}$ is the normalization and α_{esc} is a power-law index. Such a power-law scaling is consistent with population-averaged relations found by hydrodynamic RT simulations (e.g. Paardekooper, Khochfar & Dalla Vecchia 2015; Kimm et al. 2017; Lewis et al. 2020), although there is currently no consensus on the values of the escape fraction. $f_{\text{esc}}(M_h)$ is also limited to be between 0 and 1.

We account for an inhomogeneous X-ray background by parametrizing the typical emerging X-ray spectral energy distribution of high- z galaxies, using the integrated soft-band ($< 2 \text{ keV}$) luminosity per SFR (in units of $\text{erg s}^{-1} M_\odot^{-1} \text{ yr}$),

$$L_{X<2\text{keV}}/\text{SFR} = \int_{E_0}^{2\text{keV}} dE_e L_X/\text{SFR}, \quad (8)$$

where E_0 is the X-ray energy threshold below which photons are absorbed by their host galaxies. Note that we assume the specific X-ray luminosity follows a power law in photon energy, i.e. $L_X \propto E^{-\alpha_X}$ and take $\alpha_X = 1$, consistent with models of High-Mass X-ray

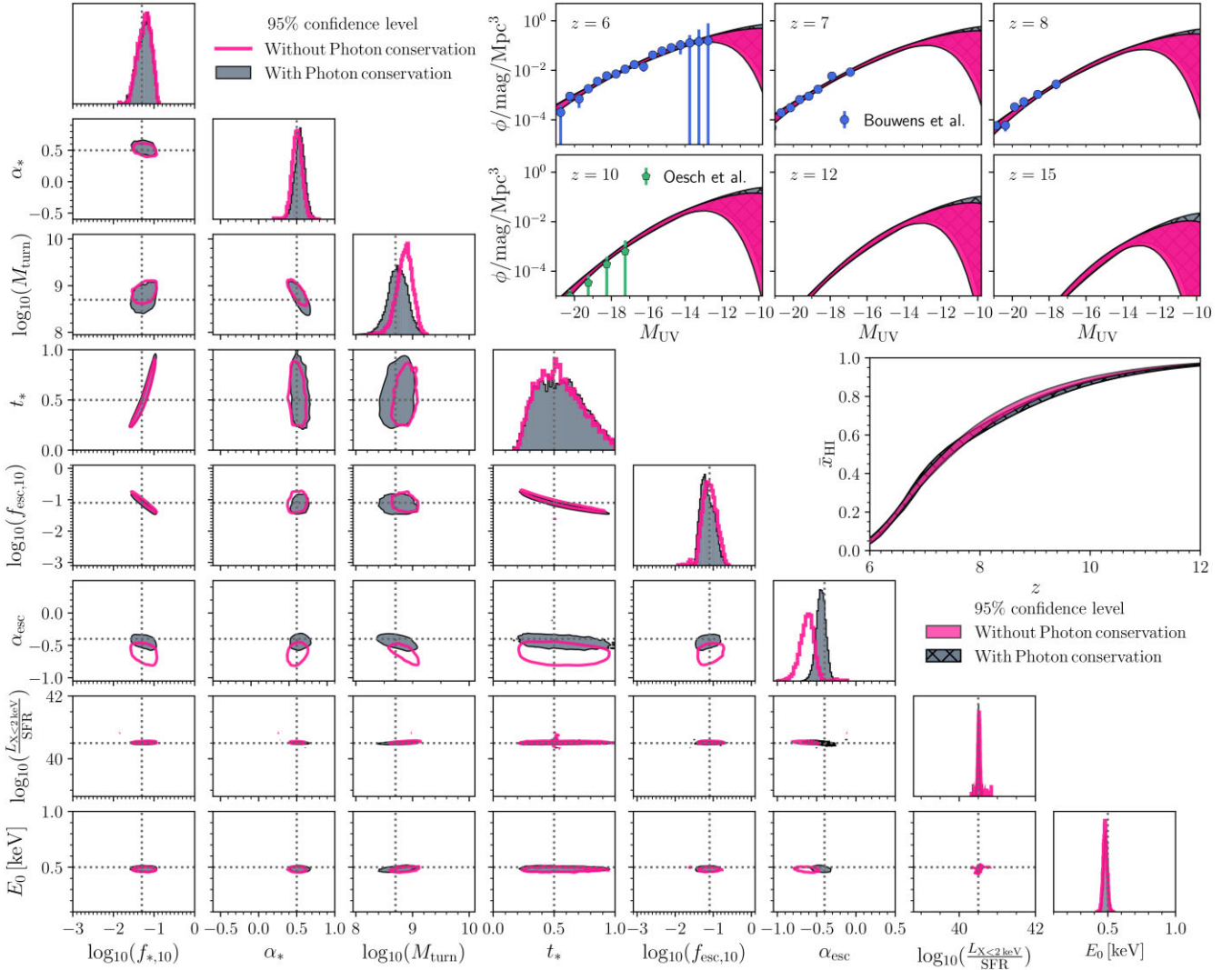


Figure 4. Marginalized posteriors with and without using our photon conservation correction (see the legend). The mock 21-cm signal was generated with photon conservation, using the fiducial parameters denoted by the dotted lines in the corner plot. The bottom left shows the 1D and 2D parameter constraints, while in the top right we show the UV LFs and the EoR history. The shaded regions represent 95 per cent credible intervals C.I.

Binaries in primordial galaxies (e.g. Mineo, Gilfanov & Sunyaev 2012; Fragos et al. 2013; Das et al. 2017). We then compute the ionization and heating rates by integrating the angle-averaged specific X-ray intensity (in units of $\text{erg s}^{-1} \text{keV}^{-1} \text{cm}^{-2} \text{sr}^{-1}$) back along the light-cone in each simulation cell. In a similar fashion, the Lyman series background is computed by integrating back along the light-cone for each cell. For both X-rays and Lyman series, we compute the corresponding attenuation accounting for hydrogen and helium in a multiphase IGM (Mesinger et al. 2011). It is important to highlight that these radiation fields are not computed via excursion set. Excursion set is only applicable for bimodal fields driven by short mean free path photons, like the patchy EoR. Photon conservation for X-ray and Lyman series photons is implicitly insured by integration along the light-cone.

In summary, our galaxy model has eight free parameters:

(i) $f_{*,10}$: the normalization of the star formation efficiency, evaluated for a halo mass of $10^{10} M_{\odot}$. We allow the parameter to vary in the range of $\log_{10}(f_{*,10}) = [-3.0, 0.0]$.

(ii) α_* : the power-law scaling of f_* with halo mass. We allow the parameter to vary in the range of $\alpha_* = [-0.5, 1.0]$.

(iii) $f_{\text{esc},10}$: the normalization of ionizing escape fraction, evaluated for the halo mass of $10^{10} M_{\odot}$. We allow the parameter to vary in the range of $\log_{10}(f_{\text{esc},10}) = [-3.0, 0.0]$.

(iv) α_{esc} : the power-law scaling of f_{esc} with halo mass. We allow the parameter to vary in the range of $\alpha_{\text{esc}} = [-1.0, 0.5]$.

(v) M_{turn} : the halo mass scale below which the abundance of star-forming galaxies is exponentially suppressed. We allow the parameter to vary in the range of $\log_{10}(M_{\text{turn}}) = [8.0, 10.0]$.

(vi) t_* : the star formation time-scale as a fraction of the Hubble time, $H^{-1}(z)$. We allow the parameter to vary in the range of $t_* = (0.0, 1.0]$.

(vii) $L_X < 2 \text{ keV}/\text{SFR}$: the soft-band X-ray luminosity per unit star formation. We allow the parameter to vary in the range of $\log_{10}(L_X < 2 \text{ keV}/\text{SFR}) = [38.0, 42.0]$.

(viii) E_0 : the minimum X-ray photon energy capable of escaping the galaxy. We allow the parameter to vary in the range of $E_0 = [0.1, 1.5]$.

4.2 The data

To compute the likelihood, we use several existing observations. These include: (i) the electron scattering optical depth to the CMB, $\tau_e = 0.059 \pm 0.006(1\sigma)$ (Pagano et al. 2020), (ii) the dark fraction of pixels in QSO spectra, $x_{\text{HII}} < 0.06 + 0.05(1\sigma)$ at $z = 5.9$ (McGreer et al. 2015), and current rest-frame UV LFs at $z \sim 6, 7, 8$, and 10 (Bouwens et al. 2016, 2015a; Oesch et al. 2017). We assume Gaussian or truncated Gaussians for the likelihood shapes, depending on whether the observations are upper limits or not. For more details on these implementations see Park et al. (2019).

In addition to these existing observations, we also make a mock detection of the 21-cm PS. Once detected, the 21-cm signal will provide much more constraining power than current observations. This combination of data will give us the tightest posterior foreseeable in the near future, allowing us to more easily identify the biases introduced by neglecting photon non-conservation.

We compute the mock cosmic 21-cm signal using the latest version (Murray et al. 2020) of 21cmFAST (Mesinger & Furlanetto 2007; Mesinger et al. 2011). The simulation box is 500 Mpc on a side with a 256^3 grid, downsampled from 1024^3 initial conditions.⁸ We use the fiducial parameters listed in Table 1. These are consistent with current observational constraints, as we show in the following section. Note that when creating the mock observation we include our correction of photon non-conservation.

We use the spherically averaged 21-cm power spectra (PS) to characterize the 21-cm light-cone. To calculate these PS, we split the simulated light-cone of our mock signal along the redshift/frequency axis in equal comoving volumes equivalent to the sampled light-cones within 21cmMC following Greig & Mesinger (2018). This results in twelve independent ‘chunks’ for which we compute the 3D PS. We consider a mock observation with HERA (Beardsley et al. 2015) consisting of 331 14-m dishes. To construct the thermal noise and cosmic variance on our mock observation of the PS we use 21CM-SENSE (Pober et al. 2013, 2014). We assume a 1000 hour observation and adopt the moderate foreground removal strategy from Pober et al. (2014). This consists of removing contaminated Fourier modes from within the foreground wedge. We also include a 20 per cent ‘model uncertainty’, crudely accounting for inaccuracies in our numerical implementation (Zahn et al. 2011; Ghara, Choudhury & Datta 2015; Hutter 2018).

In Fig. 3, we show the mock 21-cm observation. The bottom panels correspond to the PS (black solid line) calculated from the twelve independent ‘chunks’ with 1σ noise assuming a 1000 hour observation with HERA (shaded grey region). In each panel, the hatched regions outside of $k = 0.1\text{--}1 \text{ Mpc}^{-1}$ represent excluded regions in our likelihood calculation to avoid foreground contamination and shot noise, respectively. We also show 2D slices extracted from the 3D light-cone (top) and PS (bottom) generated from the smaller 250^3 Mpc^3 box sampled within 21cmMC with and without photon conservation, *computed for the same galaxy parameters*. As expected the differences between two models become larger as reionization proceeds. As we saw in the previous section, correcting for the lost ionizing photons shifts the EoR to earlier times, accelerating its evolution. This shift in timing means that *at a fixed redshift*, the uncalibrated model has a smaller Q_{HII} and smaller H II regions, as well as a slower/delayed PS evolution.

⁸For computational efficiency, we generate the forward-modelled simulations in a volume that is smaller by a factor of 2^3 , but keeping the same resolution of $\sim 2 \text{ Mpc}$. We also use a different initial cosmic seed. See Greig & Mesinger (2015) for more details.

It is important to note that the comparison in Fig. 3 is done for a fixed set of galaxy parameters (the fiducial ones in Table 1). As we will see in the next section, the impact of non-conservation can be compensated by varying the galaxy parameters. Therefore, although one can recover the observations even ignoring photon non-conservation, the corresponding inferred galaxy parameters will be biased (see also Paranjape & Choudhury 2014; Hutter 2018).

We also note that Choudhury & Paranjape (2018) found that large-scale 21-cm power is resolution dependent for excursion set approaches. In practice, any numerical simulation is resolution dependent, because the subgrid recipes themselves depend on the scales that are resolved (recombinations, non-linear densities). Here we briefly compare the 21cm power spectra from our fiducial model (computed on a 128^3 grid) to a lower resolution simulation (computed on a 64^3 grid). We find that, regardless of whether or not photon conservation is used, the difference of large-scale 21-cm power between the two resolution simulations is at the level of \sim few per cent: well within our assumed 20 per cent modeling uncertainty. This is also considerably smaller than the expected observational uncertainty, and so would have a negligible impact on inference.

5 PARAMETER BIAS DUE TO IONIZING PHOTON NON-CONSERVATION

Using our observational data sets, including the mock 21-cm PS, we perform two Bayesian inferences: with and without accounting for photon non-conservation. We use the public MCMC sampler of 3D EoR simulations 21cmMC⁹ (Greig & Mesinger 2015, 2017b; Greig & Mesinger 2018). At each MCMC step, 21cmMC calls 21cmFAST to generate the light-cone of the cosmic 21-cm signal, the UV LFs, as well as other observables.

Our two posteriors are shown in Fig. 4 with grey/red corresponding to with/without photon conservation. The 1D and 2D marginalized PDFs are shown in the lower left, while the UV LFs and inferred EoR histories are shown in the top right. Shaded regions correspond to 95 per cent credible intervals (C.I.). The recovered median and 68 per cent C.I. for all parameters are also listed in Table 1.

In general, we find the biases in the inferred galaxy properties to be modest. The largest impact is in the recovery of α_{esc} , which is underestimated by ~ 0.2 (corresponding to $\sim 2.4\sigma$). This is understandable since in our parametrization, α_{esc} is most responsible for the speed of reionization, given that the star formation rate parameters are constrained by the UV LFs (c.f. Park et al. 2019). Thus a decrease in α_{esc} can mimic the effect of ionizing photon conservation. Indeed we see that the recovered EoR histories are very similar in the two models.

We also find the recovered median value of $\log_{10}(M_{\text{turn}})$ from the model without photon conservation to be larger than that of the model with photon conservation by 1.5 per cent (35 per cent on a linear scale). This parameter also can impact the speed of reionization, though current UV LFs provide an upper limit. However, this bias is very small and the recovered 95 per cent limits are fully consistent between the two posteriors.

6 CONCLUSIONS

While excursion set approaches to generate the 3D ionization structure are well tested and efficient, the ionizing photon number is

⁹<https://github.com/21cmfast/21CMMC>

not conserved. Here, we introduce a new recipe for approximately correcting the photon non-conservation that can be applied on-the-fly within 21cmFAST. Our approach preserves the well-tested reionization topology of excursion set methods, simply shifting the ionizing field in redshift to compensate for photon non-conservation. We find that this calibration is quite insensitive to astrophysical parameters, resulting in shifts of order $\Delta z(Q_{\text{HII}}) \sim 0.1$. It is also fast, facilitating its application to high-dimensional Bayesian inference.

Using current observations and a mock 21-cm detection, we quantify the bias in astrophysical parameter recovery if photon non-conservation is ignored. For our galaxy model and fiducial parameter choice, we find this bias to be very small for most parameters. The major exception is for the power-law scaling of the ionizing escape fraction with halo mass, α_{esc} . Without correcting for photon conservation, the recovered values of the power-law index are biased low by 0.2, corresponding to 2.4σ . However, the EoR histories are largely consistent in the two models, demonstrating that a bias in α_{esc} is sufficient to compensate for ionizing photon non-conservation of excursion set models.

Our approximate correction for ionizing photon non-conservation has been publicly available since v3.0 of 21cmFAST (through the flag `PHOTON_CONS' -> True`).

ACKNOWLEDGEMENTS

JP was supported by a KIAS Individual Grant (PG078702) at Korea Institute for Advanced Study. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Alvarez M. A., Busha M., Abel T., Wechsler R. H., 2009, *ApJ*, 703, L167
 Barkana R., Loeb A., 2001, *Phys. Rep.*, 349, 125
 Beardsley A. P., Morales M. F., Lidz A., Malloy M., Sutter P. M., 2015, *ApJ*, 800, 128
 Behroozi P. S., Silk J., 2015, *ApJ*, 799, 32
 Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, *MNRAS*, 488, 3143
 Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, *MNRAS*, 479, 1055
 Bosman S. E. I., Āurovčiková D., Davies F. B., Eilers A.-C., 2021, *MNRAS*, 503, 2077
 Bouwens R. J. et al., 2015a, *ApJ*, 803, 1
 Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015b, *ApJ*, 811, 140
 Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2016, *ApJ*, 843, 129
 Bouwens R. J. et al., 2021, *AJ*, 162, 47
 Choudhury T. R., Paranjape A., 2018, *MNRAS*, 481, 3821
 Choudhury T. R., Haehnelt M. G., Regan J., 2009, *MNRAS*, 394, 960
 Das A., Mesinger A., Pallottini A., Ferrara A., Wise J. H., 2017, *MNRAS*, 469, 1166
 Dayal P., Ferrara A., Dunlop J. S., Pacucci F., 2014, *MNRAS*, 445, 2545
 Fragos T. et al., 2013, *ApJ*, 764, 41
 Friedrich M. M., Mellema G., Alvarez M. A., Shapiro P. R., Iliev I. T., 2011, *MNRAS*, 413, 1353
 Furlanetto S. R., Oh S. P., 2016, *MNRAS*, 457, 1813

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, *ApJ*, 613, 1
 Geil P. M., Wyithe J. S. B., 2008, *MNRAS*, 386, 1683
 Ghara R., Choudhury T. R., Datta K. K., 2015, *MNRAS*, 447, 1806
 Giroux M. L., Sutherland R. S., Shull J. M., 1994, *ApJ*, 435, L97
 Greig B., Mesinger A., 2015, *MNRAS*, 449, 4246
 Greig B., Mesinger A., 2017a, *MNRAS*, 465, 4838
 Greig B., Mesinger A., 2017b, *MNRAS*, 472, 2651
 Greig B., Mesinger A., 2018, *MNRAS*, 477, 3217
 Greig B., Trott C. M., Barry N., Mutch S. J., Pindor B., Webster R. L., Wyithe J. S. B., 2021a, *MNRAS*, 500, 5322
 Greig B. et al., 2021b, *MNRAS*, 501, 1
 Harikane Y. et al., 2018, *PASJ*, 70, S11
 Hassan S., Davé R., Finlator K., Santos M. G., 2017, *MNRAS*, 468, 122
 Hui L., Gnedin N. Y., 1997, *MNRAS*, 292, 27
 Hutter A., 2018, *MNRAS*, 477, 1549
 Hutter A., Dayal P., Yepes G., Gottlöber S., Legrand L., Ucci G., 2021, *MNRAS*, 503, 3698
 Kim H.-S., Wyithe J. S. B., Park J., Poole G. B., Lacey C. G., Baugh C. M., 2016, *MNRAS*, 455, 4498
 Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, *MNRAS*, 466, 4826
 Kuhlen M., Faucher-Giguère C. A., 2012, *MNRAS*, 423, 862
 Lewis J. S. W. et al., 2020, *MNRAS*, 496, 4342
 Ma X., Quataert E., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2020, *MNRAS*, 498, 2001
 Maity B., Choudhury T. R., 2022, *MNRAS*, 511, 2239
 Majumdar S., Mellema G., Datta K. K., Jensen H., Choudhury T. R., Bharadwaj S., Friedrich M. M., 2014, *MNRAS*, 443, 2843
 McGreer I. D., Mesinger A., D’Odorico V., 2015, *MNRAS*, 447, 499
 Mesinger A., Dijkstra M., 2008, *MNRAS*, 390, 1071
 Mesinger A., Furlanetto S., 2007, *ApJ*, 669, 663
 Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, 411, 955
 Mineo S., Gilfanov M., Sunyaev R., 2012, *MNRAS*, 419, 2095
 Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, *ApJ*, 530, 1
 Mitra S., Roy Choudhury T., Ferrara A., 2015, *MNRAS*, 454, L76
 Molaro M., Davé R., Hassan S., Santos M. G., Finlator K., 2019, *MNRAS*, 489, 5594
 Murray S., Greig B., Mesinger A., Muñoz J., Qin Y., Park J., Watkinson C., 2020, *J. Open Source Softw.*, 5, 2582
 Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, *MNRAS*, 462, 250
 Ocvirk P. et al., 2016, *MNRAS*, 463, 1462
 Oesch P. A., Bouwens R. J., Illingworth G. D., Labbe I., Stefanon M., 2017, *ApJ*, 855, 105
 Okamoto T., Gao L., Theuns T., 2008, *MNRAS*, 390, 920
 Paardekoope J.-P., Khochfar S., Dalla Vecchia C., 2015, *MNRAS*, 451, 2544
 Pagano L., Delouis J. M., Mottet S., Puget J. L., Vibert L., 2020, *A&A*, 635, A99
 Paranjape A., Choudhury T. R., 2014, *MNRAS*, 442, 1470
 Paranjape A., Choudhury T. R., Padmanabhan H., 2016, *MNRAS*, 460, 1801
 Park J., Mesinger A., Greig B., Gillet N., 2019, *MNRAS*, 484, 933
 Park J., Gillet N., Mesinger A., Greig B., 2020, *MNRAS*, 491, 3891
 Planck Collaboration, 2016, *A&A*, 596, A108
 Planck Collaboration et al., 2016, *A&A*, 594, A13
 Pober J. C. et al., 2013, *AJ*, 145, 65
 Pober J. C. et al., 2014, *ApJ*, 782, 66
 Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021, *MNRAS*, 506, 2390
 Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, *MNRAS*, 430, 2427
 Salcido J., Bower R. G., Theuns T., 2019, *MNRAS*, 491, 5083
 Salpeter E. E., 1955, *ApJ*, 121, 161
 Santos M. G., Ferramacho L., Silva M. B., Amblard A., Cooray A., 2010, *MNRAS*, 406, 2421
 Shapiro P. R., Giroux M. L., Babul A., 1994, *ApJ*, 427, 25
 Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
 Sobacchi E., Mesinger A., 2013a, *MNRAS*, 432, L51
 Sobacchi E., Mesinger A., 2013b, *MNRAS*, 432, 3340
 Sobacchi E., Mesinger A., 2014, *MNRAS*, 440, 1662
 Springel V., Hernquist L., 2003, *MNRAS*, 339, 312

- Sun G., Furlanetto S. R., 2016, *MNRAS*, 460, 417
- Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, *ApJ*, 868, 92
- The HERA Collaboration et al., 2022, *ApJ*, 924, 51
- Th  lie E., Aubert D., Gillet N., Ocvirk P., 2022, *A&A*, 658, 139
- Visbal E., Barkana R., Fialkov A., Tselikhovich D., Hirata C. M., 2012, *Nature*, 487, 70
- Xu H., Wise J. H., Norman M. L., Ahn K., O’Shea B. W., 2016, *ApJ*, 833, 84
- Yue B., Ferrara A., Xu Y., 2016, *MNRAS*, 463, 1968
- Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Dav   R., 2019, *MNRAS*, 483, 2983
- Zahn O., Zaldarriaga M., Hernquist L., McQuinn M., 2005, *ApJ*, 630, 657
- Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, *ApJ*, 654, 12
- Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, *MNRAS*, 414, 727

This paper has been typeset from a \LaTeX file prepared by the author.