Bubble mapping with the Square Kilometre Array – I. Detecting galaxies with *Euclid*, *JWST*, *WFIRST*, and ELT within ionized bubbles in the intergalactic medium at $z > 6$

Erik Zackrisson [iD],[1][*] Suman Majumdar [iD],[2] Rajesh Mondal [iD],[3] Christian Binggeli [iD],[1] Martin Sahlén [iD],[1] Tirthankar Roy Choudhury [iD],[4] Benedetta Ciardi,[5] Abhirup Datta,[2] Kanan K. Datta,[6] Pratika Dayal [iD],[7] Andrea Ferrara,[8,9] Sambit K. Giri [iD],[10] Umberto Maio,[11] Sangeeta Malhotra,[12,13] Garrelt Mellema [iD],[10] Andrei Mesinger,[8] James Rhoads,[12,13] Claes-Erik Rydberg[14] and Ikkoh Shimizu[15]

Affiliations are listed at the end of the paper

ABSTRACT

The Square Kilometre Array (SKA) is expected to provide the first tomographic observations of the neutral intergalactic medium at redshifts $z > 6$ and pinpoint the locations of individual ionized bubbles during early stages of cosmic reionization. In scenarios where star-forming galaxies provide most of the ionizing photons required for cosmic reionization, one expects the first ionized bubbles to be centred on overdensities of such galaxies. Here, we model the properties of galaxy populations within isolated, ionized bubbles that SKA-1 should be able to resolve at $z \approx 7$–10, and explore the prospects for galaxy counts within such structures with various upcoming near-infrared telescopes. We find that, for the bubbles that are within reach of SKA-1 tomography, the bubble volume is closely tied to the number of ionizing photons that have escaped from the galaxies within. In the case of galaxy-dominated reionization, galaxies are expected to turn up above the spectroscopic detection threshold of *JWST* and ELT in even the smallest resolvable bubbles at $z \leq 10$. The prospects of detecting galaxies within these structures in purely photometric surveys with *Euclid*, *WFIRST*, *JWST*, or ELT are also discussed. While spectroscopy is preferable towards the end of reionization to provide a robust sample of bubble members, multiband imaging may be a competitive option for bubbles at $z \approx 10$, due to the very small number of line-of-sight interlopers expected at these redshifts.

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

1 INTRODUCTION

In the currently favoured view of galaxy-dominated reionization, large ionized bubbles in the intergalactic medium (IGM) will first appear around overdensities of galaxies, progressively grow, and finally coalesce (for recent reviews, see Loeb & Furlanetto 2013; Barkana 2016; Mesinger 2016; Dayal & Ferrara 2018). Upcoming observations of the redshifted 21 cm signal from the neutral IGM will open a new window on this process, and existing constraints from the high-redshift galaxy luminosity function, from the cosmic microwave background radiation and from quasar absorption systems, can be used to forecast the viable range of 21 cm signals from neutral hydrogen in the reionization epoch (e.g. Kulkarni et al. 2016; Greig & Mesinger 2017a; Hassan et al. 2017; Mirocha, Furlanetto & Sun 2017).

While current interferometers are limited to detecting the 21 cm signal in a statistical sense (for instance the 21 cm power spectrum), phase one of the Square Kilometre Array (hereafter SKA-1) will be able to resolve physical scales down to 5–10 comoving Mpc in the plane of the sky and the corresponding physical scales along the

line-of-sight (frequency) direction at $z \approx 6$–10 (Mellema et al. 2015; Wyithe, Geil & Kim 2015; Datta et al. 2016; Mondal, Bharadwaj & Datta 2018; Mondal et al. 2019). This will for the first time allow tomography (three-dimensional imaging) of the 21 cm signal.

It is already well established that 21 cm data correlated with galaxy surveys can provide powerful constraints on reionization scenarios (Wyithe & Loeb 2007b; Lidz et al. 2009; Wiersma et al. 2013; Park et al. 2014; Hasegawa et al. 2016; Hutter et al. 2016; Sobacchi, Mesinger & Greig 2016; Vrbanec et al. 2016; Hutter, Trott & Dayal 2018). However, most of the studies in this field have focused on the prospects of wide-field (and therefore comparatively shallow) galaxy surveys, whereas relatively little effort has been devoted to the prospects of deep, small-field surveys that focus on uncovering the galaxy content of individual ionized bubbles [but see Beardsley et al. 2015; Geil et al. 2017, for discussions on how to combine MWA/HERA/SKA data with *Wide-Field Infrared Survey Telescope* (*WFIRST*) and *James Webb Space Telescope* (*JWST*) data this way].

The sharpness of the 21 cm profile at the edge of the ionizing region can provide information on the distribution of ionization sources within (e.g. a single quasar versus a spatially extended group of galaxies; Wyithe, Loeb & Barnes 2005; Datta, Bharadwaj & Choudhury 2007; Datta et al. 2008, 2012, 2016; Majumdar et al. 2011; Majumdar, Bharadwaj & Choudhury 2012; Malloy & Lidz

2013; Kakiichi et al. 2017; Giri, Mellema & Ghara 2018) and also on the relative contribution of X-ray and ultraviolet (UV) photons within the bubble (quasar/mini-quasar/high-mass X-ray binaries versus young stars; e.g. Tozzi et al. 2000; Wyithe & Loeb 2007a; Pacucci et al. 2014; Ghara, Choudhury & Datta 2016; Kakiichi et al. 2017).

Here, we will use relatively simple simulations to explore what one can hope to learn by combining an SKA-1 measurement of the dimensions of an individual ionized IGM bubble with a photometric/spectroscopic galaxy survey of its content using upcoming telescopes like the *JWST*, *Euclid*, *WFIRST*, or the Extremely Large Telescope (ELT). By mapping the galaxy content of individual, relatively isolated bubbles, it may be possible to assess the ionizing photon budget within these regions and constrain the role of the galaxies detected inside, in a more direct way than what can currently be done for the photon budget of ionized regions surrounding $z \gtrsim 6$ Lyman α emitters (e.g. Bagley et al. 2017; Castellano et al. 2018; Yajima, Sugimura & Hasegawa 2018), as both the total ionizing photon budget and the contribution from galaxies not exhibiting detectable Lyman α emission within such regions tends to remain ambiguous.

What galaxies are expected inside regions of the Universe that reionize early? Depending on the redshift, size, and isolation of such structures, these regions may be highly biased and could in principle contain galaxies with properties that deviate substantially from those in the average galaxy population at the same redshift. Throughout this paper, we will however adopt the conservative assumption that the galaxies clustered within 21 cm bubbles exhibit higher number densities but properties otherwise identical to those in the field population at the same epoch. This zeroth-order estimate can then serve as a benchmark for more detailed simulations in future efforts.

In Section 2, we explain how mapping the galaxy populations within ionized regions of the IGM at $z \gtrsim 7$ can provide constraints on the role of galaxies in the emergence of these structures. Using seminumerical simulations of galaxy-dominated reionization, we, in Section 3, predict the relation and scatter between the number of ionizing photons emitted from galaxies within a bubble and the resulting volume of that structure, as a function of redshift. The detection limits for galaxies within these structures are explored in Section 4. In Section 5, a number of simplifications adopted in this work are discussed. We also comment on the prospects of using populations of bubble galaxies to constrain early assembly/environmental bias and to place combined constraints on the luminosity function of bubble galaxies and on the time-integrated mean escape fraction of ionizing photons from these objects. Section 6 summarizes our findings.

2 THE PHOTON BUDGET OF IONIZED BUBBLES

Considering a spherical ionized region of comoving radius r and volume $V_{\rm ion} = (4/3)\pi r^3$, and ignoring the effect of recombinations inside it, the relationship between the comoving ionized volume and the total number of ionizing photons $N_{\rm ion, tot}$ that has ever been emitted into the IGM in this region can be expressed as

$$V_{\rm ion} \approx \frac{N_{\rm ion,tot}}{\langle n_{\rm H} \rangle}, \tag{1}$$

where $\langle n_{\rm H} \rangle$ is the average comoving number density of hydrogen atoms in the IGM. In scenarios where star-forming galaxies provide the bulk of ionizing photons required for cosmic reionization, $N_{\rm ion, tot}$ corresponds to the total number of ionizing photons that have ever escaped from galaxies within the bubble.

Equation (1) suggests that if SKA-1 is able to identify individual, highly ionized IGM bubbles and also estimate their volume ($V_{\rm ion}$), it may be possible to place a constraint on the integrated number of ionizing photons $N_{\rm ion, tot}$ emitted from galaxies within this structure. Formally, the $N_{\rm ion, tot}$ constraint inferred from equation (1) will be a lower limit, since a greater number of ionizing photons will be required once recombinations are considered. However, this $N_{\rm ion, tot}$ estimate is, for reasonable assumptions on the IGM clumping factor, expected to be accurate to within a factor of a few (e.g. McQuinn et al. 2007; Sobacchi & Mesinger 2014).

The number of ionizing photons *emitted* by the galaxy population into a specific region of the IGM is determined by the number of ionizing photons *produced*, modulo the escape fraction of these photons. Under the assumption of an invariant stellar initial mass function (IMF), the number of ionizing photons produced is, in turn, related to the total mass in stars produced in this region. However, neither the total mass in stars nor the total number of ionizing photons produced within a region are directly observable. All one can hope to detect is individual galaxies in the bright-end tail of the galaxy population within this volume. In what follows, we will explain how these quantities are related.

While IGM bubbles grow gradually, with galaxies in different mass and luminosity regimes contributing to $N_{\rm ion, tot}$ at different times, a constraint on $N_{\rm ion, tot}$ may nonetheless be converted into a rough estimate on the number of galaxies expected within that bubble at the epoch from which we detect the 21 cm signal. This is possible since the instantaneous, rest-frame 1500 Å UV luminosity $L_{\rm UV}$ (i.e. in the non-ionizing part of the UV; redshifted into the near-infrared at $z > 6$), which traces recent star formation (over the past $\lesssim 10^8$ yr) within a galaxy is predicted to be correlated with the total stellar mass ever formed in that system and in all the progenitors that have merged into it. This stems from the generic simulation prediction that $z > 6$ galaxies *on average* have star formation/accumulation rates that increase over time (e.g. Finlator, Oppenheimer & Davé 2011; Jaacks, Nagamine & Choi 2012; Dayal et al. 2013; Shimizu et al. 2014; Ma et al. 2015; Zackrisson et al. 2017).

The number of ionizing photons $N_{\rm ion, i}$ emitted from a single galaxy i into the IGM over its past star formation history up to the point in time when it is observed ($t_{\rm obs}$) can be expressed as

$$N_{\rm ion,i} = \int_0^{t_{\rm obs}} f_{\rm esc}(t) \dot{N}_{\rm ion}(t) \, {\rm d}t, \tag{2}$$

where $\dot{N}_{\rm ion}(t)$ is the production rate of the number of ionizing photons in this galaxy at time t and $f_{\rm esc}(t)$ describes the temporal evolution of the escape fraction of ionizing photons into the IGM.

If we define $\langle f_{\rm esc} \rangle$ as the $N_{\rm ion}$-weighted mean $f_{\rm esc}$ over the past history of the galaxy, equation (2) simplifies to

$$N_{\rm ion,i} = \langle f_{\rm esc} \rangle \int_0^{t_{\rm obs}} \dot{N}_{\rm ion}(t) \, {\rm d}t. \tag{3}$$

The total number of ionizing photons produced by a whole population of galaxies $N_{\rm ion, tot}$ in a volume $V_{\rm ion}$ can then be derived by integrating over galaxies of all UV luminosities, $L_{\rm UV}$:

$$N_{\rm ion,tot} = \langle f_{\rm esc} \rangle \int_{L_{\rm min}}^{L_{\rm max}} N_{\rm ion,i}(L_{\rm UV}) \Phi(L_{\rm UV}) V_{\rm ion} \, {\rm d}L_{\rm UV}, \tag{4}$$

where $\Phi(L_{\rm UV})$ describes the luminosity function of galaxies in this ionized region (in units of galaxies per volume per unit 1500 Å luminosity) – which is going to have a much higher scaling than the galaxy luminosity function in the field. In this equation, we have for simplicity assumed that all galaxies have the same $\langle f_{\rm esc} \rangle$ (this

assumption is relaxed in Section 5) and that $N_{\mathrm{ion},\,i}$ only depends on L_{UV}.

If we take $N_{\mathrm{ion},\,i}$ to be known then equation (4) indicates how an estimate on $N_{\mathrm{ion,\,tot}}$ (provided by SKA-1, via the bubble volume V_{ion} in equation 1) can be used to constrain the galaxy population ($\Phi(L_{\mathrm{UV}})$) within the bubble and the time-integrated escape fraction of ionizing photons $\langle f_{\mathrm{esc}} \rangle$ from the bubble galaxies. We may, for an individual ionized IGM bubble of a given size, conversely also provide a rough estimate on the number of galaxies that are expected to lie above some UV luminosity detection within this bubble, given an assumption on the relative shape or slope of the galaxy luminosity function within this bubble and on the likely value of $\langle f_{\mathrm{esc}} \rangle$. This allows us to assess the prospects of detecting bubble galaxies with some of the telescopes that are expected to be operational in the SKA-1 era, which we set out to do in the following sections.

In reality, $N_{\mathrm{ion},\,i}$ will vary substantially from galaxy to galaxy of the same observed L_{UV} due to differences in star formation history, metallicity, and dust attenuation, and we will in Section 4 use galaxy spectral energy distribution (SED) models coupled to galaxy simulations in an attempt to quantify the distribution of $N_{\mathrm{ion},\,i}/L_{\mathrm{UV}}$, i.e. the total number of ionizing photons produced over the momentary UV luminosity, and its impact on the relation between galaxy counts and the ionizing photon budget.

3 THE SIZES OF IONIZED BUBBLES

3.1 The smallest ionized bubbles detectable with SKA-1

As shown by e.g. Mellema et al. (2015), Wyithe et al. (2015), and Datta et al. (2016), SKA-1 should be able to identify individual ionized IGM bubbles of angular diameters down to $\gtrsim 5$ arcmin[1] at $z \approx$ 6–10, which corresponds to a spherical bubble radius of $\gtrsim 6$–7 cMpc or a spherical volume of $\gtrsim 1000$ cMpc^3. Ionized bubbles of this size are most readily detected using the matched filtering technique in the Fourier domain proposed by Datta et al. (2007, 2008, 2012, 2016) and Majumdar et al. (2011, 2012). This technique optimally combines the complete three-dimensional 21 cm signal from H I outside the bubble using a matched filter. This method also takes advantage of the fact that noise is uncorrelated in the Fourier domain, whereas it is correlated in the image domain, thereby resulting in a higher signal-to-noise ratio for a given bubble size than methods based on imaging (e.g. Mellema et al. 2015; Kakiichi et al. 2017; Giri et al. 2018) or one-dimensional 21 cm spectra (e.g. Geil et al. 2017).

3.2 Bubble simulations

To explore how tightly coupled V_{ion} may be expected to be to $N_{\mathrm{ion,\,tot}}$, we use a set of seminumerical simulations for reionization, which are identical to those presented by Mondal, Bharadwaj & Majumdar (2017). These simulations involve three major steps: (a) First, we simulate the matter distribution at different redshifts using a publicly available particle mesh N-BODY code[2] and assume that hydrogen follows this underlying matter field; (b) Next, we identify collapsed structures in this matter distribution using a publicly available halo finder[3] based on the Friends-of-Friend (FoF) algorithm (Davis et al.

1985); (c) We then assume a model for the sources of ionization hosted by these collapsed haloes and generate an ionizing photon field using a publicly available seminumerical code.[4] A general assumption in our ionizing source model is that the number of ionizing photons that are produced by these sources is proportional to their host halo mass M_{halo}. We use the constant of proportionality n_{ion} (dimensionless) as a parameter for our simulations. This quantity (also known as the ionization efficiency) combines a number of reionization parameters e.g. the star formation efficiency, the fraction of ionizing photons escaping into the IGM, the number of ionizing photons per baryons produced, etc. For a detailed discussion on this, we refer the readers to section 2.3 of Choudhury, Haehnelt & Regan (2009). Finally, we use this ionizing photon field and the matter density field under an excursion set formalism (Furlanetto, Zaldarriaga & Hernquist 2004) to identify ionized regions within the hydrogen distribution (e.g. Mesinger & Furlanetto 2007; Zahn et al. 2007). Our method of simulating the ionization fields during reionization is similar to that of Choudhury et al. (2009), Majumdar et al. (2014), Mondal et al. (2015), and Mondal, Bharadwaj & Majumdar (2016).

The N-body simulation that we use here has a comoving volume of $V = [215\,\mathrm{cMpc}]^3$, corresponding to $\sim 1.3°$ on the sky for $7 < z < 10$, with a 3072^3 grid of spacing 0.07 cMpc and a particle mass of $1.09 \times 10^8\,\mathrm{M_\odot}$. Thus the smallest dark matter halo that we can resolve is $1.09 \times 10^9\,\mathrm{M_\odot}$ (assuming a minimum of 10 particles required to form a halo). Once we have identified the haloes, we then map the matter and the ionizing photon density fields on a grid which is eight times coarser than our original N-body simulation resolution (i.e. on a 384^3 grid). These coarser fields are then used to implement the excursion set formalism. We identify a grid point as neutral or ionized at a certain stage of reionization, by smoothing and comparing the hydrogen density and the photon density fields using spheres of different radii starting from a minimum radius of R_{min} (the coarse grid spacing) to R_{mfp} (mean free path of the ionizing photons). A specific grid point is considered to be ionized if for any smoothing radius R ($R_{\mathrm{min}} \leq R \leq R_{\mathrm{mfp}}$) the photon density exceeds the neutral hydrogen density at that grid point. For the simulation shown here we have used $n_{\mathrm{ion}} = 23.21$ and $R_{\mathrm{mfp}} = 20\,\mathrm{Mpc}$ (which is consistent with Songaila & Cowie 2010) at all redshifts. These values of the parameters ensure that reionization ends at $z \approx 6$ and we obtain a Thomson scattering optical depth $\tau = 0.057$, which is consistent with Planck Collaboration XLVII (2016). We have used the Planck + WP best-fitting values of cosmological parameters $\Omega_{\mathrm{m}} = 0.3183$, $\Omega_{\Lambda} = 0.6817$, $\Omega_{\mathrm{b}}h^2 = 0.022032$, $h = 0.6704$, $\sigma_8 = 0.8347$, and $n_{\mathrm{s}} = 0.9619$ (Planck Collaboration XVI 2014).
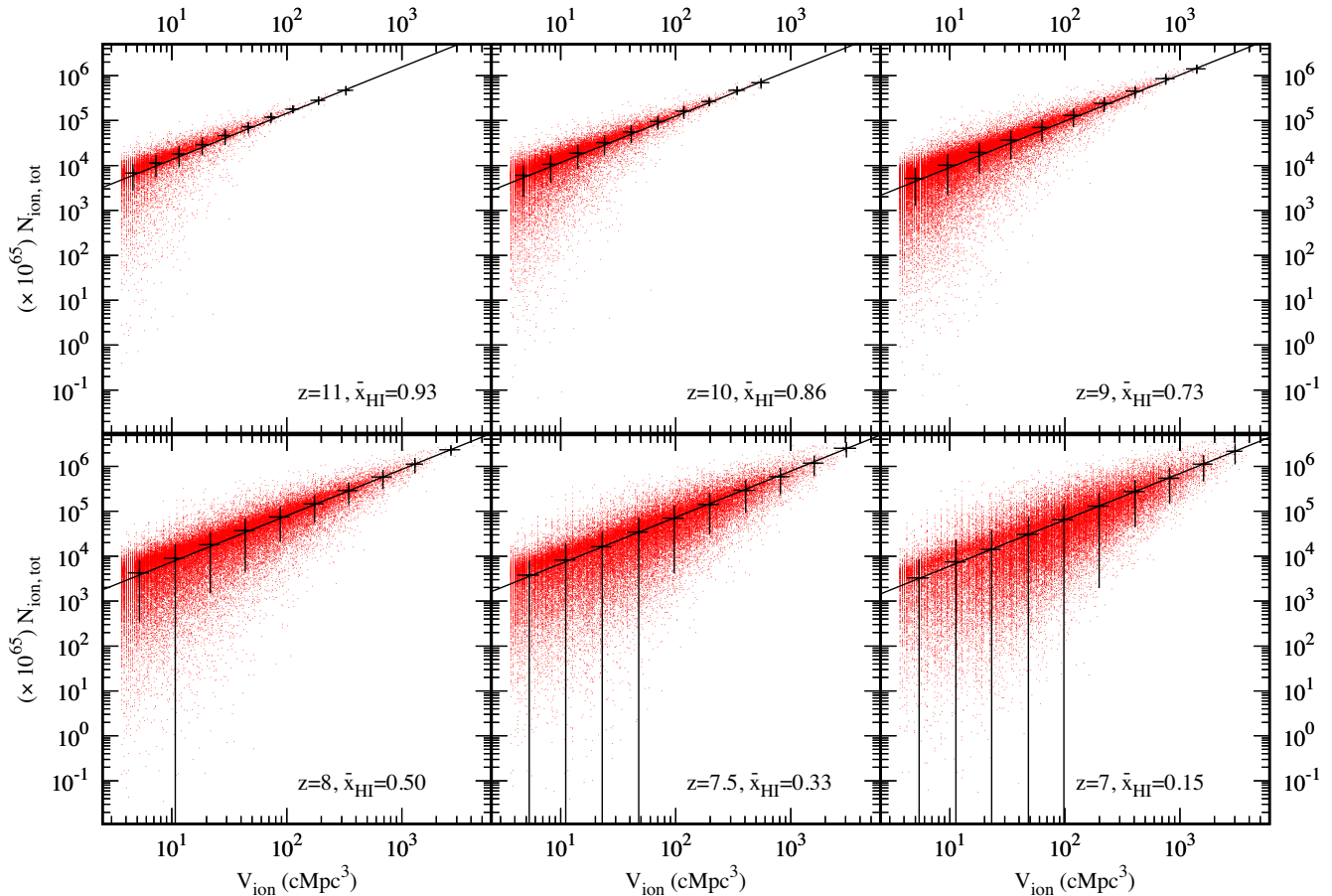
Once ionization maps have been generated at a set of redshifts, we once again make use of an FoF algorithm on these gridded ionization maps to identify individual ionized regions. In this FoF algorithm, we identify any cell having a neutral fraction $x_{\mathrm{HI}} \leq 10^{-4}$ as ionized. This reionization model and numerical machinery results in several tens to hundreds of ionized IGM bubbles above the SKA-1 tomographic limit (volume $\gtrsim 1000$ cMpc^3) at $z \approx 7$–9 within our simulated volume. Rescaling these bubble counts to the volume covered by the planned 100 deg^2 deep SKA1-LOW survey (Koopmans et al. 2015) would result in $\approx 7 \times 10^4$, 1×10^4, and 1×10^3 such bubbles per $\Delta(z) = 1$ at $z \approx 7$, 8, and 9, respectively. Hence, deep surveys with SKA-1 has the potential to detect substantial numbers of such bubbles up to fairly highly redshifts, although we stress that

---

[1]This is similar to the largest SKA1-LOW beam full width at half-maximum that minimizes the point spread function near-in sidelobe noise in full-track mode at 0.13–0.18 GHz

[2]https://github.com/rajeshmondal18/N-body

[3]https://github.com/rajeshmondal18/FoF-Halo-finder

[4]https://github.com/rajeshmondal18/ReionYuga

Figure 1. Relation between the volumes of ionized IGM bubbles V_{ion} at $z = 11$–7 and the number of ionizing photons $N_{ion, tot}$ that have escaped from galaxies and into the IGM within each such region. The black crosses indicate the size of the 1σ scatter in each bin. The different panels feature the neutral IGM fraction of our default reionization scenario along with the best-fitting N_{ion}–V_{ion} relation at each redshift. For bubbles sufficiently large to be resolved by SKA-1 ($V_{ion} \gtrsim 1000$ cMpc^3 in the case of spherical bubbles), the 1σ range in N_{ion} at fixed V_{ion} is always limited to a factor of <4.

the exact numbers would depend on the details of the reionization scenario.

In Fig. 1, we plot the number of ionizing photons $N_{ion, tot}$ that have gone into various ionized bubbles of volume V_{ion} at $z = 7$–11 in our simulations. Due to large density fluctuations on small scales, there is substantial variation (by more than one order of magnitude) in the number of ionizing photons that have been used to produce the smaller bubbles ($V_{ion} \sim 10^1$–10^2 cMpc^3). However, as bubbles approach the SKA-1 resolution limit ($V_{ion} \gtrsim 10^3$ cMpc^3), the effects of density fluctuations tend to even out, leaving a 1σ range that corresponds to a factor of <4 in the required $N_{ion, tot}$ for a fixed comoving V_{ion}. This suggests that SKA-1 measurements of the volumes of ionized bubbles may relatively tight limit on the number of ionizing photons that have been emitted into the IGM within these regions, thereby allowing for constraints on the properties of the galaxy populations within these structures. Section 5.3 features a brief discussion on how these results are affected by different assumptions on the ionization efficiency.

The best-fitting $N_{ion, tot}$–V_{ion} relation varies slightly between the different redshift snapshots, but combining the simulation data from all snapshots with significant numbers of $V_{ion} \gtrsim 1000$ cMpc^3 bubbles

gives the average relation:

$$N_{ion,tot} \approx 8 \times 10^{67} \left(\frac{V_{ion}}{cMpc^3} \right)^{1.03}. \tag{5}$$

We will adopt this relation in the following sections to predict the number of galaxies required to produce a bubble of a given volume.

Note that in an exact inside-out reionization scenario, analytically one would expect the power index in equation (5) to be ~1, which is consistent with the results from our simulations. The scatter in the power-law index from panel to panel of Fig. 1 is mainly due to the spatial fluctuations in the hydrogen number density, clustering of the sources, and non-conservation of the ionizing photon numbers in the later part of the reionization (Choudhury & Paranjape 2018).

4 HOW THE IONIZING PHOTON BUDGET WITHIN IGM BUBBLES IS TIED TO PROPERTIES OF BUBBLE GALAXIES

The results presented in Section 3 suggest that the volumes of the isolated bubbles that SKA-1 will be able to resolve are strongly coupled to the number of ionizing photons that have been emitted throughout the previous history of these regions. In Section 5.3 we also demonstrate that this number is relatively insensitive to how

the production of ionizing photons is distributed across the halo population.

In our fiducial simulations, ionized bubbles of the smallest size that SKA-1 can hope to resolve ($V_{ion} \sim 10^3$ cMpc^3) include $\sim$1000 dark matter haloes of mass $\gtrsim 10^9$ M$_\odot$, which represents a reasonable ballpark estimate of the *total* number of galaxies expected within these structures (but please note that only a very small fraction of these will be sufficiently bright to be detected). However, the exact number of haloes or galaxies needed to produce the required number of ionizing photons will depend on how efficient these are in emitting ionizing photons into the IGM. If the galaxies produce very few ionizing photons (e.g. because of intermittent star formation) or if only a small fraction of the ionizing photons enter the IGM (due to low $\langle f_{esc} \rangle$), then only very extreme matter overdensities, with more haloes and more galaxies, will be able to produce resolvable bubbles. Vice versa, if galaxies are highly efficient in emitting ionizing photons into the IGM, then resolvable bubbles will contain fewer haloes and galaxies.

We note that, under the assumption of an invariant stellar IMF, the $N_{ion, tot}$ parameter is closely tied to the total mass M_{stars} locked up in stars. For the set of simulated galaxies and the spectral evolutionary model adopted in this paper (see Section 4.1), the approximate relation is

$$M_{stars} \approx 1 \times 10^{11} \left(\frac{N_{ion,tot}}{1 \times 10^{71}} \right) \left(\frac{\langle f_{esc} \rangle}{0.1} \right)^{-1} \text{M}_\odot. \tag{6}$$

In principle, this stellar mass could be locked up within a single galaxy, but this would require a very extreme scenario. If we assume $\langle f_{esc} \rangle \leq 0.1$ and consider a galaxy that starts forming stars somewhere in the $z \approx 15$–20 range and manages to do so at a constant star formation rate (SFR) thereafter, then the $\approx 1 \times 10^{71}$ ionizing photons required to produce a $V_{ion} \approx 10^3$ cMpc^3 bubble (equation 5) by $z \approx 10$ would correspond to a total stellar mass of $\geq 1 \times 10^{11}$ M$_\odot$, an SFR ≥ 500 M$_\odot$ yr^{-1}, and a dust-free UV luminosity $M_{UV} \lesssim -25.0$. Such bright, high-mass galaxies are not yet known at $z \gtrsim 8$, and to bring such objects in agreement with the brightest galaxies known in this redshift range (Calvi et al. 2016; Stefanon et al. 2019) would require >2 mag of UV dust attenuation. In this section, we will therefore assume that the ionized IGM bubbles that SKA-1 can resolve contain a population of galaxies, rather than a single object that somehow formed in isolation, and proceed to discuss the details of how the ionizing photon budget of a bubble translates into estimates of the number of detectable galaxies within this structure.

4.1 Past production of ionizing photons tied to the rest-frame UV luminosity

By aiming a telescope with near-IR capabilities (e.g. *Euclid*, *JWST*, *WFIRST*, ELT) at the same area of the sky surveyed by SKA-1 for 21 cm emission at $z > 6$, we can detect the rest-frame UV[5] ($\lambda \gtrsim 1216$ Å) light from galaxies in these structures. Throughout this paper, we will quantify the UV luminosity L_{UV} of $z > 6$ galaxies using the monochromatic luminosity or flux at a rest-frame of 1500 Å. The UV luminosity measured this way reflects the recent SFR over the past $\sim$10–100 Myr (e.g. Boquien, Buat & Perret 2014). For star formation histories stretching over several billions years, as in the case of low-redshift galaxies, this would not be a good proxy for the total stellar mass or the total number of ionizing photons ever produced by this

object, since the prior SFR could have either been much higher or much lower than in the epoch from which we detect its light.

However, simulations of reionization-epoch galaxies generically predict that $z > 6$ galaxies should experience semicontinuous star formation, often with SFRs increasing over time for the more massive ones (e.g. Finlator et al. 2011; Jaacks et al. 2012; Dayal et al. 2013; Shimizu et al. 2014; Ma et al. 2015). Semicontinuous star formation, coupled to the limited time span since the onset of star formation (a few hundred Myr) in the $z > 6$ galaxy population, limits the variations one can expect in the ratio between N_{ion}, the cumulative number of ionizing photons a galaxy has produced in the past (either *in situ* or within smaller galaxies that have merged into this galaxy by the redshift at which it is observed), and L_{UV}. Low-mass galaxies may well experience more stochastic star formation activity (e.g. Mutch et al. 2016; Ma et al. 2018), and consequently larger variations between N_{ion} and L_{UV}, but the greater number density of such objects also means that such variations may largely average out over a bubble population that contains large numbers of galaxies. For the interested reader, Appendix A features a more thorough description of how this N_{ion}/L_{UV} parameter is tied to the prior star formation history.

By combining the Shimizu et al. (2016) simulations for $z = 7$ and $z = 10$ galaxies with the stellar population spectra produced with the STARBURST99 model (Leitherer et al. 1999) under the assumption of the Kroupa (2001) universal IMF and Geneva stellar evolutionary tracks with high mass-loss, Calzetti et al. (2000) dust attenuation, and nebular emission as in Zackrisson et al. (2017), we predict the distribution of N_{ion}/L_{UV} ratios as a function of total stellar mass $M_{stars} \geq 10^{6.5}$ M$_\odot$ in Fig. 2 for $z = 7$ and $z = 10$. For models with a standard stellar IMF, the LyC escape fraction f_{esc} has no significant impact on the 1500 Å luminosity and here it has been set to $f_{esc} = 0$.

As seen in Fig. 2, the N_{ion}/L_{UV} ratio displays galaxy-to-galaxy variations by factors of a few at the highest masses, but varies by more than two orders of magnitude among the lowest mass galaxies resolved ($\log_{10}(M_{stars}/M_\odot) \approx 6.5$) due to large temporal fluctuations in star formation activity within these objects. Such low-mass galaxies are expected to contribute most to the ionizing photon budget within an ionized IGM bubble, but are also present in larger numbers, which means that summing the fluctuating $N_{ion, i}$ contributions for the whole population of bubble galaxies still results in a fairly well-constrained $N_{ion, tot}$.

The mean N_{ion}/L_{UV} ratio (solid line) also evolves slightly with M_{stars} and reaches its highest value for the smallest M_{stars} due to the increasingly stochastic SFRs of such objects. High N_{ion}/L_{UV} ratios are produced by galaxies that have experienced a high SFR in the past, but are observed in a phase when the star formation activity is very low, leading to a near-constant N_{ion} set by the prior activity and a fading L_{UV} due to the ageing stellar population.

To put these N_{ion}/L_{UV} ratios into context, a very bright $M_{UV} \approx -20$ ($L_{UV} \approx 6 \times 10^{40}$ erg s^{-1} Å^{-1}) galaxy at $z = 10$ with $N_{ion}/L_{UV} \approx 6 \times 10^{28}$ photons erg^{-1} s Å would have produced 4×10^{69} ionizing photons over its lifetime, which – by itself – is insufficient (by more than an order of magnitude) to produce an ionized bubble that SKA-1 can resolve (requires $\sim 10^{71}$ ionizing photons) even in the case of $\langle f_{esc} \rangle \approx 1$. For $\langle f_{esc} \rangle \approx 0.1$, it would take ≈ 300 such galaxies to produce a detectable bubble. However, given the shape of the halo mass function or the galaxy luminosity function, it is far more likely that an even larger number of much fainter galaxies is present within these structures.

While it is possible that N_{ion}/L_{UV} ratios even larger than seen in Fig. 2 may be relevant for galaxies below the resolution limit of the simulation used, the N_{ion}/L_{UV} ratio cannot fluctuate without bounds. In the absence of stellar IMF variations, the lower limit

---

[5]In the case of *JWST*, also the rest-frame optical will be within reach.
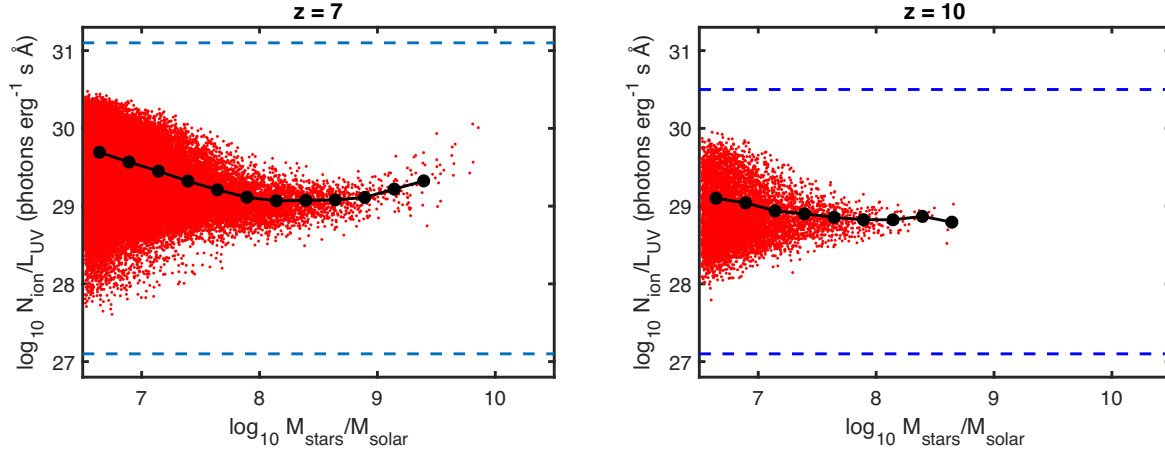
Figure 2. Ratio between the cumulative number of ionizing photons produced by a galaxy and its momentary rest-frame UV 1500 Å luminosity, as a function of its total stellar mass at $z = 7$ (left) and $z = 10$ (right). The red dots represent galaxies from the Shimizu et al. (2016) simulations. The solid lines with filled circles indicate how the arithmetic mean evolves with galaxy mass and the dashed horizontal lines represent the minimum and maximum N_{ion}/L_{UV} ratios theoretically allowed at this redshift.

would be set by a newborn stellar population (age ≈ 1 Myr) which for a STARBURST99, $Z = 0.004$ stellar population of the type adopted here is $\log_{10} N_{ion}/L_{UV} \approx 27.1$ photons erg^{-1} s Å, whereas the upper limit would be set by an instantaneous-burst population (a.k.a. a single or simple stellar population) with an age equal to the age of the Universe. At $z = 7$ and $z = 10$, this limit would be at $\log_{10} N_{ion}/L_{UV} \approx 30.5$ and 31.1 photons erg^{-1} s Å, respectively. These theoretical limits are indicated by dashed lines in Fig. 2.

In the following text, however, we will assume that the N_{ion}/L_{UV} ratio follows a base-10 lognormal function with $\mu \approx 29.34$ (28.89), $\sigma \approx 0.46$ (0.34), and arithmetic means $N_{ion}/L_{UV} \approx 3.7 \times 10^{29}$ (1.1×10^{29}) photons erg^{-1} s Å at $z = 7$ ($z = 10$). While this approach fails to capture the evolution of the mean N_{ion}/L_{UV} ratio with mass evident from Fig. 2, this does not have any substantial impact on the distribution unless assumptions on $\langle f_{esc} \rangle$ places very large weight on galaxies in some particular mass range (see Section 5). At $z = 7$, the mean N_{ion}/L_{UV} ratio varies by a factor of ≈ 4 across the range of galaxy masses considered, and limiting the contribution to the bubble ionization to some very narrow current mass range could in principle alter the result by up to this factor. At $z = 10$, the corresponding factor is ≈ 2. Given the large total number of galaxies ($\gtrsim 1000$; see Section 4.2) that in our model are expected to inhabit the ionized IGM bubbles that SKA-1 can detect, the impact of the mass evolution of the object-to-object scatter in N_{ion}/L_{UV} around the mean is negligible compared to the mass evolution of the mean N_{ion}/L_{UV} itself.

4.2 Total number of galaxies per bubble

To predict galaxy number counts within individual IGM bubbles, we will adopt the simplifying assumption that galaxies within an ionized IGM bubble exhibit higher number densities but are otherwise similar to field galaxies at the same redshift (see Section 5.1 for a discussion on this). We adopt the relative shape of the $z \approx 7$ and $z \approx 10$ UV luminosity functions by Bouwens et al. (2015), extended down to $M_{UV} = -14$, and randomly sample the scatter predicted in the case of the Shimizu et al. (2016) simulation in Fig. 2. We then calculate the number of galaxies necessary to produce the $N_{ion, tot} \approx 1 \times 10^{71}$ photons (equation 5) required to obtain a $V_{ion} \approx 10^3$ cMpc^3 ionized bubble by $z \approx 7$ and $z \approx 10$. Following this procedure, we obtain $N_{galaxies} \langle f_{esc} \rangle \approx 260$ galaxies at $z \approx 7$ and ≈ 1300 at $z \approx 10$. Hence, for

$\langle f_{esc} \rangle \approx 0.1$, we would expect a total of ≈ 2600 galaxies in a bubble resolvable by SKA-1 at $z \approx 7$, and $\approx 13\,000$ galaxies at $z \approx 10$. The value is higher at $z \approx 10$ due to a combination of lower N_{ion}/L_{UV} and differences in the luminosity function. The scatter in N_{ion}/L_{UV} just affects these estimate at the ≈ 10 per cent level compared to adopting a constant N_{ion}/L_{UV} throughout the whole galaxy population.

While we have here adopted $M_{UV} = -14$ as the faint cut-off of the $z = 7$–10 luminosity function, observations of lensed fields have indicated that it may in fact extend several magnitudes fainter than this before turning over (e.g. Bouwens et al. 2017; Livermore, Finkelstein & Lotz 2017). The effect of adopting a fainter cut-off limit would boost the total number of bubble galaxies, thus further reducing the effects of galaxy-to-galaxy scatter in N_{ion}/L_{UV}. For instance, assuming that bubble galaxies are forming down to $M_{UV} = -10$ (while keeping the same luminosity function shape) would boost the *total* number of galaxies by a factor of ≈ 30, but has a much smaller effect on the number of detectable galaxies, as will be demonstrated in the next section. The factor of ≈ 30 is smaller than would be expected from a simple extrapolation of the luminosity function to fainter magnitudes, since this extension alters the ionizing photon flux budget and requires a different absolute scaling of the luminosity function.

4.3 Galaxy detection limits

Only a small fraction of the galaxies present within an ionized IGM bubble ($\lesssim 1$ per cent by number) are likely to appear above the detection threshold of near-IR telescopes within the foreseeable future.

To provide quantitative estimates for the number of detectable galaxies, we consider both photometric detections with *Euclid*, *WFIRST*, and *JWST* plus spectroscopic detections with ELT/MOSAIC and *JWST*/NIRSpec. The pros and cons of these two detection methods are described in more detail in Section 4.5, but the basic difference is that spectroscopic surveys are less prone to line-of-sight interlopers, whereas photometric surveys in principle can probe further down the galaxy luminosity function within the bubble. Below, we describe the various detection limits we consider in the discussion on detectability of bubble galaxies. When assessing the detection of emission lines, we have chosen to be conservative and

therefore ignore the Ly α line. Even though the ionized IGM in SKA-selected bubbles may well allow a favourable transmission factor of Ly α photons through the IGM, scattering and extinction within the galaxies may still render this line very weak for many of these objects.

Euclid[6] is a 1.2 m telescope scheduled for launch in 2022 with optical and near-IR imaging capabilities that can also do 1.1–2.0 micron slitless spectroscopy (resolution $\lambda/\Delta(\lambda) = 250$). While *Euclid* will provide a survey of 15 000 deg², *Euclid* deep fields of about 40 deg² in total will also be observed, with 5σ broad-band detection limits in the optical of $m_{AB} \approx 27$ mag and $m_{AB} \approx 26$ in the *YJH* bands. Galaxy candidates at $z > 6$ can be singled out through drop-out criteria in multiband surveys of this type, by requiring these candidates to be undetected in all filters that sample their spectra at wavelengths shortward of the redshifted Ly α break, yet detected in one or several filters on the longward side of the break. Throughout this paper, we will assume that a sufficient drop-out criterion is met if an object is undetected at the 2σ level shortward of the Ly α break, yet detected at 5σ or more in at least one filter on the other side. For the multiband imaging surveys considered in this paper, we neglect any minor variation in flux detection thresholds among the different near-IR bands, and therefore simply adopt the 5σ limit as the effective drop-out detection threshold ($m_{AB} \approx 26$ in the case of *Euclid*). The line detection limit for *Euclid* is estimated at $\approx 5 \times 10^{-17}$ erg cm^{-2} s^{-1} (Marchetti, Serjeant & Vaccari 2017). For galaxies at $z \gtrsim 7$, *Euclid* can cover lines in the rest-frame UV up to $\lambda \leq 2500$ Å, which basically covers He II (1640 Å), C IV (1549 Å), O III] (1666 Å), C III] (1909 Å). However, these lines are in general expected to remain undetectable at $m_{AB} \approx 26$ mag for a $\approx 5 \times 10^{-17}$ erg cm^{-2} s^{-1} spectroscopic detection limit (e.g. Shimizu et al. 2016), which means that *Euclid* will only detect $z \gtrsim 7$ objects as faint as $m_{AB} \approx 26$ mag through imaging. Hence, we only consider an $m_{AB} \approx 26$ mag photometry threshold of this telescope.

WFIRST[7] is a 2.4 m telescope scheduled for launch in the mid-2020s, which is envisioned to be equipped with imaging capabilities in the 0.48–2.0 micron range and a slitless spectroscopy mode covering 1.00–1.95 micron. *WFIRST* will carry out wide-field surveys (2200 deg² in the high-latitude survey), but also deep field (≈ 20 deg²) observations with expected imaging and spectroscopy detection limits of $m_{AB} \approx 28$ mag and $\approx 1 \times 10^{-17}$ erg cm^{-2} s^{-1}, respectively. Here, we consider only the imaging detection limit, since the line flux detection limit will effectively be much brighter than the $m_{AB} \approx 28$ mag limit (as predicted by the simulations of Shimizu et al. 2016).

James Webb Space Telescope[8] (*JWST*), scheduled for launch in 2021, is a 6.5 m telescope that will be able to do extremely deep imaging and spectroscopy in the 0.6–5 micron range and will hence have access to the rest-frame optical lines from $z \approx 7$–13 galaxies that neither *Euclid* nor *WFIRST* will. The downside is the much smaller field of view, which is 3.6 arcmin × 3.4 arcmin for the *JWST*/NIRSpec spectrograph and 4.4 arcmin × 2.2 arcmin for *JWST*/NIRCam (imaging or spectroscopy), which implies ≈ 2–3 fields to cover just the smallest ionized bubbles that SKA-1 can resolve (≈ 5 arcmin across). If we consider a total of ≈ 20 (≈ 100) h of exposure time to cover a single ionized bubble with either photometry or spectroscopy, we arrive at a point-source detection limit of $m_{AB} \approx 29$ (30) mag for NIRCam imaging in two short-wavelength channel (0.6–2.2 μm) filters across two fields (although

photometry in two long-wavelength channel filters at 2.5–5 μm would also be achieved simultaneously). The corresponding limits for spectroscopy are redshift-dependent, because the most suitable emission line ([O III] (5007 Å) at $z = 7$ is out of *JWST*/NIRSpec range at $z = 10$. Using the Shimizu et al. (2016) model to predict the emission-line strengths for galaxies with a given UV continuum flux, we instead base the $z = 10$ limit on the [O II] (3727 Å) line at $z = 10$. For observing programmes of either ≈ 20 h or ≈ 100 h across two fields in resolution $R \approx 1000$ mode, this results in S/N ≈ 5 line detection limits of 3×10^{-19} erg s^{-1} cm^{-2} or 1.3×10^{-19} erg s^{-1} cm^{-2}, which corresponds to galaxies of $m_{AB} \approx 28.3$ or ≈ 29.2 mag at $z = 7$, but $m_{AB} \approx 26.5$ or $m_{AB} \approx 27.5$ at $z = 10$. The continuum detection limits are approximately the same if the C III] (1909 Å) line is targeted instead of [O II] at $z = 10$.

Extremely Large Telescope[9] (*ELT*) is the largest ground-based optical/near-IR telescope under construction and will have first light around 2025. The currently planned ELT instrumentation does not allow for wide-field imaging, so we do not consider this option in this paper. However, the planned MOSAIC instrument, which is expected to be operational towards the end of the 2020s, is expected to be capable of multi-object spectroscopy at 0.9–1.8 micron over a 7 arcmin diameter field. Hence, MOSAIC can cover the smallest ionized bubbles detected by SKA-1 in just one field and should, in a total of 40 h of observing time, be able to detect rest-frame UV lines at $z \approx 7$ with S/N = 5 at $\approx 1 \times 10^{-19}$ erg cm^{-2} (Evans et al. 2015). By targeting the C III] (1909 Å) line at $z = 7$ and the C IV (1549 Å) at $z = 10$, this corresponds to galaxies with $m_{AB} \approx 29.0$ mag at $z = 7$ and ≈ 28.25 mag at $z = 10$, based on predictions from the Shimizu et al. (2016) models. However, we stress that pre-imaging at this depth will be required to select the spectroscopic targets for ELT/MOSAIC, and this imaging cannot easily be performed by ELT itself given currently planned instrumentation.

4.4 Detectable galaxies per bubble

In Fig. 3, we show the number of galaxies expected above these various detection limits for three different options concerning the time-integrated, photon-number-weighted LyC escape fraction $\langle f_{esc} \rangle$. Here, we have assumed that the bubble galaxies follow a luminosity function with the same relative shape (but different scaling) as the Bouwens et al. (2015) $z \approx 7$ and $z \approx 10$ UV luminosity function, extended down to $M_{UV} = -14$. Under the assumption that $\langle f_{esc} \rangle$ has no mass/luminosity dependence (see Section 5.6 for a discussion on this), the number of galaxies expected above the various detection threshold is given by the differently stripes, for $\langle f_{esc} \rangle = 0.05, 0.1$, and 0.2.

How would these results change if we assume that the luminosity function retains its shape faintward of $M_{UV} = -14$? Our computational machinery indicates that, at fixed $\langle f_{esc} \rangle$, one expects to detect a factor of ≈ 2 fewer galaxies at $z = 7$ (a factor of ≈ 3 at $z = 10$) within an ionized bubble if the luminosity function is extended down to $M_{UV} = -10$. The conversion can simply be done by shifting all the plotted galaxy counts down by this factor. It should however be noted that this parametrization assumes that galaxies all the way down to $M_{UV} = -10$ display $N_{ion, tot}/L_{UV}$ ratios that follow a lognormal distribution with parameters similar to those presented in Section 4.1. However, galaxies as faint as $M_{UV} = -10$ may have total stellar masses as low as $M_{stars} \sim 10^5 \, M_\odot$, which is significantly below the resolution limit of our simulations. The intermittent star
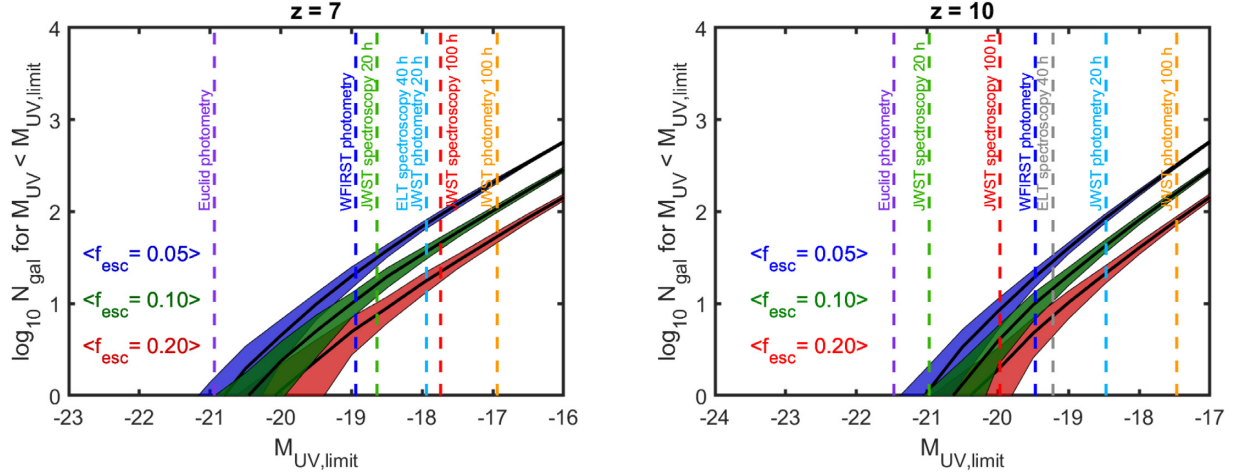
[6]http://sci.esa.int/euclid/

[7]https://www.nasa.gov/wfirst

[8]https://JWST.nasa.gov/

[9]https://www.eso.org/sci/facilities/eelt/

Figure 3. Number of galaxies expected above different limiting rest-frame UV absolute magnitude within a $V_{ion} \approx 1000$ cMpc^3 ionized bubble (set by the requirement that $\approx 1 \times 10^{71}$ ionizing photons need to be emitted into the IGM) at $z = 7$ (left) and $z = 10$ (right), for $\langle f_{esc} \rangle = 0.05$ (blue stripes), 0.1 (green stripes), or 0.2 (red stripes). The width of the stripes is set by the predicted standard deviation in the galaxy number counts between individual bubbles, caused by the random sampling of the adopted luminosity function with faint cut-off at $M_{UV} = -14$. Even in the most pessimistic case ($\langle f_{esc} \rangle = 0.2$), one expects to detect tens of galaxies above the deepest detection limits. However, even the most optimistic predictions ($\langle f_{esc} \rangle = 0.05$) indicate that *Euclid* could well be blind to galaxies in the smallest ionized IGM bubbles that SKA-1 may resolve.

formation episodes expected in such low-mass systems may well cause a shift in the mode of the distribution, which can be explored with higher resolution simulations in the future.

There are a few things to note from Fig. 3. For reasonable values of $\langle f_{esc} \rangle$ (≈ 0.05–0.2), considerable numbers of potentially detectable galaxies are expected within each bubble at both $z = 7$ and $z = 10$, and this number scales with $1/\langle f_{esc} \rangle$, since a higher $\langle f_{esc} \rangle$ means that fewer galaxies are required to provide the ionizing photons needed to form the bubble. Even in the most pessimistic case shown ($\langle f_{esc} \rangle = 0.2$), one expects to detect a handful of galaxies at $z = 7$ with *WFIRST* photometry and several tens of galaxies with either *JWST* photometry programme. Spectroscopy with ELT or *JWST* can also produce $\sim$10 bubble galaxy detections.

At $z = 10$, *JWST* spectroscopy fares somewhat worse than ELT spectroscopy because the intrinsically brighter [O III] line that gave *JWST* an edge at $z = 7$ have been redshifted out of *JWST* range. The pessimistic limits at $z = 10$ places several tens of galaxies above the detection limit of the *JWST* imaging surveys, and a few galaxies above the threshold of either *WFIRST* imaging or ELT spectroscopy.

The detection prospects for *Euclid* are considerably worse, and we conclude that *Euclid* may largely be blind to drop-out galaxies in the smallest ionized IGM bubbles that SKA-1 can resolve.

The difference between the $z = 7$ and $z = 10$ cases is that, due to the lower N_{ion}/L_{UV} at $z = 10$ (see Section 4.1) and a slightly different shape of the adopted luminosity function, a larger number of galaxies is required at $z = 10$ than at $z = 7$ to produce a bubble of a given size, provided that $\langle f_{esc} \rangle$ is kept fixed. The detection limits are also slightly shifted to brighter UV luminosities at $z = 10$.

Because $\langle f_{esc} \rangle$ is here assumed to be independent of UV luminosity (see Section 5 for a discussion on this), only a minor fraction (e.g. ≈ 7–20 per cent for *WFIRST*, but up to ≈ 20–50 per cent for the deepest *JWST* photometry limits) of the ionizing photons that have contributed to the ionization of the IGM in the bubble are accounted for by galaxies above the detection limits in the case where the luminosity function is truncated at $M_{UV} = -14$.

As an alternative to the procedure used to generate the galaxy count predictions of Fig. 3, in which the relative shape of the
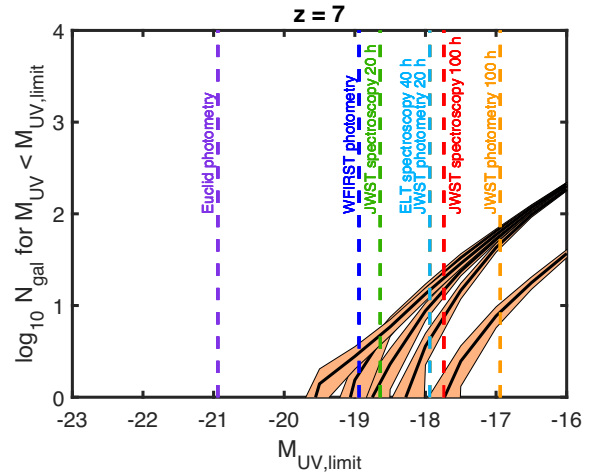
Figure 4. Same as the left-hand panel of Fig. 3, but with predictions based on halo catalogues from five different realizations of ≈ 1000 cMpc^3 bubbles drawn from the seminumerical simulations of Section 3.2, coupled to an analytical recipe for coupling halo mass to UV luminosity. Each orange stripe corresponds to the predicted galaxy counts for one bubble, with the stripe width representing the standard deviation in galaxy number counts stemming from the scatter in the relation between halo mass and UV luminosity. This procedure gives rise to a luminosity function for the bubble galaxies that differs in shape from the one assumed in Fig. 3, with fewer galaxies that would be detectable above the brighter detection limits but mostly a similar number at the faintest detection limit as the predictions for $\langle f_{esc} \rangle \approx 0.2$ in Fig. 3. However, one of the bubbles has managed to reach ≈ 1000 cMpc^3 volume with a significantly lower number of total ionizing photons than the rest, implying much lower galaxy counts.

observationally determined field luminosity functions was used as a basis for populating IGM bubbles with galaxies until a fixed ionizing photon budget had been reached, one may instead start from the halo mass distribution. In Fig. 4, we start from the dark matter haloes predicted within the ≈ 1000 cMpc^3 bubbles predicted by the fiducial reionization simulations of Section 3.2 at $z = 7$ and then adopt the

fitting function presented by Inoue et al. (2018) for the relation (and Gaussian scatter) between halo mass and UV continuum luminosity to attach galaxy fluxes to each halo. If we furthermore adopt the N_{ion}/L_{UV} distribution of Section 4.1, we would arrive at $\langle f_{esc}\rangle \approx 0.15$ for these galaxies. However, the predicted number of bright galaxies ($M_{UV} < -19$) in Fig. 4 is several times lower than would be expected from even $\langle f_{esc}\rangle \approx 0.2$, in Fig. 3, whereas the number of galaxies at the faintest detection levels ($M_{UV} \approx -17$) in most cases are similar to what one estimate for $\langle f_{esc}\rangle \approx 0.15$ through interpolation in Fig. 4. This is primarily because the procedure of relating UV luminosities to halo masses in Fig. 4 results in a luminosity function for bubble galaxies with a different shape than that assumed in Fig. 3, with more very faint galaxies at the expense of bright ones. Moreover, this procedure also gives rise to substantial bubble-to-bubble variations in predicted galaxy counts, due to differences in halo mass distributions within bubbles of similar volume. To illustrate this, Fig. 4 features galaxy predictions for five randomly selected $\approx$1000 cMpc^3 bubbles. One of these bubbles stands out in having reached this volume despite a much smaller combined dark halo mass and ionizing photon production than the rest, resulting in a galaxy population that would only be detectable at the very faintest detection limits considered.

As further discussed in Section 5.1, more sophisticated simulations, meeting observational constraints on the properties of $z \gtrsim 7$ galaxies in overdense regions, will be required to gauge which of the two approaches, Figs 3 or 4, results in the most realistic estimates.

4.5 Photometric or spectroscopic selection?

Bubble galaxies may be identified either in an imaging/photometry survey or through spectroscopy. For a given telescope and a fixed total observing time, photometry will typically reach deeper, but drop-out criteria (or photometric redshifts based on an SED fit from multiband data) have the drawback of not allowing very accurate redshift information. For a typical broad-band drop-out criterion, the redshift error will be $\Delta(z) \approx 1$. This should be compared to the size of ionized IGM bubbles, which for a spherical 1000 cMpc^3 bubble will cover a line-of-sight depth that at $z = 6$–10 will be $\Delta(z) \approx 0.03$–0.06. Hence, an imaging survey runs the risk of misidentifying galaxies located in the foreground or background as bubble members. A spectroscopic survey, on the other hand, would only need a relatively low spectral resolution of $R = \lambda/\Delta(\lambda) \gtrsim 200$ to reach the redshift accuracy required to identify the bubble membership of a given galaxy through the detection of an identified emission line.

How substantial is the risk of misidentifications in a photometric survey? This depends on the redshift of the bubble targeted. From the detection limits in Fig. 3, we see that the number of detectable galaxies above a fixed luminosity, within a bubble of fixed size, changes by no more than a factor of $\approx$3 between $z \approx 7$ and $z \approx 10$ if a constant $\langle f_{esc}\rangle$ is assumed. At the same time, the ambient number density of galaxies above a certain threshold luminosity drops by an order of magnitude between these redshifts (Bouwens et al. 2015). This leads to a situation (schematically illustrated in Fig. 5) where the number of interlopers in a drop-out survey towards a given bubble will be much greater towards the end of reionization ($z \approx 7$) than at earlier stages ($z \approx 10$).

Another way to understand this is to note that ionizing a $\sim$1000 cMpc^3 volume requires a similar ionizing photon budget, and hence a similar collapsed mass, at both redshifts (Fig. 1). However, a much higher peak in the density field is needed for such a structure to collapse and form stars by $z \approx 10$ than at $z \approx 7$. Such high peaks are correspondingly rarer, and their overdensity compared to their environment is much greater. For a bubble with a line-of-
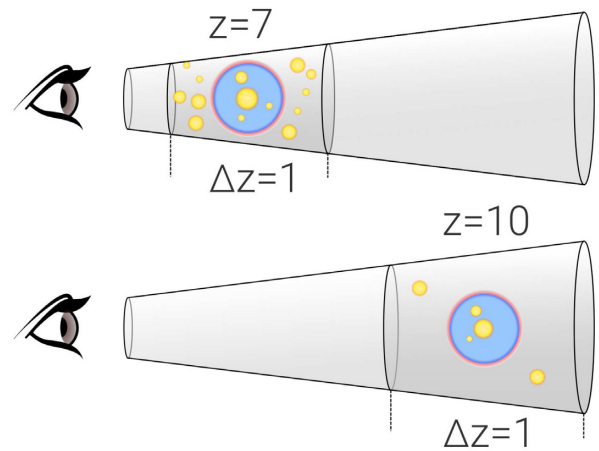
Figure 5. Schematic figure illustrating the issue of interlopers in a drop-out survey (here assumed sensitive to galaxies within a redshift interval of $\Delta(z) = 1$) towards an ionized bubble at either redshift $z = 7$ or $z = 10$. The number of galaxies inside an ionized bubble of fixed size detectable by an imaging survey are expected to change just by factors of a few between $z \approx 7$ and $z \approx 10$, whereas the number of interlopers in the line-of-sight $\Delta(z) = 1$ volume is expected be significantly higher at lower redshift. Hence, a photometric survey becomes less risky when aimed at a bubble at higher redshift (here $z \approx 10$) than when aimed at a bubble at the end of reionization ($z \approx 7$).

sight depth of $\Delta(z) \approx 0.05$, the volume probed by a broad-band imaging survey (line-of-sight resolution $\Delta(z) \approx 1$) at $z = 7$–10 will be more than 20 times larger than that of the bubble, but the average galaxy number density in this volume is also likely to be much lower, since the overdense regions tend to be the first to reionize. If we adopt the cosmic average for the number density of galaxies in the line-of-sight volume outside the bubble, the Bouwens et al. (2015) luminosity function predicts that there should at $z \approx 7$ be $\approx$25 interloper galaxies at $M_{UV} \leq -19.5$ in the $\Delta(z) \approx 1$ cylindrical volume projected against a 1000 cMpc^3 bubble. This is larger than the number of bubble galaxies for all the $\langle f_{esc}\rangle$ cases considered at $z = 7$ in Fig. 3. However, at $z = 10$, the number of interloper galaxies at $M_{UV} \leq -19.5$ (approximately the *WFIRST* photometry limit at this redshift) is $\approx$1, which means that photometric interlopers within the $\Delta(z) = 1.0$ volume selected by drop-out criteria are unlikely to be a problem, since the number of bubble galaxies predicted in Fig. 3 is expected to be several times higher than this.

A consistency check of this conclusion, based on halo statistics instead of luminosity functions, is presented in Appendix B.

We caution, however, that a purely photometric survey may also be prone to interlopers from much lower redshifts than indicated by the $\Delta(z) = 1$ volume considered above, because very strong optical emission lines can give the appearance of a Lyα break, unless the selection is based on detection on several filters shortward of the break.

5 DISCUSSION

5.1 Galaxy assembly bias in overdense regions

Throughout this paper, we have – as a first baseline approach – adopted the assumptions that the relative shape of the galaxy luminosity function at $z \geq 7$ is independent of environment, that the ionized bubbles of the IGM simply feature a scaled-up version of the field luminosity function at the same redshift, and that there are no significant differences between the properties of galaxies residing

in a matter overdensity and those in the field. However, overdense regions that reionize early are, by definition, not typical and may well have galaxy populations quite different from the cosmic average at this epoch.

At low redshifts, it is well established that the luminosity function and the ratio of red- to blue-sequence galaxies change with environment (e.g. McNaught-Roberts et al. 2014). Changes in galaxy properties with environment have observationally been traced up to $z \approx 3$ (e.g. Grützbauch et al. 2011), but exactly how early such differences get imprinted in the galaxy population remains an open question (for a review, see Overzier 2016).

In simulations, both the shape of the dark halo mass function and the properties of individual haloes of a given mass (in terms of accretion rate, spin, concentration, and shape) are predicted to be affected by the overdensity of the environment (Lee et al. 2017), and increased merger rates, galaxy interactions, and the feedback from ionizing radiation produced within overdense regions may further augment changes in galaxy properties compared to the field population.

A common expectation is that feedback from an ultraviolet background may quench star formation in low-mass dark matter haloes (e.g. Mesinger & Dijkstra 2008; Sobacchi & Mesinger 2013; Maio et al. 2016). In an ionized bubble, this could potentially alter the shape of the galaxy luminosity function at the faint end, or alternatively affect the typical ionizing emissivity of faint galaxies. To first order, this can be treated as an effective truncation of the luminosity function (below which galaxies do not contribute ionizing photons) similar to the different faint luminosity function extensions that we have considered in this paper (M_{UV} limit –14 to –10). However, if external feedback also affects the star-forming properties of massive galaxies (for a scenario of this type, see Susa 2008), then changes to the shape of the luminosity function at the bright end may also occur. Assembly bias (the notion that the statistical properties of galaxies depend on properties other than halo mass) could also manifest itself in other, more complex ways. Indeed, some simulations have indicated differences in specific SFRs, galaxy mass functions, metallicities, and dust content in overdense regions compared to field at $z \geq 6$ (Yajima et al. 2015; Sadoun et al. 2016).

As samples of $z > 7$ galaxies grow larger, it should be possible to observationally test for such environmental effects by, for instance, studying the slope of the bright end of the luminosity function as a function of clustering. A couple of notable overdensities of galaxies at $z \gtrsim 7$ have already been discovered in deep *HST* surveys – a $z \approx 7$ overdensity of 17 Lyman break galaxy candidates, out of which three are confirmed Lyman alpha emitters (two with consistent redshifts) in an area a few arcminutes across (Castellano et al. 2016, 2018), and a $z \approx 8.4$ overdensity of up to eight Lyman break galaxy candidates, out of which one is a confirmed Lyman alpha emitter, in an area just a 10 arcsec across (Ishigaki, Ouchi & Harikane 2016; Laporte et al. 2017). Provided that a fair fraction of these Lyman break galaxies are at the redshift of Lyman alpha emitters, the brightness distribution of Lyman break galaxies in these regions is roughly consistent with what our models predict for ionized regions of a scale that SKA-1 should be able to resolve. Recently, an overdensity of 12 Lyman α emitters at $z \approx 6.6$ was discovered by Harikane et al. (2019), but these appear to cover a volume that is considerably larger than covered by our simulations. By targeting such overdensities with deep photometric and spectroscopic *JWST* observations it may be possible to observationally constrain feedback and environmental effects within some of the most extreme overdensities in the reionization epoch. A few years down the line, many more such overdensities are also expected in to be uncovered

in *WFIRST* deep field observations, which are expected to cover ~100 times the solid angle of correspondingly deep *HST* surveys.

In future efforts, it would be worthwhile to study the properties of simulated galaxies drawn directly from ionized IGM regions of reionization simulations to study the effects that assembly bias and galaxy feedback are expected to have on galaxy counts within such structures. Recently, Geil et al. (2017) used such simulations to study how the sizes of ionized bubbles correlate with the luminosity of the brightest galaxy within the bubble. Their study focuses on ionized IGM bubbles at $z \approx$ 9–11 that are factors of a few smaller in radius than the SKA-1 tomographic resolution limit than we have adopted, but if we apply their best-fitting relation between bubble radius and brightest bubble galaxy at $z \approx 10$ to bubbles of radius 2–3 cMpc (similar to their mean bubble size at this redshift) to derive the UV luminosity of their brightest galaxy in these structures, and then scale the number of such galaxies ($M_{\mathrm{UV}} = -18.3$ to -19.5) up by factors of $\approx$9–30 to match the volume of our $\approx$1000 cMpc^3 bubbles, we get a good match with our $\langle f_{\mathrm{esc}} \rangle = 0.2$ predictions for this luminosity range at $z = 10$ in Fig. 3. While this indicates that there is agreement for the *average* bubble, the extreme tail of their distribution contains galaxies that are much brighter than any of our predictions in Fig. 3 – for example, their very largest bubble (radius $\approx$4.8 cMpc) at $z \approx$ 11 contains a brightest galaxy that attains $M_{\mathrm{UV}} = -22.5$ (albeit under the assumption of zero dust attenuation) which seems to completely dominate the UV luminosity of its bubble (as seen in their fig. A1), violate our assumption of a relatively well-sampled bubble luminosity function and lie closer to the monolithical galaxy case discussed in Section 4. Whether even rarer and more extreme objects could exist that produce individual ionized IGM bubbles sufficiently large to be detectable with SKA-1 warrants further study.

5.2 Population III stars and active galactic nuclei

Our current treatment assumes that ionizing photons from young stars represent the only substantial contribution to the ionizing photon budget within 21 cm bubbles. However, if objects or mechanisms other than star-forming galaxies and their associated Lyman continuum radiation contribute to bubble growth, this could of course affect the outcome.

Population III stars are expected to emit a large fraction of their radiation at Lyman continuum energies, owing both to high stellar surface temperatures at zero metallicity and potentially a top-heavy stellar IMF (e.g. Schaerer 2002). However, Population III stars in minihaloes are only likely to dominate the cosmic SFR density at $z \gtrsim 15$ (e.g. Maio et al. 2010) and are not expected to play any major, global role during late stages of cosmic reionization (e.g. Kulkarni et al. 2014). Even so, one could envision such stars to have a significant effect *locally*, in regions of delayed and highly concentrated Population III star formation.

It has been argued that Population III galaxies may form in $z < 15$ H I cooling haloes that have managed to remain chemically pristine (Stiavelli & Trenti 2010), but current simulations suggest that such systems are unlikely to attain total stellar masses higher than $\sim$10^6 M$_\odot$ (e.g. Yajima & Khochfar 2017; Inayoshi, Li & Haiman 2018). Based on the Population III galaxy SED models of Zackrisson et al. (2011), such systems cannot, over the expected lifetimes of $\sim$10^6–10^8 yr, produce more than $\sim$10^{70} ionizing photons even under the assumption of an extremely top-heavy stellar IMF (typical stellar mass $\sim$100 M$_\odot$). Since $\sim$10^{71} photons are required to produce a bubble detectable with SKA-1, such rare, exotic galaxies cannot contribute significantly to the photon budget, even if they exhibited extreme Lyman continuum leakage ($f_{\mathrm{esc}} \approx 1$).

This leaves black hole accretion as the most likely mechanism to rival the ionizing flux from star-forming galaxies within bubbles close to the SKA-1 resolution limit. A single quasar can easily produce $\sim 10^{71}$ ionizing photons in as little as $\sim 10^7$ yr (e.g. Maselli et al. 2007). An active, high-luminosity quasar would usually be readily identifiable as such from spectroscopy, but an accreting black hole that has contributed early on and then has turned dormant may not be. Is it then possible that SKA-1 may detect ionized IGM bubbles that appear completely devoid of galaxies when probed with upcoming near-IR telescope, because the quasar responsible for the bubble is in an inactive phase? This could potentially happen for the most shallow limits considered in Fig. 3 but seems unlikely for the deepest ones. Using scaling relations from Wood & Loeb (2000), a supermassive black hole radiating at the Eddington luminosity for $\sim 10^7$ yr with $f_{esc} = 1.0$ would need to have a mass of $\sim 10^8 \, M_\odot$ to seriously impact the ionizing photon budget of a 1000 cMpc^3 bubble. For a bulge-to-supermassive black hole mass ratio of $\lesssim 0.1$ at $z \leq 10$ (Targett, Dunlop & McLure 2012), this requires a host galaxy with a stellar mass of at least $\gtrsim 10^9 \, M_\odot$, which is expected to have a luminosity $M_{UV} < -19$ and therefore be in the detectable range of many of the bubble surveys considered in Fig. 3. Moreover, statistics on galaxy counts within a few bubbles of similar size can put constraints on scenarios including such transient, stochastic contributions to bubble growth; and the sharpness and morphology of the 21 cm profile of the bubble may also reveal contributions from X-ray photons that could be inconsistent with a normal stellar population (e.g. Tozzi et al. 2000; Wyithe & Loeb 2007a; Pacucci et al. 2014; Ghara et al. 2016; Kakiichi et al. 2017).

5.3 The ionization efficiency of haloes

The fiducial reionization simulations used to produce $N_{ion, tot}$–V_{ion} predictions of Fig. 1 are based on the assumption of a fixed ionization efficiency n_{ion} for all haloes above a minimum halo mass of $M_{halo, min} = 1.09 \times 10^9 \, M_\odot$. Raising (lowering) the adopted ionization efficiency alters the reionization history by shifting the completion of reionization to higher (lower) redshifts (see fig. 1 of Greig & Mesinger 2017b), but this effect can be offset by simultaneously raising (lowering) $M_{halo, min}$. Through this degeneracy, current observational constraints on cosmic reionization are consistent with ionization efficiencies that differ by an order of magnitude (Greig & Mesinger 2017a; Monsalve et al. 2018).

One could perhaps suspect that our results concerning the connection between the volume (V_{ion}) of ionized bubbles and the number of ionizing photons produced therein ($N_{ion, tot}$) would strongly depend on the overall scaling of n_{ion} or its halo mass dependence. However, while changing the relation between n_{ion} and M_{halo} does affect the reionization history of the simulations, we find that the bubble population remains strongly correlated with the overall ionization state of the Universe [i.e. the global neutral fraction ($\bar{x}_{HI}$) quoted in different panels of Fig. 1 for our standard parameter set], and that both the average $N_{ion, tot}$–V_{ion} relation and the scatter around this relation remains almost the same for $V_{ion} \geq 1000$ cMpc^3 despite rather substantial alterations of the relation between n_{ion} and M_{halo}. Specifically, by redistributing the production of ionizing photons across the halo population, while still retaining the same reionization history ($\bar{x}_{HI}(z)$) as in Fig. 1, we find that the scatter at $V_{ion} \geq 1000$ cMpc^3 is insignificantly altered when either the minimum halo mass for ionizing photon production is raised to $M_{halo, min} = 10^{10} \, M_\odot$, when a random uniform scatter that corresponds to a factor of 20 variation in ionizing efficiency is added to each halo, or when the

relation between the ionizing photons produced by a halo($\propto M_{halo}^n$) is altered from $n = 1$ to $n = 1.41$.

5.4 The impact of small-scale density variations

A shortcoming in the seminumerical simulations that we have used to obtain the relation between V_{ion} and $N_{ion, tot}$ in Section 3 is that this machinery does not take small-scale spatial variations of hydrogen density into account when calculating the recombination rate. Instead, the recombination rate is assumed to be uniform throughout the IGM (independent of density). This ignores the effect of the clumping factor and hence the relatively rapid recombinations that may take place in high-density, non-star-forming regions known as damped Lyman α absorbers (DLA) or Lyman limit systems (LLS). These DLAs and LLSs are expected to be dense enough to self-shield themselves from the ionizing background created by the galaxies and quasars and work as the sinks of ionizing radiation within an ionized bubble. Under the standard model of structure formation in our Universe, these systems are expected to be more abundant than the haloes that are capable of hosting galaxies that we assume produces majority of the ionizing photons. A failure to properly consider such small-scale, high-density systems therefore likely leads us to underestimate the $N_{ion, tot}$ required to produce bubbles of a given volume V_{ion}.

To accurately take into account the effects of these absorbers when predicting the ionization topology and the corresponding ionizing photon budget would require reioniziation simulations spanning a very large dynamic range. On one hand they should have mass resolutions of the order of Jeans scales to model the recombination processes inside these sub-Mpc objects and on the other hand they should be able to simulate large volumes (of the order of Gpc) to be able produce the bubble size distribution and the corresponding large-scale fluctuations in the signal. Recent results from theoretical modelling (Schaye 2001; Kaurov & Gnedin 2015), high-resolution but sub-Mpc size simulations (Park et al. 2016) and their sub-grid adaptation in the large-scale seminumerical simulations (Sobacchi & Mesinger 2014) suggest that non-uniform recombinations in the IGM would cost two to three or more photons per ionized hydrogen atom by the end of the reionization era. A direct and obvious implication of this on the $N_{ion, tot}$ versus V_{ion} plot shown in Fig. 1 will be an increment in the amplitude of the power-law fit at all stages of reionization. This would also increase the scatter in the plot at the low V_{ion} end of the plots. However, as we approach $V_{ion} \sim 1000$ Mpc^3 (our adopted SKA-1 detection limit), one would expect this scatter to die down substantially due to the effect of averaging over large volumes. This should be explored more carefully in future simulations.

5.5 Multiple ionizations from each ionizing photon

Our current treatment assumes that each hydrogen-ionizing (Lyman continuum) photon emitted from stars will ionize exactly one hydrogen atom. However, ionized gas may itself emit ionizing photons through free–bound and free–free transitions, effectively resulting in multiple ionizations from a single stellar Lyman continuum photon. The overall impact of this depends on the shape of the ionizing stellar continuum and the gas temperature, but is not expected to boost the effective ionizing emissivity of galaxies by more than at most a factor of 1.6 under realistic conditions (Inoue 2010).

5.6 The escape fraction of ionizing photons

Throughout this paper, we have assumed the same $\langle f_{esc} \rangle$ for all bubble galaxies, but one can also envision this parameter evolving as a function of UV luminosity, halo mass, or stellar mass (e.g. Yajima, Choi & Nagamine 2011; Kimm & Cen 2014; Paardekooper, Khochfar & Dalla Vecchia 2015; Xu et al. 2016; Sharma et al. 2017). This will affect the number of galaxies required to emit a fixed number of ionizing photons into the IGM, but the detectable number of bubble galaxies may still be tied to the time-integrated, photon-number-weighted $\langle f_{esc} \rangle$ of the whole bubble population. This population-wide $\langle f_{esc} \rangle$ quantity basically represents the ratio of the total number of ionizing photons that have ever escaped into the IGM over the total number produced within the bubble, and the results of Fig. 3 still hold as long as $\langle f_{esc} \rangle$ is interpreted this way. For example, in the case where the $z = 7$ luminosity function is considered to extend to $M_{UV} = -14$, a scenario in which the very faintest galaxies with $M_{UV} > -16$ have $\langle f_{esc} \rangle = 0.25$ and the rest have $f_{esc} = 0$ produces a population-wide $\langle f_{esc} \rangle \approx 0.1$ and the same number of detectable galaxies as predicted by $\langle f_{esc} \rangle \approx 0.1$ in that figure. Similarly, a scenario with the same luminosity function in which bright galaxies with $M_{UV} \leq -19$ have $\langle f_{esc} \rangle = 0.25$ and fainter galaxies have $\langle f_{esc} \rangle = 0.0$ corresponds to a population-wide $\langle f_{esc} \rangle \approx 0.05$ and the same number of detectable galaxies as indicated by that line.

In principle, it would seem that using SKA-1 to estimate the volume of an ionized bubble and then simply counting the galaxies brighter than a specific M_{UV} detection threshold within that structure would result in an observational constraint on the $\langle f_{esc} \rangle$ parameter. However, this would only be possible in so far as both the shape of bubble galaxy luminosity function above and below the M_{UV} detection limit and the N_{ion}/L_{UV} parameter are under control, i.e. that assembly/environmental biases of the type discussed in Section 5.1 are either negligible or can be accurately modelled. Moreover, the measurement error on the SKA-1 volume would also need to be $\lesssim 50$ per cent to make this measurement meaningful – a very difficult endeavour indeed.

5.7 Studying larger bubbles

In this paper, we have focused on ionized bubbles of size ≈ 1000 cMpc^3, which is at the limit of what SKA-1 can hope to detect. In future works, it would also be interesting to consider larger bubbles, since these would be easier to detect in the 21 cm data due to higher signal-to-noise ratio. Providing a census on the galaxy content of such structures with near-IR telescopes would on the other hand be more challenging, since the observing time required to survey the galaxy content of such bubbles to the same depth scales with the projected area of the bubble in the plane of the sky. Hence, a spherical bubble that has a volume a factor of 5 times larger requires a factor of $5^{(2/3)} \approx 3$ more observing time if the same magnitude limit is to be reached. On the other hand, larger structures are also likely to contain brighter galaxies, which would make the most extreme members detectable with telescopes wide-field survey telescopes like *Euclid*.

6 CONCLUSIONS

Our results can be summarized as follows:

(i) Our seminumerical simulations indicate that – for ionized IGM bubbles sufficiently large to be resolved by SKA-1 (volume $\gtrsim 1000$ cMpc^3) – there is a relatively tight relation between the volume of ionized IGM bubbles at $z = 7$–10 and the number of ionizing photons that have escaped from the galaxies within. For bubbles of volume $\gtrsim 1000$ cMpc^3, the 1σ bubble-to-bubble scatter in this relation is a factor of <4 (Section 3).

(ii) The total number of ionizing photons required to produce an ionized IGM bubble sufficiently large to be resolved by SKA-1 (volume $\gtrsim 1000$ cMpc^3) at $z > 6$ is $\gtrsim 1 \times 10^{71}$. This can be converted into a rough constraint on the minimum total stellar mass that has formed within such a structure (equation 6), and – using additional assumptions on the properties of the bubble galaxies – estimates on the number of galaxies detectable within that structure at the redshift where the bubble is detected (Section 4).

(iii) Using conservative assumptions on the shape of the luminosity function of bubble galaxies, the prior star formation history of these objects and their time-integrated, photon-number-weighted mean Lyman continuum escape fractions, we predict that the deepest spectroscopic surveys with *JWST* or ELT of SKA-detected ionized bubbles at $z = 7$–10 are expected to turn up at least a few bubble galaxies (and in many cases far more), even in the case of the smallest resolvable bubbles. The same also holds for purely photometric (multiband-imaging) observations with *JWST* or *WFIRST*. However, *Euclid* may only be able to detect galaxies within ionized bubble with volumes an order of magnitude higher than the SKA-1 detection limit (Section 4.4). Detailed detection predictions are presented in Figs 3 and 4.

(iv) Although spectroscopic observations (with *JWST* or ELT) may be required to survey the smallest SKA-detected bubbles towards the end of reionization ($z \approx 7$) without ending up with excessive numbers of line-of-sight interlopers, photometric surveys (with *WFIRST* or *JWST*) may be competitive at higher redshifts ($z \approx 10$) due to the smaller number of field-galaxy interlopers expected at redshifts close to that of the bubble (Section 4.5).

(v) If large numbers of bubble galaxies are detected, this could in principle allow for combined constraints on the luminosity function of bubble galaxies and the time-integrated, photon-number-weighted mean escape fraction $\langle f_{esc} \rangle$ of ionizing photons from these objects into the surrounding IGM (Section 5.6). For instance, if one adopts our baseline assumption – which seems consistent with the current observational picture of the $z \gtrsim 7$ galaxy population – that galaxies in ionized IGM bubbles are not significantly different (in terms of star formation history and relative shape of the luminosity function) from field galaxies at the same redshift, it should be possible to derive a lower limit on $\langle f_{esc} \rangle$ by adopting a very faint extension of the luminosity function when assessing the galaxy contribution to the ionizing photon budget. However, effects related to galaxy assembly bias and environmental feedback in overdensities could complicate this procedure (Section 5.1). Such effects may manifest themselves as differences between bubble and field galaxies in terms of spectral properties, luminosity function shape, and in the ratio between luminous and dark halo mass.

The detection of such signatures, which would shed new light on the properties of galaxies that form in extreme environments of the $z \gtrsim 7$ Universe, may be uncovered by the bubble galaxy observations themselves (e.g. spectral and dynamical properties uncovered through *JWST* and ELT spectroscopy; luminosity function constraints through *WFIRST* or *JWST* photometry). Early indications (prior to the detection of any ionized IGM bubbles with SKA-1) of such effects at $z \gtrsim 7$ may also come from studies with ALMA, *JWST*, and *WFIRST* of currently known overdensities of $z \gtrsim 7$ galaxies, since some of these structures are likely located in ionized IGM bubbles that SKA-1 will eventually be able to resolve though 21 cm tomography. The theoretical understanding of such differences between galaxies in overdensities and the field,

and how they affect the emergence of ionized IGM bubbles, must ultimately come from numerical simulations geared to this specific problem.

REFERENCES

Bagley M. B. et al., 2017, ApJ, 837, 11
Barkana R., 2016, Phys. Rep., 645, 1
Beardsley A. P., Morales M. F., Lidz A., Malloy M., Sutter P. M., 2015, ApJ, 800, 128
Boquien M., Buat V., Perret V., 2014, A&A, 571, A72
Bouwens R. J. et al., 2015, ApJ, 803, 34
Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, ApJ, 843, 129
Calvi V. et al., 2016, ApJ, 817, 120
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Castellano M. et al., 2016, ApJ, 818, L3
Castellano M. et al., 2018, ApJ, 863, L3
Choudhury T. R., Paranjape A., 2018, MNRAS, 481, 3821
Choudhury T. R., Haehnelt M. G., Regan J., 2009, MNRAS, 394, 960
Datta K. K., Bharadwaj S., Choudhury T. R., 2007, MNRAS, 382, 809
Datta K. K., Majumdar S., Bharadwaj S., Choudhury T. R., 2008, MNRAS, 391, 1900
Datta K. K., Friedrich M. M., Mellema G., Iliev I. T., Shapiro P. R., 2012, MNRAS, 424, 762
Datta K. K., Ghara R., Majumdar S., Choudhury T. R., Bharadwaj S., Roy H., Datta A., 2016, JA&A, 37, 27
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1
Dayal P., Dunlop J. S., Maio U., Ciardi B., 2013, MNRAS, 434, 1486
Evans C. et al., 2015, preprint (arXiv:1501.04726)
Finlator K., Oppenheimer B. D., Davé R., 2011, MNRAS, 410, 1703
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1
Geil P. M., Mutch S. J., Poole G. B., Duffy A. R., Mesinger A., Wyithe J. S. B., 2017, MNRAS, 472, 1324
Ghara R., Choudhury T. R., Datta K. K., 2016, MNRAS, 460, 827
Giri S. K., Mellema G., Ghara R., 2018, MNRAS, 479, 5596
Greig B., Mesinger A., 2017a, MNRAS, 465, 4838
Greig B., Mesinger A., 2017b, MNRAS, 472, 2651
Grützbauch R. et al., 2011, MNRAS, 418, 938
Harikane Y. et al., 2019, ApJ, 883, 142
Hasegawa K. et al., 2016, preprint (arXiv:1603.01961)
Hassan S., Davé R., Finlator K., Santos M. G., 2017, MNRAS, 468, 122

Hutter A., Dayal P., Müller V., Trott C., 2017, ApJ, 836, 176
Hutter A., Trott C. M., Dayal P., 2018, MNRAS, 479, L129
Inayoshi K., Li M., Haiman Z., 2018, MNRAS, 479, 4017
Inoue A. K., 2010, MNRAS, 401, 1325
Inoue A. K. et al., 2018, PASJ, 70, 55
Ishigaki M., Ouchi M., Harikane Y., 2016, ApJ, 822, 5
Jaacks J., Nagamine K., Choi J. H., 2012, MNRAS, 427, 403
Kakiichi K., Graziani L., Ciardi B., Meiksin A., Compostella M., Eide M. B., Zaroubi S., 2017, MNRAS, 468, 3718
Kaurov A. A., Gnedin N. Y., 2015, ApJ, 810, 154
Kimm T., Cen R., 2014, ApJ, 788, 121
Koopmans L. et al., 2015, Proc. Sci., The Cosmic Dawn and Epoch of Reionisation with SKA. SISSA, Trieste, p. PoS#1
Kroupa P., 2001, MNRAS, 322, 231
Kulkarni G., Hennawi J. F., Rollinde E., Vangioni E., 2014, ApJ, 787, 64
Kulkarni G., Choudhury T. R., Puchwein E., Haehnelt M. G., 2016, MNRAS, 463, 2583
Laporte N. et al., 2017, ApJ, 837, L21
Lee C. T., Primack J. R., Behroozi P., Rodríguez-Puebla A., Hellinger D., Dekel A., 2017, MNRAS, 466, 3834
Leitherer C. et al., 1999, ApJS, 123, 3
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Lidz A., Zahn O., Furlanetto S. R., McQuinn M., Hernquist L., Zaldarriaga M., 2009, ApJ, 690, 252
Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, ApJ, 835, 113
Loeb A., Furlanetto S. R., 2013, The First Galaxies in the Universe. Princeton Univ. Press, Princeton, New Jersey
Ma X., Kasen D., Hopkins P. F., Faucher-Giguère C.-A., Quataert E., Kereš D., Murray N., 2015, MNRAS, 453, 960
Ma X. et al., 2018, MNRAS, 478, 1694
Maio U., Ciardi B., Dolag K., Tornatore L., Khochfar S., 2010, MNRAS, 407, 1003
Maio U., Petkova M., De Lucia G., Borgani S., 2016, MNRAS, 460, 3733
Majumdar S., Bharadwaj S., Datta K. K., Choudhury T. R., 2011, MNRAS, 413, 1409
Majumdar S., Bharadwaj S., Choudhury T. R., 2012, MNRAS, 426, 3178
Majumdar S., Mellema G., Datta K. K., Jensen H., Choudhury T. R., Bharadwaj S., Friedrich M. M., 2014, MNRAS, 443, 2843
Malloy M., Lidz A., 2013, ApJ, 767, 68
Marchetti L., Serjeant S., Vaccari M., 2017, MNRAS, 470, 5007
Maselli A., Gallerani S., Ferrara A., Choudhury T. R., 2007, MNRAS, 376, L34
McNaught-Roberts T. et al., 2014, MNRAS, 445, 2125
McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043
Mellema G., Koopmans L., Shukla H., Datta K. K., Mesinger A., Majumdar S., 2015, Proc. Sci., HI tomographic imaging of the Cosmic Dawn and Epoch of Reionisation with SKA, SISSA, Trieste, PoS#10
Mesinger A., ed., 2016, Understanding the Epoch of Cosmic Reionization: Challenges and Progress (Astrophysics and Space Science Library). Springer International Publishing, Switzerland
Mesinger A., Dijkstra M., 2008, MNRAS, 390, 1071
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mirocha J., Furlanetto S. R., Sun G., 2017, MNRAS, 464, 1365
Mondal R., Bharadwaj S., Majumdar S., Bera A., Acharyya A., 2015, MNRAS, 449, L41
Mondal R., Bharadwaj S., Majumdar S., 2016, MNRAS, 456, 1936
Mondal R., Bharadwaj S., Majumdar S., 2017, MNRAS, 464, 2992
Mondal R., Bharadwaj S., Datta K. K., 2018, MNRAS, 474, 1390
Mondal R., Bharadwaj S., Iliev I. T., Datta K. K., Majumdar S., Shaw A. K., Sarkar A. K., 2019, MNRAS, 483, L109
Monsalve R. A., Greig B., Bowman J. D., Mesinger A., Rogers A. E. E., Mozdzen T. J., Kern N. S., Mahesh N., 2018, ApJ, 863, 11
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
Overzier R. A., 2016, A&AR, 24, 14
Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544
Pacucci F., Mesinger A., Mineo S., Ferrara A., 2014, MNRAS, 443, 678

Park J., Kim H.-S., Wyithe J. S. B., Lacey C. G., 2014, MNRAS, 438, 2474

Park H., Shapiro P. R., Choi J.-h., Yoshida N., Hirano S., Ahn K., 2016, ApJ, 831, 86

Planck Collaboration XVI, 2014, A&A, 571, A16

Planck Collaboration XLVII, 2016, A&A, 596, A108

Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS, 462, 893

Sadoun R., Shlosman I., Choi J.-H., Romano-Díaz E., 2016, ApJ, 829, 71

Sahlén M. et al., 2009, MNRAS, 397, 577

Schaerer D., 2002, A&A, 382, 28

Schaye J., 2001, ApJ, 559, 507

Sharma M., Theuns T., Frenk C., Bower R. G., Crain R. A., Schaller M., Schaye J., 2017, MNRAS, 468, 2176

Shimizu I., Inoue A. K., Okamoto T., Yoshida N., 2014, MNRAS, 440, 731

Shimizu I., Inoue A. K., Okamoto T., Yoshida N., 2016, MNRAS, 461, 3563

Sobacchi E., Mesinger A., 2013, MNRAS, 432, 3340

Sobacchi E., Mesinger A., 2014, MNRAS, 440, 1662

Sobacchi E., Mesinger A., Greig B., 2016, MNRAS, 459, 2741

Songaila A., Cowie L. L., 2010, ApJ, 721, 1448

Stefanon M. et al., 2019, ApJ, 883, 99

Stiavelli M., Trenti M., 2010, ApJ, 716, L190

Susa H., 2008, ApJ, 684, 226

Targett T. A., Dunlop J. S., McLure R. J., 2012, MNRAS, 420, 3621

Tozzi P., Madau P., Meiksin A., Rees M. J., 2000, ApJ, 528, 597

Vrbanec D. et al., 2016, MNRAS, 457, 666

Wiersma R. P. C. et al., 2013, MNRAS, 432, 2615

Wood K., Loeb A., 2000, ApJ, 545, 86

Wyithe J. S. B., Loeb A., 2007a, MNRAS, 374, 960

Wyithe J. S. B., Loeb A., 2007b, MNRAS, 375, 1034

Wyithe J. S. B., Loeb A., Barnes D. G., 2005, ApJ, 634, 715

Wyithe S., Geil P., Kim H., 2015, Proc. Sci., Imaging H II Regions from Galaxies and Quasars During Reionisation with SKA, SISSA, Trieste, PoS#15

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

Yajima H., Khochfar S., 2017, MNRAS, 467, L51

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

Yajima H., Shlosman I., Romano-Díaz E., Nagamine K., 2015, MNRAS, 451, 418

Yajima H., Sugimura K., Hasegawa K., 2018, MNRAS, 477, 5406

Zackrisson E., Rydberg C.-E., Schaerer D., Östlin G., Tuli M., 2011, ApJ, 740, 13

Zackrisson E. et al., 2017, ApJ, 836, 78

Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, ApJ, 654, 12

## APPENDIX A: THE $N_{\rm ION}/L_{\rm UV}$ PARAMETER AND ITS RELATION TO THE PRIOR STAR FORMATION HISTORY

In Fig. A1, we illustrate the evolution of the $N_{\rm ion}/L_{\rm UV}$ ratio for three different parametric star formation histories over a few hundred Myr (the time-scale relevant for $z > 7$ galaxies). A constant SFR leads to a rapid rise in $N_{\rm ion}/L_{\rm UV}$ during the first 100 Myr after the onset of star formation, followed by a slower increase in this ratio thereafter. Temporal variations in the SFR throughout the star formation history makes $N_{\rm ion}/L_{\rm UV}$ ratio either go up (following a drop in SFR) or down (for a boost in SFR). Galaxies for which the SFR history shows similar trends (e.g. increasing or semiconstant SFR) can be expected to exhibit similar $N_{\rm ion}/L_{\rm UV}$ ratios. This is demonstrated in Fig. A2 using the non-parametric star formation histories of three high-mass galaxies (reaching total stellar mass $\sim 10^9$–$10^{10}\,{\rm M_\odot}$ by $z \approx$ 7) from the Shimizu et al. (2016) simulations. These three galaxies all exhibit SFRs that initially increase over time (albeit with stochastic fluctuations along the way) and then stabilize at different levels of near-constant SFR (Fig. A2a). In the absence of temporal variations in the stellar IMF, the number of ionizing photons (or alternatively the total stellar mass) produced becomes a smoothly increasing function of time (Fig. A2b). The rest-frame 1500 Å luminosity (Fig. A2c) traces the SFR fairly faithfully, and the result is a $N_{\rm ion}/L_{\rm UV}$ ratio (Fig. A2d) that varies no more than a factor of a few after the large initial SFR fluctuations that occur at early times. Please note that the SFR histories in these cases include the star formation that has taken place *in situ* within these galaxies, but also within the progenitors that merged to form them.

To generate Figs A1 and A2, we have for simplicity assumed a dust-free stellar population with constant metallicity $Z = 0.004$, and stellar populations SEDs from the STARBURST99 stellar population models (Leitherer et al. 1999). These results admittedly neglect the effects of dust attenuation on the UV flux and the metallicity distribution within galaxies. Moreover, the star formation history is expected to become increasingly stochastic at lower galaxy masses due to feedback effects (e.g. Shimizu et al. 2014; Ma et al. 2018), which will result in longer spells of low star formation activity and larger galaxy-to-galaxy variations in the $N_{\rm ion}/L_{\rm UV}$ ratio as a result. All of this is, however, taken into account in the LYCAN machinery for predicting high-redshift galaxy spectra (Zackrisson et al. 2017) that is used in Section 4 of this paper.
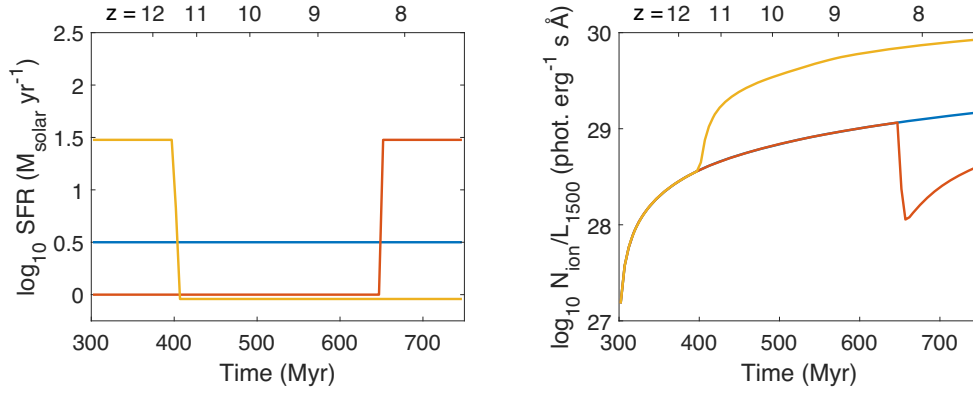
Figure A1. Demonstration of how three parametric star formation scenarios for $z \geq 7$ galaxies translate into different predictions on the evolution of the N_{ion}/L_{UV} ratio. The time axis indicates the age of the Universe from 300 ($z \approx 14$, when star formation is assumed to start) to 750 Myr ($z \approx 7$). *Left:* Star formation histories of the three models. The yellow represents a galaxy where the SFR drops by a factor of ≈ 30 after 100 Myr, the orange line a galaxy where the SFR remains constant for 350 Myr and then rises by a factor of ≈ 30, and the blue line represents a galaxy where the SFR remains constant throughout 450 Myr. *Right:* Prediction of how the N_{ion}/L_{1500} ratio (cumulative number of ionizing photons produce divided by momentary UV 1500 Å luminosity) evolves for the three star formation scenarios of the left-hand panel. In the constant SFR scenario, the N_{ion}/L_{1500} gradually increases, but drops in SFR (yellow line) lead to higher ratios whereas boosts in SFR (orange line) lead to lower ratios. In these models, a constant metallicity of $Z = 0.004$ is assumed and dust attenuation of L_{1500} is neglected.
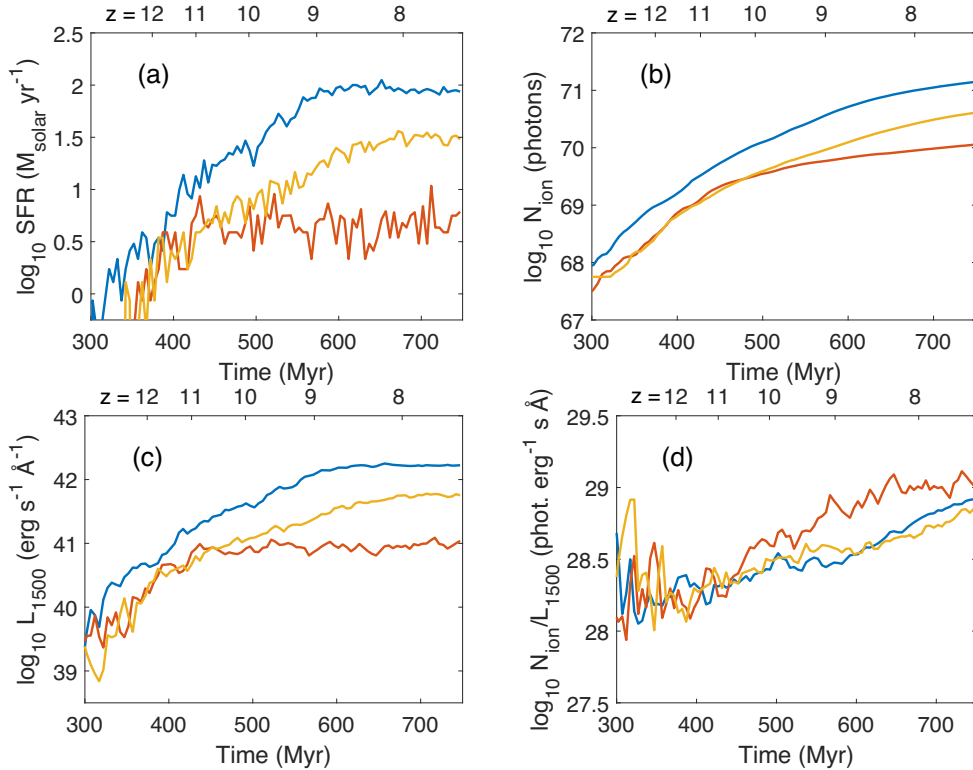
Figure A2. Connection between the prior production of ionizing photons and the momentary rest-frame UV luminosity for three simulated high-mass (total stellar mass $\gtrsim 10^9 \, M_\odot$) galaxies at $z \approx 7$. (a) Star formation activity until 750 Myr after the big bang, showing rapid growth followed with fluctuations on short time-scales followed by stabilization. (b) The corresponding cumulative production of ionizing photons. (c) The momentary rest-frame 1500 Å luminosity in the absence of dust effects. (d) The ratio between the cumulative number of ionizing photons produced and the momentary 1500 Å luminosity, showing fluctuations of factors of only factors a few at late times.

APPENDIX B: BUBBLE GALAXIES VERSUS LINE-OF-SIGHT INTERLOPERS IN PHOTOMETRIC SURVEYS

In Section 4.5, we use an argument based on luminosity functions to argue that photometric surveys in the direction of ionized IGM bubbles are less likely to pick up line-of-sight interlopers for bubbles at $z \approx 10$ compared to $z \approx 7$. However, one may also reach the same conclusion by directly comparing the number of high-mass haloes present within simulated ionized bubbles, like the one presented in Section 3, to those in the ambient field along a column spanning $\Delta(z) \approx 1$. In this case, the outcome will depend on the detailed recipe for relating ionzing photon fluxes to haloes of a given mass. In the simulations discussed in Section 3, bubbles with volume ≈ 1000 cMpc^3 contain ≈ 20–50 haloes with virial mass $\geq 10^{10}$ M$_\odot$ h^{-1} at $z \approx 7$–10. The ambient field will on the other hand contain ≈ 600 such haloes in the $\Delta(z) \approx 1$ volume along the line of sight at $z = 7$, yet only ≈ 10 at $z = 10$. This lends further support to the notion that the fraction of line-of-sight interlopers may be very high in photometric bubble surveys at $z \approx 7$, but much smaller at $z \approx 10$. This calculation of the ambient halo density is based on the halo mass function derived from the Bolshoi–Planck and MultiDark-Planck *N*-body simulation (Rodríguez-Puebla et al. 2016), and assuming a fiducial flat ΛCDM cosmology with parameter values $\Omega_m = 0.3$, $\Omega_b = 0.044$, $h = 0.7$, $\sigma_8 = 0.8$, $n_s = 0.96$. The cosmological computations are performed using a modified version of CAMB (Lewis, Challinor & Lasenby 2000; Sahlén et al. 2009).

[1]*Observational Astrophysics, Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden*
[2]*Discipline of Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology Indore, Simrol, Indore 453552, India*
[3]*Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK*
[4]*National Centre for Radio Astrophysics, TIFR, Post Bag 3, Ganeshkhind, Pune 411007, India*
[5]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1 D-85748 Garching, Germany*
[6]*Department of Physics, Presidency University, 86/1 College Street, Kolkata 700073, India*
[7]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[8]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[9]*Kavli IPMU, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa 277-8583, Japan*
[10]*Department of Astronomy, Stockholm University, Oskar Klein Center, AlbaNova, SE 106 91 Stockholm, Sweden*
[11]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str 1, D-85741 Garching, Germany*
[12]*Arizona State University, School of Earth and Space Exploration, Tempe, AZ 85287, USA*
[13]*NASAs Goddard Space Flight Center, Astrophysics Science Division, Code 660, Greenbelt, MD 20771, USA*
[14]*Zentrum für Astronomie, Institut für Theoretische Astrophysik, Universitä Heidelberg, Albert-Ueberly-Str 2, D-69120 Heidelberg, Germany*
[15]*Department of Earth and Space Science, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043, Japan*

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