Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **531,** 2943–2957 (2024) https://doi.org/10.1093/mnras/stae1325
Advance Access publication 2024 May 24

The THESAN project: connecting ionized bubble sizes to their local environments during the Epoch of Reionization

Meredith Neyer [1]★ Aaron Smith [2] Rahul Kannan [3] Mark Vogelsberger [1,4] Enrico Garaldi [5] Daniela Galárraga-Espinosa [5] Josh Borrow [1] Lars Hernquist,[7] Rüdiger Pakmor [5] and Volker Springel [5]

[1]*Department of Physics & Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2]*Department of Physics, The University of Texas at Dallas, Richardson, TX 75080, USA*
[3]*Department of Physics and Astronomy, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada*
[4]*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[5]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str 1, D-85741 Garching, Germany*
[6]*Department of Physics and Astronomy, University of Pennsylvania, 209 South 33rd Street, Philadelphia, PA, 19104, USA*
[7]*Center for Astrophysics|Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*

Accepted 2024 May 20. Received 2024 April 23; in original form 2023 September 29

ABSTRACT

An important characteristic of cosmic hydrogen reionization is the growth of ionized gas bubbles surrounding early luminous objects. Ionized bubble sizes are beginning to be probed using Lyman α emission from high-redshift galaxies, and will also be probed by upcoming 21 cm maps. We present results from a study of bubble sizes using the state-of-the-art THESAN radiation-hydrodynamics simulation suite, which self-consistently models radiation transport and realistic galaxy formation. We employ the mean free path method and track the evolution of the effective ionized bubble size at each point ($R_{\rm eff}$) throughout the Epoch of Reionization. We show that there is a slow growth period for regions ionized early, but a rapid 'flash ionization' process for regions ionized later as they immediately enter a large, pre-existing bubble. We also find that bright sources are preferentially in larger bubbles, and find consistency with recent observational constraints at $z \gtrsim 9$, but tension with idealized Lyman α damping-wing models at $z \approx 7$. We find that high-overdensity regions have larger characteristic bubble sizes, but the correlation decreases as reionization progresses, likely due to runaway formation of large percolated bubbles. Finally, we compare the redshift at which a region transitions from neutral to ionized ($z_{\rm reion}$) with the time it takes to reach a given bubble size and conclude that $z_{\rm reion}$ is a reasonable local probe of small-scale bubble size statistics ($R_{\rm eff} \lesssim 1$ cMpc). However, for larger bubbles, the correspondence between $z_{\rm reion}$ and size statistics weakens due to the time delay between the onset of reionization and the expansion of large bubbles, particularly at high redshifts.

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

1 INTRODUCTION

Hundreds of millions of years after the big bang, the first stars and galaxies began to form and emit radiation that ionized the surrounding hydrogen gas to mark the end of the cosmic dark ages and the beginning of the Epoch of Reionization (EoR; Shapiro & Giroux 1987; Barkana & Loeb 2001; Furlanetto, Oh & Briggs 2006; Wise 2019). During the EoR, which took place at redshifts of $z \approx$ 5–20, the gas in the Universe underwent a phase transition during which ionized bubbles expanded around radiation sources, eventually yielding a nearly fully ionized Universe as observed today. Central to this process is the percolation of ionized bubbles as they merged into one another causing rapid increases in bubble sizes (Furlanetto & Oh 2016). Observing galaxies and the intergalactic medium (IGM) during the EoR poses challenges due to its high redshift. However, recent and forthcoming *JWST* surveys and 21 cm telescopes promise insights into high-redshift galaxies and ionized bubble properties, respectively (Robertson 2022). A complete picture of the EoR requires understanding both the IGM and galaxy sources, motivating studies to decode ionized bubble dynamics and morphologies as well as their connection to the local environment (Gnedin & Madau 2022).

The EoR can be explored using a variety of observational techniques, including analysing spectra of Lyman α emitting galaxies at high redshift. The recently launched *JWST* is revolutionizing the field of high-redshift galaxy observations by providing unprecedented access to high-resolution rest-frame optical imaging of reionization-era galaxies. Several of the first *JWST* observations have indicated brighter galaxies at earlier times than most theoretical models predicted (Finkelstein et al. 2023; Harikane et al. 2023b, c; Robertson et al. 2023), and above extrapolations of data taken from the *Hubble Space Telescope* (Finkelstein et al. 2022). Several theoretical studies have sought to explain this discrepancy by considering modifications to galaxy formation models, and improved treatment of dust effects

★ E-mail: mneyer@mit.edu

(Kannan et al. 2023; Shen et al. 2023), different early stellar populations (Steinhardt et al. 2023), skewed mass-to-light ratios for early galaxies (Inayoshi et al. 2022), and alternative dark energy models in the early universe (Smith et al. 2022a). In addition to findings of more faint active galactic nuclei (AGNs) than expected (Bischetti et al. 2022; Harikane et al. 2023a), these early results hint at potential paradigm shifts in the role of various populations of galaxies in contributing to reionization mechanisms. Spectroscopic studies facilitated by the Near Infrared Spectrograph aboard the *JWST* have also inferred constraints on ionized bubbles sizes around high-redshift galaxies. Typically, these are based on damping-wing absorption redward of the Ly α line (Fujimoto et al. 2023; Hayes & Scarlata 2023; Hsiao et al. 2023; Jung et al. 2023; Umeda et al. 2023), but there have also been studies that use the transmission of the Ly α emission line itself to constrain the ionized bubble sizes around galaxies (Saxena et al. 2023; Witstok et al. 2024).

In parallel to galaxy observations, 21 cm radio interferometers, including the Low Frequency Array (van Haarlem et al. 2013), Hydrogen Epoch of Reionization Array (DeBoer et al. 2017; HERA Collaboration 2023), Square Kilometer Array (Mellema et al. 2013), and others, are beginning to map the neutral hydrogen in the Universe. The complementary advantage of the 21 cm observations is the ability to directly probe the IGM through line emission from the forbidden spin-flip hyperfine transition of neutral hydrogen. By looking for the redshifted signal from the EoR, these instruments will allow us to see the distribution of neutral and ionized gas in the IGM. These measurements will be critical for studying the ionized bubbles during the EoR as well as constraining cosmological parameters (McQuinn et al. 2006; Mesinger, Furlanetto & Cen 2011; Liu & Parsons 2016; Park et al. 2019; Kannan et al. 2022b).

In conjunction with the recent and upcoming observational capabilities, there has also been strong progress on developing theoretical methodologies to explore and interpret signals from the EoR. This is challenging because predictions of ionized bubble sizes will depend on the volume of the simulated box, which must be large enough to obtain a converged bubble size distribution. A variety of methods have been employed to study the processes of hydrogen reionization including perturbative evolution of density fields with ionized regions classified with excursion-set theory, furnishing rapid insights into ionized region classification and model parameter spaces (Furlanetto, Zaldarriaga & Hernquist 2004a; Mesinger et al. 2011; Park et al. 2019; Fialkov, Barkana & Jarvis 2020; Muñoz, Dvorkin & Cyr-Racine 2020; Lu et al. 2024). Beyond this, post-processing of simulated dark matter haloes with galaxy formation and radiative transfer models yields complementary insights about the physics of ionizing sources (Ciardi, Stoehr & White 2003; Iliev et al. 2007; McQuinn et al. 2009) and semi-analytical models of galaxy formation that include a variety of approximate or efficient prescriptions for including inhomogeneous reionization on the fly (Mutch et al. 2016; Davies et al. 2023; Puchwein et al. 2023). Finally, the most accurate but also computationally demanding are self-consistent radiation-hydrodynamics simulations, which couple the intricacies of galaxy formation and fully time-dependent radiative transfer physics (Xu, Wise & Norman 2013; Gnedin 2014; Pawlik et al. 2017; Rosdahl et al. 2018; Lewis et al. 2022).

In anticipation of the forthcoming 21 cm data, numerous theoretical studies have sought to characterize the properties of ionized bubbles and their connections to the dominant processes in cosmic reionization. Different reionization schemes and source models can lead to different bubble morphologies, and the distribution of bubble sizes has emerged as a distinguishing feature between different reionization models (McQuinn et al. 2007; Majumdar et al. 2016). Recent investigations indicate that quasars provide a negligible contribution to the ionizing radiation needed to reionize the Universe (Eide et al. 2020; Jiang et al. 2022), suggesting that stellar matter within galaxies is likely primarily responsible for hydrogen reionization. Discerning whether this process is driven by smaller numbers of highly luminous galaxies or the collection of numerous faint galaxies remains a critical goal of reionization research (Bera et al. 2023; Kostyuk et al. 2023; Yeh et al. 2023). The specific topological properties of the spread of ionized regions with potential to distinguish otherwise degenerate models is also a topic of current research. There has been a focus to determine the conditions for cosmological regions to ionize via an 'inside-out' scenario, where the ionizing radiation originates from within the galaxy, or an 'outside-in' scenario, where the ionizing radiation emanates from external sources and propagates into the galaxy to reionize the gas (Choudhury, Haehnelt & Regan 2009).

Topological analyses of the ionization field, redshift of reionization, and bubble morphologies have been fruitful in characterizing the distinctive signatures of reionization (Friedrich et al. 2011; Busch et al. 2020; Thélie et al. 2022; Cain et al. 2023; Elbers & van de Weygaert 2023). There are a variety of ways to characterize the topology and morphology of reionization, including the utilization of Betti numbers (Giri & Mellema 2021; Kapahtia et al. 2021), genus curves (Lee et al. 2008), and various alternative metrics (Giri, Mellema & Jensen 2020). Other works focus specifically on forecasts for the 21 cm signal (Furlanetto, Zaldarriaga & Hernquist 2004b; Zaldarriaga, Furlanetto & Hernquist 2004) or observed spectra (Gazagnes, Koopmans & Wilkinson 2021).

A variety of methods exist for the detection and analysis of ionized bubble sizes within hydrodynamical simulations and semi-analytical codes. Prominent among these are the mean free path (MFP) method, which calculates bubble sizes by tracing rays from ionized cells to their first neutral cell encounter (Mesinger & Furlanetto 2007); spherical averaging, which calculates the largest ionized sphere based on a set ionization threshold (Zahn et al. 2007); friends of friends (FOF), which links ionized cells within a given distance to form a bubble (Ivezić et al. 2014), granulometry uses a 'sieving' process to count objects fitting within a given structure (Kakiichi et al. 2017); and the watershed method identifies bubbles by filling 'catchment basins' from local minima until neighbouring basins intersect (Lin et al. 2016).

Many of these analyses have been performed using semi-numerical schemes that do not self-consistently model the impact of galaxies on hydrogen reionization. To build upon these existing studies, we perform an analysis of ionized bubble sizes and connections to their local cosmic environments within the THESAN simulations (Garaldi et al. 2022; Kannan et al. 2022a; Smith et al. 2022b; Garaldi et al. 2023), which combines the galaxy formation model of IllustrisTNG (Weinberger et al. 2017; Pillepich et al. 2018a, b) with radiative process modelling including radiative transport (AREPO-RT; Kannan et al. 2019), non-equilibrium heating and cooling, and realistic ionizing sources throughout a large box of 95.5 cMpc per side. The THESAN simulations have been used to make a wide range of EoR predictions (e.g. Kannan et al. 2022b; Qin et al. 2022; Borrow et al. 2023; Kannan et al. 2023; Xu et al. 2023; Yeh et al. 2023; Shen et al. 2024a, b).

The paper is organized as follows. In Section 2, we describe our methods, including a brief summary of the THESAN suite of simulations, the MFP methodology, and our particular implementation of the MFP bubble size calculator. In Section 3, we present our findings on the time evolution of bubble sizes, followed by an exploration of the impact of environmental factors on these sizes in Section 4. In Section 5, we focus on the utility of the redshift of reionization as a

probe of bubble sizes. Finally, we present our overall conclusions in Section 6. Supplementary discussions on grid resolution, smoothing scale, and simulation convergence are provided in Appendices A, B, and C, respectively. All further mentions of 'reionization' refer specifically to hydrogen reionization.

2 SIMULATION AND ANALYSIS METHODS

We briefly describe the THESAN reionization simulations in Section 2.1, focusing primarily on the ionization field as it pertains to this study. We then provide additional details about our procedure for determining bubble sizes using the MFP method in Section 2.2, and our MFP implementation and data products in Section 2.3.

2.1 THESAN simulations

The THESAN project (Garaldi et al. 2022; Kannan et al. 2022a; Smith et al. 2022b; Garaldi et al. 2023) is a suite of large-volume ($L_{\rm box}$ = 95.5 cMpc) radiation-magneto-hydrodynamic simulations that simultaneously resolve large-scale structure for reionization and realistic galaxy formation using the IllustrisTNG model (Pillepich et al. 2018a), which is an updated version of the model used in Illustris (Vogelsberger et al. 2013, 2014a, b). THESAN provides state-of-the-art resolution and physics for a reionization simulation of this volume and presents a unique opportunity to investigate connections between the topology of reionization and the galaxies responsible for reionizing the Universe. Photons from sources including stars and AGNs are tracked self-consistently in three energy bins (13.6–24.6, 24.6–54.4, and ≥54.4 eV). Stellar population properties including luminosities and spectral energy densities are determined by using the Binary Population and Spectral Synthesis library (Eldridge et al. 2017). THESAN also incorporates non-equilibrium thermochemistry for tracking cooling by hydrogen and helium, as well as equilibrium cooling by metals. The high-resolution, fiducial simulation, THESAN-1, has dark matter and baryonic mass resolutions of 3.1×10^6 and 5.8×10^5 $M_\odot$, respectively. Haloes are resolved down to masses of $M_{\rm halo} \sim 10^8\, h^{-1}\, M_\odot$ (Garaldi et al. 2023). The simulations use the efficient quasi-Lagrangian code AREPO-RT (Kannan et al. 2019), an extension of the moving mesh code AREPO (Springel 2010; Weinberger, Springel & Pakmor 2020) that includes radiative transport, to solve the fluid dynamics equations on an unstructured Voronoi mesh produced by approximately following the flow of the gas. The radiative transport equations are solved using a moment-based approach assuming the M1 closure condition (Levermore 1984; Dubroca & Feugeas 1999). The gravity solver uses the hybrid Tree-PM method, which estimates short-range gravitational forces using a hierarchical oct-tree algorithm (Barnes & Hut 1986). Long-range gravitational potentials are calculated solving the Poisson equation using the Fourier method.

The THESAN simulations output 81 snapshots of particle positions and properties spanning redshifts from 20 to 5.5. These particle snapshots are converted into Cartesian grids with varying resolutions: 128, 256, 512, and 1024 cells per side. These Cartesian grid representations encapsulate volume-weighted properties, including the ionized fraction. In this paper, we use the renders with 512 cells per side. A discussion of convergence across different grid resolutions is provided in Appendix A, but we note that we demonstrate satisfactory convergence of bubble size statistics. The resolution effects are minimal between global neutral fractions of 0.1 and 0.9, i.e. ∼10 (∼30) per cent agreement for 256 (128) cells per side compared to the fiducial 512 resolution. The deviation is more significant at the very beginning and end of reionization, which is expected due to the increased importance of small-scale structures during these phases.

Large-scale environmental properties such as the dark matter overdensity, $\delta \equiv (\rho - \bar{\rho})/\bar{\rho}$, where $\bar{\rho}$ is the mean density, are most meaningful after smoothing to a given filter scale. Bubble statistics are sensitive to small-scale features, so it is preferable to smooth out potentially transient or local features. Therefore, we also construct smoothed versions of the Cartesian outputs. Specifically, the particles are first binned at 1024^3 resolution and then a periodic mass-conserving Gaussian smoothing kernel is applied with smoothing scales defined by standard deviations of 0 ckpc (no smoothing), 125 ckpc, 250 ckpc, 500 ckpc, and 1 cMpc. We emphasize that lower resolution grids are coarse-grained versions of the high-resolution one with volume and mass weights propagated correctly. Results for bubble size evolution and comparison to the redshift of reionization are reported using the unsmoothed outputs, whereas the environmental analysis is performed using a smoothing scale of 125 ckpc chosen to be above the grid resolution and extend up to typical galaxy separations such that the bubble sizes reflect the sources and environmental properties. The effects of the varying smoothing scales are discussed in Appendix B, including potential sensitivities to oversmoothing of the ionization field. As the degree of smoothing increases, so does the median bubble size at each instance. This correlation is an expected outcome of the blurring of ionized regions induced by smoothing. The smallest bubbles tend to be approximately the same size as the smoothing scale (or the cell size in the unsmoothed case). The influence of smoothing scales on bubble sizes diminishes as reionization progresses, leading to substantial agreement once the majority of the gas becomes ionized and bubble sizes significantly exceed the smoothing scales.

2.2 Mean free path methodology

In this paper, we exclusively employ the MFP method for determining characteristic bubble sizes. We have chosen the MFP method because of its close correspondence with physical parameters, such as the MFP of photons emitted from the central cell. Lin et al. (2016) assert that both the MFP and watershed methods stand out as the most meaningful metrics for determining bubble sizes. Notably, the MFP determination of bubble sizes yields size distributions which are not biased toward artificially large or small bubble sizes in well-defined test cases (Lin et al. 2016), signifying that the peak of the bubble size distribution (BSD) aligns with the 'correct' radius for a binary field with a single spherical ionized region. Unlike the watershed approach, the MFP method allows us to assign each cell a characteristic bubble size according to the ionization states of the surrounding gas. This corresponds to the effective bubble size 'seen' by a region in the simulation. For example, if a small bubble begins to merge with a larger one, methods such as FOF and watershed might classify all the ionized cells as constituents of one extensive bubble with identical size characteristics. In contrast, the MFP method would retain more information about the progenitor bubbles because the MFP bubble size would be smaller for cells within the smaller bubble compared to the larger one. This distinction is useful for studying environmental effects on bubble sizes, as well as in drawing comparisons with photon MFP and line-of-sight results.

In our analysis, we identify bubble sizes by extending 192 rays in isotropically determined directions, following the HEALPIX prescription of equal area segmentations of a sphere (Gorski et al. 1999). This ensures uniform sampling of the directions of the rays. The number of HEALPIX rays was chosen to account for the complex bubble morphology, while keeping computation time low. We find

less than 5 per cent deviation in median bubble sizes compared to using 768 rays per cell (and variations $\lesssim 10$ per cent compared to 48 HEALPIX rays). We calculate the length of each ray between the starting point and the first instance that it intersects a cell possessing an ionization fraction below a threshold value of 0.5. We utilize second-order ray tracing with gradient limiters to maintain values in the range [0, 1] and avoid overshooting neighbouring values. For any neutral cells where the ionized fraction is $x_{\rm H\,II} < 0.5$, the MFP bubble size is assigned a value of zero. This choice of threshold, $x_{\rm H\,II} = 0.5$ is arbitrary, and median bubble sizes differ by $\lesssim 10$ per cent compared to thresholds of $x_{\rm H\,II} = 0.1$ and 0.9, for most of the EoR. There are larger relative variations ($\lesssim 30$ per cent) at high neutral fraction ($x_{\rm H\,I} \gtrsim 0.8$), when bubbles are small. The algorithm we employ is similar to the method utilized by Tools21cm (Giri et al. 2020), specifically adapted to account for periodic boundary conditions and systematic sampling of every ionized cell with uniformly distributed ray directions according to the HEALPIX prescription. Each ray is followed through the ionization field which has been linearly interpolated between neighbouring cells, again with second-order ray tracing. The ray's length is recorded once it surpasses the neutral threshold value. We allow a maximum ray length of twice the box size (191 cMpc) due to the periodicity of the grid.

2.3 Mean free path data products

We apply the MFP method to the Cartesian outputs with a resolution of 512 cells per side such that the cell sizes are $L_{\rm cell} = 186.5$ ckpc. We compute the *effective* MFP bubble size, denoted by $R_{\rm eff}$, for each cell by averaging multiple second-order ray traces along 192 HEALPIX directions. Positioning these bubbles as 'centred' on the cell itself permits $R_{\rm eff}$ to be considered as a distinct property of each cell, facilitating direct comparisons with other spatial attributes. Note that we consider $R_{\rm eff}$ to be the average of the lengths of the radial rays, not as the effective radius of the volume sampled by the 192 rays. These statistics are, by construction, volume-weighted as each equal-volume Cartesian grid cell gives an equal contribution to population statistics (e.g. BSDs and median bubble sizes).

Additionally, we generate a BSD histogram encompassing all ray lengths originating from ionized cells throughout the simulation box. For the smoothed renders we also compile two-dimensional histograms of ray lengths and the local dark matter overdensity δ and other environmental quantities of the starting cell to explore environmental effects. The global BSD histogram is saved with a bin resolution of $L_{\rm cell}/8$, representing a characteristic scale on which the ionized fraction may change. For the unsmoothed case, we perform the MFP analysis on all 81 of the THESAN snapshots to achieve maximum redshift resolution. For the smoothed outputs, we report results from the MFP bubble size analysis for nine snapshots, chosen such that the global volume-weighted neutral fraction increases from $x_{\rm H\,I} = 0.1$ to 0.9 in increments of 0.1.

In Fig. 1, we illustrate the density ρ coloured by the ionized fraction $x_{\rm H\,II}$, redshift of reionization $z_{\rm reion}$, and effective bubble size $R_{\rm eff}$ in the left, middle, and right columns, respectively. The snapshot corresponds to a redshift when the global neutral fraction is one half. The middle row displays the entire THESAN box ($L_{\rm box} = 95.5$ cMpc), while the top and bottom rows offer detailed views of two zoomed-in regions. The leftmost column shows neutral gas in blue and ionized gas in yellow, with variations in colour intensity denoting the density, with lighter (darker) shades corresponding to more (less) dense regions. The left column is a high-resolution (2048^2) projection through 10 per cent of the box, whereas the middle and right columns show the equivalent slice through the 512 cells per side render box. This figure illustrates the complexity of interconnected ionized regions along with the richness of the large-scale structure. The redshift of reionization retains the structure of the bubble evolution and the instantaneous bubble size ($R_{\rm eff}$) shows the morphology at a neutral fraction of 0.5.

3 EVOLUTION OF BUBBLE SIZES

As the Universe becomes reionized by the emission of ionizing radiation from astrophysical sources, the ionized gas bubbles grow and merge in a process called percolation. As neighbouring regions of ionized gas run into one another, percolation causes sudden and dramatic increases in bubble sizes within a short period of time. The top panel of Fig. 2 illustrates the BSDs with each line showing the distribution for neutral fractions between 0.1 and 0.9 at select redshifts between 6 and 11, while the bottom panel quantifies the cumulative distribution functions reflecting the probability that an ionized bubble has an effective radius above the given R). The distributions are volume-weighted as each Cartesian grid cell contributes an equal number of MFP measurements. The bubbles tend to grow in size throughout reionization with the distributions being unimodal and most strongly peaked during the early ($x_{\rm H\,I} \gtrsim 0.9$, yellow line) and late ($x_{\rm H\,I} \lesssim 0.1$, dark purple line) phases of the EoR. In the former case, individual bubbles are forming and growing in isolation, yielding a fairly consistent characteristic size. Conversely, towards the end of the EoR, most of the ionized gas resides within extensive bubbles that have formed through the percolation of many smaller bubbles. The unimodality of the distributions indicates that there continues to be a well-defined characteristic bubble size throughout reionization, despite the complex growth and percolation processes. In addition, since we are considering volume-weighted statistics, the largest bubble will be primarily responsible for the distributions on the largest scales, while contributions from smaller bubbles are systematically repressed.

At higher redshifts when the first ionized bubbles are beginning to grow around ionizing sources, these structures tend to be small, on scales of a few hundred comoving kiloparsecs (ckpc). By a redshift of $z \sim 9$, the characteristic bubble size surpasses the threshold of $R_{\rm eff} = 1$ cMpc. Fig. 3 shows the median, mean, and log mean bubble sizes over time, with the shaded region representing the 16th–84th percentile range in the top panel. The global neutral fraction is plotted on the lower horizontal axis, while corresponding redshifts are marked on the upper horizontal axis. Predicted line-intensity mapping results from Kannan et al. (2022b) are shown in light blue and are consistent with the MFP estimates of bubble sizes found here, especially at lower redshifts. Additionally, the analytic model assuming no bubble overlap from Furlanetto & Oh (2005), shown as the black dash–dotted line, also agrees well with our bubble size measurements.

In the bottom panel of Fig. 3, we show median bubble sizes weighted by the number of haloes brighter than the threshold dust-attenuated UV magnitude $M_{\rm UV}$ according to the dust correction from Gnedin (2014) and Vogelsberger et al. (2020). Since brighter sources tend to reside in larger bubbles, these curves lie above the overall bubble sizes, more closely reflecting observations that are biased towards more luminous galaxies. Bubble size constraints from Lyα damping wings found in high-z galaxy spectra from *JWST* are shown with the dark blue circle (Hsiao et al. 2023) and purple squares (Umeda et al. 2023). The bubble size ($R < 0.2$ pMpc) and neutral fraction ($x_{\rm H\,I} > 0.9$) inferred by Hsiao et al. (2023) for the $z = 10.17$ triply lensed galaxy MACS0647-JD are consistent with our findings. However, the bubble sizes reported in Umeda et al.

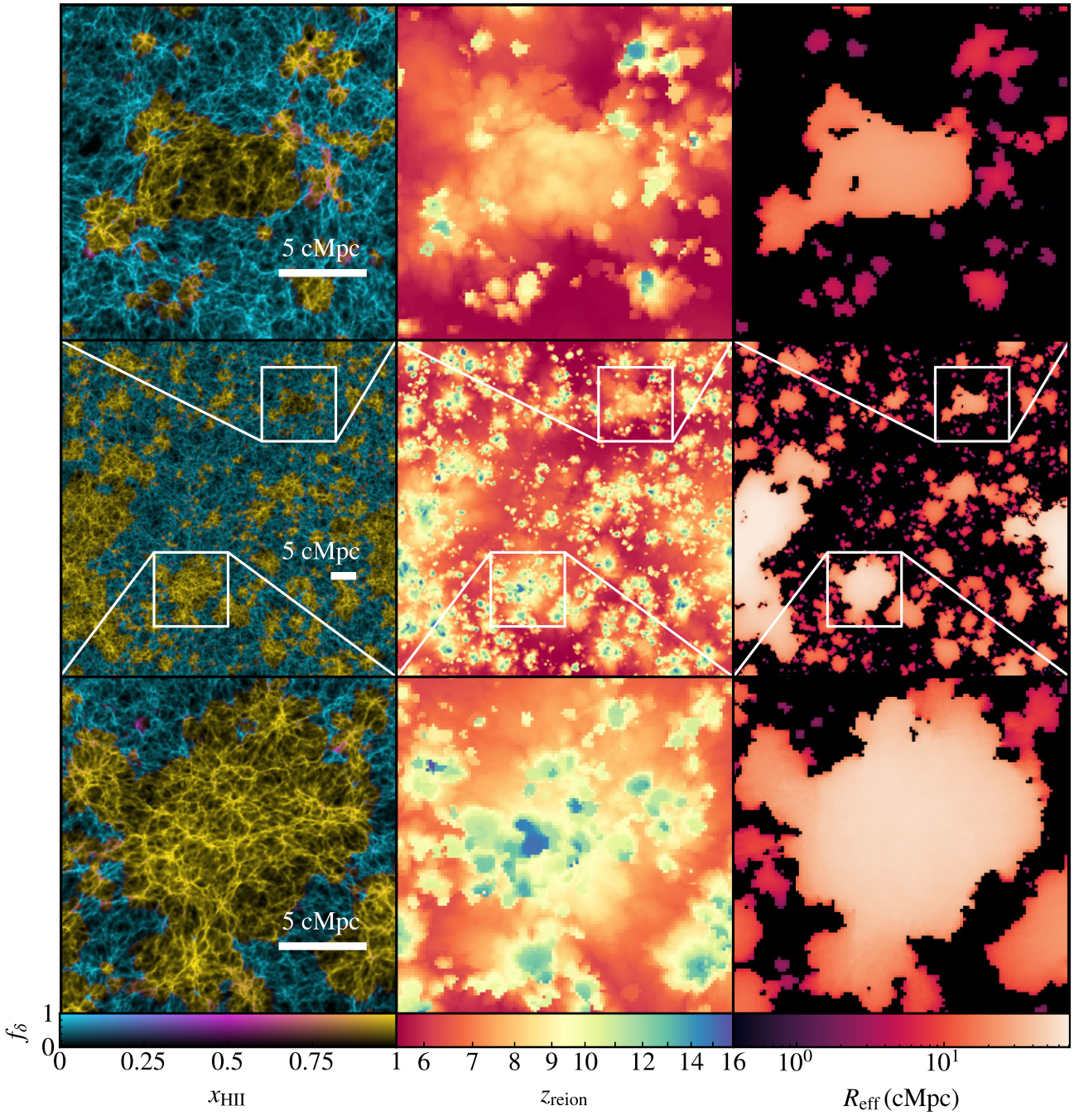

Figure 1. Image of the THESAN-1 box (centre row) and zoomed-in regions (top and bottom rows) illustrating the ionized fraction $x_{\rm H\,II}$ shaded by density ρ, redshift of reionization $z_{\rm reion}$, and effective bubble size $R_{\rm eff}$. The leftmost column shows a high-resolution projection through the central 10 per cent of the box with a two-dimensional colour bar such that the blue and yellow colours correspond to neutral and ionized gas, respectively, while the brightness scale corresponds to the logarithmic fractional overdensity, f_δ. The middle column shows an equivalent slice through the intermediate-resolution render box of the redshift of reionization, defined as the last redshift at which each cell is neutral ($x_{\rm H\,II} < 0.5$). The rightmost column shows the effective bubble size as seen by each cell, displaying sharp edges following the open or closed topology of the ionized regions. The cosmic web is not as prominent in the $z_{\rm reion}$ and $R_{\rm eff}$ images as in the density projection due to volume-weighting $x_{\rm H\,II}$ over scales larger than individual filament widths ($L_{\rm box}/512 \approx 0.2$ cMpc).

(2023) are higher than our estimates, particularly at lower redshifts toward the end of the EoR. This discrepancy is possibly due to their assumption of a homogeneous IGM, potentially leading to a systematic overestimation of neutral fractions during the EoR. Additionally, Keating et al. (2023) have found that many of these Ly α damping-wing observations can be explained with smaller bubble sizes and larger intrinsic Ly α fluxes from the host galaxies themselves. Late-time BSDs are also limited by the box size, which could be addressed by running larger-volume simulations in the future within the same framework. As a more concrete next step, we plan to investigate analytic biases with end-to-end forward modelling for a more fair comparison. Finally, with only a few observations it is difficult to simultaneously constrain the local line-of-sight bubble size and global evolution of the neutral fraction.

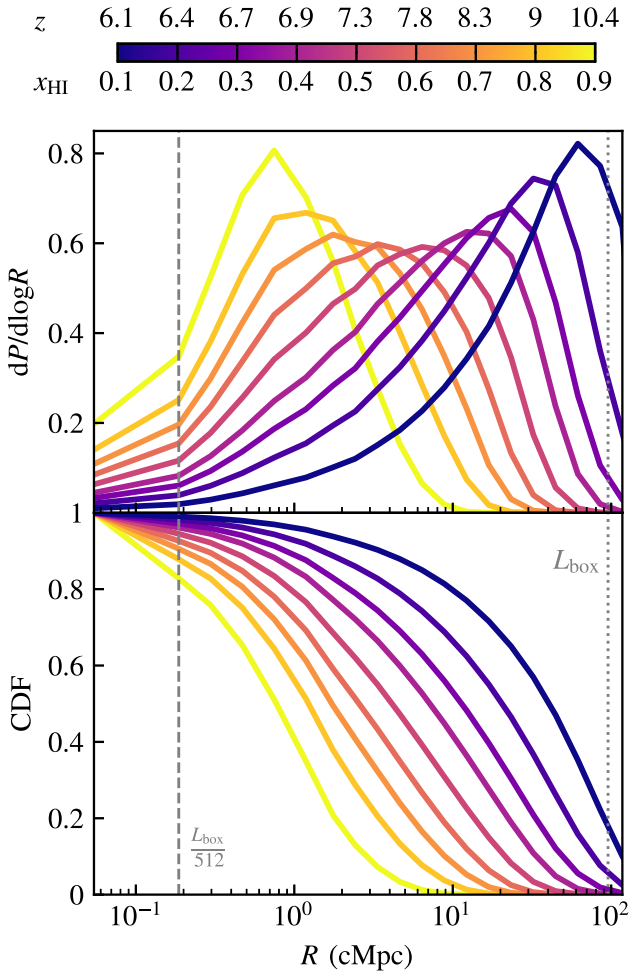

Figure 2. Ionized bubble size distributions for the fiducial THESAN-1 simulation. Each line shows the distribution at a different neutral fraction, or equivalently, redshift. All statistics are volume-weighted. The top panel shows the probability distribution function for bubble sizes and the bottom panel quantifies the cumulative distribution function for bubbles with effective radii $\geq R$. Bubble sizes grow over time as reionization progresses with the formation and growth of bubbles that eventually percolate to create much larger interconnected bubbles. There is a sharp peak at $R \sim 1$ cMpc for $x_{\rm HI} = 0.9$ (yellow line) as most bubbles have recently formed and have not yet merged. There is another sharp peak at $R \gtrsim 50$ cMpc for $x_{\rm HI} = 0.1$ (dark purple line) once most of the ionized cells are part of large regions spanning the simulation volume.

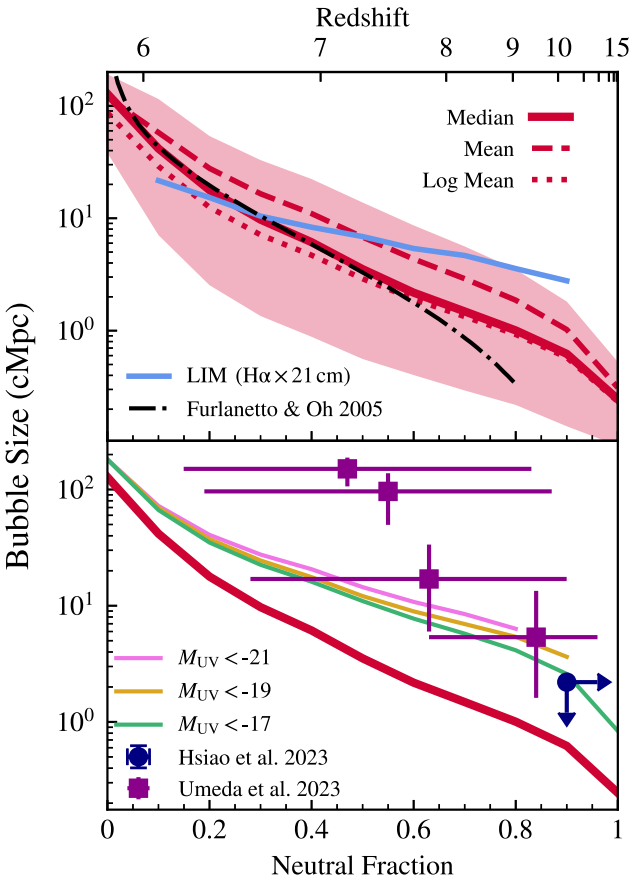

Figure 3. Redshift and global neutral fraction evolution of the characteristic bubble sizes for the fiducial THESAN-1 high-resolution simulation renderings. The median, mean, and log mean bubble sizes are shown as a function of neutral fraction and redshift in the top panel. The shaded region indicates the 16th–84th percentile range. We compare to the THESAN line intensity mapping (LIM) results from Kannan et al. (2022b) in light blue, which are fairly consistent at redshifts below about 9. We also compare with the analytic model from Furlanetto & Oh (2005) in the black dash–dotted line. The bottom panel shows median bubble sizes weighted by the number of bright haloes according to the threshold dust-corrected UV magnitude M_{1500}. These are compared with the observationally inferred bubble size constraints from Hsiao et al. (2023) in the dark blue circle and from Umeda et al. (2023) in the purple squares.

The variation in bubble size is relatively small at the beginning and end of the EoR. At very high redshifts ($z \gtrsim 10$), ionized bubbles have only recently begun to form around luminous objects and have not yet grown enough to instigate percolation, a process that rapidly increases individual bubble sizes. Consequently, nearly all of the first ionized bubbles remain below 1 cMpc in size, as most galaxies are not yet massive enough to sustain high star formation rates. As the bubbles continue to expand due to radiation continuously ionizing more and more gas around the source, the bubbles begin to percolate and coalesce. This leads to a relatively large spread in bubble sizes during the intermediate stages of the EoR mirroring the asynchronous nature of reionization ($6 \lesssim z \lesssim 10$). At even later stages ($z \lesssim 6$), as the majority of bubbles become encompassed by larger neighbouring bubbles, the sizes reflect the vast network of interconnected ionized regions. By this point near the end of the EoR, these immense bubbles housing the majority of the ionized gas all merge with each other until there is effectively one large bubble filling nearly the entire simulation box, which is dominant in volume-weighted statistics. This is indicated by the median bubble size at neutral fractions close to zero becoming larger than the side length of the box.

Regions that become ionized at different times during the EoR have dramatically different size evolution patterns. To highlight this, we explore the typical bubble growth trajectories for regions reionized at different redshifts. Categorizing by the redshift of reionization, $z_{\rm reion}$, allows us to identify different populations, each with unique bubble size histories over time. The $z_{\rm reion}$ of a simulation cell is defined as the lowest redshift at which the cell transitions from neutral ($x_{\rm HI} < 0.5$) to ionized ($x_{\rm HI} \geq 0.5$). We examine the evolution of bubble sizes for seven $z_{\rm reion}$ ranges delineated by the bin edge values: $z_{\rm reion} \in [5.5, 6, 7, 8, 10, 12, 15, 20]$. The median bubble size trajectories for these different $z_{\rm reion}$ ranges are shown in Fig. 4.

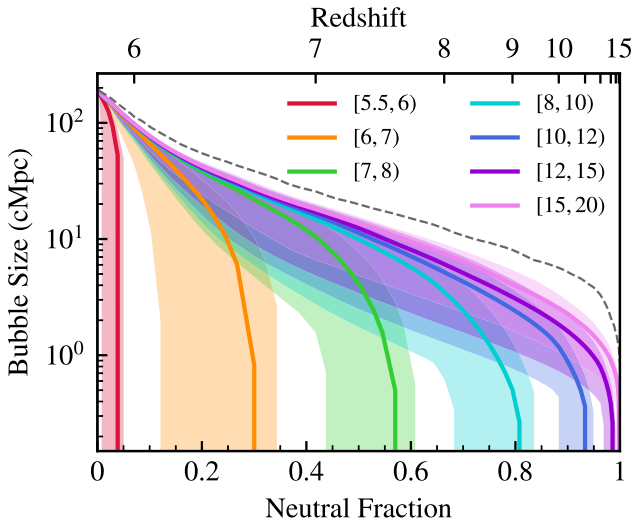

Figure 4. Bubble size trajectories over time grouped by the local redshift of reionization $z_{\rm reion}$ for the high-resolution fiducial THESAN-1 model. Each solid curve shows the median effective bubble size $R_{\rm eff}$ from all histories within each $z_{\rm reion}$ range, while the shaded regions are the 16th–84th percentile ranges. The dashed grey curve shows the largest bubble size measured from each snapshot. We define $z_{\rm reion}$ of each cell to be the lowest redshift at which the cell crossed from neutral ($x_{\rm H\,II} < 0.5$) to ionized ($x_{\rm H\,II} \geq 0.5$). By $z \sim 6$, regions are flash-ionized because they immediately join a large pre-existing bubble.

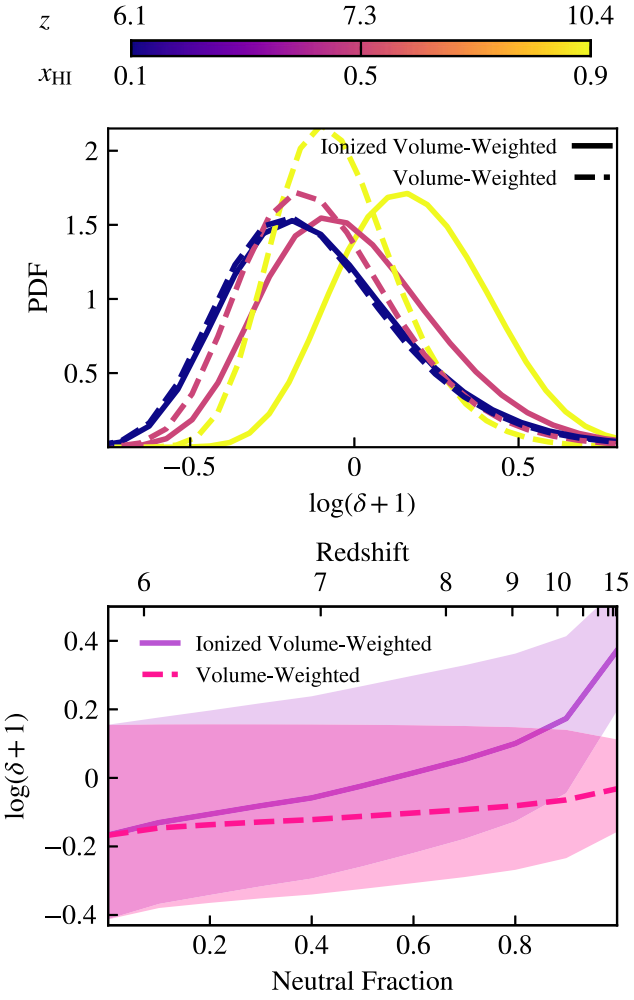

Figure 5. *Top panel*: Probability density functions of local environmental overdensity for the entire simulation volume (dashed lines) and just the ionized volume (solid lines), shown at times when the global neutral fraction is 0.1, 0.5, and 0.9. *Bottom panel*: The median overdensity and 16th–84th percentile range as functions of neutral fraction and redshift. The distributions are nearly non-overlapping at the earliest times but quickly converge as the entire simulation box becomes ionized.

Cells with $z_{\rm reion} \geq 15$ (pink line) are the first cells to become ionized and initiate the reionization process. The bubbles around these cells quickly reach an effective radius of about 1 cMpc, then begin to grow roughly exponentially with the declining neutral fraction until they reach the maximum allowed bubble size of ∼200 cMpc, marking the point when the entire box is reionized. Cells with $z_{\rm reion} \gtrsim 10$ follow a similar trajectory, indicating that most of these cells become ionized without immediately joining a pre-existing large bubble. The bubble sizes for these early-reionized cells start small (∼cMpc scales) and then exhibit a gradual growth pattern, similar to delayed versions of the first ionized regions.

Once $z_{\rm reion} \lesssim 10$, many of the newly ionized cells promptly merge with existing ionized bubbles, which have sizes ranging from a few cMpc to tens of cMpc for $7 \lesssim z_{\rm reion} \lesssim 10$ and $6 \lesssim z_{\rm reion} \lesssim 7$, respectively. Following the main mergers into large pre-existing ionized regions, the bubble sizes grow at a more subdued rate similar to the growth observed in earlier-reionized cells. This gradual assimilation process is coherent, considering that the later-ionized cells are likely integrated into the very same bubbles that the early-reionized cells began to form. Thus, this slower expansion may track a characteristic bubble growth rate persisting throughout the procession of reionization. The large vertical scatter in bubble size indicates that inside-out reionization is still occurring even at $z \lesssim 10$ (Kannan et al. 2022a; Borrow et al. 2023). New small bubbles are still forming in neutral areas even after much of the IGM has become ionized and percolation of large bubbles has become a significant cause of bubble size increases.

The most rapid growth rates are observed right at the culmination of reionization, specifically after a redshift of 6, shown in the red line in Fig. 4. These late-reionization regions are 'flash-ionized' as they immediately become part of the last large pre-existing bubble which already fills most of the simulation box. These cells represent some of the final neutral island regions in the simulation. Thus, no clear bubble boundaries are discernible by the time they become ionized, as the ionized gas essentially forms one final space-filling bubble.

4 CONNECTIONS TO THE LOCAL ENVIRONMENT

In an effort to understand the conditions that influence ionized bubble properties, we investigate the relationship between the sizes of ionized bubbles and the local environmental overdensity. We smooth the overdensities from the Cartesian grid on the scale of 125 ckpc such that the overdensities are probing the local environment as opposed to smaller scale density fluctuations. We begin by examining the distribution of overdensities within our simulated volume at different stages of reionization. These distributions, both volume-weighted and ionized volume-weighted, are shown in the top panel of Fig. 5. The ionized volume-weighted distributions are shifted towards higher overdensities early in reionization as compared to their full volume-weighted counterparts. This indicates that regions that are

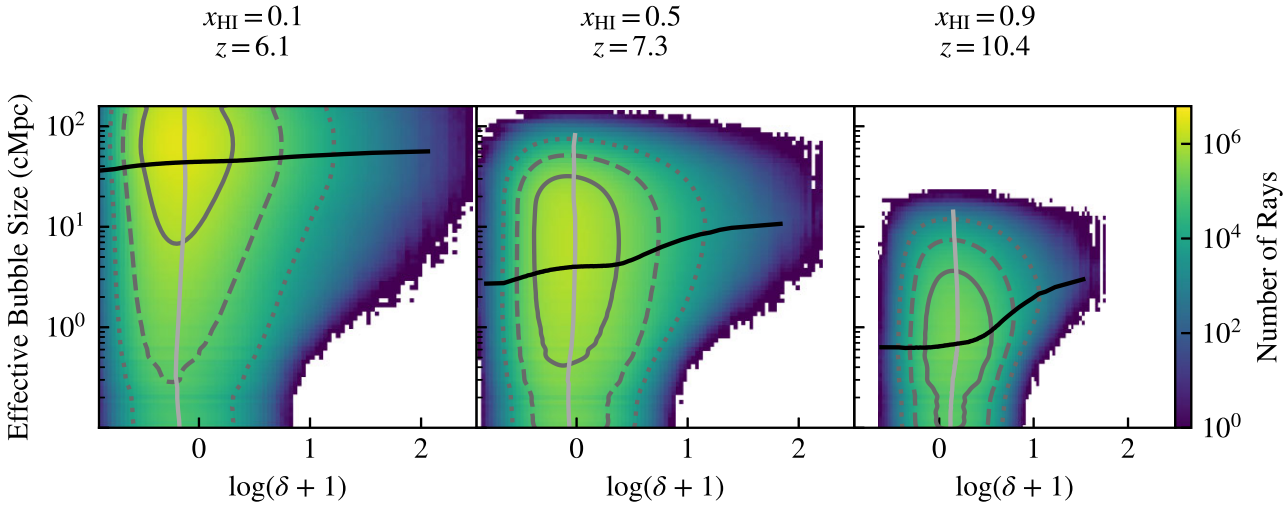

Figure 6. Characteristic ionized bubble sizes compared to overdensity for THESAN high resolution cube smoothed on 125 ckpc scales at different neutral fractions $x_{\rm H\,I}$ and their corresponding redshifts. Grey contours show the 1σ, 2σ, and 3σ regions in solid, dashed, and dotted lines, respectively. While the contours do not indicate a strong correlation between overdensity and bubble sizes, the shape of the histogram indicates that regions with high local overdensities preferentially have large bubbles. The black solid lines show the median bubble size at each overdensity, and the light grey lines show the median overdensity at each bubble size.

more overdense tend to be ionized early and that many of the first regions that are reionized reside in locally overdense environments. This is also reflected in the bottom panel of Fig. 5 which shows the median and 16th–84th percentile range of overdensity with both weighting schemes. We again see that at high redshifts, the ionized volume-weighted median of overdensity is larger than the volume-weighted median. These medians and one sigma ranges converge to the same values by the end of reionization when nearly all of the volume of the box contains ionized gas. Specifically at global neutral fractions of $x_{\rm H\,I} = \{0.9, 0.5, 0.1\}$ the bias in the median overdensity is $\log((\delta_{\rm H\,II} + 1)/(\delta + 1)) = \{0.24, 0.09, 0.02\}$.

We move on to investigate the specific correlation between local environmental overdensity and ionized bubble sizes. For consistency with the overdensity analysis, we smooth the bubble sizes on the same 125 ckpc scales to focus on the most 'significant' bubbles in the local area. In Fig. 6, we show two-dimensional histograms in the δ–$R_{\rm eff}$ plane highlighting various environmental effects, where each panel corresponds to a different neutral fraction and its corresponding redshift. The contours outline the central 68, 95, and 99.7 percentile ranges to guide the eye in discerning the volume-weighted statistics. The solid black curves show the median effective bubble size for each of the overdensity bins, and the light grey curves show the median overdensity for each of the bubble size bins. In summary, the local overdensity has the strongest impact early on during the formation phases. As reionization progresses, there is a significant deficit of small bubbles in overdense regions whereas underdense regions can still harbour neutral structures.

At early times during the EoR, when 90 per cent of the gas remains neutral, bubble sizes are generally small with $R_{\rm eff} \lesssim$ a few cMpc, but there is a slight correlation between overdensity and bubble sizes, as shown by the asymmetrical shape of the 2D histograms in Fig. 6. By examining the central 68 percentile contour in the rightmost panel of Fig. 6, we see that the smallest bubbles primarily reside in regions with $0 \lesssim \log(\delta + 1) \lesssim 0.5$. This indicates that initially these bubbles are primarily forming in overdense regions before expanding into surrounding areas with a wider range of densities, as depicted by the broader shape of the contours in the higher bubble size region of the panel. As reionization progresses and more of the gas becomes ionized, the overdensity dependence washes out, likely due to ionized regions joining pre-existing bubbles. By the time the EoR is almost over and 90 per cent of the gas is ionized, most of the bubbles are large ($R_{\rm eff} \gtrsim 10$ cMpc) and there is no significant correlation with particular overdensity values.

We focus on the relationship between bubble sizes and overdensity to determine whether bubbles predominantly form and grow within high-density regions or low-density regions. Higher-density regions typically contain more ionizing sources, but their higher density of initially neutral gas makes them more difficult to ionize. In the top panel of Fig. 7, we show the median bubble size in relation to the corresponding overdensity at which it is centred. We employ adaptive binning in overdensity to improve the statistical significance at the highest and lowest densities. Across all redshifts that we examined, a clear pattern emerges where the bubble size increases with overdensity. This indicates that the largest bubbles are more likely to originate from and surround denser regions. However, this trend becomes less pronounced as reionization progresses and bubbles undergo percolation, coalescing into fewer ionized regions. As a result of structure formation, the late-time ionized landscape is volume-biased towards being underdense.

In the bottom panel of Fig. 7, we also illustrate the relationship between the variance in logarithmic bubble sizes and overdensity. Notably, the variance is significantly higher within low-density regions compared to more overdense areas. This is likely due to void-like regions becoming ionized by pre-existing bubbles growing into them (outside-in reionization) as well as some sources creating ionized bubbles within the low-density regions themselves (inside-out reionization). Like the median bubble size, the variance in bubble sizes also flattens out over time as reionization progresses. The variance of bubble sizes in the low-density regions starts to decrease at late times while the variance in overdense regions continues to increase throughout the EoR. The increase in variance in dense regions is likely due to the increasing number of bright sources over time. Although many of the largest bubbles originate from areas of higher overdensity, the formation of new sources in neutral regions

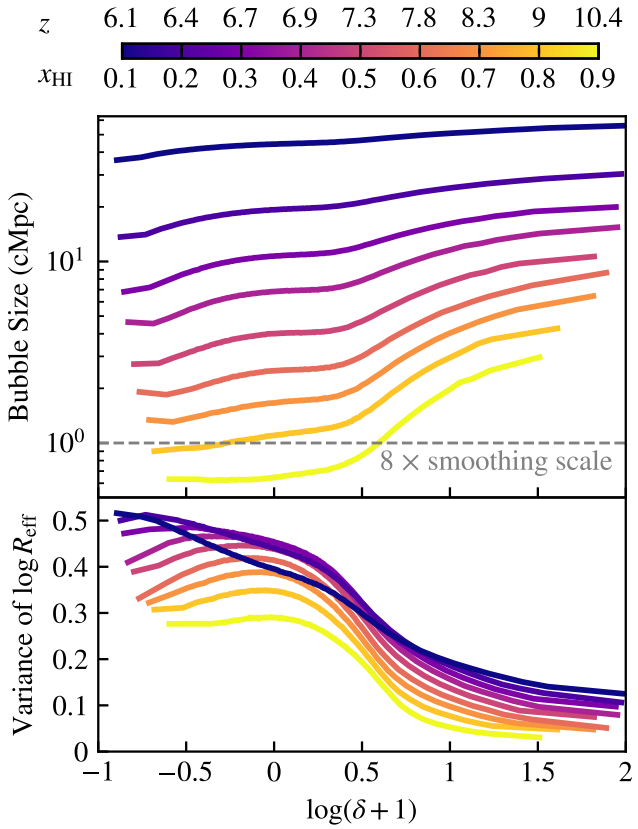

Figure 7. Median bubble size $R_{\rm eff}$ as a function of overdensity at different ionized fractions for the fiducial model smoothed on 125 ckpc scales. The grey dashed line shows eight times the smoothing scale. Larger overdensities tend to have larger bubble sizes, especially at early times when the bubbles are initially forming. At later times, when most of the gas in the simulation box is ionized, bubble sizes have a much weaker dependence on overdensity, likely due to the prevalence of few large bubbles which encompass most of the ionized volume. Overall, there is larger variance in bubble sizes at lower overdensities. The variance flattens out with time such that void-like regions start to decrease while overdense regions increase, which in the latter case is primarily driven by the increasing number of bright objects with time.

results in the simultaneous creation of many smaller bubbles as well. This can explain the continual increase in bubble size variance in overdense regions as the largest bubbles get bigger, but the small bubbles continue to form due to the high density of haloes.

Further analysis reveals that, at $z = 10$, bubbles situated in the highest overdensity regions are approximately 20 per cent larger than the global median size. However, this disparity diminishes to less than 5 per cent by $z = 7$. This is again due to ionized bubbles primarily forming in very overdense regions early on and then expanding and merging to encompass underdense regions later in the EoR.

We also investigate how bubble sizes around galaxies depend on the galaxy properties, specifically halo mass stellar mass, and UV magnitude, as these relationships are of much interest to the observational and theoretical communities (Davies et al. 2023; Hayes & Scarlata 2023). Our findings, summarized in Fig. 8, reveal that the median bubble size around different mass halos captures the correlation of these characteristics with expected bubble sizes throughout the EoR. Additionally, we examined the variance in bubble sizes for these galactic properties. The observed trends closely align with the local overdensity analysis: more massive, brighter galaxies are associated with larger bubble sizes with lower variance, especially during the earlier phases of reionization. As the EoR concludes, both the median and variance in bubble sizes flatten out, suggesting a weaker connection between galaxy properties and effective bubble sizes. This is consistent with the notion that bubbles have undergone percolation, resulting in a small number of dominant ionized regions.

We quantify the degree of correlation between the effective bubble size and halo mass by fitting a power-law model to the $R_{\rm eff}$ versus $M_{\rm halo}$ curves at each of the neutral fractions shown in Fig. 8. We note the best-fitting power-law slope only considers halo masses in the range $M_{\rm halo} \geq 10^9\,{\rm M}_\odot$. The best-fitting parameter values are listed in Table 1. The evolution of this metric throughout the EoR as shown in Fig. 9 clearly shows the strength of the early correlation and robustness of the flattening with time. We plot this power-law slope, $\mathrm{d}\log R_{\rm eff}/\mathrm{d}\log M_{\rm halo}$, as a function of the global neutral fraction for the fiducial, high-resolution THESAN-1 run (red curve) as well as the medium resolution runs THESAN-2 and THESAN-WC-2 (blue and purple curves). Specifically, THESAN-2 uses the same model as THESAN-1 but with two (eight) times coarser spatial (mass) resolution, and THESAN-WC-2 exhibits weak convergence of $x_{\rm H\,I}(z)$ by increasing the birth cloud escape fraction from 0.37 to 0.43 to compensate for additional unresolved low-mass haloes. Each of the THESAN-2 models has been analysed on a Cartesian grid with 256 cells per side, so we also show the small difference by re-calculating THESAN-1 on this lower resolution grid for consistency (red dashed curve). More details of the different models can be found in Kannan et al. (2022a) and Garaldi et al. (2023). Our analysis reveals that the higher-resolution THESAN-1 simulation exhibits the steepest power-law slope between bubble size and halo mass. We attribute this to the abundant low-mass haloes and dense structures of clumps and filaments that are less resolved in the lower-resolution runs. Further discussion of the convergence of bubble size statistics across the different resolution THESAN runs is contained in Appendix C.

5 CONNECTIONS TO THE REDSHIFT OF REIONIZATION

The redshift of reionization, denoted as $z_{\rm reion}$, serves as a powerful probe of the spatiotemporal distribution of ionizing radiation during the EoR. In this work, we define $z_{\rm reion}$ for each point in space as the latest redshift at which the Cartesian grid cell transitions from a neutral to an ionized state, specifically when the ionized hydrogen fraction reaches $x_{\rm H\,II} \geq 0.5$. To obtain an accurate estimate of $z_{\rm reion}$, we employ linear interpolation between the simulation snapshots, each separated by a time cadence of $\sim$10 Myr.

To investigate the relationship between $z_{\rm reion}$ and $R_{\rm eff}$, we calculate the last redshift at which each cell has an effective bubble size that crosses a specified threshold radius, denoted as z ($R_{\rm eff} \geq R_0$). As a first glance, we present the distributions of $z_{\rm reion}$ and z ($R_{\rm eff} \geq R_0$) for various threshold radii in Fig. 10. Our analysis reveals a reasonably good agreement between the distributions of $z_{\rm reion}$ and the redshift at which bubble sizes surpass a threshold radius when $R_{\rm eff} \lesssim 1$ cMpc.

As the threshold for bubble size increases, two notable shifts occur in the distribution: the peak moves slightly towards lower redshifts, and the high-redshift tail contracts. At high redshifts, bubbles are predominantly small, leading to the omission of many of the earliest ionized regions when higher bubble size thresholds are applied. This results in the attenuation of the slow-rising high-redshift tail observed in $z_{\rm reion}$. Additionally, the distributions are slightly shifted towards lower redshifts compared to $z_{\rm reion}$ due to the time it takes between a cell becoming ionized and the subsequent growth of a bubble to the

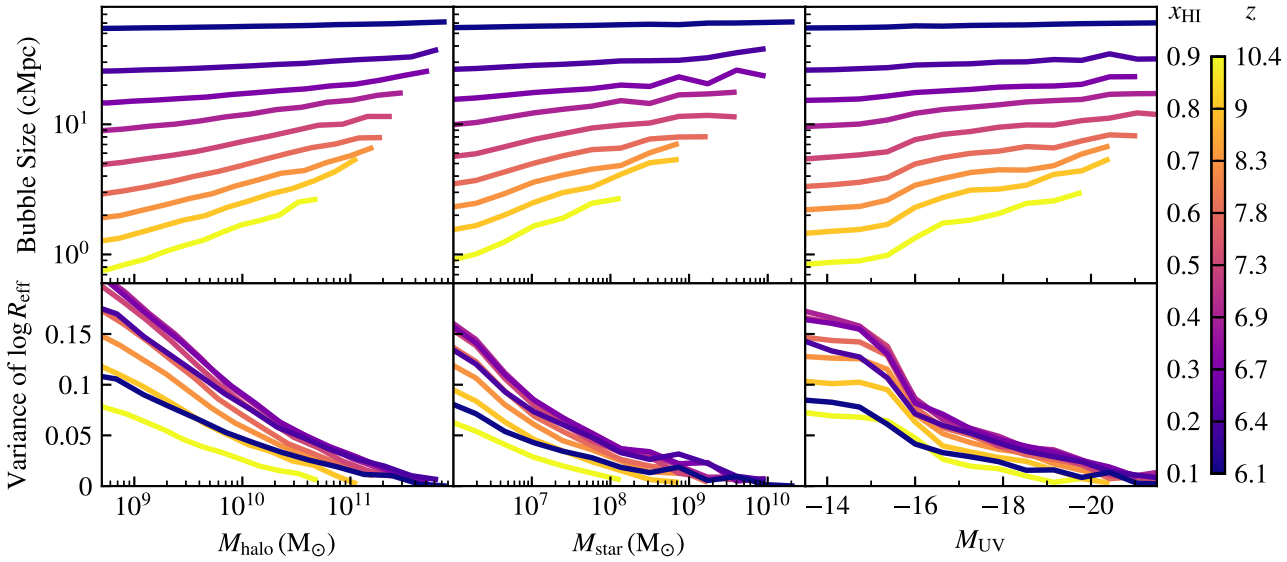

Figure 8. Similar to Fig. 7, we compare median bubble sizes and variances as functions of galactic halo mass, stellar mass, and UV magnitude. Haloes are resolved down to masses of $M_{\rm halo} \sim 10^8\,h^{-1}\,{\rm M}_\odot$, so resolution effects should not be significant in the range considered. We find similar trends to those for overdensity but as these haloes host the sources of reionization the dependence extends across the entire resolved halo range, exhibiting a characteristic power-law behaviour.

Table 1. Power-law fits to the bubble size–halo mass relationship (shown in the top left panel of Fig. 8) in the form $R_{\rm eff}/{\rm cMpc} = R_0\,(M_{\rm halo}/{\rm M}_\odot)^\alpha$ for each neutral fraction, $x_{\rm H\,I}$ (and corresponding redshift, z) for THESAN-1. The power-law slope parameter, α, is plotted in Fig. 9 along with those from lower resolution simulations.

$x_{\rm H\,I}$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
z	6.1	6.4	6.7	6.9	7.3	7.8	8.3	9.0	10.4
α	0.016	0.050	0.084	0.11	0.15	0.19	0.22	0.27	0.29
R_0	39	9.0	2.5	0.87	0.21	0.067	0.021	0.0051	0.0020

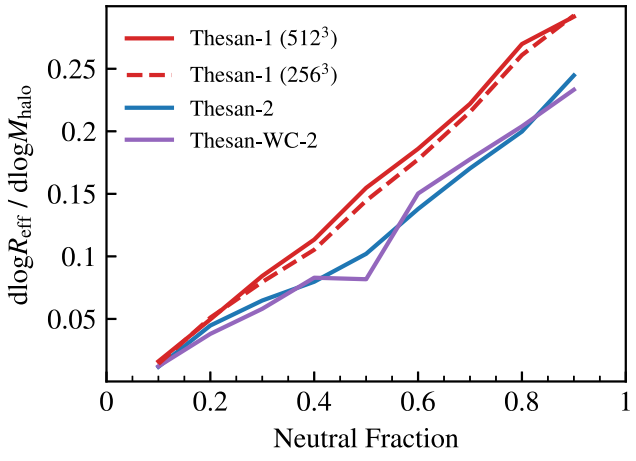

Figure 9. Power-law slope for $R_{\rm eff}$ versus $M_{\rm halo}$ over neutral fraction for the fiducial, high-resolution THESAN-1 run (red line) as well as the medium resolution runs THESAN-2 (blue line) and THESAN-WC-2 (purple line). We also compare to THESAN-1 on a grid with 256 cells per side (red dashed line) for consistency with the THESAN-2 grid resolution. For all runs, the power law is steepest (highest $\mathrm{d}R_{\rm eff}/\mathrm{d}M_{\rm halo}$) early in the reionization process driven by the cosmic structure formation but flattens as bubbles expand into void regions.

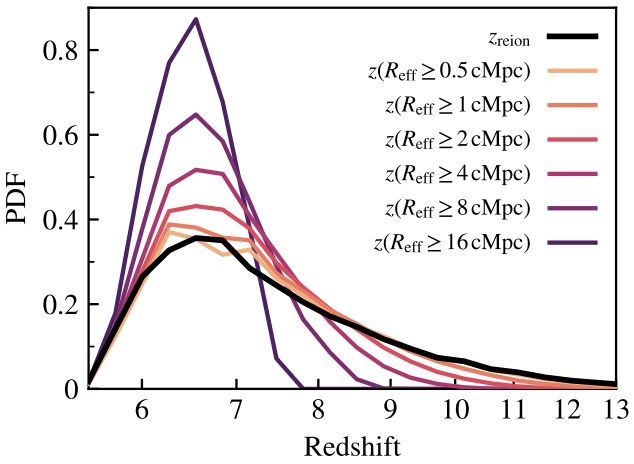

Figure 10. Probability density functions for $z_{\rm reion}$ (in black) and the redshift at which the bubble size surpasses various thresholds. We see that $z_{\rm reion}$ is a good tracer of bubble sizes below about 1 cMpc, but time delays between reionization and bubble growth weaken the correspondence between $z_{\rm reion}$ and larger bubble size statistics.

threshold size. The time delay between reionization and growing to bubble sizes larger than ~ 1 cMpc, washes out the correspondence between $z_{\rm reion}$ and the redshift at which a bubble grows to that size.

In Fig. 11, we more closely examine the time lags between cell reionization and the attainment of specific bubble sizes. We find that larger bubble size thresholds directly translate to longer time lags between these two events, while $z_{\rm reion}$ is a good tracer of small bubble sizes below approximately 1 cMpc due to much shorter time lags. For all thresholds, this relative redshift difference with respect to $z_{\rm reion}$ is more pronounced earlier in the EoR, likely due to the initially small sizes of bubbles early on in reionization. By the time a large fraction of the simulation volume becomes ionized, this time lag becomes

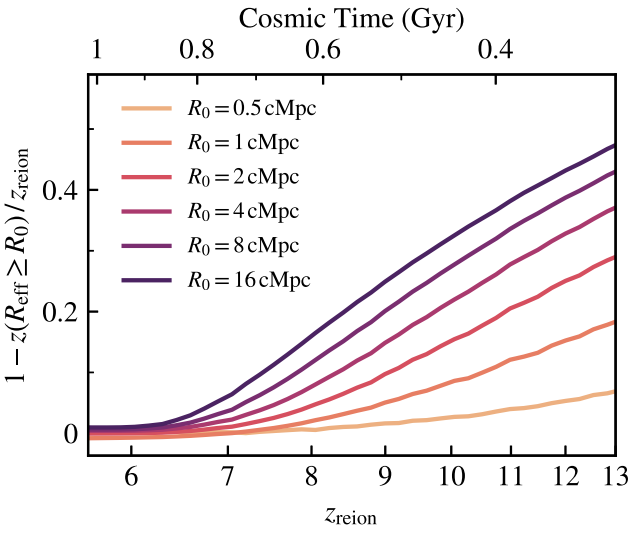

Figure 11. Relative difference between redshift of reionization and redshift at which the bubble size exceeds some threshold size R_0 plotted against $z_{\rm reion}$. This illustrates the delay between reionization and bubble growth throughout the EoR, with the largest time lags corresponding to higher redshifts.

much smaller on average. This is because newly ionized regions are more likely to immediately coalesce into large, pre-existing bubbles or otherwise flash ionize, thereby reducing the time required to reach the specified $R_{\rm eff}$ values.

To further quantify the bubble growth after the initial reionization of the region, we study the times and rates at which bubbles double in radius. Bubbles that have become ionized at different redshifts during the EoR have unique bubble growth properties. We specifically look at the doubling time (time for the bubble radius to double in size), growth rate (change in effective volume over doubling time), and effective speed of ionization fronts (change in effective radius over doubling time) in Fig. 12. Each colour point indicates the radius change we are considering (i.e. the dark purple shows the doubling times and growth rates for bubbles growing from 8 cMpc in radius to 16 cMpc). The horizontal axis is $z_{\rm reion}$ allowing us to view how different populations' bubbles grow differently, whether they were ionized early in reionization or later. We note that both growth rate and ionizing front speed are calculated assuming spherical bubbles; thus, they may not fully represent the morphological complexities present once bubbles coalesce.

By inspection of Fig. 12, we see that the latest ionized regions on the left side of the plots have the shortest doubling times and, consequently, the fastest growth rates. This is consistent with our picture of 'flash ionization' of regions late in the EoR as they quickly join large bubbles after becoming ionized, i.e. runaway percolation. The doubling times for regions ionized before $z \approx 12$ are similar and quite high, around 100 Myr, even for the different bubble radii. This could be partially due to the long time between local reionization in the region and the end of the EoR, but may also indicate exponential behaviour in the growth rate of the radius, as the characteristic doubling time appears to be independent of the bubble radius.

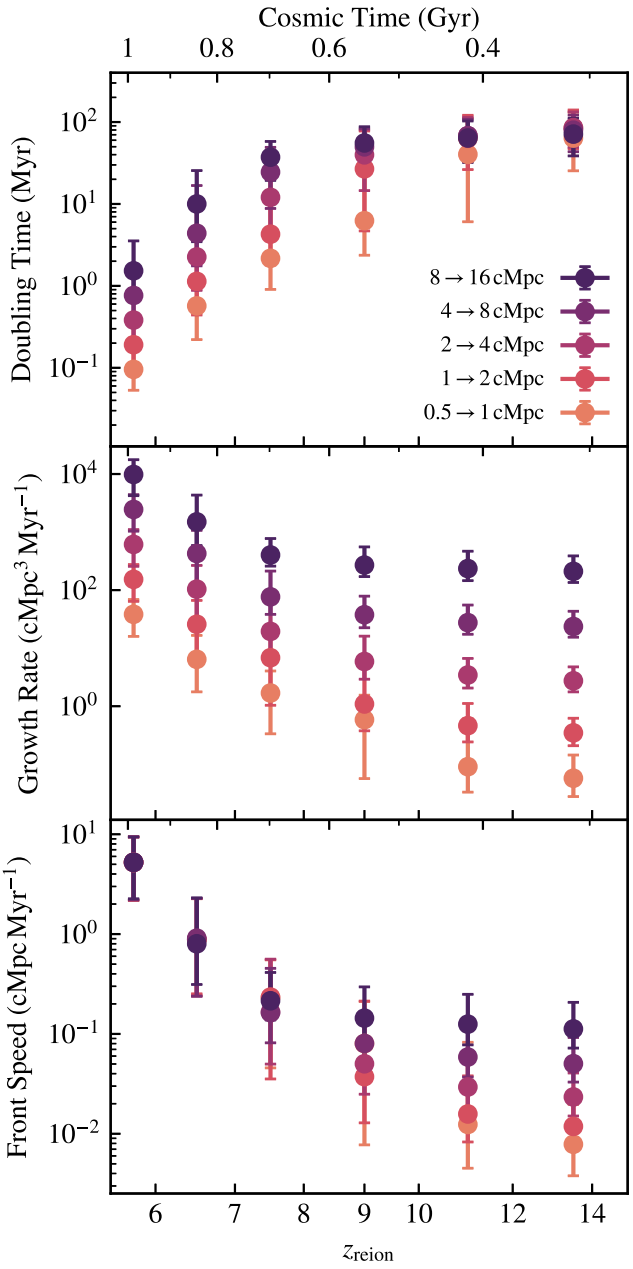

Figure 12. Top panel: Median and 16th–84th percentile range of doubling time grouped by local $z_{\rm reion}$. Middle panel: Median and 16th–84th percentile range of growth rate of bubbles (change in effective volume over doubling time) in the same $z_{\rm reion}$ ranges. Bottom panel: Median and 16th–84th percentile range of the speed of ionization fronts in the same $z_{\rm reion}$ ranges, assuming spherical bubbles. Each colour corresponds to corresponding bubble size thresholds; e.g. dark purple points show the doubling times and growth rates for bubbles growing from 8 to 16 cMpc in radius. The later a region is locally reionized, the longer the doubling time, and the faster the growth rate. This is consistent with the picture of slow bubble growth early in reionization and 'flash ionization' and runaway percolation at lower redshifts.

6 CONCLUSIONS

One of the most promising methods to study the large-scale processes during the EoR is the topological analysis of ionized bubbles. A significant body of work has been dedicated to developing procedures for identifying and characterizing these bubbles from idealized calculations, semi-analytical models, and fully coupled radiative transfer simulations. In this context, this work aims to study ionized bubble size statistics during reionization as modelled by the THESAN simulations. We primarily employ the MFP method, which calculates bubble sizes by extending rays from ionized cells until they intersect with neutral cells, as an effective technique for determining

ionized bubble sizes. We examined the time evolution of the bubble sizes, connections between sizes and environmental properties, and the utility of $z_{\rm reion}$ as an effective proxy for tracking bubble size characteristics.

Our detailed analysis of the THESAN simulations has led to several insights into the nature of ionized bubbles during the EoR. We summarize our main findings as follows:

(i) Bubbles begin to form coincident with the first significant ionizing sources around redshift $z \sim 15$ and exhibit a growth trajectory from $R_{\rm eff} \sim 100\,$ckpc to $R_{\rm eff} \sim 100\,$cMpc by redshift $z \sim 5.5$, indicative of the progressive completion of cosmic reionization within the simulation volume.

(ii) Bubble sizes are notably larger around bright galaxies. This suggests that observational measurements of bubble sizes around high-redshift sources are likely skewed towards larger sizes, and accounting for this bias helps reconcile the discrepancy between predicted and observed bubble sizes.

(iii) We find redshift-dependent variations in bubble growth rates. Specifically, regions that undergo reionization early in the EoR display relatively slow bubble expansion. In contrast, regions that are reionized later in the EoR experience 'flash ionization' and rapidly join larger, pre-existing bubbles.

(iv) Bubbles in high-density regions are generally larger than those in low-density regions, particularly when $x_{\rm H\,I} \gtrsim 0.5$ in the earlier stages of the EoR. This correlation is most pronounced at very high overdensities before the bubbles have expanded and merged enough to wash out correlations with environment. A similar trend exists for bubbles surrounding massive galaxies. In both cases, we characterize the median and variance in bubble sizes, and show that the power-law slope $\mathrm{d}\log R_{\rm eff}/\mathrm{d}\log M_{\rm halo}$ flattens as reionization progresses.

(v) We comment on the utility of the redshift of reionization ($z_{\rm reion}$) as a viable theoretical tracer for small bubble sizes, particularly for radial sizes below $\lesssim 1\,$cMpc. This provides a convenient local metric for studying non-local phenomena and capturing the temporal history of ionized regions. We note that at high redshifts, there exists a substantial time lag between local reionization and large bubble growth, but by $z \lesssim 6.5$, this delay significantly shortens.

In summary, this work enhances our understanding of the topological and evolutionary properties of ionized bubbles during the EoR, including connections between bubble sizes and both small-scale galaxy and large-scale environmental properties.

There is also a wealth of new observational data available from high-z galaxy spectra from the *JWST*. These data can be harnessed to constrain both ionized bubble sizes and the neutral fraction through the analysis of Ly$\,\alpha$ damping wings. Thus, it would be optimal to identify bubble sizes through a synthetic observation framework that models the spectra and data reduction pipeline currently in use. It would be natural to extend this study to incorporate similar systematics to the observations and facilitate more complete physical connections to theoretical topological metrics.

Further follow-up studies to this could also include performing a more extensive analysis of bubble sizes across the other simulations in the THESAN suite, which include different dark matter models, escape fractions, and dominant sources of ionizing radiation. Such comparative studies would elucidate the distinct signatures imprinted by different physical processes within the ionized bubbles. Moreover, we plan to perform a complementary topological examination of $z_{\rm reion}$ to analyse the formation and evolution of ionized regions, which will provide invaluable information about the mechanisms by which the Universe became ionized. Follow-up work will aim to continue to draw conclusions about the interplay of reionization and galaxy formation, including both local and environmental information.

ACKNOWLEDGEMENTS

We thank the anonymous referee for their insightful comments and suggestions. We thank Koki Kakiichi and Laura Keating for insightful discussions related to this work. AS acknowledges support under the Institute for Theory and Computation Fellowships at the Center for Astrophysics|Harvard & Smithsonian. MV acknowledges support through NASA ATP grants 19-ATP19-0019, 19-ATP19-0020, and 19-ATP19-0167, and NSF grants AST-1814053, AST-1814259, AST-1909831, and AST-2007355. EG acknowledges support from the CANON Foundation Europe through the Canon Fellowship programme during part of the work presented in this paper. The authors gratefully acknowledge the Max Planck Computing and Data Facility (https://www.mpcdf.mpg.de/) for support in hosting data and releasing them to the public, as well as the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Centre (www.lrz.de). Additional computing resources were provided by the Engaging cluster supported by the Massachusetts Institute of Technology. We are thankful to the community developing and maintaining software packages extensively used in our work, namely MATPLOTLIB (Hunter 2007), NUMPY (Walt, Colbert & Varoquaux 2011), and SCIPY (Jones et al. 2001).

DATA AVAILABILITY

All data produced within the THESAN project are fully and openly available at https://thesan-project.com, including extensive documentation and usage examples (Garaldi et al. 2023). We invite inquiries and collaboration requests from the community. In conjunction with this paper, we add effective bubble size grid calculations to our online repository. Furthermore, a minimal version of our parallelized C++ ray-tracing MFP code is accessible at https://github.com/meredithneyer/mfp-bubbles-thesan.

REFERENCES

Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Barnes J., Hut P., 1986, Nature, 324, 446
Bera A., Hassan S., Smith A., Cen R., Garaldi E., Kannan R., Vogelsberger M., 2023, ApJ, 959, 2
Bischetti M. et al., 2022, Nature, 605, 244
Borrow J., Kannan R., Garaldi E., Smith A., Vogelsberger M., Pakmor R., Springel V., Hernquist L., 2023, MNRAS, 525, 5932
Busch P., Eide M. B., Ciardi B., Kakiichi K., 2020, MNRAS, 498, 4533
Cain C., D'Aloisio A., Gangolli N., McQuinn M., 2023, MNRAS, 522, 2047
Choudhury T. R., Haehnelt M. G., Regan J., 2009, MNRAS, 394, 960
Ciardi B., Stoehr F., White S. D. M., 2003, MNRAS, 343, 1101
Davies J. E., Bird S., Mutch S., Ni Y., Feng Y., Croft R., Matteo T. D., Wyithe J. S. B., 2023, MNRAS, 525, 2553
DeBoer D. R. et al., 2017, PASP, 129, 045001
Dubroca B., Feugeas J., 1999, Academie des Sciences Paris Comptes Rendus Serie Sciences Mathematiques, 329, 915, https://ui.adsabs.harvard.edu/abs/1999CRASM.329..915D/abstract
Eide M. B., Ciardi B., Graziani L., Busch P., Feng Y., Di Matteo T., 2020, MNRAS, 498, 6083
Elbers W., van de Weygaert R., 2023, MNRAS, 520, 2709
Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, PASA, 34, e058
Fialkov A., Barkana R., Jarvis M., 2020, MNRAS, 491, 3108

Finkelstein S. L. et al., 2022, ApJ, 940, L55
Finkelstein S. L. et al., 2023, ApJ, 946, L13
Friedrich M. M., Mellema G., Alvarez M. A., Shapiro P. R., Iliev I. T., 2011, MNRAS, 413, 1353
Fujimoto S. et al., 2023, preprint (arXiv:2308.11609)
Furlanetto S. R., Oh S. P., 2005, MNRAS, 363, 1031
Furlanetto S. R., Oh S. P., 2016, MNRAS, 457, 1813
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004a, ApJ, 613, 1
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004b, ApJ, 613, 16
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181
Garaldi E., Kannan R., Smith A., Springel V., Pakmor R., Vogelsberger M., Hernquist L., 2022, MNRAS, 512, 4909
Garaldi E. et al., 2024, MNRAS, 530, 3765
Gazagnes S., Koopmans L. V. E., Wilkinson M. H. F., 2021, MNRAS, 502, 1816
Giri S. K., Mellema G., 2021, MNRAS, 505, 1863
Giri S., Mellema G., Jensen H., 2020, J. Open Source Softw., 5, 2363
Gnedin N. Y., 2014, ApJ, 793, 29
Gnedin N. Y., Madau P., 2022, Living Rev. Comput. Astrophys., 8, 3
Gorski K. M., Wandelt B. D., Hansen F. K., Hivon E., Banday A. J., 1999, preprint(astro-ph/9905275)
HERA Collaboration, 2023, ApJ, 945, 124
Harikane Y. et al., 2023, ApJ, 959, 18,
Harikane Y., Nakajima K., Ouchi M., Umeda H., Isobe Y., Ono Y., Xu Y., Zhang Y., 2024, ApJ, 960, 22
Harikane Y. et al., 2023c, ApJS, 265, 5
Hayes M. J., Scarlata C., 2023, ApJ, 954, L14
Hsiao T. Y.-Y. et al., 2023, preprint (arXiv:2305.03042)
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Iliev I. T., Pen U.-L., Bond J. R., Mellema G., Shapiro P. R., 2007, ApJ, 660, 933
Inayoshi K., Harikane Y., Inoue A. K., Li W., Ho L. C., 2022, ApJ, 938, L10
Ivezić Ž., Connelly A. J., VanderPlas J. T., Gray A., 2014, Statistics, Data Mining, and Machine Learning in Astronomy, Princeton University Press
Jiang L. et al., 2022, Nat. Astron., 6, 850
Jones E., Oliphant T., Peterson P. et al., 2001, SciPy: Open source scientific tools for Python, accessed 2023, http://www.scipy.org/
Jung I. et al., 2024, ApJ, 967, 14
Kakiichi K. et al., 2017, MNRAS, 471, 1936
Kannan R., Vogelsberger M., Marinacci F., McKinnon R., Pakmor R., Springel V., 2019, MNRAS, 485, 117
Kannan R., Garaldi E., Smith A., Pakmor R., Springel V., Vogelsberger M., Hernquist L., 2022a, MNRAS, 511, 4005
Kannan R., Smith A., Garaldi E., Shen X., Vogelsberger M., Pakmor R., Springel V., Hernquist L., 2022b, MNRAS, 514, 3857
Kannan R. et al., 2023, MNRAS, 524, 2594
Kapahtia A., Chingangbam P., Ghara R., Appleby S., Choudhury T. R., 2021, J. Cosmol. Astropart. Phys., 2021, 026
Keating L. C., Bolton J. S., Cullen F., Haehnelt M. G., Puchwein E., Kulkarni G., 2023, preprint (arXiv:2308.05800)
Kostyuk I., Nelson D., Ciardi B., Glatzle M., Pillepich A., 2023, MNRAS, 521, 3077
Lee K.-G., Cen R., Gott J. Richard I., Trac H., 2008, ApJ, 675, 8
Levermore C. D., 1984, J. Quant. Spec. Radiat. Transf., 31, 149
Lewis J. S. W. et al., 2022, MNRAS, 516, 3389
Lin Y., Oh S. P., Furlanetto S. R., Sutter P. M., 2016, MNRAS, 461, 3361
Liu A., Parsons A. R., 2016, MNRAS, 457, 1864
Lu T.-Y., Mason C. A., Hutter A., Mesinger A., Qin Y., Stark D. P., Endsley R., 2024, MNRAS, 528, 4872
Majumdar S. et al., 2016, MNRAS, 456, 2080
McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815
McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, MNRAS, 381, 75
McQuinn M., Lidz A., Zaldarriaga M., Hernquist L., Hopkins P. F., Dutta S., Faucher-Giguère C.-A., 2009, ApJ, 694, 842
Mellema G. et al., 2013, Exp. Astron., 36, 235
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Muñoz J. B., Dvorkin C., Cyr-Racine F.-Y., 2020, Phys. Rev. D, 101, 063526
Mutch S. J. et al., 2016, MNRAS, 463, 3556
Park J., Mesinger A., Greig B., Gillet N., 2019, MNRAS, 484, 933
Pawlik A. H., Rahmati A., Schaye J., Jeon M., Dalla Vecchia C., 2017, MNRAS, 466, 960
Pillepich A. et al., 2018a, MNRAS, 473, 4077
Pillepich A. et al., 2018b, MNRAS, 475, 648
Puchwein E. et al., 2023, MNRAS, 519, 6162
Qin W., Schutz K., Smith A., Garaldi E., Kannan R., Slatyer T. R., Vogelsberger M., 2022, Phys. Rev. D, 106, 123506
Robertson B. E., 2022, ARA&A, 60, 121
Robertson B. E. et al., 2023, Nat. Astron., 7, 611
Rosdahl J. et al., 2018, MNRAS, 479, 994
Saxena A. et al., 2023, A&A, 678, A68
Shapiro P. R., Giroux M. L., 1987, ApJ, 321, L107
Shen X., Vogelsberger M., Boylan-Kolchin M., Tacchella S., Kannan R., 2023, MNRAS, 525, 3254
Shen X. et al., 2024a, preprint (arXiv:2402.08717)
Shen X. et al., 2024b, MNRAS, 527, 2835
Smith T. L., Lucca M., Poulin V., Abellan G. F., Balkenhol L., Benabed K., Galli S., Murgia R., 2022a, Phys. Rev. D, 106, 043526
Smith A., Kannan R., Garaldi E., Vogelsberger M., Pakmor R., Springel V., Hernquist L., 2022b, MNRAS, 512, 3243
Springel V., 2010, ARA&A, 48, 391
Steinhardt C. L., Kokorev V., Rusakov V., Garcia E., Sneppen A., 2023, ApJ, 951, L40
Thélie E., Aubert D., Gillet N., Ocvirk P., 2022, A&A, 658, A139
Umeda H., Ouchi M., Nakajima K., Harikane Y., Ono Y., Xu Y., Isobe Y., Zhang Y., 2023, preprint (arXiv:2306.00487)
Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031
Vogelsberger M. et al., 2014a, MNRAS, 444, 1518
Vogelsberger M. et al., 2014b, Nature, 509, 177
Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nat. Rev. Phys., 2, 42
Walt S. v. d., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Weinberger R. et al., 2017, MNRAS, 465, 3291
Weinberger R., Springel V., Pakmor R., 2020, ApJS, 248, 32
Wise J. H., 2019, Contemp. Phys., 60, 145
Witstok J. et al., 2024, A&A, 682, A40
Xu H., Wise J. H., Norman M. L., 2013, ApJ, 773, 83
Xu C. et al., 2023, MNRAS, 521, 4356
Yeh J. Y. C. et al., 2023, MNRAS, 520, 2757
Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, ApJ, 654, 12
Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622
van Haarlem M. P. et al., 2013, A&A, 556, A2

APPENDIX A: IMPACT OF GRID RESOLUTION

In this appendix, we examine the effects of simulation resolution on ionized bubble size estimates for resolutions of 128, 256, and 512 cells per side. The THESAN simulations and AREPO-RT code employ a Voronoi tessellation constructed from a particle representation of the underlying gas distribution. Cartesian grid outputs are subsequently produced by translating these particle locations and properties on to a three-dimensional cubic grid with 1024 cells per side. For the purposes of this study, we downsample this grid to produce lower resolution outputs at 512, 256, and 128 cells per side. Due to computational constraints, we utilize the 512 cells per side output as the fiducial highest resolution for this work. We demonstrate convergence of bubble size estimates between the lower resolution runs and the 512 cells per side resolution, thereby confirming that the effects of resolution on our findings are negligible.

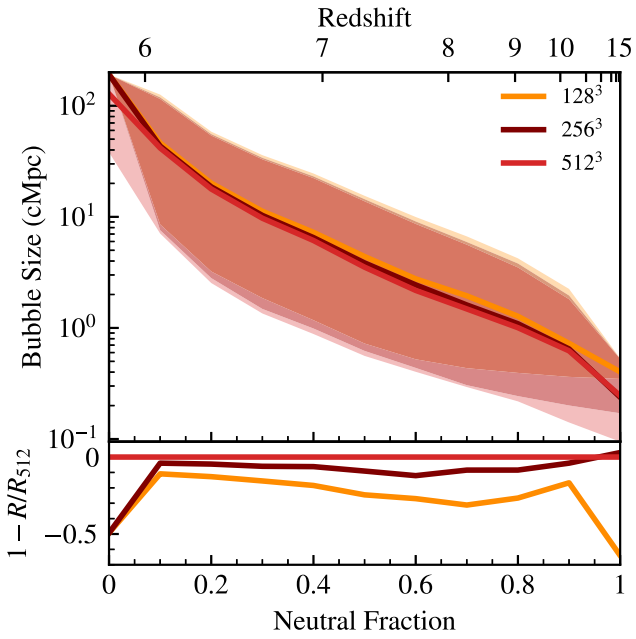

Figure A1. *Upper panel:* MFP determined median bubble size for THESAN-1 for three resolutions: $128^3, 256^3, 512^3$. The shaded areas show one standard deviation ranges. *Lower panel:* Relative difference compared to the highest resolution run, $1 - R_N/R_{512}$, showing the bubble sizes are converged to within 10 per cent accuracy.

The bubble size distributions for each of the resolutions look qualitatively similar. Specifically, the bubble sizes grow from ∼100 ckpc scales at the onset of reionization to ∼100 cMpc scales by the end of reionization.

We compare the time evolution of the characteristic bubble sizes across different resolution outputs to check convergence. Fig. A1 shows the median bubble sizes as a function of neutral fraction for resolutions of 128, 256, and 512 cells per side. We find that bubble sizes are converged to within approximately 10 and 30 per cent accuracy for 256 and 128 cells per side resolutions, respectively. The bubble sizes of the lower resolution simulations tend to be larger than the bubble sizes for the 512 cells per side run because the cells themselves are larger and artificially expand the bubbles.

The bubble sizes primarily diverge at the very beginning and end of reionization. This is expected due to the resolution-scale effects becoming important when there are only small patches of ionized and neutral gas at $x_{\rm H\,I} \gtrsim 0.9$ and $x_{\rm H\,I} \lesssim 0.1$, respectively. At the beginning of reionization, small pockets of ionized gas are either washed out or blurred into an artificially large bubble on the scale of the cell size, which will lead to more divergence from the high resolution model than intermediate times when the scales of neutral and ionized gas regions are much larger than the cell sizes. Similarly, at the end of reionization when the last 10 per cent of gas is becoming ionized, there are small-scale regions of neutral gas remaining in the high resolution 512 cell per side box, but those have been washed out by the resolution effects leading to a larger bubble size by the end of reionization.

APPENDIX B: IMPACT OF GRID SMOOTHING SCALE

We also investigate the effects of smoothing on the derived bubble sizes. The smoothing we use in this work, primarily for environmental effects is volume-weighted. The smoothing is performed using a Gaussian kernel applied to the simulation outputs before converting to a Cartesian grid.

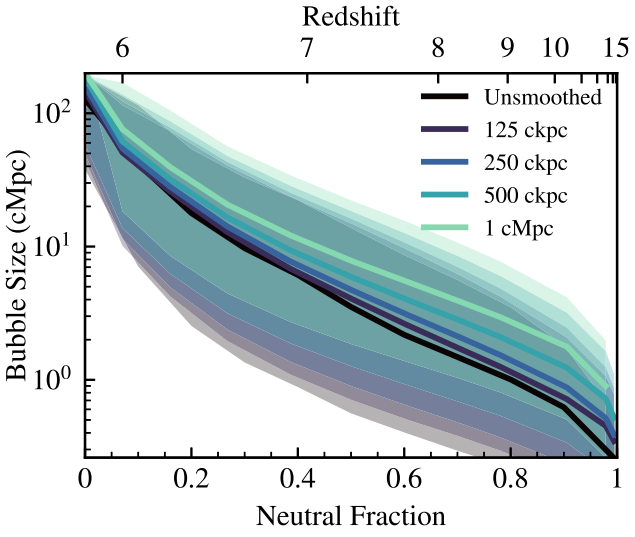

Figure B1. Median bubble sizes over time as characterized by neutral fraction and redshift for different smoothing scales. The smoothing scales do not seem to change bubble sizes much at later times, but bubbles are biased larger for larger smoothing scales, early on. This is partly due to the blurring of small ionized regions during the smoothing process that artificially increases bubble sizes, likely of the order of the smoothing scale. Once many bubbles are larger than the smoothing scale, the bias introduced by the smoothing becomes negligible.

We study the behaviour of the characteristic bubble sizes for several different smoothing scales: unsmoothed, 125 ckpc, 250 ckpc, 500 ckpc, and 1 cMpc. The comparison of their median bubble sizes and one sigma regions are shown in Fig. B1. As the smoothing scale increases, so does the median bubble size. This is likely due to the blurring of ionized regions with the smoothing leading to a larger observed bubble size. At the beginning of reionization when the universe is still over 90 per cent neutral, the median bubble sizes depend heavily on the smoothing scale. When the first ionized bubbles begin to form, they are immediately blurred by the smoothing kernel leading to the smallest bubble sizes being roughly the size of the smoothing scale (and for the unsmoothed case, the smallest bubbles are approximately the size of one cell). These smoothing-dependent effects appear to wash out as reionization progresses, likely due to most ionized cells being part of bubbles with sizes much larger than the smoothing scales. By the time reionization is ending, smoothing effects on the median bubble size seem to have nearly completely disappeared as the entire simulation box becomes ionized and part of one large bubble, as expected.

We explore the effects of smoothing for two main reasons: the environmental analysis and future computational convenience. In order to quantify the effect of the environment on one simulation cell's bubble size, we need to smooth the density to get information about the surrounding medium in the simulation. Checking that the bubble sizes are consistent under this smoothing ensures that our environmental results reflect the conditions in the local surrounding areas of the bubble centres without artificially impacting the bubble size measurements.

We also check the effects of smoothing on bubble sizes to inform future work that may take advantage of the computational convenience of having a smoother field to analyse. Since many

topological analyses require smooth fields, it is important for future work to be able to quantify the effects of this smoothing on bubble sizes.

APPENDIX C: IMPACT OF SIMULATION RESOLUTION

In this appendix, we demonstrate the convergence of bubble sizes with simulation resolution. Specifically, in Fig. C1, we analyse the median and mean $R_{\rm eff}$ evolution from the THESAN-2 and THESAN-WC-2 medium resolution simulations in comparison with the results from the flagship THESAN-1 simulation. Recall from Section 4 that these are both from the same initial conditions as THESAN-1 but with two (eight) times lower spatial (mass) resolution, and that THESAN-WC-2 has a increases the birth cloud escape fraction from 0.37 to 0.43 compensate for lower star formation. We note that these calculations employ a grid resolution of 256^3, which has been shown in Fig. A1 to be converged with the 512^3 results within 10 per cent at all times. The impact on the median and mean statistics is most pronounced in the initial stages of reionization, primarily due to the abundance of small bubbles surrounding the lowest-mass haloes captured by the THESAN-1 simulation, as well as variations in escape fraction (see Yeh et al. 2023). As these bubbles merge to form larger structures, the statistical agreement improves.

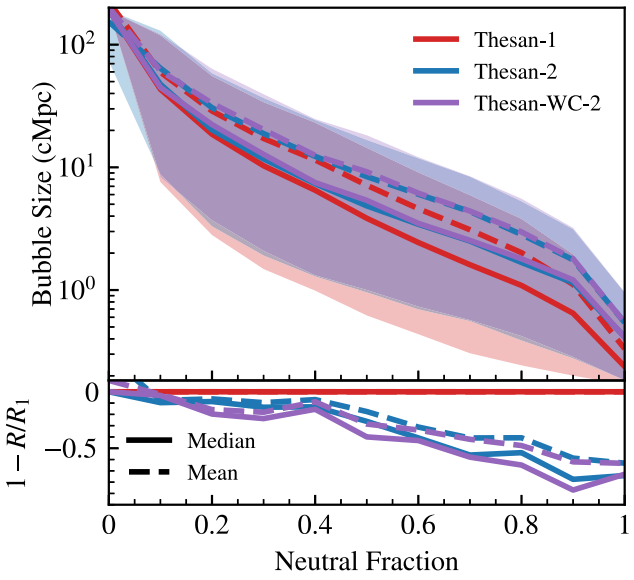

Figure C1. Median (solid) and mean (dashed) bubble sizes over time as characterized by neutral fraction and redshift for different simulation resolutions. The statistics are affected most strongly during the first half of reionization ($x_{\rm H\,II} \lesssim 0.5$) due to the relatively numerous small bubbles around the lowest-mass haloes resolved by the simulation and escape fraction differences. The agreement improves as bubbles coalesce into larger ones.

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