

Mapping reionization bubbles in JWST era

I. Empirical edge detection with Lyman alpha emission from galaxies

Ting-Yi Lu^{1,2,*}, Charlotte A. Mason^{1,2}, Andrei Mesinger³, David Prelogović^{3,4}, Ivan Nikolic³,
Anne Hutter^{1,2}, Samuel Gagnon-Hartman³, Mengtao Tang⁵, Yuxiang Qin⁶, and Koki Kakiichi^{1,2}

¹ Cosmic Dawn Center (DAWN), Copenhagen N, Denmark

² Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen N, Denmark

³ Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

⁴ Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, 34136 Trieste, Italy

⁵ Steward Observatory, University of Arizona, 933 N Cherry Ave, Tucson, AZ 85721, USA

⁶ Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia

Received 6 November 2024 / Accepted 21 March 2025

ABSTRACT

Context. Ionized bubble sizes during reionization trace physical properties of the first galaxies. JWST’s ability to spectroscopically confirm and measure Lyman-alpha ($Ly\alpha$) emission in sub- L^* galaxies makes it possible to map to map ionized bubbles in 3D. However, existing $Ly\alpha$ -based bubble measurement strategies rely on constraints from single galaxies, which are limited by the large variability in intrinsic $Ly\alpha$ emission.

Aims. As a first step, we present two bubble-size-estimation methods using $Ly\alpha$ spectroscopy of ensembles of galaxies, enabling us to map ionized structures and marginalize over $Ly\alpha$ emission variability. We tested our methods using gigaparsec-scale reionization simulations of the intergalactic medium (IGM).

Methods. To map bubbles in the plane of the sky, we developed an edge detection method based on the asymmetry of $Ly\alpha$ transmission as a function of spatial position. To map bubbles along the line of sight, we develop an algorithm using the tight relation between $Ly\alpha$ transmission and the line-of-sight distance from galaxies to the nearest neutral IGM patch.

Results. Both methods can robustly recover bubbles with a radius ≥ 10 comoving Mpc, sufficient for mapping bubbles even in the early phases of reionization, when the IGM is ~ 70 – 90% neutral. These methods require ≥ 0.002 – 0.004 galaxies/cMpc³, a 5σ $Ly\alpha$ equivalent width upper limit of $\lesssim 30$ Å for the faintest targets, and redshift precision $\Delta z \leq 0.015$, which is feasible with JWST spectroscopy. Shallower observations will provide robust lower limits on bubble sizes. Additional constraints on IGM transmission from $Ly\alpha$ escape fractions and line profiles will further refine these methods, paving the way to our first direct understanding of ionized bubble growth.

Key words. intergalactic medium – dark ages, reionization, first stars

1. Introduction

Between $20 \gtrsim z \gtrsim 5$, the first luminous sources reionized neutral hydrogen in the intergalactic medium (IGM), forming ionized bubbles around them which eventually merged (e.g., see Mesinger et al. 2016, for a recent review). Understanding the properties, formation, and evolution of these first sources and how they drove the reionization process is a frontier in astronomy. While significant progress has been made in the last two decades in measuring the volume-averaged timeline of reionization (e.g., Fan et al. 2006; Stark et al. 2010; Treu et al. 2012; Pentericci et al. 2014; Mason et al. 2018a; Davies et al. 2018; Jung et al. 2020; Qin et al. 2021; Bosman et al. 2022; Bolan et al. 2022; Umeda et al. 2024; Nakane et al. 2024; Tang et al. 2024a), we currently have no robust measurements of the typical sizes of ionized bubbles and how they evolve over time.

The first large bubbles are predicted to form around early galaxy overdensities, thus the sizes of ionized regions and their host overdensities can provide key insights into the reionization process. In individual regions this would enable ionizing

photon “accounting”, where visible galaxies’ ionizing photon output can be compared to that required to grow the bubble, providing insight into their ionizing photon escape fractions and star formation histories, and the contribution of galaxies below even JWST’s detection limits (e.g., Shapiro & Giroux 1987; Haiman 2002; Mason & Gronke 2020; Endsley & Stark 2022; Whittle et al. 2024; Torralba-Torregrosa et al. 2024). On a global scale, the distribution of ionized region sizes at a fixed redshift – the reionization “morphology” – is sensitive to the halo mass scale of the dominant ionizing sources (e.g., McQuinn et al. 2007a; Mesinger et al. 2016; Seiler et al. 2019; Hutter et al. 2023; Lu et al. 2024), with models where low-mass halos ($M_h \sim 10^8 M_\odot$) dominate the ionizing-photon-budget-predicting median bubble radii of ~ 10 – 30 cMpc (~ 4 – 12 arcmin) during the mid-stages of reionization. In the next decade, there will be substantial observational efforts to identify and start characterizing the sizes of ionized regions over scales large enough to capture multiple bubbles, via upcoming 21 cm experiments to observe the neutral hydrogen directly (e.g., Yatawatta et al. 2013; Bowman et al. 2013) and wide-area near-IR spectroscopic surveys to observe galaxies that should trace ionized regions, for example, with Euclid, Roman, Subaru/Prime Focus Spectrograph (PFS), and Very Large Telescope (VLT)/Multi

* Corresponding author.

Object Optical and Near-infrared Spectrograph (MOONS) (van Mierlo et al. 2022; Euclid Collaboration: Scaramella et al. 2022; Wang et al. 2022; Greene et al. 2022; Maiolino et al. 2022; Gagnon-Hartman et al. 2025; Mesinger et al. 2024).

Lyman-alpha ($\text{Ly}\alpha$) absorption by neutral hydrogen in high- z objects, such as galaxies, quasars (QSOs), and gamma-ray bursts, is currently our best tool to identify ionized regions, as $\text{Ly}\alpha$ suffers strong damping wing absorption due to neutral hydrogen in the foreground of the objects (Miralda-Escude & Rees 1998; Mesinger & Furlanetto 2008). Although considerably fainter than QSOs, galaxies have the benefit of being more numerous and less biased tracers of the reionization morphology (Shen et al. 2007; Eilers et al. 2024), such that they can be used to sample the IGM. Spectroscopic and narrow-band surveys in the last decades have identified a number of fields showing an excess of $\text{Ly}\alpha$ emission from galaxies at $z \sim 6-9$, implying large-scale ($R > 1$ pMpc) ionized regions (e.g., Castellano et al. 2016; Tilvi et al. 2020; Endsley & Stark 2022; Wold et al. 2022; Napolitano et al. 2024; Tang et al. 2024a). Upcoming wide-area near-IR spectroscopy with Euclid, Roman, PFS, and MOONS will be able to identify more candidate ionized regions by detecting bright ($\gtrsim 1 \times 10^{-17}$ erg s $^{-1}$ cm $^{-2}$) $\text{Ly}\alpha$ emission, and photometric overdensities of UV-bright galaxies. However, as we subsequently demonstrate, these instruments neither have the sensitivity to sample nor the wavelength coverage to spectroscopically confirm sufficient numbers of galaxies in these ionized regions to fully map bubbles and characterize their sizes (Hutter et al. 2018; Zackrisson et al. 2020), thus deeper follow-up imaging and spectroscopy will be required.

JWST has provided unique capabilities to make mapping these ionized regions, and the overdensities they surround, feasible. Its ability to spectroscopically confirm UV faint $z > 6$ galaxies, even down to $M_{\text{UV}} \approx -18$ at $z \sim 14$ (Curtis-Lake et al. 2023; Carniani et al. 2024), has allowed us to precisely map $z > 5$ galaxy distributions in 3D for the first time (e.g., Helton et al. 2024; Meyer et al. 2024; Chen et al. 2024; Tang et al. 2024a; Witstok et al. 2024). The sensitivity of NIRSpec (Jakobsen et al. 2022) has enabled the detection $\text{Ly}\alpha$ emission from UV-faint galaxies (e.g., Saxena et al. 2023; Tang et al. 2023; Chen et al. 2024; Jones et al. 2025), which is infeasible from the ground, and can be used to comprehensively trace ionized regions. JWST's ability to measure $\text{Ly}\alpha$ escape fractions from Balmer lines also provides a promising new indicator of highly ionized regions, where the fraction of $\text{Ly}\alpha$ escaping from galaxies through the IGM can reach close to unity (e.g., Chen et al. 2024). Such power makes it possible to map ionized bubbles. This potential has been demonstrated empirically in early JWST results. For example, the CEERS/EGS field, which contains $\gtrsim 50\%$ of all $\text{Ly}\alpha$ emission known to date above $z > 7$, is likely the largest known ionized region at this epoch (Oesch et al. 2015; Roberts-Borsani et al. 2016; Stark et al. 2017; Tilvi et al. 2020; Jung et al. 2022; Larson et al. 2022). Several studies using NIRSpec spectroscopy have detected new high equivalent width ($\text{EW} \gtrsim 100 \text{ \AA}$) $\text{Ly}\alpha$ emission, and measured $\text{Ly}\alpha$ escape fractions to confirm several galaxies in the region likely transmit $>50\%$ of their $\text{Ly}\alpha$ through the IGM (Tang et al. 2023, 2024a; Jung et al. 2024; Napolitano et al. 2024; Chen et al. 2024). Chen et al. (2024) and Tang et al. (2024a) demonstrate $z \sim 7-8$ galaxies in EGS appear clustered in several narrow redshift bins around $\text{Ly}\alpha$ emitting galaxies, hinting at several proper-megaparsec-scale bubbles along the line of sight (LOS).

However, a quantitative link from discrete galaxy observations such as these to the sizes of ionized regions has so far been challenging due to the intrinsic variability of $\text{Ly}\alpha$ emission. With

an estimate of the fraction of $\text{Ly}\alpha$ transmitted through the IGM, it should be possible to infer the LOS distance of a galaxy to neutral gas (e.g., Mesinger & Furlanetto 2008; Mason & Gronke 2020), providing we can estimate the $\text{Ly}\alpha$ emitted by the source, *before* attenuation by the IGM. In the past few years, significant progress has been made in understanding the properties of $\text{Ly}\alpha$ emerging from post-reionization ($5 \lesssim z \lesssim 6$) galaxies (e.g., Prieto-Lyon et al. 2023; Tang et al. 2024b; Bolan et al. 2024; Lin et al. 2024). In particular, JWST observations provide key information about the relations between galaxies' $\text{Ly}\alpha$ emission (rest-frame equivalent widths, $\text{Ly}\alpha$ escape fractions $f_{\text{esc},\text{Ly}\alpha}$, and $\text{Ly}\alpha$ velocity offsets and line-profiles) and galaxy UV and optical properties such as UV continuum magnitude, UV slopes and [OIII]+H β emission (e.g., Saxena et al. 2023, 2024; Prieto-Lyon et al. 2023; Chen et al. 2024; Tang et al. 2024b; Lin et al. 2024). These observations can be used as baselines to infer the impact of the neutral IGM on $\text{Ly}\alpha$ during reionization. However, even in the ionized IGM, $\text{Ly}\alpha$ properties such as EW, escape fraction and lineshape can vary by orders of magnitude in galaxies with similar physical properties (e.g., Stark et al. 2010; Oyarzún et al. 2017; Jaskot et al. 2019; Hayes et al. 2023), which radiative transfer simulations predict is largely due to sightline variability in neutral gas and dust distributions in the ISM and CGM (e.g., Smith et al. 2017; Blaizot et al. 2023).

The first attempts to infer the sizes of ionized regions have focused on individual galaxies, using the approach described above to infer the relation between $\text{Ly}\alpha$ transmission and the distance of the galaxy to the neutral IGM (e.g., Mason & Gronke 2020; Hayes & Scarlata 2023; Umeda et al. 2024; Witstok et al. 2024; Torralba-Torregrosa et al. 2024). However, the intrinsic variability of $\text{Ly}\alpha$ emission makes estimates of the sizes of ionized bubbles based on single galaxies highly uncertain. To provide robust estimates of the sizes of ionized regions thus requires combining information from *ensembles* of galaxies to overcome this variability.

In this work, we thus aim to provide simple approaches to map ionized bubbles using observations of ensembles of galaxies, given the new capabilities of JWST. Here we focus on empirical methods for detecting bubble edges via integrated $\text{Ly}\alpha$ observations (i.e. EW, escape fractions). In a companion paper, Nikolić et al. (2025), we will present a framework for inferring the centers and radii of individual ionized regions based on $\text{Ly}\alpha$ line profiles. In Section 2 we introduce our IGM simulations and forward modeling of $\text{Ly}\alpha$ emission. We develop methods to map ionized bubbles on the plane of the sky (Section 3) and along sightline skewers (Section 4) using galaxy $\text{Ly}\alpha$ emission observations, and describe the observational requirements for robust bubble size recovery. Our methods make use of the almost one-to-one relation between the IGM damping wing absorption profile and the distance of galaxies to the nearest neutral IGM. In Section 5, we discuss prospect for ionized bubble mapping with current and future facilities and surveys. We present our conclusions in Section 6.

Throughout this paper, we use a flat Λ CDM cosmology with $\Omega_{\text{m}} = 0.31$, $\Omega_{\Lambda} = 0.69$, and $h = 0.68$ and magnitudes are in the AB system.

2. IGM simulations and forward modeling galaxy observables

We create simulations of the IGM with forward-modeled galaxy observables to test our bubble mapping methods. In this section, we describe how we simulate ionization cubes (Section 2.1), model emitted $\text{Ly}\alpha$ lines from galaxies (Section 2.2) and infer

the fraction of Ly α flux transmitted through the IGM for each galaxy (Section 2.3), which is the key input to our bubble mapping methods.

2.1. IGM simulations

We use 21cmFASTv2¹ (Mesinger et al. 2011, 2016; Sobacchi & Mesinger 2014) to create realistic, large volume reionization IGM simulations. 21cmFASTv2 is a semi-numerical code that can quickly create reionization simulations for coeval density cubes (Mesinger et al. 2011).

We create a (300 cMpc)³ density cube with (1.5 cMpc)³ resolution, smaller than the characteristic bubble size ($\gtrsim 10$ cMpc) when IGM neutral fraction is <90% (Lu et al. 2024). We generate ionization cubes of $\bar{x} = 0.4$ – 0.9 at $z = 8$ by changing the ionization efficiency of galaxies, similar to the approach described by Lu et al. (2024). For simplicity, we use a bright-galaxy-driven reionization model, in which $M_h > 10^{11} M_\odot$ halos can ionize the IGM. The broad bubble size distribution of our bright-galaxy-driven reionization model allows us to test bubble size recovery for a wide range of bubble size. We find our results do not depend significantly on \bar{x} (which is the dominant driver of bubble sizes) because our methods primarily rely on the change of IGM transmission at the edges of bubbles. Thus we also do not expect our results to change significantly if we had assumed another reionization model, which will change the characteristic bubble size. Moreover, since most of the IGM during reionization has only two phases, bubble edges remain equally distinct in most reionization models.

We extract a catalog of halo masses and positions at $z = 8$ from the density cube using extended Press-Schechter theory (Sheth et al. 2001) and a halo-filtering method developed by Mesinger & Furlanetto (2007). The minimum halo mass extracted in this work is $\sim 4.2 \times 10^9 M_\odot$. As described in Lu et al. (2024), we use the UV luminosity – halo mass relation from Mason et al. (2015, 2023), adding a scatter of 0.5 mag following Ren et al. (2018), Whitler et al. (2020), to map from halo masses to UV magnitude. Lu et al. (2024) demonstrated this reproduces the observed UV LFs very well at $z \sim 7$ – 10 , though at the faint end ($M_{UV} \gtrsim -18$) of LF in our simulation slightly underpredicts the number density of galaxies. In later sections we will investigate bubble mapping given different galaxy number densities by setting different M_{UV} detection limits for mock observations with our simulations. When discussing the corresponding detection limits in real observations, we will quote M_{UV} limits which match our simulated number densities using the LFs in Bouwens et al. (2021).

2.2. Forward modeling Ly α emission

We will use galaxies’ Ly α emission properties to recover transmissions at their positions in a bubble. To obtain these properties, we forward model the observed Ly α emission lines following the approach by Mason et al. (2018a). We assume the intrinsic Ly α line shape is a single Gaussian, and model the observed line after transfer through interstellar medium (ISM), circumgalactic medium (CGM), and IGM.

We generate the Ly α damping wing optical depth due to the IGM for every halo (selecting $\sim 10^5$ halos with $M_{UV} \leq -17$, or $M_h \gtrsim 10^{9.6} M_\odot$) along a sightline in the + z -direction from the

galaxy to a distance of 200 cMpc away², using periodic boundary conditions. We calculate the Ly α damping wing optical depth profile, $\tau_{IGM}(\Delta v_{Ly\alpha})$ (e.g., Miralda-Escude 1998) at Ly α velocity offsets from systemic ($\Delta v_{Ly\alpha}$) between -300 km/s and 5700 km/s by summing the contributions of LOS neutral hydrogen in the ionization cube as described in Mesinger et al. (2015).

Our Gaussian Ly α emission line model is described by two quantities: the velocity offset from systemic ($\Delta v_{Ly\alpha}$) and line width (σ_v). Following Mason et al. (2018a), we assume the velocity offset of Ly α emission depends on a galaxy’s mass. Velocity offsets are observed to be positively correlated with quantities which scale with galaxy mass, likely because of scattering in the ISM and outflows (e.g., Yang et al. 2017; Garel et al. 2021; Blaizot et al. 2023; Hayes et al. 2023). We sample $\Delta v_{Ly\alpha}$ from the $P(\Delta v_{Ly\alpha}|M_h)$ distributions in Mason et al. (2018a) for each galaxy. The line width, σ_v , also depends on scattering in ISM and is thus related to $\Delta v_{Ly\alpha}$ (Neufeld 1991). Motivated by the close relationship between $\Delta v_{Ly\alpha}$ and Ly α FWHM observed by Verhamme et al. (2015) we model it using $\sigma_v = 20$ km/s + $\Delta v_{Ly\alpha}/2.355$ (adding a small constant to avoid zero dispersion). We assume no redshift evolution in the M_h – $\Delta v_{Ly\alpha}$ relationship. If the dependence of $\Delta v_{Ly\alpha}$ on M_h weakens at $z > 6$, that is, $\Delta v_{Ly\alpha}$ of galaxies are lower than our model predicts, then the amount of transmitted Ly α flux will reduce. This would, in turn, reduce the detectability of Ly α emission from high- z galaxies.

To account for absorption due to infalling residual neutral gas in the ionized IGM or CGM, which is expected to suppress Ly α flux redward of systemic at $z > 5$ (Santos 2004; Dijkstra et al. 2011; Mason et al. 2018a; Park et al. 2021), we cut off Ly α emission blueward of the circular velocity ($v_{circ} = [10GM_h H(z)]^{1/3}$). We define the emitted Ly α line profile, $J_\alpha(\Delta v_{Ly\alpha}, M_h, v)$, as the Ly α line after CGM absorption and assume ISM and CGM properties do not evolve significantly with redshift at $z \sim 6$ – 9 . As our methods rely on the relative transmission between sources at the same redshift, redshift evolution of Ly α emission or CGM properties should not significantly impact our results.

We define the integrated Ly α transmission, \mathcal{T} , for each galaxy as the ratio between the observed and emitted line equivalent width (EW):

$$\mathcal{T} = \frac{EW_{Ly\alpha,obs}}{EW_{Ly\alpha,em}} \equiv \frac{\int_{-\infty}^{\infty} dv J_\alpha(\Delta v_{Ly\alpha}, M_h, v) e^{-\tau_{IGM}(\Delta v_{Ly\alpha})}}{\int_{-\infty}^{\infty} dv J_\alpha(\Delta v_{Ly\alpha}, M_h, v)}. \quad (1)$$

In the following sections, we will refer to this integrated Ly α transmission as “Ly α transmission”.

To model the emitted Ly α EW ($EW_{Ly\alpha,em}$) we use an empirically determined relation between galaxy M_{UV} and observed Ly α equivalent width (EW_{obs}) at $z \sim 5$ – 6 , derived by Tang et al. (2024a). These distributions are based on >700 $z \sim 5$ – 6 Lyman-break galaxies with ground-based Ly α spectroscopy and JWST photometry, having UV properties similar to the reionization era galaxies from recent JWST observations (e.g., Endsley et al. 2024; Topping et al. 2024), and should be representative of the Ly α emission before significant IGM damping wing attenuation, including Ly α transmission in the ISM and CGM, that is, including the impact of local damped Ly α absorbers (e.g., Shapley et al. 2003; Reddy et al. 2016; Heintz et al. 2025). We draw the emitted Ly α equivalent width, EW_{em} , from the $P(EW_{em}|M_{UV})$ distributions derived by Tang et al. (2024b), using the M_{UV} for each halo as described in Section 2.1.

² Beyond this distance, HI contributes negligibly to damping wing absorption (Mesinger & Furlanetto 2008).

¹ <https://github.com/andreimesinger/21cmFAST>

For each galaxy we predict the observed EW, EW_{obs} . In the following section we describe how we infer the Ly α transmission from EW_{obs} . We assume the observed Ly α flux has been corrected for slit losses, as, for example, the JWST MSA may not cover the full Ly α and UV flux (Bhagwat et al. 2024; Ning et al. 2024), Tang et al. (2024b) demonstrates NIRSspec recovers $\sim 80\%$ of ground-based Ly α flux.

2.3. Estimating Ly α transmission from (mock) observations

Our forward model of galaxy Ly α spectra enables us to create mock observations. From these we now seek to *recover* the Ly α transmission, \mathcal{T} , for each galaxy to use as a probe of the IGM. We can use EW_{obs} of a galaxy to recover \mathcal{T} given Equation (1). Using Bayes' theorem, the inferred probability distribution of \mathcal{T} for a galaxy with an observed M_{UV} and EW_{obs} can be recovered as

$$P(\mathcal{T}|M_{\text{UV}}, EW_{\text{obs}}) \propto P(EW_{\text{obs}}|\mathcal{T}, M_{\text{UV}})P(\mathcal{T}), \quad (2)$$

where $P(EW_{\text{obs}}|\mathcal{T}, M_{\text{UV}})$ is the likelihood of obtaining EW_{obs} given a transmission fraction \mathcal{T} and galaxy's UV magnitude M_{UV} . $P(\mathcal{T})$ is the prior on Ly α transmission. Technically, \mathcal{T} can depend on M_{UV} because of environmental effects, for example, brighter galaxies may be more likely to reside in overdense, and thus more ionized, regions (Mason et al. 2018a,b). However, as we are interested in galaxies which are colocated spatially, we can assume they have similar environments, and thus we do not include any dependence of \mathcal{T} on M_{UV} . We will explore a more complete treatment in future work. We can build the likelihood $P(EW_{\text{obs}}|\mathcal{T}, M_{\text{UV}})$ using $P_0(EW)$, our 'intrinsic' EW distribution at $z \sim 5-6$ (see Section 2.2 and Tang et al. 2024b): $P(EW_{\text{obs}}|\mathcal{T}, M_{\text{UV}}) \propto P_0(EW_{\text{obs}}/\mathcal{T}|M_{\text{UV}})$. Here we assume no redshift evolution of the intrinsic EW distribution $z \sim 5-6$ and $z > 6$. If the intrinsic EW of $z > 6$ galaxies are stronger, we will overestimate \mathcal{T} . Here we focus on providing a simplistic inference method and leave advancing the transmission inference framework to a future work.

To account uncertainties or upper limits in EW_{obs} measurements, we marginalize over EW_{obs} (e.g., Schenker et al. 2014; Mason et al. 2018a; Tang et al. 2024a). For a galaxy with detected $EW_{\text{obs,noise}}$ and uncertainty, σ_{EW} , we assume a Gaussian EW_{obs} distribution, such that the inferred \mathcal{T} distribution can be written as

$$P(\mathcal{T}|M_{\text{UV}}, EW_{\text{obs,noise}}) \propto P(\mathcal{T}) \times \int_0^{\infty} \frac{\exp\left[-\frac{(EW_{\text{obs}} - EW_{\text{obs,noise}})^2}{2\sigma_{EW}^2}\right]}{\sqrt{2\pi}\sigma_{EW}} P(EW_{\text{obs}}|\mathcal{T}, M_{\text{UV}}) dEW_{\text{obs}}. \quad (3)$$

For a galaxy with Ly α equivalent width below the 5σ detection limit, the inferred \mathcal{T} distribution can be written

$$P(\mathcal{T}|M_{\text{UV}}, EW_{\text{obs,noise}} < 5\sigma_{EW}) \propto P(\mathcal{T}) \times \frac{1}{2} \int_0^{\infty} \text{erfc}\left(\frac{EW_{\text{obs}} - 5\sigma_{EW}}{\sqrt{2}\sigma_{EW}}\right) P(EW_{\text{obs}}|\mathcal{T}, M_{\text{UV}}) dEW_{\text{obs}}. \quad (4)$$

We assume a uniform prior on $p(\mathcal{T})$ between 0 and 1. We note that Ly α escape fractions estimated from Balmer line observations could be used as a prior on $p(\mathcal{T})$, which would further improve our estimates of the IGM transmission (e.g., Nikolić et al. 2025), but for simplicity here we consider only

measurements of the Ly α EW. We do not account for the uncertainty in M_{UV} here, as it is a subdominant factor and is typically well constrained (Tang et al. 2024b). However, in principle, this uncertainty could be marginalized over.

Due to the large dispersion of the emitted EW distribution (Tang et al. 2024b), the inferred $P(\mathcal{T}|M_{\text{UV}}, EW_{\text{obs,noise}})$ for individual galaxies can be very broad. By making use of ensembles of galaxies, our method marginalizes over this intrinsic variability to recover bubble sizes by locating positions where many galaxies' observed Ly α properties are consistent with high \mathcal{T} .

As our method relies on measuring changes in transmission relative to that expected in an ionized IGM this sets an approximate maximum required Ly α EW sensitivity limit comparable to the median EW for galaxies in a fully ionized IGM ($\approx 30 \text{ \AA}$, e.g., Stark et al. 2011; Jung et al. 2020; Tang et al. 2023), that is, such that we can test if the observed EW distribution is consistent with $\mathcal{T} \approx 1$ (inside a large bubble) or if \mathcal{T} is significantly lower (in small bubbles or the neutral IGM).

3. Mapping bubbles in the plane of the sky

Here we describe our method of making bubble maps in the plane of the sky, using an edge-detection technique based on the asymmetric distribution of Ly α transmission around the edges of ionized bubbles (Section 3.1). We detail the suggested parameter setup for our method in Section 3.2 and the observational requirements in Section 3.3.

3.1. Transmission asymmetry score

In the plane of the sky, we expect galaxies inside an ionized bubble to have higher Ly α transmission compared to galaxies outside the bubble, that is, bubbles should be signposted by strong Ly α -emitters. To map the bubbles requires identifying the edges, which should be marked by a steep drop in Ly α transmission. At a point on the sky we can calculate how 'asymmetric' Ly α transmission is as a function of transverse distance, and identify bubble edges as places where this asymmetry is maximized. In principle, we can also define a threshold in transmission to define bubble edges (as suggested by Kakiichi et al. 2023), but this transmission threshold will depend on the size of the bubbles and thus the average IGM neutral fraction, \bar{x} . Thus here we consider the *asymmetry* of transmission as it should be more independent of bubble sizes and the reionization history (as we discuss below), making it a more flexible technique.

In Figure 1 we show a toy model for our method assuming a spherical ionized bubble. If we estimate the transmissions of galaxies in a redshift slice (the green window in the left panel), and plot transmission as a function of galaxy position in one direction on the plane of the sky, we will get a bell-shaped transmission profile: the peak of transmission corresponds to galaxies in the center of the bubble, and the bubble edges are where the transmission drop to almost zero. When measuring the transmission profile through a small window (the orange windows in the left panel), the level of symmetry of the transmission profile around the window center indicates the position of bubbles.

If the transmission profile on the right and left of the window center are both low and flat (window A in the right panel of Figure 1), the window is outside a bubble. If the transmission profiles on both sides of a window are high and symmetric (window B in the right panel), the window is inside a bubble and close to the bubble center. If one side of a window has a flat and low transmission profile, and the other side has a rising transmission

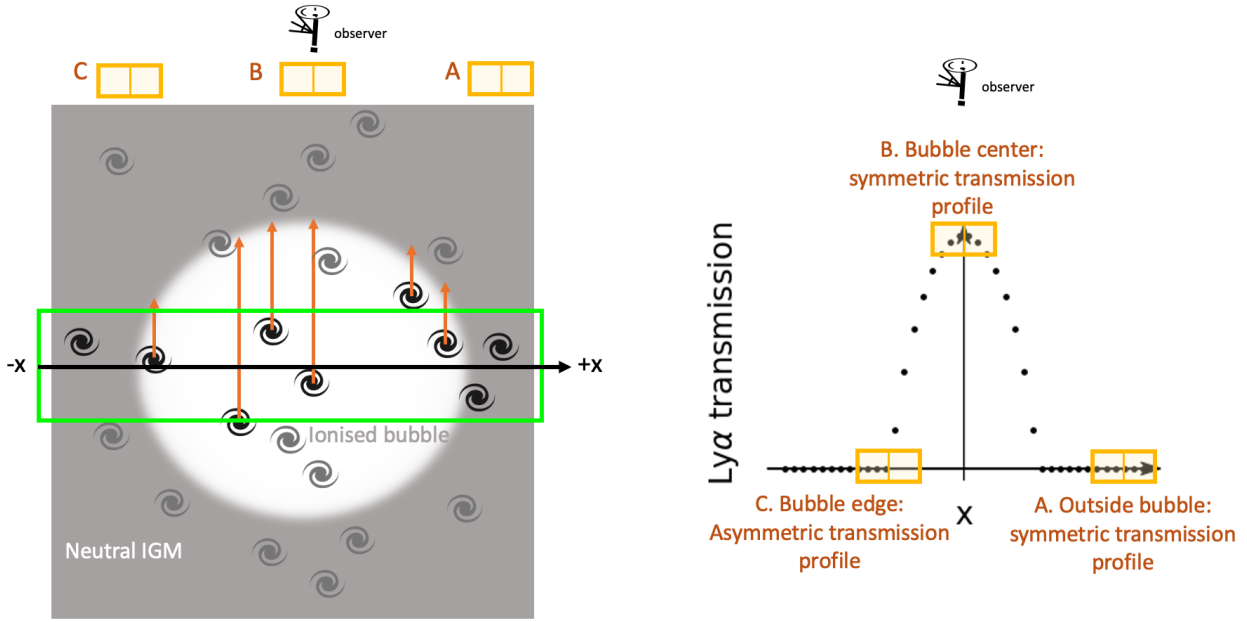


Fig. 1. Cartoon illustrating the Ly α transmission profile of galaxies in a plane of the sky around an ionized bubble. *Left:* Galaxies in and around a spherical bubble with the observer at the top. *Right:* The integrated Ly α transmissions, \mathcal{T} as a function of transverse distance (x) measured using galaxies in the green box in the left panel. Transmission profiles $\mathcal{T}(x)$ viewed in different windows (A, B, and C) are different. *Window A:* Galaxies are all outside the bubble and all have $\mathcal{T} \approx 0$, resulting in a flat transmission profile in the window. *Window B:* Galaxies are located around the center of the bubble, therefore \mathcal{T} in the window are all high and the transmission profile is symmetric around the window center. *Window C:* Half of the galaxies are outside the bubble and have $\mathcal{T} \approx 0$. The other half of the galaxies are inside the bubble, having rising transmission toward the bubble center. The transmission profile around the center of window C is therefore asymmetric. The level of asymmetry can therefore be a probe of bubble edges.

profile (window C in the right panel), the window is at the edge of a bubble. By defining and measuring an asymmetry score for these transmission profiles in windows at different positions in a field, it should be easy to detect bubble edges by identifying high asymmetry score regions.

We define an asymmetry vector for each position, \mathbf{r} , in an observed field within a circular window using

$$\mathbf{A}(\mathbf{r}) = \frac{1}{N_{\text{gal,window}}} \sum_{i=1}^{N_{\text{gal,window}}} \frac{\mathcal{T}_i \cdot \mathbf{r}_i}{|\mathbf{r}_i|^2}, \quad (5)$$

where $N_{\text{gal,window}}$ is the number of galaxies in the window centered at \mathbf{r} , \mathcal{T}_i is the Ly α transmission of i -th galaxy, and $\mathbf{r}_i = (\mathbf{r}_{\text{gal},i} - \mathbf{r})/R_{\text{window}}$ is the position of the i -th galaxy with respect to the window center, normalized by the window radius (R_{window}). In practice, we calculate the asymmetry vector along the horizontal and vertical transverse directions. Looking at the center position of window C in the right panel of Figure 1, we see how Equation (5) reflects asymmetry along the x -axis. At the edge of the bubble, only galaxies in the $+x$ direction have high transmission \mathcal{T}_i , thus the asymmetry vector will be positive with large magnitude. Moving into the center of the bubble, galaxies have high transmission equally in the $-x$ and $+x$ directions, thus $\mathbf{A}(x) \rightarrow 0$. We define the asymmetry score as the magnitude of the asymmetry vector ($|\mathbf{A}|$).

Figure 2 shows an example ionization map (left panel) and the asymmetry score map (middle panel) generated using the Ly α transmission of galaxies in that field (assuming perfect recovery of transmission). In addition, we show the asymmetry vector map (right panel). The asymmetry score map shows that at bubble edges, we recover high asymmetry scores as required. Asymmetry scores in bubble centers and neutral regions are low

because of the flat transmission profile. The directions of asymmetry vectors point toward bubble edges. In this paper, for simplicity, we only use the asymmetry score to recover bubble maps. We do not use the direction information of the asymmetry vector in this work, but it is important to note that it could be used to further refine bubble finding algorithms.

Figure 2 shows that high asymmetry score areas match real bubble edges. However, to define bubble edges we must calibrate a threshold in asymmetry. To do this, we convert both the input ionization map and asymmetry map into binary maps (neutral or ionized), and compare how similar the two maps are. Detailed descriptions of binary map conversion and calibration can be found in Appendix A.1. We find the best asymmetry score threshold, A_{thresh} , to be $A_{\text{thresh}} = 0.3$ – 0.4 , and that it is independent of the IGM mean neutral fraction \bar{x} , meaning it is possible to map bubbles using this method irrespective of the surrounding IGM conditions or bubble sizes.

In what follows, we calculate the asymmetry score using the peak value of the Ly α transmission posterior (i.e. the maximum likelihood value) estimated for each galaxy, described in Section 2.3.

3.2. Suggested parameter setup

Our method has three free parameters: asymmetry score threshold (A_{thresh}), window radius (R_{window}), and window depth. Here we first discuss how altering each parameter changes the result, then recommend the choice of each parameter, finding a single set of parameters which describe bubbles well.

Asymmetry score threshold, A_{thresh} , is used to convert asymmetry maps into binary bubble maps. A_{thresh} affects the segmentation and the size of the recovered bubbles. A lower A_{thresh} leads to bigger bubble sizes and smoother bubble shapes. When A_{thresh}

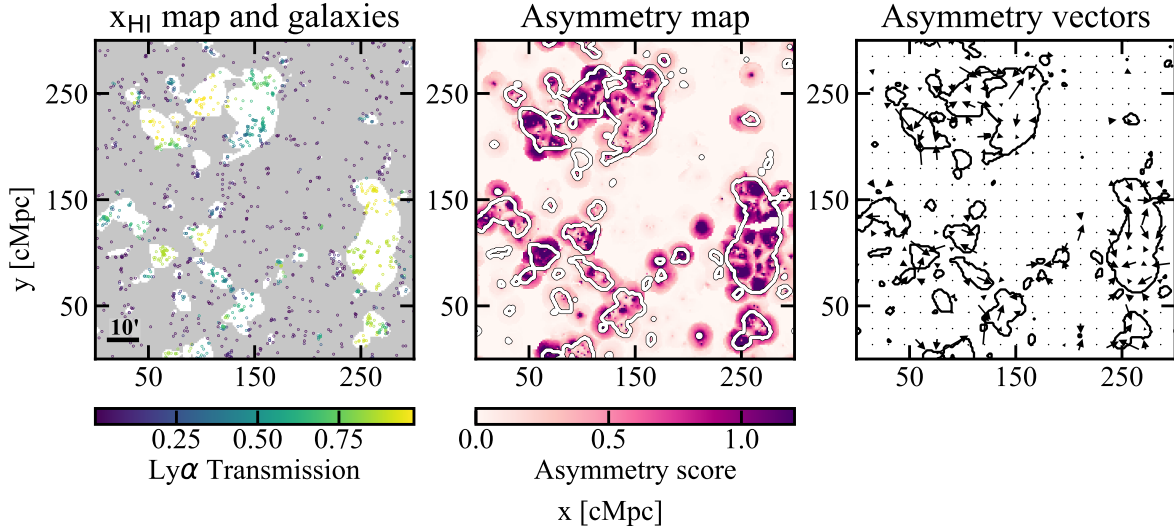


Fig. 2. Example of an asymmetry score map made using galaxies’ Ly α transmission. *Left:* Simulated ionization map (4 cMpc slice from our simulation with mean neutral fraction $\bar{x} = 0.8$) and galaxies used to calculate asymmetry score. White regions are ionized, gray are neutral. The galaxies are color-coded by their modeled Ly α transmission, where we see galaxies with highest transmission are in the centers of bubbles. In the lower left corner we display a 10’ ruler (at $z = 8$) for gauging bubble size in angular scale. *Middle:* Asymmetry score map (red map). We plot the input ionization map (white contours) on top of the map. As expected, bubble edges have higher asymmetry score. *Right:* Asymmetry vector map where we see asymmetry vectors, color-coded by asymmetry score, point toward bubble edges, black contours mark the input ionization map.

is set too high, the resulting bubbles no longer represent ionized regions, but the regions around the edges of bubbles.

Window radius, R_{window} , in our bubble finding algorithm determines the minimum recovered bubble size and the noise on bubble edge recovery, it should be roughly comparable to the minimum bubble radius. The asymmetry score map produced using a smaller R_{window} reflects the ionization fluctuations on a smaller scale, thus it can be good for finding small bubbles. On the other hand, a big bubble may be split into several small bubbles using a small R_{window} . Different reionization morphologies may require different optimal R_{window} . However, this mostly impacts the recovery of very small bubbles (≤ 10 cMpc). To decide on R_{window} , we also need to consider the number of galaxies available for calculating the asymmetry score in each window. A minimum R_{window} can be estimated using the surface density of galaxies Σ_{gal} . The window radius for a desired average number of galaxies per window, N_{window} , will be

$$R_{\text{window}} = \sqrt{\frac{N_{\text{window}}}{\pi \Sigma_{\text{gal}}}}. \quad (6)$$

Window depth is the depth of the window in the redshift direction. It affects the minimum recovered bubble size and the number of galaxies in a field. A deeper window depth will include galaxies in a wider redshift range for the asymmetry map calculation, but will bury the signal from small bubbles.

In Appendix A we test the best A_{thresh} , R_{window} , and window depth combination for $\bar{x} = 0.4$ – 0.9 (i.e. before significant overlap of ionized regions, Miralda-Escudé et al. 2000). Encouragingly, we find the best A_{thresh} (0.3), window radius (12 cMpc), and window depth (4 cMpc, $\Delta z \sim 0.015$ at $z = 8$) remain unchanged across the tested \bar{x} range, meaning our method works generically for a broad range of bubble sizes. We also test the minimum number of galaxies required to calculate the asymmetry score in a window, finding excluding windows with $N_{\text{window}} < 2$ reduces noise in the recovered maps.

3.3. Observational requirements for mapping bubbles in the plane of the sky

We now consider what observational conditions are required to map ionized bubbles in the plane of the sky. We test the number density of sources sampled, and the maximum Ly α EW³ limit required to recover bubble edges. Full details of our tests are described in Appendix B and we summarize the key results here. We find we need to observe ≥ 0.004 galaxies/cMpc³ (corresponding to $M_{\text{UV}} \leq -17.2$ at $z = 8$) and a 5σ Ly α equivalent limit ≤ 30 Å for the faintest galaxies to obtain the best recovery of our simulated bubble maps: our algorithm can recover bubbles with $R \geq 10$ cMpc with $< 30\%$ bubble size precision in this case, but that we can still trace bubbles well with slightly lower number density (≥ 0.002 galaxies/cMpc³, corresponding to $M_{\text{UV}} \leq -17.8$ at $z = 8$).

To fully map bubbles we need the field of view to be larger than bubble diameters. Typical bubble sizes depend on the reionization timeline and ionizing sources, which we discuss further in Section 5.1, but are expected to be $R \sim 2$ – 30 cMpc ($R \sim 2$ – 20 arcmin) during the first half of reionization (Lu et al. 2024). To accurately find bubble edges the field of view must also be larger than our window radius (12 cMpc ≈ 5 arcmin at $z = 8$). These areas are feasible with wide JWST surveys, particularly in the early stages of reionization.

Figure 3 shows the asymmetry maps made using Ly α transmission in our $z = 8$ simulations at $\bar{x} = 0.5$ – 0.9 . We show the maps obtained using galaxies with ≈ 0.002 galaxies/cMpc³ (corresponding to $M_{\text{UV}} \leq -17.8$ at $z = 8$) and Ly α flux limit corresponding to a 5σ Ly α equivalent limit ≤ 30 Å for the faintest galaxies in the map (i.e. brighter galaxies have lower EW limits). We generate the maps by first using the Bayesian method in Section 2.3 to infer $p(\mathcal{T})$ for each galaxy given their observed EW, the uncertainty of EW, and M_{UV} , then calculating asymmetry score at each sky position using the maximum likelihood of

³ We use Ly α EW limit instead of flux sensitivity because in our case the technique is most sensitive to the EW limit, as we need to detect deviations from the EW distribution for the ionized IGM (Section 2.3).

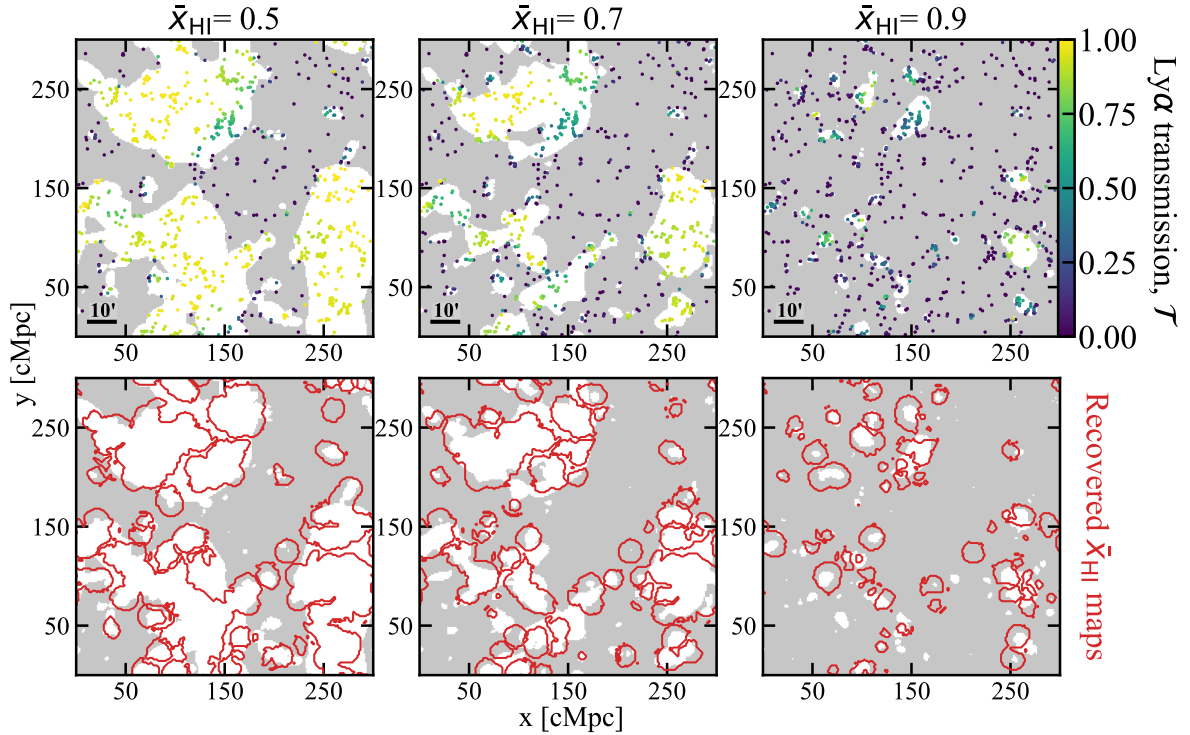


Fig. 3. *Top panels:* Ionization maps and galaxies colored by their Ly α transmissions, at $\bar{x}_{\text{HI}} = 0.5, 0.7,$ and 0.9 for our $z = 8$ simulation. *Lower panels:* Recovered bubble edges (red contours) using the approach described in Section 3 and Appendix A.1 on top of the input ionization maps (white – ionized, gray – neutral). For our mock observations we use a number density of $n_{\text{gal}} = 0.002/\text{cMpc}^3$ (corresponding to a detection limit of $M_{\text{UV}} < -17.8$ at $z = 8$) and a Ly α 5σ equivalent width limit of $\geq 30 \text{ \AA}$ for the faintest galaxies. We see the bubble edges are recovered well at all neutral fractions.

\mathcal{T} of galaxies around the position⁴. We set the asymmetry score to 0 for positions with fewer than 2 galaxies in the window to reduce noise in the recovered map.

In addition to the number density and flux limit requirements, we also require precise systemic redshift resolution. The window depth must be comparable to the diameter of typical bubbles – in the first half of reionization the median bubbles have radii of $\lesssim 10\text{--}30 \text{ cMpc}$ (Lu et al. 2024). Thus, to capture these bubbles, asymmetry maps should be made using galaxies in a thin redshift window (window depth = 4 cMpc or $\Delta z \sim 0.015$ at $z = 8$, as described above in Section 3.2), requiring spectroscopic observations (i.e. narrow-bands are too broad to cover single bubbles during most of reionization, see also Sobacchi & Mesinger 2015, for similar conclusions).

Finally, to map entire ionized bubbles in the plane of the sky requires a field of view (FoV) that is larger than bubble diameters. For reference, the 1% largest ionized bubble diameter at $\bar{x} = 0.5$ (0.9) is ~ 300 (60) cMpc or ~ 100 (20) arcmin for reionization driven by high mass halos (see Figure 10, also Lu et al. 2024). In the earliest stages of reionization ($\bar{x} \gtrsim 0.8$), it should thus be feasible to map entire ionized regions with deep JWST observations.

Upcoming wide-area imaging and spectroscopy surveys, such as the Euclid Deep survey (Euclid Collaboration: Scaramella et al. 2022; van Mierlo et al. 2022; Marchetti et al.

⁴ We use the maximum likelihood of \mathcal{T} rather than sampling from the inferred $P(\mathcal{T})$ distributions because when the EW limit is shallow, $P(\mathcal{T})$ is very broad, which we found leads to many false bubbles in the recovered maps. Using the maximum likelihood \mathcal{T} means that galaxies where the transmission is well constrained to $\mathcal{T} \approx 0$ or $\mathcal{T} \approx 1$ contribute most to the asymmetry score, improving the recovery of the bubbles. This is further demonstrated in Appendix B.2.

2017), Roman High Latitude Spectroscopy Survey (HLSS; Wang et al. 2022), Subaru PFS (Greene et al. 2022), and VLT MOONS (Maiolino et al. 2020; Gagnon-Hartman et al. 2025; Mesinger et al. 2024), will observe thousands of reionization-era galaxies on $>1 \text{ sq. deg}$ scales required to cover multiple ionized regions in the mid-stages of reionization and observe their Ly α . However, these surveys will not be deep enough to recover bubble maps by themselves, using the method we describe here, as they will not reach the required galaxy number density and Ly α EW limits. However the bright Ly α emitters detected in these surveys should signpost large ionized regions and would provide ideal targets for deep spectroscopic follow-up with JWST. We will discuss observational prospects for mapping ionized regions in more detail in Section 5.1.

4. Mapping bubbles along the line-of-sight

JWST’s field of view (spanning $\approx 10 \text{ cMpc}$ at $z \sim 8$) can capture entire bubbles in the early stages of reionization, but JWST can also efficiently map ionized structures along the line-of-sight (LOS) direction at all stages of reionization. The median bubble size at the mid-point of reionization spans $\Delta z \sim 0.1\text{--}0.3$. JWST can detect galaxies over such a wide redshift window with precise systemic redshift resolution ($R \sim 1000$ corresponding to $\Delta z \lesssim 0.001$ with NIRSpec medium resolution grating and NIRCам grism spectroscopy, e.g. Tang et al. 2023; Oesch et al. 2023; Meyer et al. 2024). Here we describe a novel algorithm to recover bubble maps in redshift skewers using the LOS transmission profile.

Instead of the bell-shaped Ly α transmission profile for bubbles in the plane of the sky (Figure 1), the transmission profile in

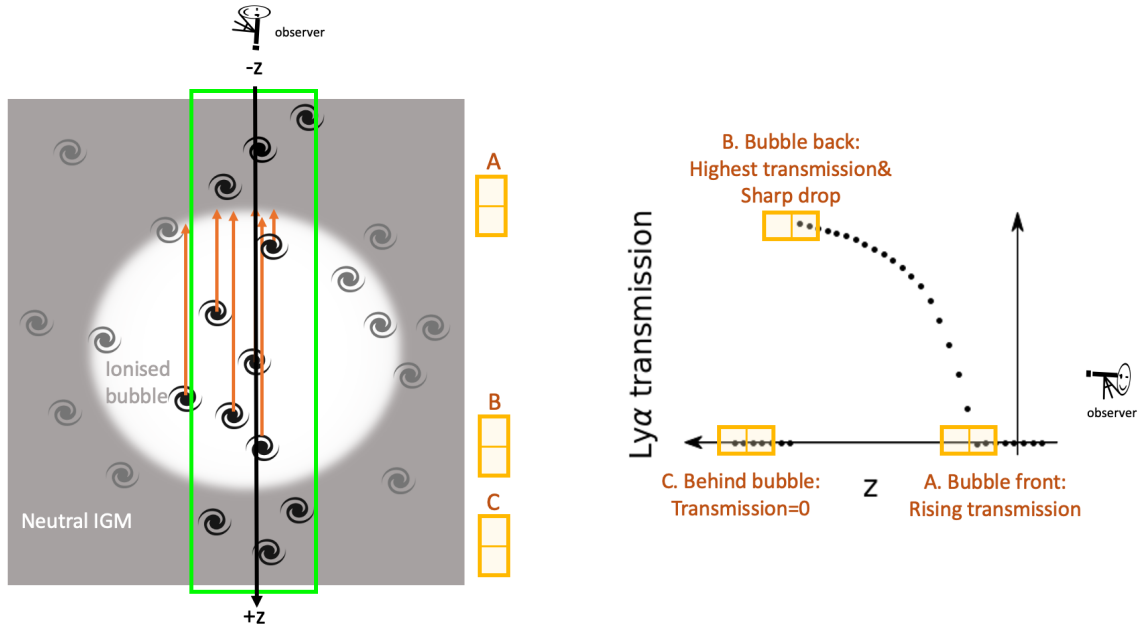


Fig. 4. Cartoon illustrating the transmission profile of galaxies along the LOS direction around a bubble. *Left:* Galaxies in and around a spherical bubble. *Right:* The integrated Ly α transmissions as a function of the LOS distance (d) measured using galaxies in the green box in the left panel. Transmission profiles (T)(z) viewed in different windows (A, B, and C) are different. *Window A:* Galaxies in front of the bubble are in fully neutral gas and thus experience a strong Ly α damping wing and have almost zero transmission. Stepping back into the bubble the transmission increases because of the increased path length of ionized IGM in front of galaxies. *Window B:* Galaxies located at the back of the bubble have the highest transmission because they have the longest path length through the ionized IGM. *Window C:* Galaxies on the far side of the bubble are in fully neutral IGM and have almost zero transmission. Galaxies' integrated Ly α transmission is tightly correlated with the distance between the galaxy and the nearest neutral IGM patch along the LOS. We can use this correlation to map ionized regions along the LOS (Section 4).

the LOS direction has a shark-fin-like shape (see Figure 4 for a cartoon illustration). As we move along the LOS away from the observer, the Ly α transmission changes as follows: in front of a bubble (closest to the observer), we expect almost no transmitted Ly α flux from galaxies. Stepping into the bubble, the closer a galaxy is to the back end of the bubble, the higher the Ly α transmission, because of the increasing path length of ionized IGM in front of galaxy. Galaxies at the back of the bubble have the highest Ly α transmission. When stepping out of the back size of the bubble, we again expect almost no transmitted Ly α flux from galaxies. Thus, the points of nonzero Ly α transmission should mark the edges of the bubble.

There is a tight correlation between a galaxy's integrated Ly α transmission \mathcal{T} and the line-of-sight distance from the galaxy to the first neutral patch, d_{IGM} , which we can use to map the bubble. This is because the damping wing optical depth is dominated by the first neutral IGM along the line-of-sight (Mesinger & Furlanetto 2008). We derive $\mathcal{T}(d_{\text{IGM}})$ and its 16–84 percentile range numerically from our simulations by measuring the \mathcal{T} of galaxies with the same d_{IGM} along 10^5 sightlines (Section 2.2). Figure 5 shows $\mathcal{T}(d_{\text{IGM}})$ in our simulation at $z \approx 8$ for $\bar{x} = 0.5, 0.7, 0.9$. We find $\mathcal{T}(d_{\text{IGM}})$ has negligible dependence on \bar{x} at fixed redshift. We also compare $\mathcal{T}(d_{\text{IGM}})$ for galaxies with emitted Ly α velocity offset $\Delta v_{\text{Ly}\alpha} > 200$ km/s and $\Delta v_{\text{Ly}\alpha} < 200$ km/s, finding $\mathcal{T}(d_{\text{IGM}})$ has only a small dependence on velocity offset once a galaxy is at least > 10 cMpc from a neutral patch (see also, e.g. Endsley & Stark 2022; Prieto-Lyon et al. 2023). We see the variance (coming from the distribution of additional neutral patches along sightlines) of the $\mathcal{T}(d_{\text{IGM}})$ curve is also small, resulting in less than 10 cMpc under- or over-estimate of d_{IGM} when estimating d_{IGM} from \mathcal{T} .

We note that the relation shown in Figure 5 is only valid for $z = 8$, as the damping wing optical depth increases with increasing redshift due to increasing gas density (Gunn & Peterson 1965), but the minimal dependence on \bar{x} and velocity offset for > 10 cMpc bubbles is qualitatively true at other redshifts.

Figure 6 shows a sightline in our simulation at $\bar{x} = 0.8$, selected in a sky area corresponding to 3 JWST NIRSpec footprints (42.8 arcmin^2 at $z = 8$), collapsed along the y direction. We show ionized regions and the positions of all $M_{\text{UV}} < -17.5$ galaxies. We also plot the integrated Ly α transmissions and Ly α EW of these galaxies as a function of LOS distance, along with our averaged $\mathcal{T}(d_{\text{IGM}})$ profile based on the positions and sizes of the recovered bubbles in the skewer. The Ly α transmissions of galaxies in each bubble follow our averaged profile very well, with all Ly α transmissions falling along or below the curve.

The scatter of transmission below the $\mathcal{T}(d_{\text{IGM}})$ curve mainly comes from bubbles which are clumpy along the line of sight, so galaxies at the same redshift within the field of view are not necessarily at the same distance to neutral gas. A minor cause of the scatter of transmissions ($\sigma_{\mathcal{T}} < 0.05$) is the Ly α emission line velocity offset. A strongly red-shifted Ly α emission line (e.g. > 300 km/s) will suffer less from the damping wing absorption and therefore have higher transmission than the expected value in our $\mathcal{T}(d_{\text{IGM}})$ relation.

In this paper we focus on estimating the maximum bubble diameter along the LOS. Therefore we want to find a $\mathcal{T}(d_{\text{IGM}})$ curve for each bubble which envelopes the observed transmission spikes, and thus marks the front and back of the bubble. In the next section we describe a simple algorithm to do this.

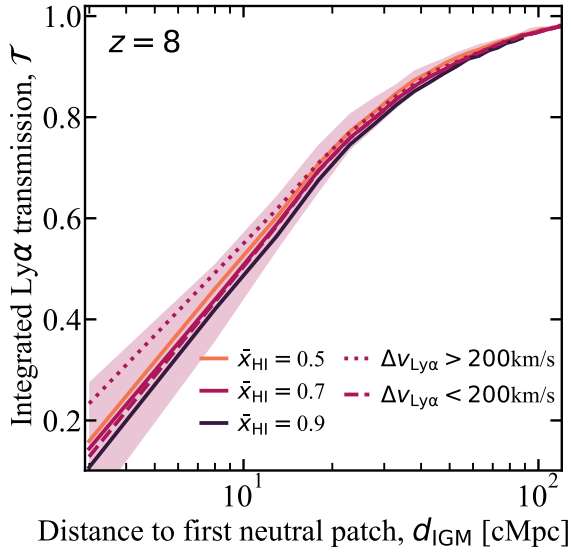


Fig. 5. Integrated Ly α transmission as a function of the distance between a galaxy and the first neutral IGM patch the galaxy light encounters at $z = 8$. Solid lines in orange, maroon, and black curves show $\mathcal{T}(d_{\text{IGM}})$ for all the galaxies at $\bar{x} = 0.5, 0.7, 0.9$, respectively. The shaded region shows the 16–84 percentile range of the $\bar{x} = 0.7$ curve, for $d_{\text{IGM}} > 10$, the 16–84 percentile range is $\Delta\mathcal{T} < 0.1$. Maroon dashed and dotted lines show $\mathcal{T}(d_{\text{IGM}})$ at $\bar{x} = 0.7$, made using galaxies with emitted Ly α velocity offsets $\Delta v_{\text{Ly}\alpha} > 200$ km/s and $\Delta v_{\text{Ly}\alpha} < 200$ km/s respectively. None of these curves differ significantly, meaning $\mathcal{T}(d_{\text{IGM}})$ does not depend strongly on \bar{x} or $\Delta v_{\text{Ly}\alpha}$, especially for $d_{\text{IGM}} > 10$ cMpc. The curves are for $z = 8$ only, but the qualitative trends hold at other z .

4.1. An algorithm for finding bubbles along the LOS

To detect bubbles along a LOS using Ly α transmission estimates, we develop a simple algorithm. For each galaxy we require an estimate of the integrated Ly α transmission (e.g. via comparing the observed Ly α EW to the predicted EW based on $z \sim 5$ –6 observations, as described in Section 2.3), a precise systemic redshift ($\sigma(z) < 0.015$) and sky coordinates so that we can make a 3D map of galaxy positions. The steps of the algorithm are as follows and illustrated in Figure 7:

1. Starting from the galaxy farthest away from the observer (galaxy with the highest redshift), g_i , we invert the $\mathcal{T}(d_{\text{IGM}})$ curve (Figure 5) to estimate the distance from g_i to the front of the bubble, D_i using the estimated Ly α transmission of g_i .
2. If g_i is the first galaxy in a bubble, we set the redshift of the back of the bubble to the redshift of g_i , $z_{\text{back}} = z_i$. We set the LOS bubble diameter $D_{\text{bubble}} = D_i$, and mark all the galaxies within a distance D_{bubble} in front of g_i to be in the same bubble as g_i (panel a of Figure 7).
3. A galaxy’s Ly α transmission may not follow exactly $\mathcal{T}(d_{\text{IGM}})$ due to an irregular bubble shape or over-predicted emergent Ly α line, and thus may fall below the $\mathcal{T}(d_{\text{IGM}})$ line. To make sure all the galaxies in the bubble have transmission falling on or below the $\mathcal{T}(d_{\text{IGM}})$ curve, we move to the next galaxy (galaxy with the second highest redshift), g_{i+1} , a distance $d_{i,i+1}$ from g_i . We estimate the distance of g_{i+1} from the neutral IGM, D_{i+1} by inverting $\mathcal{T}(d_{\text{IGM}})$ as above.
4. If $D_{i+1} + d_{i,i+1} > D_{\text{bubble}}$, we update $D_{\text{bubble}} \rightarrow D_{i+n} + d_{i,i+n}$ and again mark galaxies within the updated bubble (panel b of Figure 7).
5. By iterating through all the galaxies in the sightline, we can identify all the ionized bubbles (panel c of Figure 7).

To account for the uncertainty in Ly α transmission estimates we use the following approach. For each mock galaxy, we use the lower 16% bound on transmission inferred using the Bayesian approach described in Section 2.3 as the input \mathcal{T} . We use this approach rather than the full inferred $P(\mathcal{T})$ or maximum likelihood \mathcal{T} because of the exponential-like relation between distance to HI and Ly α transmission (Figure 5), this would result in a bias to overestimated bubble sizes in our algorithm. We find using the lower bounds of \mathcal{T} always provides lower bound on bubble size that converges to the true bubble size when we include sufficient galaxies in the sightline (see Section 4.2).

4.2. Observational requirements for mapping bubbles along the line-of-sight

To detect bubbles along the line-of-sight requires (1) a good estimate of Ly α transmission for each galaxy, which depends on the Ly α EW limit, and (2) a sufficient number of galaxies in the line-of-sight to sample the region. To test the required observational constraints we apply our algorithm to sightlines extracted from our simulations, assuming different spectroscopic completeness and Ly α EW limits.

We sample 900 10 cMpc \times 10 cMpc \times 300 cMpc sightlines from our $\bar{x} = 0.8$ simulation cube, corresponding to individual JWST NIRSpec pointings spanning $\Delta z \approx 1$ at $z \sim 8$. For each galaxy in each sightline, we infer $p(\mathcal{T})$ using the Bayesian approach in Section 2.3. We apply our algorithm to detect bubbles along these sightlines given galaxy positions and the lower limit of the inferred $p(\mathcal{T})$ (see discussion in Section 4.1 above).

In Figure 8 we plot the median recovered bubble diameter, D_{recover} , from our algorithm versus true bubble diameter, D_{true} , for ~ 5000 bubbles. We calculate the true bubble positions and diameters in each sightline by collapsing the IGM skewer in both spatial directions, measuring the mean neutral fraction at each redshift position in the 1D skewer, and classify regions that have mean neutral fraction below $x_{\text{HI}} = 0.5$ as ionized. We match the recovered bubbles to the true bubbles by iterating over the true bubbles and searching for recovered bubbles that overlap with them. If a true bubble has no corresponding recovered bubble, we record the size of the recovered bubble as 0 cMpc. If more than one recovered bubble overlaps with a true bubble, we count both recovered bubbles. We compare the results assuming a number density ≈ 0.004 galaxies/cMpc 3 in each skewer (corresponding to complete spectroscopy of $M_{\text{UV}} \leq -17.2$ galaxies at $z \approx 8$ in our simulations) using Ly α flux limits corresponding to 5σ EW limits of 10–80 Å for the faintest galaxies (i.e. brighter galaxies will have lower EW limits). For >80 cMpc bubbles, the median recovered bubble size drops because some sightlines pass through big ionized bubbles but lack sufficient galaxies to sample the space, resulting in big ionized bubbles being recovered to be spatially close smaller bubbles. We note that this ‘bubble splitting’ issue can be reduced if the estimation of \mathcal{T} is improved.

We see our algorithm can robustly recover the sizes of bubbles along sightlines extremely well, to $\lesssim 5\%$ for $D_{\text{true}} \approx 10$ –80 cMpc bubbles, providing the flux limit corresponds to a Ly α EW limit of $\lesssim 30$ Å, roughly the median EW expected in a fully ionized IGM (see Section 2.3). Even with an EW limit of $\lesssim 50$ Å we can obtain a lower limit on the bubble size within 30% of the true size. For true bubbles with $D_{\text{true}} \lesssim 15$ cMpc, we recover larger bubble diameters because we set the mean x_{HI} threshold to 0.5 when identifying true bubbles in 1D skewers, which captures only the most ionized region at the center of small bubbles.

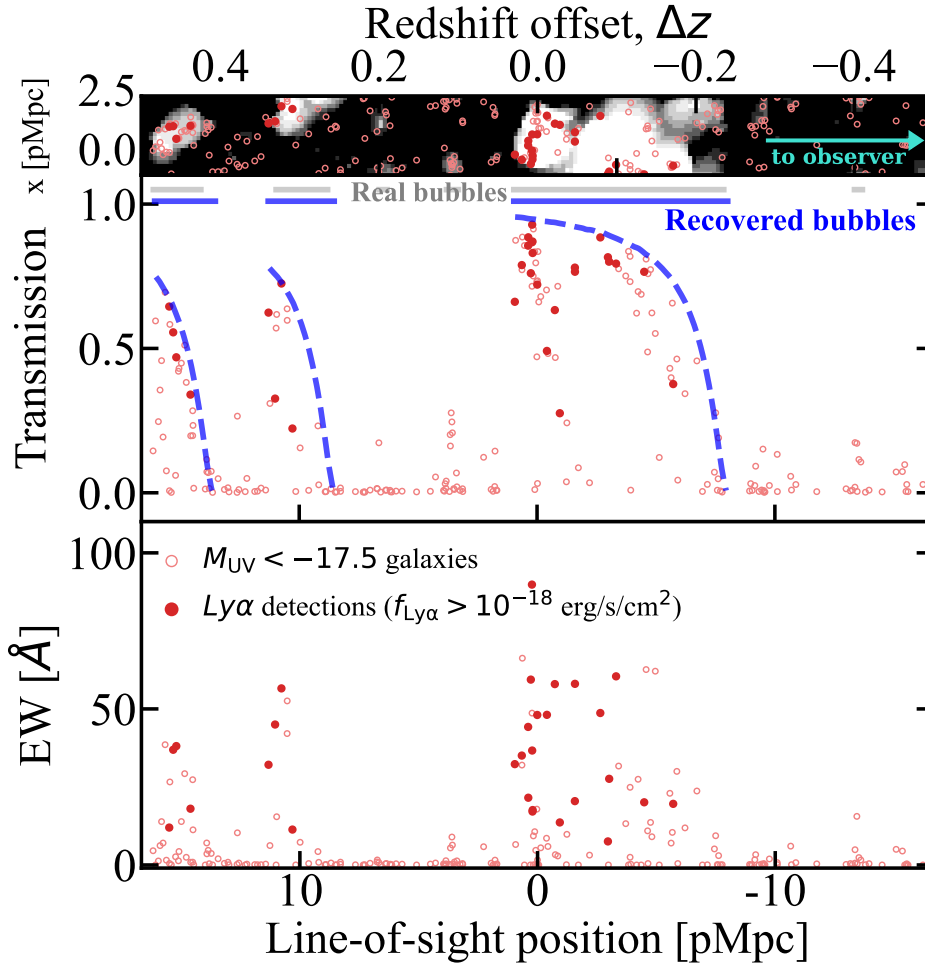


Fig. 6. Example sightline skewer showing galaxies’ $Ly\alpha$ transmissions and EW as a function of the LOS position, where the zero point of the LOS position is arbitrary. The *top panel* shows the ionization map in the x - z plane (white ionized, black neutral gas). Red empty and solid dots are galaxies with $M_{UV} > -17.5$ in the slice. Red solid points are galaxies with $f_{Ly\alpha} > 10^{-18}$ erg/s/cm². The *middle panel* shows the $Ly\alpha$ transmission for each galaxy as a function of LOS position. Galaxies in bubbles show high transmission while galaxies in neutral IGM having transmission ≈ 0 . Horizontal Gray lines show the spans of the simulated ionized bubbles along the sightline. Blue dashed lines are the $\mathcal{T}(d_{IGM})$ curves derived using the damping wing optical depth (Figure 5), galaxy transmissions follow this curve very well. The blue horizontal lines show the span of the recovered bubbles using our method (Section 4.1). The *bottom panel* shows the observed $Ly\alpha$ EW of the galaxies, highlighting that strong $Ly\alpha$ EW ≥ 50 Å likely signpost large ionized regions.

We find the number density of galaxies along a sightline is the dominant factor for the accuracy of the recovered bubble size. Figure 9 shows the ratio between recovered bubble size and real bubble size versus galaxy number density along sightlines, obtained by sampling recovered bubbles in our simulations varying the $M_{UV,limit}$ detection limit between -19 and -17 to sample a range of observed galaxy number densities. We find we can recover the true bubble size within $<30\%$ when the observed number density of galaxies reaches $\geq 0.004/\text{cMpc}^3$ and $D_{recover}/D_{true}$ converges to ≈ 1 when $\geq 0.005/\text{cMpc}^3$. We recover a slightly larger (1.1 – $1.2\times$) median bubble size compared to the ‘true’ bubble size because of the irregular shapes of bubbles, where small ionized projections at the start or end of bubbles are identified via strong $Ly\alpha$ emission by our mapping algorithm but can be missed in our calculation of ‘true’ bubble positions in the 1D skewers. Future works carefully accounting for the 3D shapes of bubbles will improve the bubble size comparison.

We also test the effect of the systemic redshift uncertainty and find it has a minor impact on the recovered bubble size as long as it is smaller than the bubble diameter. We find our method requires a redshift uncertainty of $\sigma_z \lesssim 0.015$ (~ 5 cMpc

at $z = 8$). This requirement is achievable with JWST NIRSpec grating and NIRCам grism spectroscopy (which can measure $\Delta z \approx 0.001$).

5. Discussion

JWST’s ability to spectroscopically confirm and detect $Ly\alpha$ from $M_{UV} \lesssim -17$ galaxies opens a new window on reionization, providing the potential to map ionized regions in four dimensions for the first time. This is a substantial step beyond the volume-averaged view we have had of the $z > 6$ IGM prior to JWST (e.g., Stark et al. 2010; Schenker et al. 2014; Pentericci et al. 2014; Greig et al. 2017; Davies et al. 2018; Mason et al. 2018a; Jung et al. 2020; Bolan et al. 2022). We have developed ionized bubble mapping methods which, for the first time, use information from ensembles of galaxies (see also our companion paper, Nolić et al., in prep.). In doing so, our methods mitigate the effects of intrinsic variability of $Ly\alpha$ transmission to enable robust estimates of the positions and sizes of ionized bubbles. In this section we discuss optimal strategies for mapping individual ionized regions to understand reionization on a local scale

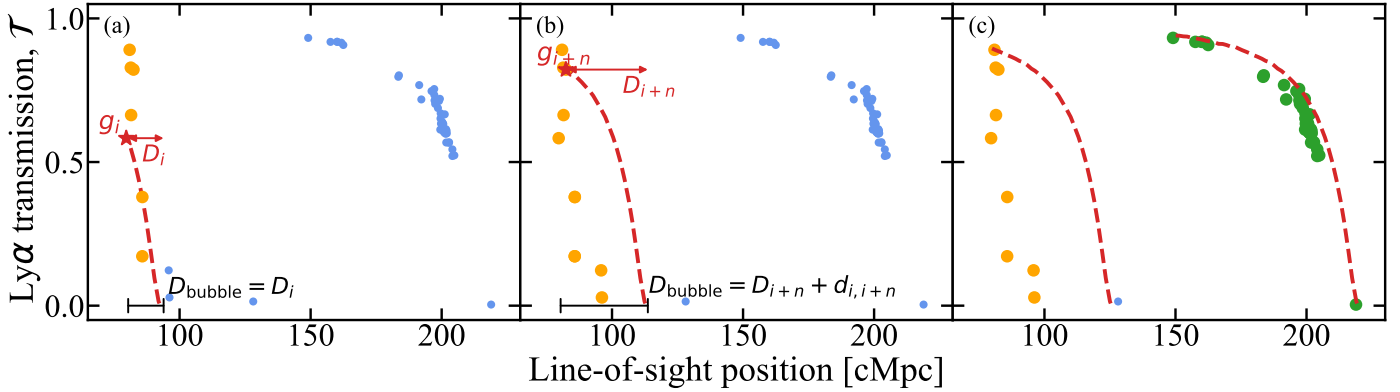


Fig. 7. Demonstration of our LOS-bubble-finding algorithm, where the zero point of the LOS position is arbitrary. (a) Starting from the galaxy farthest away from the observer, g_i , we invert $\mathcal{T}(d_{\text{IGM}})$ (Figure 5, the red dashed line) to estimate the distance between the galaxy and the nearest neutral IGM patch, D_i . We set the LOS bubble diameter, $D_{\text{bubble}} = D_i$, and the redshift of the back of the bubble, $z_{\text{back}} = z_i$, that is, the redshift of the first galaxy. We then mark all galaxies in front of g_i within D_{bubble} (orange dots) to be inside the same bubble as g_i . (b) We then step through the bubble toward the observer to update our estimate of D_{bubble} . For a galaxy g_{i+n} which is a distance $d_{i,i+n}$ from galaxy g_i we estimate the distance of galaxy g_{i+n} from neutral IGM, D_{i+n} . If $D_{i+n} + d_{i,i+n} > D_{\text{bubble}}$ we update $D_{\text{bubble}} \rightarrow D_{i+n} + d_{i,i+n}$. (c) Once we have iterated through all galaxies contained within D_{bubble} and D_{bubble} converges, we continue iterating through all the galaxies in the line-of-sight in the same manner to find all the bubbles. In this example sightline we find two bubbles and their associated galaxies (orange galaxies on the left and green galaxies on the right, the remaining blue point is a galaxy between the two bubbles and is not considered to be inside a bubble).

(Section 5.1) and prospects for constraining the bubble size distribution (Section 5.2).

5.1. Optimal observational strategies for mapping ionized regions

In Sections 3 and 4 we presented two simple methods for mapping ionized regions in the plane of the sky and along the line of sight. To map an ionized region and estimate its size requires observing a volume sufficient to detect the edges of the bubble, a sufficient number density of spectroscopically confirmed galaxies to trace the volume, $\geq 0.002\text{--}0.005$ galaxies/cMpc³, systemic redshifts (i.e. from nonresonant lines such as [OIII]5007) with high redshift precision ≤ 0.015 to map galaxies on scales smaller than the expected bubble sizes, and deep Ly α observations, reaching 5σ rest-frame EW limits of ≈ 30 Å for the faintest ($M_{\text{UV}} \sim -17$) surveyed galaxies, to estimate the transmission of Ly α through the IGM, within reach of deep JWST observations (e.g., Jones et al. 2024; Tang et al. 2024a).

The required integration times to estimate Ly α transmission and determine the spectroscopic redshifts of galaxies are easily achievable for $z \sim 7$ galaxies and deep, but feasible, at $z \sim 8$. To achieve a $5\text{-}\sigma$ detection limit of 30 Å for Ly α emission from $M_{\text{UV}} \sim -18.5$ (-17.2) galaxies at $z \sim 7$ ($z \sim 8$) using the G140M filter on JWST, the integration time required is 2 hours (40 hours). For redshift confirmation using [O III] emission in $M_{\text{UV}} \sim -18.5$ (-17.2) galaxies at $z \sim 7$ ($z \sim 8$) with the G395M filter on JWST, the required integration time is 6 hours (130 hours), assuming that more than 90% of galaxies with $M_{\text{UV}} \leq -18$ have [O III] equivalent widths exceeding 120 Å (Endsley et al. 2024). For $z \sim 8$ sources, targeting galaxy overdensities would achieve the necessary galaxy number density with more reasonable integration times (e.g., Meyer et al. 2024; Tang et al. 2024a; Nikolić et al. 2025).

To understand the required volumes for mapping ionized bubbles in Figure 10 we show the diameter of ionized bubbles (median and 68% range) predicted by the simulations by Lu et al. (2024), in both the plane of the sky in arcmin (left axis) and along the line-of-sight in Δz (right axis) at $z = 8$.

We show the predictions for two reionizing source models, a gradual reionization driven by galaxies in low mass ($M_{\text{h}} > 5 \times 10^8 M_{\odot}$) halos, versus a more rapid reionization driven by galaxies in higher mass halos ($M_{\text{h}} > 10^{11} M_{\odot}$). These models should span the expected range of bubble sizes. We compare the bubble sizes of ionized regions to the field of view of various JWST Cycle 1–3 spectroscopy surveys (assuming square survey areas): JADES GOODS-N+GOODS-S (Bunker et al. 2024), CEERS (Finkelstein et al. 2022) and COSMOS-3D (Kakiichi et al. 2024). We note only CEERS and JADES obtained Ly α spectra. As noted in Section 3.3, upcoming wide-area spectroscopic surveys with, for example, Roman, Euclid, PFS, MOONS (Cirasuolo et al. 2012; van Mierlo et al. 2022; Euclid Collaboration: Scaramella et al. 2022; Wang et al. 2022; Greene et al. 2022; Maiolino et al. 2020; Gagnon-Hartman et al. 2025; Mesinger et al. 2024) will span >1 sq. deg and contain many bubbles signposted by bright Ly α -emitters, but are not sensitive enough by themselves to detect the high number density of galaxies we need for mapping. Thus deep follow-up with JWST will be essential.

From Figure 10 it is clear that mapping whole ionized regions in the plane of the sky with JWST becomes most feasible when $\bar{x} \gtrsim 0.5$. In the mid-stages of reionization ($\bar{x} \sim 0.4\text{--}0.6$) the field of view of the marked JWST surveys may be covered by a single ionized region. This is consistent with the high field-to-field variation in the Ly α fraction measured from current ~ 100 sq. arcmin JWST fields (Napolitano et al. 2024; Tang et al. 2024a). Wide area JWST spectroscopy in fields with high Ly α detection fractions should be able to map whole bubbles in the mid-to-early stages of reionization. Figure 10 demonstrates that mapping bubbles along the line of sight is an efficient approach for estimating bubble sizes with JWST at all stages of reionization, as the largest bubbles at $\bar{x} \gtrsim 0.5$ span $\Delta z < 0.7$, which can be easily mapped with the wide wavelength coverage of JWST NIRSpec and NIRCам grism.

In the past few years, ground-based observations have found evidence for clustered Ly α emitters at $z \sim 7$ (e.g., Vanzella et al. 2011; Tilvi et al. 2020; Jung et al. 2022), implying the presence of ionized regions, but have been limited in the ability to detect Ly α from UV faint ($M_{\text{UV}} \gtrsim -20$) galaxies. Excitingly,

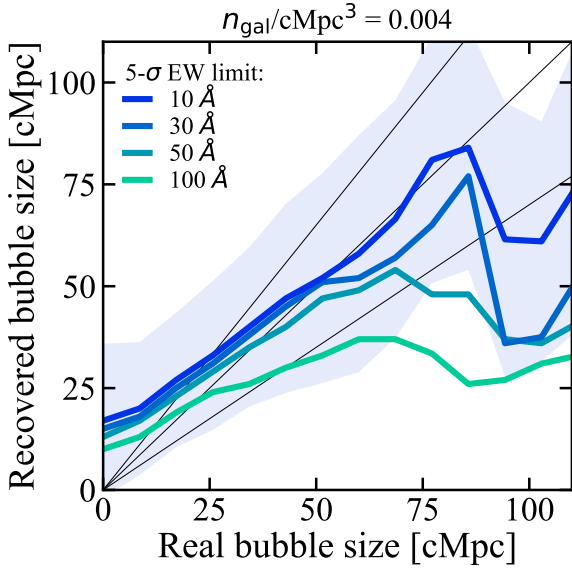


Fig. 8. Median recovered bubble size versus real bubble size for LOS bubbles assuming a number density of $0.004 \text{ galaxies/cMpc}^3$ (corresponding to a complete spectroscopic sampling of galaxies down to $M_{\text{UV,limit}} = -17.2$ at $z \approx 8$) for $5\sigma \text{ Ly}\alpha$ equivalent width limits = $[10, 30, 50, 100] \text{ \AA}$ for the faintest galaxies (colored lines). The shaded region shows the 68% range of recovered bubble sizes for the $\text{Ly}\alpha$ equivalent width limit = 10 \AA line. When the $\text{Ly}\alpha$ equivalent width limit is deeper than 50 \AA we can estimate \mathcal{T} relatively well for each galaxy and thus we can recover bubble sizes $\sim 10\text{--}80 \text{ cMpc}$ to within 30% of their true values ($1\text{--}1$ and $1\text{--}1 \pm 0.3$ relations marked by solid gray lines). The dip at $R_{\text{real}} \sim 100 \text{ cMpc}$ is due to sightlines that pass through a big bubble but without enough galaxies to sample the space, causing us to recover a few smaller bubbles instead of one big bubble. This could be improved in future work when \mathcal{T} can be more precisely estimated, for example, with a prior from the $\text{Ly}\alpha$ escape fraction (see Section 2.3).

early JWST studies have demonstrated that comprehensively mapping bubbles along the line of sight is now finally feasible. Chen et al. (2024), Witstok et al. (2024) and Tang et al. (2024a) recently showed that with JWST spectroscopy it is possible to start mapping out the 3D distribution of galaxies and their $\text{Ly}\alpha$ around high EW ($\geq 50 \text{ \AA}$) $\text{Ly}\alpha$ -emitting galaxies, which likely signpost ionized regions. JWST has vastly increased the number of detectable sources compared to ground-based studies. Our line-of-sight bubble mapping method requires a density of $\geq 0.004 \text{ galaxies/cMpc}^3$, corresponding to complete spectroscopic observations of $M_{\text{UV}} \gtrsim -18$ (-17.2) galaxies at mean density at $z \approx 7$ (8) (e.g., Bouwens et al. 2021), but these densities could be obtained with a shallower detection limit in an overdense sightline, to obtain an estimate of the expected bubble size. The outlook for identifying and obtaining spectra of these required galaxy number densities is excellent with JWST: for example, around one of the strongest known $z > 7 \text{ Ly}\alpha$ emitters, JADES-13682 ($z = 7.3$, $\text{EW} \approx 400 \text{ \AA}$, Saxena et al. 2023), estimates of a factor $\sim 7\text{--}25\times$ overdensity have been reported using NIRCcam imaging and grism spectroscopy (Endsley et al. 2023; Witstok et al. 2024; Tang et al. 2024a). Reaching $\text{EW} \approx 30 \text{ \AA}$ for $M_{\text{UV}} = -18$ (-17.2) galaxies requires ≈ 8 (45) h with JWST NIRSpec (e.g., Jones et al. 2024; Tang et al. 2024a), feasible with deep JWST programs.

The methods we have presented here could be further refined in several ways in future work. As described in Section 2.3, estimates of $\text{Ly}\alpha$ escape fractions could be used as a prior on $p(\mathcal{T})$, which would further improve our estimates of the IGM trans-

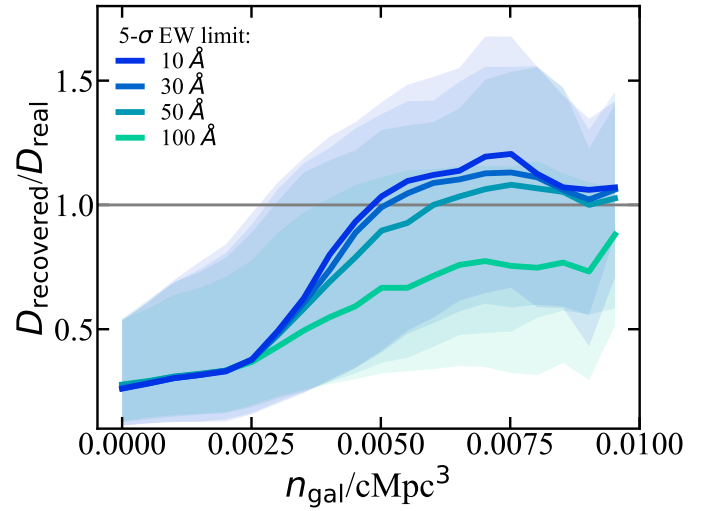


Fig. 9. Median recovered bubble size to real bubble size ratio ($D_{\text{recovered}}/D_{\text{real}}$) versus galaxy number density for $5\sigma \text{ Ly}\alpha$ equivalent limits = $[10, 30, 50, 100] \text{ \AA}$ (solid lines) binned by galaxy number density. Shaded regions show the 16–84 percentile of the recovered size ratio. The method converges to the correct bubble size as we reach $\geq 0.004 \text{ galaxies/cMpc}^3$.

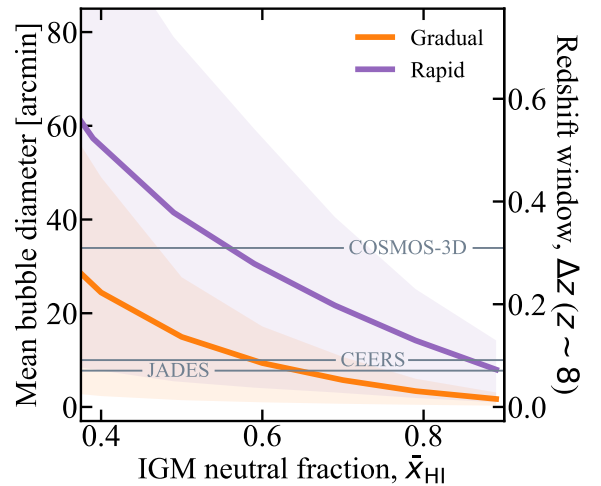


Fig. 10. Mean bubble diameter as a function of \bar{x} . We adopt the characteristic bubble size measured using the mean-free-path method in Lu et al. (2024) and assume $z = 8$ (there is minimal impact on the physical sizes over the epoch of reionization, $z \sim 6\text{--}10$). We show bubble diameter in arcmin on the plane-of-the-sky on the left axis and along the line-of-sight in redshift width on the right axis. Gray lines mark the area of existing and upcoming spectroscopic surveys with JWST.

mission and thus bubble maps. Additionally, with deep high-resolution ($R > 2000$) spectroscopy to resolve the $\text{Ly}\alpha$ lineshape at high S/N, it could be possible infer the wavelength-dependent damping wing attenuation (Mason & Gronke 2020), rather than just the integrated transmission. This would provide a more precise estimate of $\text{Ly}\alpha$ transmission for each galaxy and thus of the bubble size and position (Nikolić et al. 2025). Furthermore, in this work we have not considered the impact of self-shielding systems on $< \text{cMpc}$ scales inside ionized regions, which would add additional $\text{Ly}\alpha$ opacity (e.g., Bolton & Haehnelt 2013; Mesinger et al. 2015; Mason & Gronke 2020; Nasir et al. 2021). However, these absorbers are predicted to be small ($\approx 1\text{--}20 \text{ ckpc}$) and not uniformly distributed in the IGM and thus the probability

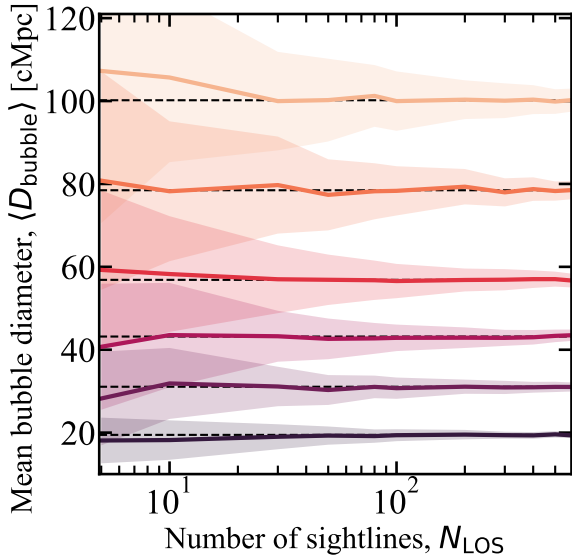


Fig. 11. Mean bubble diameter recovered from measuring bubbles along N_{LOS} sightlines using our bubble mapping algorithm (Section 4), for simulations with different bubble-size distributions. The solid and shaded regions show the median and 68% range of the recovered mean bubble size from 150 realizations. With ≥ 15 –30 sightlines it would be possible to distinguish different bubble size distributions.

of the majority of galaxies intersecting an absorber along the line of sight is low. We note that [Chen et al. \(2024\)](#) demonstrated that strong small-scale (< 100 pkpc) galaxy overdensities may contain dense neutral gas which attenuates $\text{Ly}\alpha$, though this is on much smaller scales than the expected sizes of bubbles. Future systematic observations of $\text{Ly}\alpha$ visibility as a function of overdensity will be important for understanding systematics in our methods due to residual neutral gas in ionized regions. However, as our methods utilize ensembles of galaxies and rely on ‘edges’ in $\text{Ly}\alpha$ transmission at fixed redshift, rather than absolute transmission, our approach should not be strongly impacted by the inclusion or redshift evolution of local small-scale absorbers in the ionized IGM or CGM.

The prospects for obtaining larger samples of ionized regions along multiple sightlines are promising. Upcoming wide-area imaging and spectroscopic surveys as described above, and on-going $\text{Ly}\alpha$ narrow-band surveys (e.g., [Ouchi et al. 2018](#); [Hu et al. 2019](#)) will provide thousands of bright $z \gtrsim 7$ galaxy candidates and $\text{Ly}\alpha$ -emitters (flux $\gtrsim 1 \times 10^{-17}$ erg s $^{-1}$ cm $^{-2}$). At $z \gtrsim 7$ these bright galaxies and $\text{Ly}\alpha$ -emitters are likely to trace overdensities and sit in ionized regions. These regions would be ideal for spectroscopic follow-up with JWST to spectroscopically confirm faint sources ($M_{\text{UV}} \gtrsim -18$) to map overdensities in 3D and obtain $\text{Ly}\alpha$ spectroscopy to infer the bubble sizes. A key prediction of reionization simulations is that reionization starts in overdensities (e.g., [Iliev et al. 2008](#); [Mesinger & Furlanetto 2007](#); [Trac & Gnedin 2011](#); [Ocvirk et al. 2016](#); [Hutter et al. 2021](#)), with galaxy properties playing a role in shaping the size distribution of ionized regions (e.g., [McQuinn et al. 2007b](#); [Seiler et al. 2019](#); [Lu et al. 2024](#)). Larger samples would enable the first systematic tests of bubble sizes as a function of overdensity and galaxy properties to enable an understanding of reionization on a local scale as we could directly compare bubble sizes to the estimated ionizing emission from the detected galaxies in the region. This could provide new insight in the ionizing photon escape fraction, star formation histories of the visible galaxies, and the contribution of galaxies below JWST’s detection limits.

5.2. Prospects for estimating the size distribution of ionized regions

With a sufficiently large survey it should be possible to infer the bubble size *distribution* as a function of redshift. This distribution is sensitive to the halo mass scale of the dominant ionizing sources with reionization driven by higher mass halos producing rarer, larger bubbles at fixed \bar{x} compared to low mass halos, (e.g., [McQuinn et al. 2007b](#); [Seiler et al. 2019](#); [Lu et al. 2024](#)) and thus offers important insights in the drives of reionization.

Since mapping bubbles along sightlines does not require wide continuous survey areas, we could efficiently build up a ‘bubble size distribution’ using multiple JWST fields (with spatially separated pointings to avoid having to model correlations). We note this will require appropriate definition of bubble sizes for comparison between models and observations as our line-of-sight method compresses 3D information effectively into 1D skewers where we trace the distribution of ionized ‘lengths’ in the IGM.

To explore how many fields are necessary to estimate the bubble size distribution we perform a Monte Carlo simulation using our IGM simulations. We take $10 \text{ cMpc} \times 10 \text{ cMpc} \times 300 \text{ cMpc}$ (corresponding to single JWST NIRSpec pointings spanning $\Delta z \approx 1$ at $z \sim 8$) from our simulation cubes, as described in Section 4.2. We use simulations spanning $\bar{x} = [0.4-0.9]$ to capture a range of bubble size distributions. We use our sightline bubble mapping technique to estimate the sizes of the ionized regions in each skewer.

In Figure 11 we plot the recovered mean bubble diameter as a function of the number of sightlines used, for simulations with different bubble size distributions, obtained from sampling 150 realizations of our sightlines. We see that we can distinguish differences in mean bubble diameter of $\gtrsim 10$ cMpc with just 15–30 independent sightlines. The prospects for building up samples of this number of sightlines is very feasible, for example, JWST pure-parallel NIRCcam imaging and slitless spectroscopy surveys are observing $\gtrsim 200$ pointings in Cycles 1–3 which would be ideal for spectroscopic follow-up (e.g., [Williams et al. 2021](#); [Morishita et al. 2023](#); [Egami et al. 2024](#)), and $\text{Ly}\alpha$ spectroscopy has already been obtained in four independent sightlines in Cycles 1–2 ([Tang et al. 2023, 2024a](#); [Jung et al. 2024](#); [Jones et al. 2024](#); [Napolitano et al. 2024](#)).

6. Conclusions

JWST provides our first opportunity to chart the reionization process in four dimensions. In this paper, we develop two simple methods to map ionized bubbles using ensembles of galaxies, allowing us to average over variation in intrinsic $\text{Ly}\alpha$ emission. Our main conclusions are as follows:

1. We develop an edge detection algorithm to map ionized bubbles in the plane of the sky, based on the asymmetry of $\text{Ly}\alpha$ transmission at each position on the sky. We test our method over an IGM neutral fraction range of $x_{\text{HI}} = 0.4-0.9$. Our method is independent of \bar{x} and can robustly map bubbles $\gtrsim 10$ cMpc.
2. Mapping bubbles in the plane of the sky requires a galaxy number density of $\gtrsim 0.002$ galaxies/cMpc 3 (corresponding to $M_{\text{UV}} \lesssim -18.5$ (-17.8) at $z \sim 7$ (8), e.g. [Bouwens et al. 2021](#)) to trace the density field (with the most precise recovery for $n_{\text{gal}} \gtrsim 0.004/\text{cMpc}^3$), survey areas greater than the typical sizes of ionized regions (~ 25 –3600 arcmin 2 for $\bar{x} \sim 0.9$ –0.4), and accurate galaxy redshift determination ($\Delta z \lesssim 0.015$). We require a 5σ $\text{Ly}\alpha$ equivalent width limit

of $\lesssim 30 \text{ \AA}$ for the faintest galaxies. These sensitivity requirements are not within reach of currently planned wide-area spectroscopic surveys (PFS, MOONS, Euclid-Deep, Roman HLS), however these surveys will find bright Ly α -emitters which likely signpost large ionized regions and are thus ideal for deeper follow-up with JWST. Covering the areas required for mapping bubbles will be possible in the early stages of reionization with wide-area JWST spectroscopy.

3. An efficient strategy to map bubbles at all stages of reionization with JWST is along the line-of-sight, which can be performed even in a single NIRSpect pointing, as bubble diameters are expected to span a redshift range $\Delta z \sim 0.05\text{--}0.6$ for $\bar{x} \sim 0.9\text{--}0.4$. We develop an algorithm which makes use of the nearly one-to-one relation between integrated Ly α transmission and distance from a galaxy to the first neutral IGM patch, the relation is calibrated using inhomogeneous large-scale IGM simulations (Lu et al. 2024).
4. The spectroscopic and imaging capabilities of JWST already provide the sensitivity to map bubbles along the line-of-sight. We find we require a number density of ≥ 0.004 galaxies/cMpc³ and EW limit of $\approx 30 \text{ \AA}$ to accurately recover bubbles ≥ 10 cMpc along the line-of-sight, which is attainable with deep spectroscopy of $M_{UV} \lesssim -18$ (-17.2) sources at $z \approx 7$ (8), and/or in overdensities of brighter sources. For individual galaxy detections, we require integration times of ~ 2 (40) hours to detect Ly α and ~ 6 (130) hours to detect [OIII] for $M_{UV} \sim -18.5$ (-17.2) galaxies at $z \sim 7$ (8). Given these requirements, mapping bubbles at $z \approx 7$ is feasible with current spectroscopic efforts, while for $z \approx 8$, targeting overdensity fields is necessary to reduce integration time constraints and improve survey completeness. Shallower observations should still provide robust lower limits on bubble sizes.

Early JWST studies have already demonstrated the potential of JWST spectroscopy to map the 3D distribution of galaxies in ionized regions (Chen et al. 2024; Witstok et al. 2024; Tang et al. 2024a). By constraining the sizes of these ionized regions based on estimates of galaxies' Ly α transmission we will be able to link galaxies directly to the growth of ionized regions, enabling a local understanding of reionization. Additional constraints on IGM transmission from Ly α escape fractions and line profiles will further improve bubble mapping methods, potentially reducing constraints on galaxy number densities (Nikolić et al. 2025). Accurate bubble maps will provide the opportunity for ionized photon 'accounting': enabling constraints on the escape fraction at $z > 6$ and the ionizing contribution of galaxies below JWST's detection limits. Future observations, sampling a range of environments, could start to estimate the size distribution of ionized regions and quantitatively establish how overdensity is linked to the growth of ionized regions.

Data availability

Our code to create ionized bubble maps in the plane of the sky and along sightlines given galaxy positions, M_{UV} , Ly α flux, and redshifts is publicly available here: https://github.com/ting-yi-lu/bubble_mapping_paper

Acknowledgements. The authors thank the anonymous referee for insightful comments. We thank Dan Stark for useful discussions. TYL, CAM and AH acknowledge support by the VILLUM FONDEN under grant 37459. CAM acknowledges support from the Carlsberg Foundation under grant CF22-1322. AM acknowledges support from the Italian Ministry of Universities and Research (MUR) through the PRIN project "Optimal inference from radio

images of the epoch of reionization", the PNRR project "Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing". The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant DNR140. This work has been performed using the Danish National Life Science Supercomputing Center, Computerome.

References

- Bhagwat, A., Napolitano, L., Pentericci, L., Ciardi, B., & Costa, T. 2024, ArXiv e-prints [arXiv:2408.16063]
- Blaizot, J., Garel, T., Verhamme, A., et al. 2023, *MNRAS*, 523, 3749
- Bolan, P., Lemaux, B. C., Mason, C., et al. 2022, *MNRAS*, 517, 3263
- Bolan, P., Bradáč, M., Lemaux, B. C., et al. 2024, *MNRAS*, 531, 2998
- Bolton, J. S., & Haehnelt, M. G. 2013, *MNRAS*, 429, 1695
- Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, *MNRAS*, 514, 55
- Bouwens, R. J., Oesch, P. A., Stefanon, M., et al. 2021, *AJ*, 162, 47
- Bowman, J. D., Cairns, I., Kaplan, D. L., et al. 2013, *PASA*, 30, e031
- Bunker, A. J., Cameron, A. J., Curtis-Lake, E., et al. 2024, *A&A*, 690, A288
- Carniani, S., Hainline, K., D'Eugenio, F., et al. 2024, *Nature*, 633, 318
- Castellano, M., Dayal, P., Pentericci, L., et al. 2016, *ApJ*, 818, L3
- Chen, Z., Stark, D. P., Mason, C., et al. 2024, *MNRAS*, 528, 7052
- Cirasuolo, M., Afonso, J., Bender, R., et al. 2012, *SPie Conf. Ser.*, 8446, 84460S
- Curtis-Lake, E., Carniani, S., Cameron, A., et al. 2023, *Nat. Astron.*, 7, 622
- Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018, *ApJ*, 864, 142
- Dijkstra, M., Mesinger, A., & Wyithe, J. S. B. 2011, *MNRAS*, 414, 2139
- Egami, E., Fan, X., Sun, F., et al. 2024, *SAPPHIRES: Slitless Areal Pure-Parallel High-Redshift Emission Survey, JWST Proposal. Cycle 3*, 6434
- Eilers, A.-C., Mackenzie, R., Pizzati, E., et al. 2024, *ApJ*, 974, 275
- Endsley, R., & Stark, D. P. 2022, *MNRAS*, 511, 6042
- Endsley, R., Stark, D. P., Whitler, L., et al. 2023, *MNRAS*, 524, 2312
- Endsley, R., Stark, D. P., Whitler, L., et al. 2024, *MNRAS*, 533, 1111
- Euclid Collaboration (Scaramella, R., et al.) 2022, *A&A*, 662, A112
- Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, *AJ*, 132, 117
- Finkelstein, S. L., Bagley, M. B., Arrabal Haro, P., et al. 2022, *ApJ*, 940, L55
- Gagnon-Hartman, S., Davies, J., & Mesinger, A. 2025, *A&A*, submitted [arXiv:2502.20447]
- Garel, T., Blaizot, J., Rosdahl, J., et al. 2021, *MNRAS*, 504, 1902
- Greene, J., Bezanson, R., Ouchi, M., Silverman, J., & the PFS Galaxy Evolution Working Group 2022, ArXiv e-prints [arXiv:2206.14908]
- Greig, B., Mesinger, A., Haiman, Z., & Simcoe, R. A. 2017, *MNRAS*, 466, 4239
- Gunn, J. E., & Peterson, B. A. 1965, *ApJ*, 142, 1633
- Haiman, Z. 2002, *ApJ*, 576, L1
- Hayes, M. J., & Scarlata, C. 2023, *ApJ*, 954, L14
- Hayes, M. J., Runholm, A., Scarlata, C., Gronke, M., & Rivera-Thorsen, T. E. 2023, *MNRAS*, 520, 5903
- Heintz, K. E., Brammer, G. B., Watson, D., et al. 2025, *A&A*, 693, A60
- Helton, J. M., Sun, F., Woodrum, C., et al. 2024, *ApJ*, 962, 124
- Hu, W., Wang, J., Zheng, Z.-Y., et al. 2019, *ApJ*, 886, 90
- Hutter, A., Trott, C. M., & Dayal, P. 2018, *MNRAS*, 479, L129
- Hutter, A., Dayal, P., Yepes, G., et al. 2021, *MNRAS*, 503, 3698
- Hutter, A., Heneka, C., Dayal, P., et al. 2023, *MNRAS*, 525, 1664
- Iliev, I. T., Shapiro, P. R., McDonald, P., Mellema, G., & Pen, U.-L. 2008, *MNRAS*, 391, 63
- Jaccard, P. 1912, *New Phytologist*, 11, 37
- Jakobsen, P., Ferruit, P., Alves de Oliveira, C., et al. 2022, *A&A*, 661, A80
- Jaskot, A. E., Dowd, T., Oey, M. S., Scarlata, C., & McKinney, J. 2019, *ApJ*, 885, 96
- Jones, G. C., Bunker, A. J., Saxena, A., et al. 2024, *A&A*, 683, A238
- Jones, G. C., Bunker, A. J., Saxena, A., et al. 2025, *MNRAS*, 536, 2355
- Jung, I., Finkelstein, S. L., Dickinson, M., et al. 2020, *ApJ*, 904, 144
- Jung, I., Finkelstein, S. L., Larson, R. L., et al. 2022, *ApJ*, submitted [arXiv:2212.09850]
- Jung, I., Finkelstein, S. L., Arrabal Haro, P., et al. 2024, *ApJ*, 967, 73
- Kakiichi, K., Hennawi, J. F., Ono, Y., et al. 2023, *MNRAS*, 523, 1772
- Kakiichi, K., Egami, E., Fan, X., et al. 2024, *COSMOS-3D: A Legacy Spectroscopic/Imaging Survey of the Early Universe, JWST Proposal. Cycle 3*, 5893
- Larson, R. L., Finkelstein, S. L., Hutchison, T. A., et al. 2022, *ApJ*, 930, 104
- Lin, X., Cai, Z., Wu, Y., et al. 2024, *ApJS*, 272, 33
- Lu, T.-Y., Mason, C. A., Hutter, A., et al. 2024, *MNRAS*, 528, 4872
- Maiolino, R., Cirasuolo, M., Afonso, J., et al. 2020, *The Messenger*, 180, 24
- Marchetti, L., Serjeant, S., & Vaccari, M. 2017, *MNRAS*, 470, 5007
- Mason, C. A., & Gronke, M. 2020, *MNRAS*, 499, 1395
- Mason, C. A., Trenti, M., & Treu, T. 2015, *ApJ*, 813, 21
- Mason, C. A., Treu, T., Dijkstra, M., et al. 2018a, *ApJ*, 856, 2
- Mason, C. A., Treu, T., de Barros, S., et al. 2018b, *ApJ*, 857, L11
- Mason, C. A., Trenti, M., & Treu, T. 2023, *MNRAS*, 521, 497

- McQuinn, M., Hernquist, L., Zaldarriaga, M., & Dutta, S. 2007a, *MNRAS*, **381**, 75
- McQuinn, M., Lidz, A., Zahn, O., et al. 2007b, *MNRAS*, **377**, 1043
- Mesinger, A., & Furlanetto, S. 2007, *ApJ*, **669**, 663
- Mesinger, A., & Furlanetto, S. R. 2008, *MNRAS*, **385**, 1348
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, **411**, 955
- Mesinger, A., Aykotalp, A., Vanzella, E., et al. 2015, *MNRAS*, **446**, 566
- Mesinger, A., Greig, B., & Sobacchi, E. 2016, *MNRAS*, **459**, 2342
- Mesinger, A., Ciardi, B., Davies, J. E., Gagnon-Hartman, S., & D'Odorico, V. 2024, *The Messenger*, **193**, 24
- Meyer, R. A., Oesch, P. A., Giovanazzo, E., et al. 2024, *MNRAS*, **535**, 1067
- Miralda-Escude, J. 1998, *ApJ*, **501**, 15
- Miralda-Escude, J., & Rees, M. J. 1998, *ApJ*, **497**, 21
- Miralda-Escudé, J., Haehnelt, M., & Rees, M. J. 2000, *ApJ*, **530**, 1
- Morishita, T., Mason, C., Trenti, M., et al. 2023, *A NIRCAM Pure-Parallel Imaging Survey of Galaxies Across the Universe, JWST Proposal. Cycle 2*, 3990
- Nakane, M., Ouchi, M., Nakajima, K., et al. 2024, *ApJ*, **967**, 28
- Napolitano, L., Pentericci, L., Santini, P., et al. 2024, *A&A*, **688**, A106
- Nasir, F., Cain, C., D'Aloisio, A., Gangolli, N., & McQuinn, M. 2021, *ApJ*, **923**, 161
- Neufeld, D. A. 1991, *ApJ*, **370**, L85
- Nikolić, I., Mesinger, A., Mason, C. A., et al. 2025, *A&A*, submitted [arXiv:2501.07980]
- Ning, Y., Cai, Z., Lin, X., et al. 2024, *ApJ*, **963**, L38
- Ocvirk, P., Gillet, N., Shapiro, P. R., et al. 2016, *MNRAS*, **463**, 1462
- Oesch, P. A., Dokkum, P. G. V., Illingworth, G. D., et al. 2015, *ApJ*, **804**, L30
- Oesch, P. A., Brammer, G., Naidu, R. P., et al. 2023, *MNRAS*, **525**, 2864
- Ouchi, M., Harikane, Y., Shibuya, T., et al. 2018, *PASJ*, **70**, S13
- Oyarzún, G. A., Blanc, G. A., González, V., Mateo, M., & Bailey, J. I. 2017, *ApJ*, **843**, 133
- Park, H., Jung, I., Song, H., et al. 2021, *ApJ*, **922**, 263
- Pentericci, L., Vanzella, E., Fontana, A., et al. 2014, *ApJ*, **793**, 113
- Prieto-Lyon, G., Mason, C., Mascia, S., et al. 2023, *ApJ*, **956**, 136
- Qin, Y., Mesinger, A., Bosman, S. E. I., & Viel, M. 2021, *MNRAS*, **506**, 2390
- Reddy, N. A., Steidel, C. C., Pettini, M., Bogosavljević, M., & Shapley, A. E. 2016, *ApJ*, **828**, 108
- Ren, K., Trenti, M., & Mutch, S. J. 2018, *ApJ*, **856**, 81
- Roberts-Borsani, G. W., Bouwens, R. J., Oesch, P. A., et al. 2016, *ApJ*, **823**, 143
- Santos, M. R. 2004, *MNRAS*, **349**, 1137
- Saxena, A., Robertson, B. E., Bunker, A. J., et al. 2023, *A&A*, **678**, A68
- Saxena, A., Bunker, A. J., Jones, G. C., et al. 2024, *A&A*, **684**, A84
- Schenker, M. A., Ellis, R. S., Konidaris, N. P., & Stark, D. P. 2014, *ApJ*, **795**, 20
- Seiler, J., Hutter, A., Sinha, M., & Croton, D. 2019, *MNRAS*, **487**, 5739
- Shapiro, P. R., & Giroux, M. L. 1987, *ApJ*, **321**, L107
- Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, *ApJ*, **588**, 65
- Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, *AJ*, **133**, 2222
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*, **323**, 1
- Smith, A., Bromm, V., & Loeb, A. 2017, *MNRAS*, **464**, 2963
- Sobacchi, E., & Mesinger, A. 2014, *MNRAS*, **440**, 1662
- Sobacchi, E., & Mesinger, A. 2015, *MNRAS*, **453**, 1843
- Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M., & Bunker, A. 2010, *MNRAS*, **408**, 1628
- Stark, D. P., Ellis, R. S., & Ouchi, M. 2011, *ApJ*, **728**, L2
- Stark, D. P., Ellis, R. S., Charlot, S., et al. 2017, *MNRAS*, **464**, 469
- Tang, M., Stark, D. P., Chen, Z., et al. 2023, *MNRAS*, **526**, 1657
- Tang, M., Stark, D. P., Topping, M. W., Mason, C., & Ellis, R. S. 2024a, *ApJ*, **975**, 208
- Tang, M., Stark, D. P., Ellis, R. S., et al. 2024b, *MNRAS*, **531**, 2701
- Tilvi, V., Malhotra, S., Rhoads, J. E., et al. 2020, *ApJ*, **891**, L10
- Topping, M. W., Stark, D. P., Endsley, R., et al. 2024, *MNRAS*, **529**, 4087
- Torralba-Torregrosa, A., Matthee, J., Naidu, R. P., et al. 2024, *A&A*, **689**, A44
- Trac, H. Y., & Gnedin, N. Y. 2011, *Adv. Sci. Lett.*, **4**, 228
- Treu, T., Trenti, M., Stiavelli, M., Auger, M. W., & Bradley, L. D. 2012, *ApJ*, **747**, 27
- Umeda, H., Ouchi, M., Nakajima, K., et al. 2024, *ApJ*, **971**, 124
- van Mierlo, S. E., Caputi, K. I., Ashby, M., et al. 2022, *A&A*, **666**, A200
- Vanzella, E., Pentericci, L., Fontana, A., et al. 2011, *ApJ*, **730**, L35
- Verhamme, A., Orlitová, I., Schaerer, D., & Hayes, M. 2015, *A&A*, **578**, A7
- Wang, Y., Zhai, Z., Alavi, A., et al. 2022, *ApJ*, **928**, 1
- Whitler, L. R., Mason, C. A., Ren, K., et al. 2020, *MNRAS*, **495**, 3602
- Whitler, L., Stark, D. P., Endsley, R., et al. 2024, *MNRAS*, **529**, 855
- Williams, C. C., Oesch, P., Barrufet, L., et al. 2021, *PANORAMIC – A Pure Parallel Wide Area Legacy Imaging Survey at 1–5 Micron, JWST Proposal. Cycle 1*, 2514
- Witstok, J., Smit, R., Saxena, A., et al. 2024, *A&A*, **682**, A40
- Wold, I. G. B., Malhotra, S., Rhoads, J., et al. 2022, *ApJ*, **927**, 36
- Yang, H., Malhotra, S., Rhoads, J. E., & Wang, J. 2017, *ApJ*, **847**, 38
- Yatawatta, S., de Bruyn, A. G., Brentjens, M. A., et al. 2013, *A&A*, **550**, A136
- Zackrisson, E., Majumdar, S., Mondal, R., et al. 2020, *MNRAS*, **493**, 855

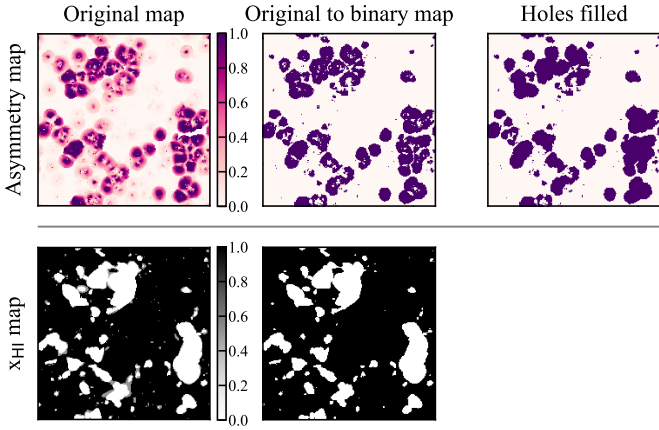


Fig. A.1. Example of converting an asymmetry score map into a binary ionization map (top panel), and converting an ionization map into a binary ionization map (bottom panel). The left column shows the original maps. The second column shows the maps converted into binary maps by setting all the pixels with asymmetry score (\bar{x}) above (below) a threshold ionized. The third column shows the binary map after we use `scipy.ndimage.binary_fill_holes`, to fill the holes, which correspond to areas inside bubbles, in the second column.

Appendix A: Optimizing bubble map recovery from asymmetry maps

Here we describe how we calibrate the parameters used for our asymmetry mapping technique (Section 3). We describe how we compare the recovered maps to the simulated ionization maps in Appendix A.1, our comparison metric in Appendix A.2 and the shape of the window function for creating asymmetry maps in Appendix A.3.

A.1. Creating binary \bar{x} and asymmetry maps

The asymmetry map corresponds to the positions of bubble edges, not the positions of ionized regions. To compare the recovered asymmetry maps to the original ionization maps, we need to fill the recovered bubble edges. Here we describe how we convert the asymmetry and the original ionization maps into binary maps, in which 0 (1) in the map means neutral (ionized).

We convert our simulated ionization maps into binary maps to compare with our recovered maps. We first collapse the 3D ionization cubes into 2D binary maps by taking the mean neutral fraction over the line-of-sight direction. We then convert the neutral fraction at each sky position into a binary value by setting the position to be ionized if its mean neutral fraction is less than 0.5. The lower panel in Figure A.1 shows the mean neutral fraction for the ionization slice (first column) and the binary ionization map.

To convert the asymmetry maps to binary maps we first set an asymmetry score threshold (A_{thresh}). Pixels with asymmetry score above (below) the threshold are marked as ionized (neutral). We find $A_{\text{thresh}} \approx 0.3 - 0.4$ provides the best map recovery (see below). In the upper panel of Figure A.1, we show an original asymmetry map in the first column and the binary map in the second column.

Secondly, since high asymmetry scores reflect bubble edges, we use the `scipy.ndimage` function, `binary_fill_holes`, to fill the closed outlines in the binary asymmetry map. The final binary asymmetry map is shown in the upper right corner in Figure A.1.

Figure A.2 shows the asymmetry maps made using different A_{thresh} for a window radius of 12 cMpc and depth of 4 cMpc (see Section 3.1 for a description of the window). As A_{thresh} increases, the recovered bubbles become more fractured. This is because the asymmetry scores around bubble edges are not constant. Only a few areas around bubble edges have very high asymmetry scores. If we only connect these high asymmetry score points, the recovered bubble becomes small and fractured. We find $A_{\text{thresh}} > 0.5$ will result in too many small bubbles.

We test for the best A_{thresh} using our $z = 8$ simulations, assuming we can detect all the galaxies with $M_{\text{UV}} \leq -17.8$ (corresponding to 0.002 galaxies/cMpc³) and a 5σ Ly α EW detection limit of 30 Å (an optimistic observational requirement, see Appendix B.1). We vary A_{thresh} between 0.15 and 0.4 and use a combination of visual and the Jaccard index (see Appendix A.2) to find the best $A_{\text{thresh}} \approx 0.3$. We find this $A_{\text{thresh}} \approx 0.3$ is independent of the IGM neutral fraction for $0.4 \leq \bar{x} \leq 0.9$, and does not depend strongly on the window size (Appendix A.3), making it a robust, model-independent, choice for making bubble maps.

A.2. Comparison metrics

To evaluate the recovery of the bubble maps relative to the input ionization map we use the Jaccard index (\mathcal{J} , Jaccard 1912). This is defined as the fraction of correctly recovered ionized and neutral pixels in the following map:

$$\mathcal{J} \equiv \frac{\text{number of correctly recovered neutral or ionized pixels}}{\text{number of pixels in the asymmetry map}}. \quad (\text{A.1})$$

The Jaccard index directly reflects how good our recovered map is compared to the input one. $\mathcal{J} = 0$ ($\mathcal{J} = 1$) means the recovered map is completely different from (the same as) the input map. By visual inspection of a range of maps at different \bar{x} from our simulations, we found $\mathcal{J} > 0.8$ indicates a good recovery of bubble maps.

In Figure A.3 we show the values of Jaccard index for different window radius and depth combinations at $\bar{x}=0.4-0.9$. The Jaccard index clearly traces the optimum parameters until $\bar{x} = 0.9$, when the mean bubble size becomes small and most of the field is neutral. Thus we use a combination of the Jaccard index with visual inspections of the maps to establish the best parameters for our method.

A.3. Optimal window shape

To determine the best window radius and depth combination for the asymmetry score (Equation 5), we calculate the Jaccard index for test window radii and depths for simulations at $\bar{x}=0.4-0.9$ (Figure A.3). We fix the $A_{\text{thresh}} = 0.4$. We find the best recovery is for window radius < 30 cMpc and window depth < 25 cMpc. For window depth < 30 cMpc the best window radius and depth are anticorrelated. This is because as the window depth increases, the true bubbles in the stacked ionization map become smaller, a small window radius is optimal to resolve small true bubbles. The window depth needs to be thin enough to prevent ionized bubbles from being washed out in the stacked ionization maps, but thick enough to include sufficient to sample the space. By visual inspection, we find the best parameters for our entire test \bar{x} range are [window radius, window depth]=[12, 4]cMpc.

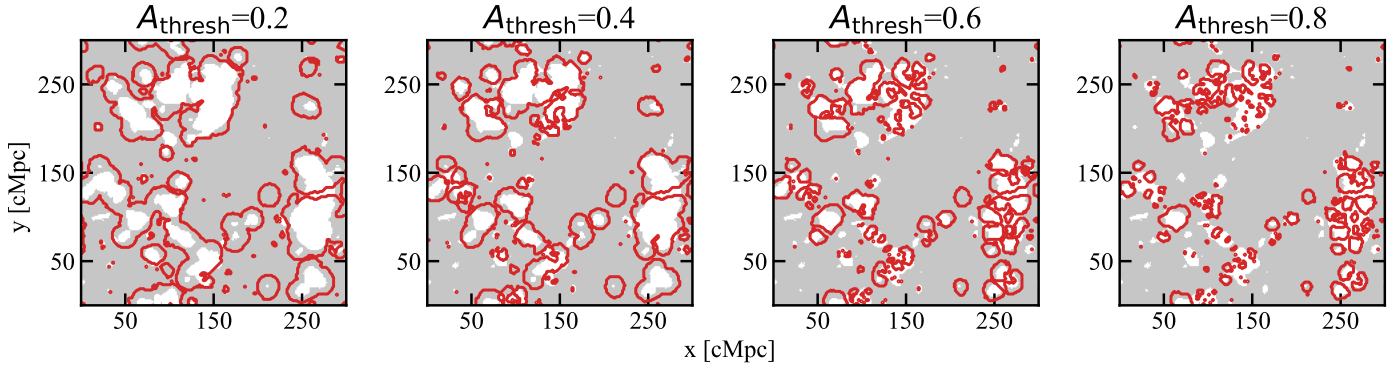


Fig. A.2. Recovered bubble maps using different A_{thresh} . From left to right column are the maps using A_{thresh} of 0.2, 0.4, 0.6, 0.8, respectively. When $A_{\text{thresh}} > 0.5$, the recovered bubbles break into too many small pieces.

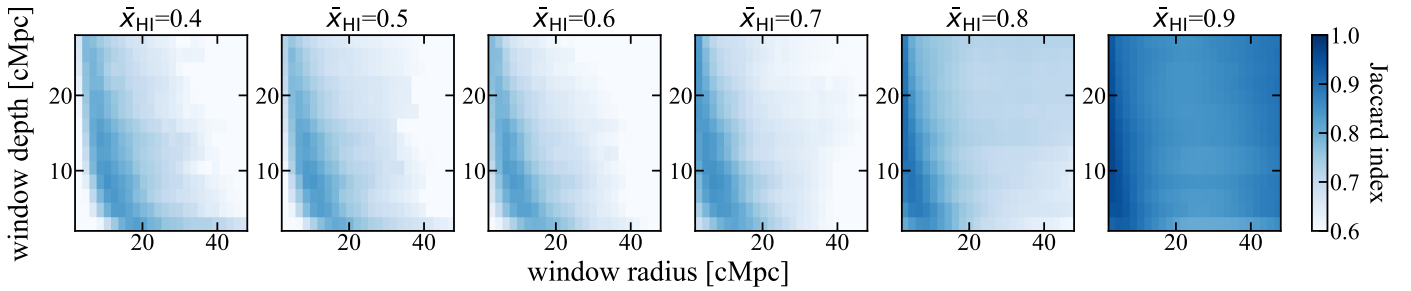


Fig. A.3. Jaccard index for different window radius and depth combinations as a function of \bar{x} .

Appendix B: Tests of observational requirements for asymmetry maps

Here we describe the number density of sources and Ly α EW limit required to map bubbles in the plane of the sky.

B.1. Required galaxy number density

To test the required number density of sources for creating robust asymmetry maps, we create a series of mock surveys at $z = 8$ varying the UV magnitude limit, $M_{\text{UV,limit}}$, of detectable sources, assuming a 5σ Ly α EW detection limit of 30 \AA for the faintest sources. We infer Ly α transmissions of individual galaxies using the Bayesian method as described in Section 2.2, using the maximum likelihood value of \mathcal{T} . We use a combination of visual inspection and the Jaccard index to judge how well the input bubble maps are recovered.

Figure B.1 shows the recovered bubble maps for $n_{\text{gal}} \approx 0.0040, 0.0025, 0.0015, 0.0009, \text{ and } 0.0005/\text{cMpc}^3$ at $\bar{x} = 0.7$. We find excellent recovery of bubbles for $n_{\text{gal}} \geq 0.002/\text{cMpc}^3$. Lower galaxy number densities do not sufficiently sample the space, resulting in recovered bubbles only centered around the strongest galaxy overdensities.

B.2. Required EW limit

To estimate the EW limit requirements for mapping bubbles using the asymmetry method, we create mock surveys at a range of \bar{x} now varying the Ly α EW limit, $EW_{\text{Ly}\alpha,\text{limit}}$ for the faintest observable galaxies (i.e. we vary the flux limit, so brighter galaxies will have lower EW limits), assuming $n_{\text{gal}} = 0.004/\text{cMpc}^3$ ($M_{\text{UV,limit}} = -17.2$ at $z = 8$). Figure B.2 shows example asymmetry maps at $\bar{x} = 0.7$ using $EW_{\text{Ly}\alpha,\text{limit}} = 10 - 100 \text{ \AA}$ for the faintest galaxies.

As expected, we can recover bubble edges very well for Ly α EW limits deeper than the median pre-IGM damping Ly α EW ($\approx 30 \text{ \AA}$, see Section 2.3). The shallower the Ly α EW limit is, the more likely we break big ionized bubbles into small bubbles around high EW sources, as we lose information about \mathcal{T} .

We test the bubble size recovery by randomly selecting positions where both the recovered bubble map and the input bubble map are ionized, measuring the sizes of bubbles around the points using the mean free path method (Mesinger & Furlanetto 2007), and comparing the recovered bubble sizes to the simulated bubble sizes. We find our asymmetry method can recover bubble sizes to within 30% of the true size when $n_{\text{gal}} \geq 0.004$ for 5σ Ly α EW detection limit $\leq 30 \text{ \AA}$ for $10 \text{ cMpc} \leq R \leq 100 \text{ cMpc}$ bubbles (Figure B.3). We slightly overestimate $R \leq 10 \text{ cMpc}$ bubbles because the spatial resolution is restricted by our window radius. We slightly underestimate $R \geq 60 \text{ cMpc}$ bubbles as very large bubbles can be split by our algorithm - further refinement of A_{thresh} could improve the segmentation of bubbles.

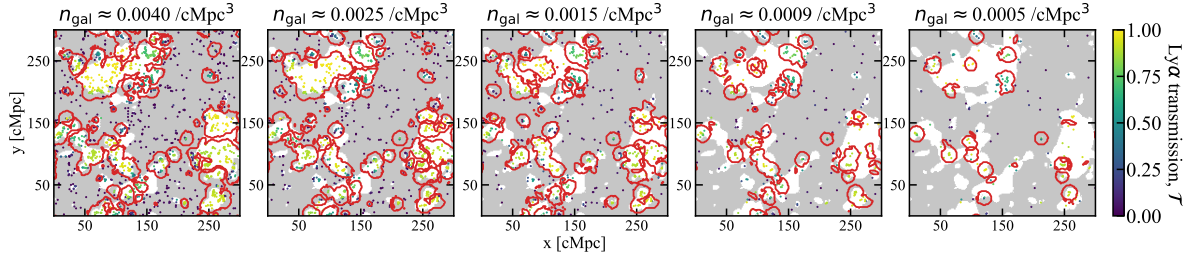


Fig. B.1. Recovered bubble maps corresponding to different galaxy number densities $n_{\text{gal}} \approx 0.0040 - 0.0005/\text{cMpc}^3$ at $\bar{x} = 0.7$. To achieve these number densities we vary $M_{\text{UV,limit}}$ in our $z = 8$ simulations, and assume flux limit corresponding to a 5σ EW limit of 30 \AA for the faintest galaxies. Dots show the detectable galaxies, color-coded by their true $\text{Ly}\alpha$ transmissions. As number density of galaxies increases, bubble recovery improves. With $n_{\text{gal}} \gtrsim 0.002/\text{cMpc}^3$ we can recover bubbles well: lower n_{gal} is not sufficient to sample the space.

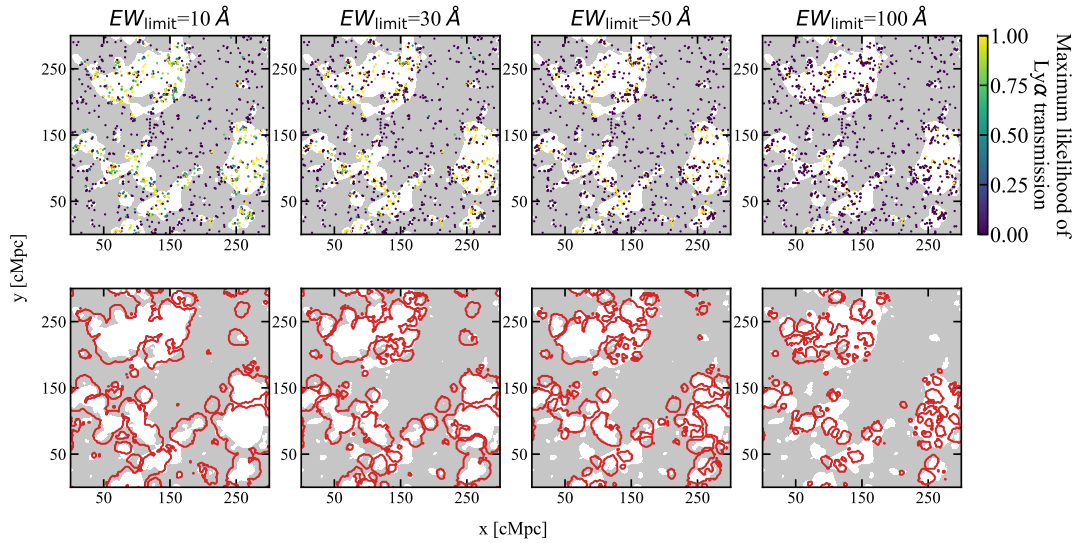


Fig. B.2. *Upper panel:* Input ionization map (gray neutral, white ionized) at $\bar{x} = 0.7$ and $M_{\text{UV}} \leq -17.2$ galaxies (dots) used to calculate asymmetry maps at $z = 8$. Galaxies are color-coded by their peak $\text{Ly}\alpha$ transmission (\mathcal{T}) inferred from the observed EW and uncertainty of EW. From left to right we show results using 5σ EW limit, $EW_{\text{Ly}\alpha,\text{limit}} = 10, 30, 50,$ and 100 \AA for the faintest galaxies. *Lower panel:* Asymmetry maps (red contours) made using the galaxies in the upper panel. We can recover bubble edges very well when $EW_{\text{Ly}\alpha,\text{limit}}$ is at least the median emergent EW of galaxies ($\approx 30 \text{ \AA}$). When $EW_{\text{Ly}\alpha,\text{limit}}$ is too shallow ($\gg 30 \text{ \AA}$), we cannot get useful information about $\text{Ly}\alpha$ transmission from our observations, thus the recovered bubbles only trace the regions around galaxies with EW greater than the detection limit.

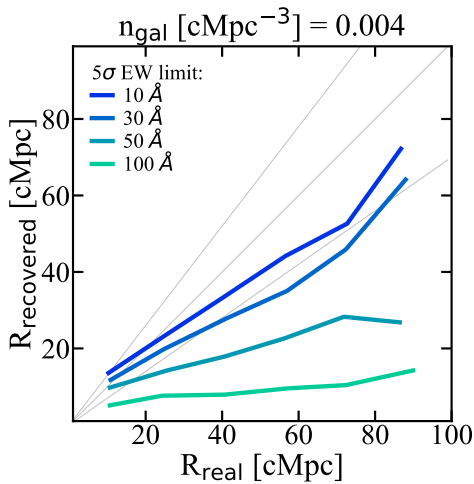


Fig. B.3. Median recovered bubble size versus real bubble size for 5σ $\text{Ly}\alpha$ equivalent width limit = $[10, 30, 50, 100] \text{ \AA}$, galaxy number density $\approx 0.004/\text{cMpc}^3$ in our asymmetry method. Gray lines from top to bottom show $+30\%$, 0% , and -30% recover bubble size error, respectively. We can recover bubble size to within 30% error when galaxy number density is greater than $0.004/\text{cMpc}^3$.