SCUOLA
NORMALE
SUPERIORE

Classe di Scienze

Corso di perfezionamento in
Fisica

XXXVI ciclo

Statistical Signatures of Cosmic Dawn Galaxies

Settore Scientifico Disciplinare FIS/05

Candidato
dr. Ivan NIKOLIĆ

Relatore

Prof. Andrei Albert MESINGER

SCUOLA
NORMALE
SUPERIORE

Anno accademico 2024/2025

Contents

Abstract

Radiation from the first stars, black holes and galaxies heated and ionized the pervasive neutral hydrogen in the intergalactic medium (IGM), culminating in the final phase change of our Universe: the Epoch of Reionization (EoR). This important epoch is the observational frontier, with current telescopes providing limited data spanning the first billion years of our Universe. However, the EoR is an inhomogeneous and complicated process with current telescopes being able to only detect the very brightest of EoR galaxies. As a result, there are many open questions, including: *When did the EoR happen? What are the sources driving it? What were the properties of those sources?*

This thesis focuses on the inference techniques and modeling improvements that are key to answering these questions. Several large-scale probes can be used for EoR inference, including the Lyman-α forest in high-redshift QSO observations, CMB optical depth, galaxy observations including the galaxy luminosity function, Lyman-α observations, and the patchy kinetic Sunyaev-Zel'dovich effect (a secondary anisotropy of the CMB sourced by free electrons and the velocity field during EoR). We combine these probes using Bayesian inference and semi-empirical scaling relations to infer a late-ending EoR. However, we demonstrate that some results can also depend on the galaxy-to-galaxy scatter around galaxy scaling relations. We investigate how such stochasticity in astrophysical relations impacts various quantities that are key for EoR science. We find that the scatter in the star-formation relation is key for modeling emissivities, UV luminosity functions, and the global EoR history. This stochasticity can also be learned from data, and we present methods to do so. Answering questions about what sources drove the EoR is intrinsically linked to the patchy nature of the process, as reflected in the morphology of the ionization field. We introduce a method to do this through the damping wing absorption of neutral hydrogen in the observed Lyman alpha emission lines of galaxies, showing that it will allow us to infer ionized bubbles surrounding groups of JWST-observed galaxies. By combining inferences across global and local scales, current and future telescopes are poised to uncover the details of the EoR and the first galaxies that drove it.

1

Introduction

1.1 Cosmology Basics and Galaxy Formation

Our understanding of the Universe has been continually evolving over the past 100 years. Discoveries at the start of the twentieth century put galaxies at the forefront of astrophysical research, setting up the Universe beyond our Milky Way (Lemaître, 1927; Hubble, 1929). The discovery that galaxies recede from us put astrophysical objects in the context of evolving spacetime. Together with observations, the previous century started with the emergence of a theory able to explain the evolving Universe, the theory of General Relativity (GR) (Einstein (1915), see also Wald (1984)). Our understanding of the Universe – from the initial conditions to the galactic web we see now – still relies on the Einsteins tensor equation. We start here by an introduction to general cosmology, following it up with the theory of structure formation needed to understand the formation and evolution of galaxies.

1.1.1 *Friedman Equations*

General relativity describes how matter and spacetime influence each other and co-evolve. This is described by the Einstein equation:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{1.1}$$

where $g_{\mu\nu}$ is the metric of the spacetime, $R_{\mu\nu}$ is the Ricci tensor and R is the curvature scalar representing the curvature of the spacetime, Λ is the cosmological constant, and $T_{\mu\nu}$ is the energy-momentum tensor. Solving the full second-order differential tensor equation is difficult, but we can gain insight using an approximation of the homogeneous Universe which is valid for large scales. In this approximation the energy-momentum tensor becomes diagonal and constant over space coordinates. From there we can describe the evolution of the Universe using the scale factor $a(t)$ to describe how spatial scales evolve with time. That is, we can write the spacetime metric as:

$$\mathrm{d}s^2 = -c^2 \mathrm{d}t^2 + a^2(t) \left[\frac{\mathrm{d}R^2}{1 - KR^2} + R^2 \mathrm{d}\Omega^2 \right]. \tag{1.2}$$

Using the homogeneous, isotropic approximation of the energy-momentum tensor, we can obtain the evolution of the scale factor using the Friedman equations (Friedmann, 1922):

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + 3\frac{P}{c^2} \right) + \frac{\Lambda c^2}{3} \tag{1.3}$$

and

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{K c^2}{a^2} + \frac{\Lambda c^2}{3}. \tag{1.4}$$

In the above equations, K is the curvature signature, 0 for a flat spacetime, -1 for negatively curved spacetime, and $+1$ for positively curved spacetime; ρ and P are the mass and pressure densities respectively. Knowing the composition of the Universe in the form of the equation of state ($\omega = \frac{P}{\rho}$) one gets the evolution of density and temperature with respect to time. We can get some insight by evaluating the evolution for different constituents of the Universe. First, in the case where the equation of state of a particular constituent is time-independent, its density evolves as:

$$\rho \propto a^{-3(1+\omega)}. \tag{1.5}$$

For radiation, $\omega = 1/3$, so we get $\rho \propto a^{-4}$ (corresponding to a dilution as the space expands and increase of wavelength due to redshifting), while for regular matter $\omega \approx 0$ so $\rho \propto a^{-3}$, i.e. density drops as the volume increases. Dark energy has a negative equation of state ($\omega = -1$), so its density doesn't evolve with time, i.e. $\rho =$const. Of course, our Universe is composed of all three things mentioned above, so we can find the generalized equation of the scale factor. First, we rewrite the Eq. 1.4 defining the Hubble parameter:

$$H = \frac{\dot{a}}{a} = H_0 \cdot E(z) \tag{1.6}$$

where z is the redshift ,$1 + z \equiv \frac{1}{a}$. Redshift is the standard quantity that tracks the evolution of the Universe in the cosmological literature, and $E(z)$ is

$$E(z) = \left[\Omega_{\Lambda,0} + (1 - \Omega_0)(1+z)^2 + \Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\mathrm{r},0}(1+z)^4\right]^{1/2}. \tag{1.7}$$

Here, $\Omega_i = \rho_i/\rho_{\mathrm{crit}} = \frac{8\pi\rho_i G}{3H^2}$ is the density parameter for each component (ratio of the density of component i to the critical density of the Universe: $\rho_{\mathrm{crit}} = \frac{3H^2}{8\pi G}$), while Ω_0 is the total density in units of critical density: $\Omega_0 = \Omega_{\mathrm{m},0} + \Omega_{\Lambda,0} + \Omega_{\mathrm{r},0}$. Equation 1.7 shows that the evolution of the scale factor can be directly deduced by knowing the constituents of the Universe at the present time. Quantifying the energy budget of each of the components has been one of the goals of many cosmological observations in the past few decades. Observations of Type Ia supernovae in distant galaxies showed that the Universe is accelerating, necessitating a non-zero dark energy component (Riess et al., 1998; Perlmutter et al., 1999). Observations of the cosmic microwave background allowed us a glimpse of the Universe when photons and baryons decoupled, further helping us constrain the ingredients of the Universe. Experiments targeting the CMB, such as WMAP (D. N. Spergel et al., 2007) and later Planck (Planck Collaboration, Aghanim, Akrami, et al., 2020), helped establish the concordance cosmology, which we call ΛCDM cosmology. These two datasets jointly point to a Universe that is dominated by dark energy. Even more recently, DESI observations gave indications that dark energy has an evolving equation of state (so-called $\omega_0\omega_a$CDM) (Adame et al., 2025). We are still lacking the theoretical framework for describing this major constituent of the Universe, and more observations at different redshifts will help us improve our understanding of it.

Observations also point to the fact that matter is dominated by elusive dark matter, whose particle nature is still unknown. Its presence was known already in the twentieth century from the observations of the rotation curves of galaxies (Rubin, Ford, and Thonnard, 1980). CMB observations (D. N. Spergel et al., 2007; Planck Collaboration, Aghanim, Akrami, et al., 2020, which also constrain $\Omega_{\mathrm{b},0}$, i.e., the baryon fraction), together with galaxy observations (Abbott et al., 2022; Adame et al., 2025), point to $\Omega_{\mathrm{m},0} \approx 0.3$, roughly six times more abundant than baryons. Additionally, the probes mentioned above indicate that the current density of the Universe is very close to $\Omega_0 \approx 1$, indicating a flat Universe. Radiation density at the present time is negligible, obtained from the CMB photon density and integrated starlight. Throughout this thesis, we will make use of the concordance cosmology parameters obtained by (Planck Collaboration, Aghanim, Akrami, et al., 2020): ($\Omega_{m,0}$, $\Omega_{b,0}$, $\Omega_{\Lambda,0}$, h, σ_8, $n_s = 0.321, 0.049, 0.679, 0.67, 0.81, 0.963$). h is defined as $h = \frac{H_0}{100\ \mathrm{km/s/Mpc}}$, and σ_8 and n_s will be defined later.

1.1.2 *Structure Formation* Despite the uncertainties in the parameters and the unknown physics of some of the components, ΛCDM is successful in explaining the growth of structures from the "initial conditions" to the present day. These initial conditions represent the Universe when it was homogeneous on large scales ($z \gtrsim 100 - 1000$), with fluctuations on the order of 10^{-6} (as we observe with the CMB; see (D. N. Spergel et al., 2007; Komatsu et al., 2011; Planck Collaboration, Aghanim, Akrami, et al., 2020)). These initial fluctuations eventually become the structures we observe. First, we have to see how these fluctuations grow with time. We write the fluctuations with respect to the average density: $\rho = \overline{\rho}(1 + \delta)$. Knowing the evolution of matter with the scale factor, how the scale factor changes with time, and how matter evolves under the influence of gravity, we can write the fluid equation as:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - 4\pi G\overline{\rho}_0 a^{-3}\delta = 0, \tag{1.8}$$

where δ is the perturbation in the density. In the matter-dominated epoch, which is a valid approximation after recombination, the equation above has two solutions:

$$\delta = \frac{3}{5}t^{2/3} + \frac{2}{5}t^{-1}, \tag{1.9}$$

The first part represents the growing mode, which scales as $\delta \propto a$, and it dominates structure formation over the decaying mode.

This shows that perturbations grow linearly with the scale factor. The next question is: how do these fluctuations become structures? That is, how do these fluctuations collapse into gravitationally bound structures, i.e., halos? Under the assumption of spherical collapse, it can be shown that a sphere whose initial density perturbation is δ_i will collapse if it satisfies:

$$(1 + \delta_i) > \Omega_i^{-1} \tag{1.10}$$

where $\Omega_{m,i}$ is the density parameter. In our Universe (see previous subsection), the condition in Eq. 1.10 will be satisfied for all overdensities in the limit of $t \to \infty$ (unless dark energy is dynamically evolving). Again, under the assumption of energy conservation, such a density will collapse at redshift $z_{\rm coll}$:

$$(1 + z_{\rm coll}) = 0.356\delta_i(1 + z_i), \tag{1.11}$$

where z_i is the initial redshift at which the overdensity is evaluated. Taking into account Eq. 1.9, we can write:

$$(1 + z_{\rm coll}) = \frac{\delta_0}{1.686}, \tag{1.12}$$

where δ_0 is the overdensity linearly extrapolated to $z = 0$. Eq. 1.12 has important implications for structure formation, indicating that overdensities collapse once their linearly extrapolated value reaches 1.686 (see Eke, Shaun Cole, and Frenk (1996) for modifications for different values of Ω_Λ). The calculation above assumes a single perturbation in the Universe evolving independently. Calculating the evolution of perturbations throughout the Universe is more complex. We can start by writing the joint probability distribution function of perturbations, $\mathrm{P}[\delta(\vec{x}), t]$ (following, e.g., Padmanabhan (1993)). In the very early Universe, we can assume that perturbations evolve independently in Fourier space:

$$P[\{\delta(\vec{k})\}, t] = \prod g_k[\delta(k), t], \tag{1.13}$$

where g_k is the probability distribution of a single mode in Fourier space having the amplitude and phase: $\delta_k = r_k e^{i\phi_k}$. Given the approximation of an uncorrelated Gaussian random field, g_k is defined by its mean and variance. By construction, the mean is 0, while $\langle|\delta_k|^2\rangle = \sigma_k^2$. In real space, this probability density for some density contrast q takes the shape:

$$P[q] = \frac{1}{\sqrt{2\pi\Delta_x^2}} e^{-\frac{q^2}{2\Delta_x^2}}, \tag{1.14}$$

where the real-space variance can be written, under the assumption of a Gaussian field, as an integral over k-space:

$$\Delta_x^2 = \int_0^\infty \Delta_k^2 \frac{\mathrm{d}k}{k}, \tag{1.15}$$

where

$$\Delta_k^2 \equiv \frac{1}{2\pi^2 V} k^3 \sigma_k^2. \tag{1.16}$$

Equation 1.16 defines the normalized power spectrum in the space of overdensities. We can also perform the integrals in the more physical and observable mass space. Since the mass enclosed in a region is an integral of the overdensity over some region:

$$\delta M(\vec{x}) = \bar{\rho}\int_V \delta(\vec{r} + \vec{x})\mathrm{d}\vec{r} \to \bar{\rho}\int_{\text{all space}} W(\vec{r})\delta(\vec{r} + \vec{x})\mathrm{d}\vec{r}, \tag{1.17}$$

it also follows the statistics of a Gaussian random field. On the right-hand side, we generalized the equation for an arbitrary window function $W(\vec{r})$ that defines the mass. From here, we can write the mass variance:

$$\sigma_M^2(R) \equiv \langle\delta_M^2\rangle = \frac{1}{V_W^2}\int_0^\infty \Delta_k^2 W_k^2 \frac{\mathrm{d}k}{k}, \tag{1.18}$$

where W_k is the Fourier transform of the window function defining the limits of integration for the given enclosed mass, and V_W is the volume defined by such window functions. The choice of the window function depends on the use case, but here we will stick with the spherical top-hat filter, as implied by the left-hand side of Eq. 1.17. In order to compute the distribution of halos in the Universe, we need to specify the matter power spectrum, which is usually expressed as:

$$\sigma_k^2 = A k^{n_s} T^2(k) D^2(z,k). \tag{1.19}$$

In Eq.1.19, A is the amplitude of the fluctuation, usually normalized at $8, h^{-1}$Mpc and computed with the spherical top-hat filter. This is an important cosmological quantity, constrained by the CMB (Planck Collaboration, P. A. R. Ade, et al., 2016; Planck Collaboration, Aghanim, Akrami, et al., 2020), galaxy clustering (Philcox and Ivanov, 2022), and galaxy lensing (Abbott et al., 2022). Furthermore, we have the initial power spectrum set by inflation physics (the k^{n_s} term) (Guth, 1981), whose power-law index is constrained by CMB observations (Planck Collaboration, Aghanim, Akrami, et al., 2020). $T(k)$ is the transfer function that alters the initial power spectrum according to the details of baryon physics (see, for example, (Seljak and Zaldarriaga, 1996)), and finally, we have the growth factor $D(z,k)$ that describes the growth of linear perturbations. From here, we can proceed by defining the collapsed fraction as the fraction of halos that have collapsed into structures greater than a given mass at a given scale. Following the definition of collapsed structures from Eq.1.12, we have:

$$f_{\rm coll}(> M_h, z) = \int_{\delta_c}^{\infty} P(\delta_M, R, z)\, d\delta_M = \frac{1}{\sqrt{2\pi}\sigma_M(R,z)} \int_{\delta_c=1.686}^{\infty} e^{-\frac{\delta_M^2}{2\sigma_M^2(R,z)}}\, d\delta_M\,. \tag{1.20}$$

where the scale R is set by the mass, and the mass variance is further defined by the matter power spectrum (i.e., Eq.1.18). However, the equation above doesn't account for halos whose mass might not be enough on the scale k to collapse into a structure, but that are part of a larger halo that is able to collapse. From the perspective of a random walk, this fact can be easily taken into account by realizing that, for a Gaussian random field, the probability of such halos at a given scale is equal to the right-hand side of Eq.1.20 (the cloud-in-cloud problem; see Bond, S. Cole, et al. (1991)). Therefore, the actual collapsed fraction is:

$$\begin{aligned}
f_{\rm coll}(> M_h, z) &= \int_{\delta_c}^{\infty} P(\delta_M, R, z)\, d\delta_M + \int_{-\infty}^{\delta_c} C(\delta_M, \delta_c, R, z)\, d\delta_M && (1.21)\\
&= 2 \times \int_{\delta_c}^{\infty} P(\delta_M, R, z)\, d\delta_M && (1.22)\\
&= 2 \times \frac{1}{2}\,\mathrm{erfc}\left[\frac{\delta_c(z)}{\sqrt{2}\sigma_M(R)}\right] && (1.23)\\
&= \mathrm{erfc}\left[\frac{\delta_c(z)}{\sqrt{2}\sigma_M(R)}\right]. && (1.24)
\end{aligned}$$

where C corresponds to the probability of a halo being embedded in a larger halo, but which, at that scale, doesn't collapse by itself. After differentiating, one gets:

$$\frac{df_{\rm coll}\,(> M, z)}{dM} = \sqrt{\frac{2}{\pi}}\frac{\delta_c(z)}{\sigma_M^2(M)}\left|\frac{d\sigma_M(M)}{dM}\right| \exp\left[-\frac{\delta_c^2(z)}{2\sigma_M^2(M)}\right] \tag{1.25}$$

From this, we get the usual shape of the halo mass function (HMF Press and Schechter, 1974):

$$\frac{dn(> M)}{dM} = \frac{\bar{\rho}_0}{M}\frac{df_{\rm coll}(> M, z)}{dM} \tag{1.26}$$

In a similar manner, we can construct conditional mass functions, i.e., the mass function of regions whose linear overdensity is $\delta_{\rm bias}$ and mass scale is $M_{\rm bias}$ (Bond, S. Cole, et al., 1991; C. Lacey and Shaun Cole, 1993; Somerville and Kolatt, 1999). The formalism outlined above is an elegant way to describe the growth of structures, and it's called the excursion set formalism, while the halo mass function obtained using this method is the Press–Schechter halo mass function (Press and Schechter, 1974).

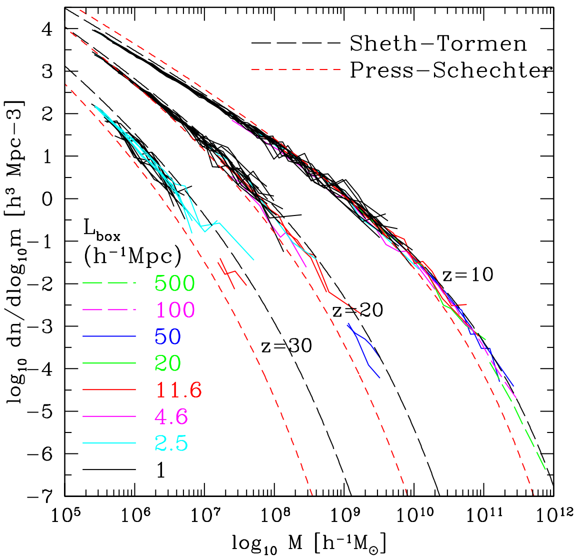

Figure 1.1: Comparison of different halo mass functions. Colored solid lines show results for N-body simulations with different sizes from Reed et al. (2007), corrected for cosmic-variance and finite-size. This is compared to two popular halo mass functions, Press-Schechter (red dashed line, Press and Schechter, 1974) and Sheth-Tormen (black long-dashed line, Sheth, Mo, and Tormen, 2001). Taken from (Reed et al., 2007).

This formalism works very well when compared to N-body simulations, though it underpredicts the abundance of massive halos while overpredicting the abundance of small ones. Instead, one can calculate the halo mass function directly from N-body simulations, assuming some form for the mass function. One of the popular forms is the one by Sheth and Tormen (Sheth and Tormen, 1999; Sheth, Mo, and Tormen, 2001):

$$\frac{dn(>M,z)}{dM} = -\frac{\bar{\rho}_0}{M}\frac{\partial \ln \sigma_M}{\partial M}\sqrt{\frac{2}{\pi}}A\left(1+\frac{1}{\hat{\nu}^{2p}}\right)\hat{\nu}\exp\left[-\frac{\hat{\nu}^2}{2}\right], \tag{1.27}$$

where $\hat{\nu} \equiv \sqrt{a}\delta_c(z)/\sigma_M$. This shape is motivated by the random walk, where the barrier is replaced by the prediction from the ellipsoidal collapse model:

$$\delta_c(M,z) = \sqrt{a}\delta_c(z)\left[1+b\left(\frac{\sigma_M^2(M)}{a\delta_c^2(z)}\right)^c\right], \tag{1.28}$$

(see also Bond and Myers (1996)). By fitting to numerical simulations, we obtain the Sheth–Tormen mass function, where the free parameters are given by $A = 0.353$, $p = 0.175$, $a = 0.707$, $b = 0.34$, and $c = 0.81$ (see also Jenkins et al., 2001). There are other fits in the literature, tailored to specific numerical simulations (Reed et al., 2007; Watson et al., 2013), mostly based on numerical data. We show in Fig. 1.1 a result from Reed et al. (2007) for different sizes of numerical simulations, compared to the Sheth–Tormen and Press–Schechter mass functions at different redshifts. Numerical simulations are more consistent with each other, but there is significant scatter on a simulation-to-simulation basis (see Chapter 4).

There are a number of modifications tailored to specific simulations, though a correct halo mass function that works on all scales and redshifts is still an open question.

1.1.3 *Baryon Physics* The formalism outlined above serves as a background for galaxy evolution. In order to study how galaxies evolve, we need to understand the evolution of baryons, which give galaxies their observable properties.

Baryons follow a different evolution than dark matter. At redshifts $z > 1000$, protons and electrons cannot form neutral atoms due to their coupling with photons. This also prohibits gravitational collapse, causing the baryons and photons to follow the perturbations set by inflation. At $z \approx 1100$, the photon–baryon medium becomes cold enough to form neutral hydrogen, leading to cosmic recombination. This is associated with the decoupling of photons from baryons, forming the cosmic microwave background we observe today (Dodelson, 2003). Recombination also allowed baryons to evolve independently under gravity. Since dark matter began to undergo gravitational collapse earlier, baryons fell into the gravitational wells of structures formed by dark matter. This process dominated the history of the Universe until

$z \sim 100$, when the first density perturbations reached non-linear scales. To model when that occurred, we return to Eq. 1.8.

The difference between baryons and matter can be modeled by introducing a pressure term ($\vec{F} = -a^{-1}\nabla p$). In the simplest case, such as an ideal fluid, this pressure term is proportional to the density. Adding this to Eq. 1.8, we get:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = 4\pi G\bar{\rho}_0 a^{-3}\delta_{\mathrm{DM}+b} + a^{-2}c_s^2\nabla^2\delta \tag{1.29}$$

In this equation, gravity is now written as a function of $\delta_{\mathrm{DM}+b}$, which is the overdensity of both dark matter and baryons, since gravity acts on both components jointly. The last term is the pressure term, with $c_s \equiv \sqrt{\frac{\partial p}{\partial \rho}}$ denoting the sound speed, which depends on the equation of state of the baryons. Analogously to the condition in Eq. 1.10, we can define the scales at which baryon perturbations are stable. Setting the time derivatives to zero and evaluating the equation, we obtain the following condition:

$$0 = \left[4\pi G\bar{\rho}_0 a^{-3} - a^{-2}c_s^2|\mathbf{k}|^2\right]\delta_k \tag{1.30}$$

which sets a characteristic scale for the collapse of baryonic structures at a given time – the Jeans scale:

$$\lambda_j = \sqrt{\frac{\pi c_s^2}{G\bar{\rho}_0 a^{-1}}} \tag{1.31}$$

that are able to collapse. Scales smaller than this are pressure-supported, while scales larger than that begin their collapse. Analogously, we can define the Jeans mass, describing the mass of structures capable of collapsing:

$$M_J \equiv \frac{4\pi}{3}\left(\frac{\lambda_j}{2}\right)^3\bar{\rho}_0 = \left(\frac{\pi^{5/2}}{6G^{3/2}}\bar{\rho}_0^{-1/2}\right)\quad\left(c_s^2 a\right)^{3/2}. \tag{1.32}$$

From here, we can trace the evolution of the Jeans mass throughout cosmic history. Before recombination, baryons behave as a relativistic fluid ($c_s^2 = c^2/3$), so at recombination ($z \sim 1100$), we get $M_J \sim 10^{18}M_\odot$. All smaller scales are pressure-supported. Again, it is important to note that dark matter had already begun its gravitational collapse. After cosmic recombination, baryons start to behave as an ideal fluid, though they remain thermally coupled to photons (i.e., the CMB). In that regime, $c_s^2 \propto T \propto a^{-1}$, so the Jeans mass becomes approximately constant, $M_J \sim 10^5 M_\odot$. These are the perturbations that first begin to collapse. After baryons decouple from photons ($z \approx 200$), they cool adiabatically ($T \propto a^{-2}$), so the Jeans mass evolves as

$$M_J \sim 7 \times 10^3\left(\frac{1+z}{10}\right)^{3/2} M_\odot. \tag{1.33}$$

The situation changes again during and after the Epoch of Reionization, as we will see later.

To describe the first structures, we need to understand their temperature evolution during collapse. Unlike dark matter, which virializes into structures, baryons can radiate their thermal energy, allowing them to collapse further and form the first stars. The evolution of the first stars and galaxies depends on the thermal evolution of gravitational structures. As mentioned before, dark matter forms halos hierarchically. Baryons fall into these structures and are gravitationally heated, since the temperature of the medium outside halos is lower than the virial temperature of the halos (R. Barkana and A. Loeb, 2001):

$$T_{\mathrm{vir}} \approx 10^4\ \mathrm{K}\left(\frac{\mu}{0.6}\right)\left(\frac{M_h}{10^8 M_\odot}\right)^{2/3}\left(\frac{1+z}{10}\right)\left[\frac{\Omega_{m,0}}{0.3}\frac{1}{\Omega_m(z)}\frac{\Delta_c}{18\pi^2}\right], \tag{1.34}$$

where Δ_c is the average density of the halo and μ is the molecular weight of the primordial gas. Again, due to the electromagnetic coupling between baryons and photons, baryonic gas can radiate away its energy. How efficiently it can do so depends on its atomic and molecular composition.

The most significant cooling mechanism in primordial gas comes from the cooling of atomic hydrogen through recombination and collisional de-excitation. The same cooling channel is also present for atomic helium. However, this cooling channel becomes inefficient below 10^4 K for hydrogen and 10^5 K for helium (see Fig. 1.2). Halos that cool efficiently through this mechanism have virial masses of $\sim 10^8, M_\odot$ at $z > 10$. However, the first stars formed in smaller

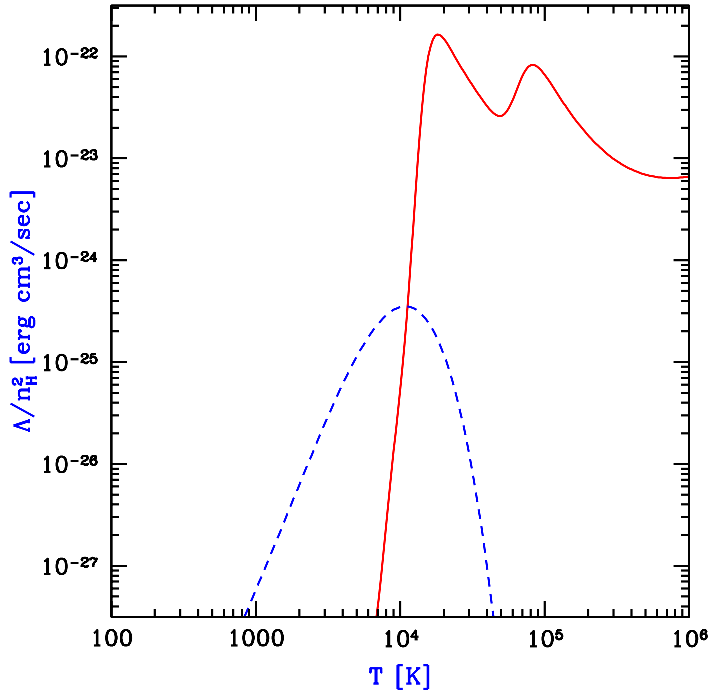

Figure 1.2: Cooling rate normalized by the square of the hydrogen density for primordial gas with $n_H = 0.045\text{cm}^{-3}$. The red line shows the atomic cooling (hydrogen and helium), while the blue line corresponds to additional molecular cooling. Adapted from R. Barkana and A. Loeb (2001).

halos due to the hierarchical growth of structure—provided there was a way to cool below 10^4 K. Given the dearth of metals, the first stars could only cool using molecular hydrogen, H_2 (or less abundant isotopic equivalents such as HD), allowing temperatures to drop to 10^3 K. After cosmic recombination, the abundance of H_2 was negligible, but potential formation channels opened via the creation of H^- ions within dense halos (Zoltan Haiman, Martin J. Rees, and Abraham Loeb, 1996), enabling the cooling of smaller halos. However, the H_2 molecule is easily dissociated by Lyman-Werner photons (photons in the energy range 11.26–13.6 eV). The processes governing the formation and destruction of these molecules on halo scales are complex. This complexity makes it difficult to determine from first principles when the first stars appeared (Tegmark et al., 1997; Bromm and R. B. Larson, 2004; Jemma Wolcott-Green, Zoltán Haiman, and Greg L. Bryan, 2021).

The details of the subsequent collapse of such objects depend heavily on their chemical composition, the surrounding radiation field, and the nature of momentum transport. Virialized gas begins to collapse isothermally, aided by molecular cooling. At this point, temperature and sound speed are constant, so $M_J \propto \rho^{-1/2}$. The Jeans mass decreases as the gas core contracts, leading to fragmentation into sub-cores, which then collapse further to form stars or black holes. However, if the collapse transitions to an adiabatic regime, the Jeans mass begins to increase. Thus, the moment when collapse changes from isothermal to adiabatic defines the mass scale of the fragments that become stars (see also reviews in Milosavljević and Safranek-Shrader (2016) and Klessen and Glover (2023)).

How massive are these fragments? In the local Universe, the initial mass function that describes stellar masses peaks at $\sim 0.1, M_\odot$ (Salpeter, 1955; Chabrier, 2003). However, the local Universe is metal-enriched, and metals provide many additional cooling channels, promoting fragmentation of molecular clouds. In contrast, cooling in the early Universe was inefficient. We expect the first stars (also known as Population III stars) to be massive, potentially reaching a few $\times 100, M_\odot$. More detailed numerical simulations that include additional physical processes—such as magnetic fields, shocks, and turbulence—predict lower masses for the first stars, pointing to a broad mass distribution (Nakamura and Umemura, 2001; Hirano et al., 2014). If the fragments are even more massive, they may collapse directly into black holes, potentially forming the seeds of the supermassive black holes we observe in galaxies today. Therefore, the epoch of first star formation—marking the Cosmic Dawn—had a profound impact on subsequent galaxy evolution.

1.2 Galaxy Evolution

The first stars in the Universe dramatically impacted cosmic evolution. Their radiation altered the IGM, while their metal production influenced future star formation and many galaxy properties. However, the most significant wave of galaxy formation began with the appearance of atomic cooling halos—halos with sufficient mass for atomic hydrogen to cool the gas efficiently. These first galaxies looked very different from those in the local Universe, so it is worthwhile to examine their properties in more detail.

1.2.1 *UV and X-ray Production* As discussed above, the first stars differed markedly from those in our local environment. Due to their high temperatures and low metal content, they burned hydrogen more quickly and at higher temperatures. This led to intense UV and X-ray emission.

First, the Lyman-Werner (LW) background built up by the first stars influenced subsequent generations of stars. This effect was especially significant in the vicinity of the first stars, impacting nearby halos as well (Dijkstra, Zoltán Haiman, et al., 2008). The LW background may have allowed pockets of pristine gas to remain unpolluted, enabling Population III star formation to occur even at later times. These events could be observable with JWST (Fujimoto, Naidu, et al., 2025).

Additionally, the first stars and galaxies are expected to have produced a large amount of X-rays, primarily through the formation of high-mass X-ray binaries (T. Fragos et al., 2013; Pacucci, Andrei Mesinger, et al., 2014). This had important implications for the thermal history of the IGM. Since X-rays have low cross-sections ($\sigma \propto E^{-\beta}$, where β ranges from -3 to -3.5, e.g., (Madau and Tassos Fragos, 2017)), they propagated through the Universe, heating it (and producing minor ionization). The total amount of heating remains uncertain. Upper limits from 21cm experiments suggest that heating from massive binaries was stronger at $z \gtrsim 15$ compared to lower redshifts (The HERA collaboration et al., 2023).

1.2.2 *Metal Enrichment and Abundances* The first supernovae enriched the ISM of the earliest halos, as well as the surrounding intergalactic medium (IGM). This metal enrichment from the first stars altered all subsequent star formation. Metal cooling enabled lower temperatures in the ISM, leading to what we refer to as Population II star formation. With the emergence of atomic cooling halos, these structures began to resemble the galaxies we observe in the local Universe. However, early galaxies still contained very low amounts of metals. Variations in metal abundances led to many distinct observable properties.

The amount of metals in galaxies is quantified by metallicity, which can be expressed in various ways. The presence of metals has a dramatic impact on the spectral energy distributions (SEDs) of galaxies. Metals produce emission lines in stellar spectra, as well as in the ISM via nebular emission. A strong nebular continuum is now routinely observed with JWST (G. Roberts-Borsani, Tommaso Treu, et al., 2024). JWST is already detecting non-negligible metal content in the most distant spectroscopically confirmed galaxies. For example, the most distant galaxy known at the time of writing, JADES-GS-z14-0 (Carniani, Hainline, et al., 2024), has also been detected by ALMA through the [O$_{\rm III}$] 88μm emission line (Carniani, D'Eugenio, et al., 2025). Meanwhile, GS-z12 has been detected via the C III]$\lambda\lambda 1907, 1909$ emission line (D'Eugenio et al., 2024). On the other hand, the remarkable capabilities of JWST allow us to detect extremely metal-poor galaxies late in the EoR (Cullen et al., 2025).

Local observations of metallicity show a tight relation between stellar mass, star formation rate, and metallicity (F. Mannucci et al., 2010; Curti, Filippo Mannucci, et al., 2020a). Whether this relation still holds at the highest redshifts remains an open question (Curti, Maiolino, et al., 2024; Nakajima et al., 2023; Chemerynska et al., 2024). One of JWST's key science goals is to trace the evolution of metals in the Universe out to the highest redshifts.

1.2.3 *Burstiness* One of the most important quantities describing galaxies is their star formation rate (SFR). Star formation results from a balance between the accretion and consumption of gas, and feedback from previous generations of stars. In the local Universe, this balance is achieved through the formation of molecular clouds, whose dynamics and collapse are regulated by the interplay between the supernova feedback timescale and the dynamical timescale of the gas accreting onto galaxies (see reviews in Kennicutt, 1998; R. B. Larson, 2003; McKee and Ostriker, 2007). The situation during the earliest epochs of galaxy evolution was markedly different. The Universe during Cosmic Dawn was denser than it is today. Moreover, timescales for star formation (and for feedback from supernovae) were shorter due to the younger stellar populations, while the free-fall time was longer ($t_{\rm ff} \propto (1+z)^{3/2}$) (Kimm, Cen, et al., 2015; Faucher-Giguère, 2018). In addition, galaxies were less massive, so feedback from star formation had a stronger impact than in the local Universe. Taken together, these effects suggest that star formation in the early Universe proceeded in a bursty manner: intense episodes of star formation were followed by suppression due to strong feedback (see also Steven R. Furlanetto and Mirocha, 2022). This burstiness was driven by two key processes.

First, the initial episodes of star formation consumed the virialized gas that had cooled. Further star formation then relied on the accretion of pristine gas from the IGM. At high redshift, the Universe was still neutral and dense, providing ample fuel for star formation. This led to increased gas accretion (Faucher-Giguère, Kereš, and C.-P. Ma, 2011). However, supernovae could halt both star formation and gas accretion in the surrounding regions. This feedback was especially effective in smaller halos. The combination of increased accretion and violent feedback led to stochastic star formation in high-redshift galaxies.

Second, early galaxies merged more frequently (as also implied by the excursion set formalism described above; Neistein, van den Bosch, and Dekel, 2006). Mergers had a significant impact on galaxy evolution. They supplied additional gas that boosted star formation but also generated shocks, making the process stochastic and difficult to model from first principles. Intense merger episodes redistributed metals within galaxies and changed their morphology (Duan et al., 2025). Mergers also influenced black hole growth (Volonteri et al., 2005) and the escape of photons (Kostyuk and Benedetta Ciardi, 2024; Yuan et al., 2024), although some results remain debated (see also Mascia, Pentericci, Llerena, et al., 2025).

This stochasticity in star formation had many consequences. First of all, bursts of star formation increased galaxies' UV luminosity, potentially outshining older stellar populations. These bursts also enhanced nebular emission. Variability in UV emission has important implications for the detectability of these galaxies with JWST and ALMA—both in photometry and spectroscopy—especially at very high redshifts ($z \gtrsim 10$). This stochasticity is one of the leading explanations for the unexpectedly large number of galaxies observed in UV photometry with JWST (C. A. Mason, Trenti, and Tommaso Treu, 2023; Mirocha and Steven R. Furlanetto, 2023a; Shen et al., 2023; A. Pallottini and A. Ferrara, 2023; Ciesla et al., 2024; Gelli, C. Mason, and Hayward, 2024, see more on this in subsequent chapters).

1.3 The Theory of the Epoch of Reionization

As discussed previously, the first galaxies produced large amounts of ionizing photons, powered by massive stars and intense episodes of star formation. These ionizing photons strongly impacted the ISM, ionizing regions that were not self-shielded and producing nebular emission (see above). As a result, large areas of these galaxies were left open for ionizing photons to escape into the surrounding medium. The abundance of these photons was sufficient to trigger a phase transition in the Universe: the Epoch of Reionization (EoR) (see reviews in Zaroubi, 2013; Andrei Mesinger, 2016).

Evidence for reionization began to accumulate toward the end of the 20th century. Spectra of very distant quasars showed transmission blueward of the rest-frame Lyman-α line, which would not be expected in a fully neutral Universe (James E. Gunn and Peterson, 1965). Observations from SDSS pushed the frontier of the ionized Universe beyond $z \gtrsim 5$. In recent decades, an increasing number of cosmological probes have targeted the EoR. Still, the exact timing and duration of this epoch remain uncertain. In this section, we outline the physical concepts required to understand the process.

The EoR was driven by the cumulative production and escape of ionizing photons into the IGM. The rate at which these photons escaped can be quantified by the ionizing emissivity, $\dot{N}_{\rm ion}$:

$$\dot{N}\mathrm{ion} \equiv \rho\mathrm{UV}, \xi_{\rm ion}, f_{\rm esc}, \tag{1.35}$$

where $\rho_{\rm UV}$ is the UV luminosity density, $\xi_{\rm ion}$ is the production efficiency of ionizing photons per unit UV luminosity, and $f_{\rm esc}$ is the escape fraction of these photons. This decomposition allows us to connect the relatively well-constrained $\rho_{\rm UV}$ to $\dot{N}$ion, via the two less constrained parameters, ξion and $f_{\rm esc}$.

The quantity $\xi_{\rm ion}$, the number of ionizing photons per unit UV luminosity density, is the first major uncertainty in modeling reionization. On a galaxy-by-galaxy basis, it is defined as $\xi_{\rm ion} = \frac{N(H^0)}{L_{\rm UV}}$, where $N(H^0)$ is the ionizing photon production rate, and $L_{\rm UV}$ is the UV luminosity, typically measured at 1500 Å. Canonically, $N(H^0)$ is linked to dust-corrected Hα luminosity via the relation: $L(\mathrm{H}\alpha), [\mathrm{erg}, \mathrm{s}^{-1}] = 1.36 \times 10^{-12}, N(H^0), [\mathrm{s}^{-1}]$ (Leitherer and Heckman, 1995). More accurate estimates can be obtained via spectral fitting, assuming an initial mass function (IMF), but this remains challenging. JWST observations are beginning to suggest higher $\xi_{\rm ion}$ values at higher redshifts (Marco Castellano et al., 2023; Llerena et al., 2024).

However, not all ionizing photons escape into the IGM. Photons first ionize the surrounding ISM and only escape along optically thin or low-density channels. Ionizing photons absorbed by the ISM are reprocessed into nebular lines and continuum emission. This makes modeling ionizing photon escape extremely complex, likely depending on star formation history, geometry, and other poorly constrained galactic properties (Paardekooper, Khochfar, and Dalla Vecchia, 2015; X. Ma, Kasen, et al., 2015). For these reasons, the escape fraction $f_{\rm esc}$ remains one of the greatest uncertainties in reionization modeling.

At lower redshifts, when the Universe is already ionized, ionizing photons can be observed directly (Flury et al., 2022). These observations show that escape fractions depend on stellar age, burstiness of star formation, stellar mass, metallicity, dust content, and more. Various diagnostics for estimating $f_{\rm esc}$ have been proposed (X. Xu et al., 2023; Jaskot et al., 2024; Roy et al., 2024), but extrapolating them to earlier epochs where direct measurements are impossible remains difficult.

Nevertheless, we know that ionizing photons do escape. These photons travel through the less dense IGM, ionizing neutral hydrogen along their path. The cumulative ionization creates distinct ionized regions, or bubbles, in the early stages of reionization. The first such bubbles likely appeared during the Cosmic Dawn, around the rare earliest galaxies. As more galaxies underwent episodes of star formation, the ionized bubbles grew and merged, eventually percolating through the Universe. Since galaxies are the main source of ionizing photons, it is expected that most regions of the IGM were reionized by local sources – an "inside-out" scenario of reionization.

Reionization of the Universe progressed by the accumulation of ionized photons, counteracted by the recombination of the ionized atoms beyond the ionization shocks. The number of recombinations affects the end stages of reionization. It can be written as follows:

$$\dot{N}_{\rm rec} = \alpha_{AB} C \langle n_{\rm H} \rangle^2, \tag{1.36}$$

where α_{AB} is the recombination coefficient (case A or B, depending on the ionized region's properties) and C is the clumping factor, $C = \frac{\langle n^2 \rangle}{\langle n \rangle^2}$. Equation 1.36 shows that recombination depends on how clustered the ionized gas is after ionization fronts pass. How much recombination occurs—and on which timescales and spatial scales—is a major uncertainty in modeling EoR (Emanuele Sobacchi and Andrei Mesinger, 2014; F. B. Davies, Sarah E. I. Bosman, and Steven R. Furlanetto, 2024). Current estimates of the clumping factor range from low values ~ 2 to high values ~ 12 (McQuinn, Oh, and Faucher-Giguère, 2011; Cain et al., 2023). Complicating matters further are measurements of short mean free paths of ionizing photons at $z \sim 6$ (G. D. Becker, D'Aloisio, et al., 2021; Zhu et al., 2023). Self-shielded absorption systems, or "photon sinks," can absorb ionizing photons, contributing to a "soft landing" of the neutral fraction. Combining Eq.1.36 with Eq.1.35, we can write the growth of the ionized fraction:

$$\frac{{\rm d}Q}{{\rm d}t} = \dot{N}_{\rm ion} - \dot{N}_{\rm rec} \tag{1.37}$$

Equation 1.37 describes the growth of individual ionized regions. Taking the ensemble average yields the evolution of the volume filling factor of ionized hydrogen:

$$\frac{dQ_{\rm HII}}{dt} = f_{\rm esc}\xi_{\rm ion}N_{\rm UV,b}\frac{df_{\rm coll}\ (> M_{\rm min}, z)}{dt} - \alpha_{\rm AB}\left\langle n_{\rm H}\right\rangle Ca^{-3}Q_{\rm HI}, \tag{1.38}$$

where $N_{\rm UV,b}$ is the number of ionizing photons produced per baryon (Steven R. Furlanetto, Zaldarriaga, and Hernquist, 2004) (see Steven R. Furlanetto, Zaldarriaga, and Hernquist (2004)). Eq. 1.38 encodes the complex nature of EoR. Starting from the smallest scales of star-formation in galaxies through the largest structures of ionized bubbles, solving Eq. 1.38 on spatial scales is challenging.

Despite a well-developed theoretical framework and growing observational constraints, many uncertainties remain. Before exploring them, it's important to note why this process is so important. As the last major phase transition in the Universe, EoR impacted most of the baryons and left long-lasting consequences. For instance, feedback from ionization fronts raised the Jeans mass (J. Miralda-Escudé and M. J. Rees, 1994; D'Aloisio et al., 2019a), suppressing star formation in low-mass halos and even affecting higher-mass systems. Moreover, the baryonic structure of the IGM – we observe for example in Lyman-α forest of distant quasars – is shaped by its temperature and ionization state. A thorough understanding of EoR is thus essential to interpret such observations and even explore physics beyond the standard model (Viel et al., 2013). Going back in the past, EoR is the most important signature of the first galaxies. Description of the IGM requires knowledge of all of the galaxies, many which are beyond the detection limit with current and even future telescopes. Turning the problem on the other side, describing the galaxy-IGM connection will help us

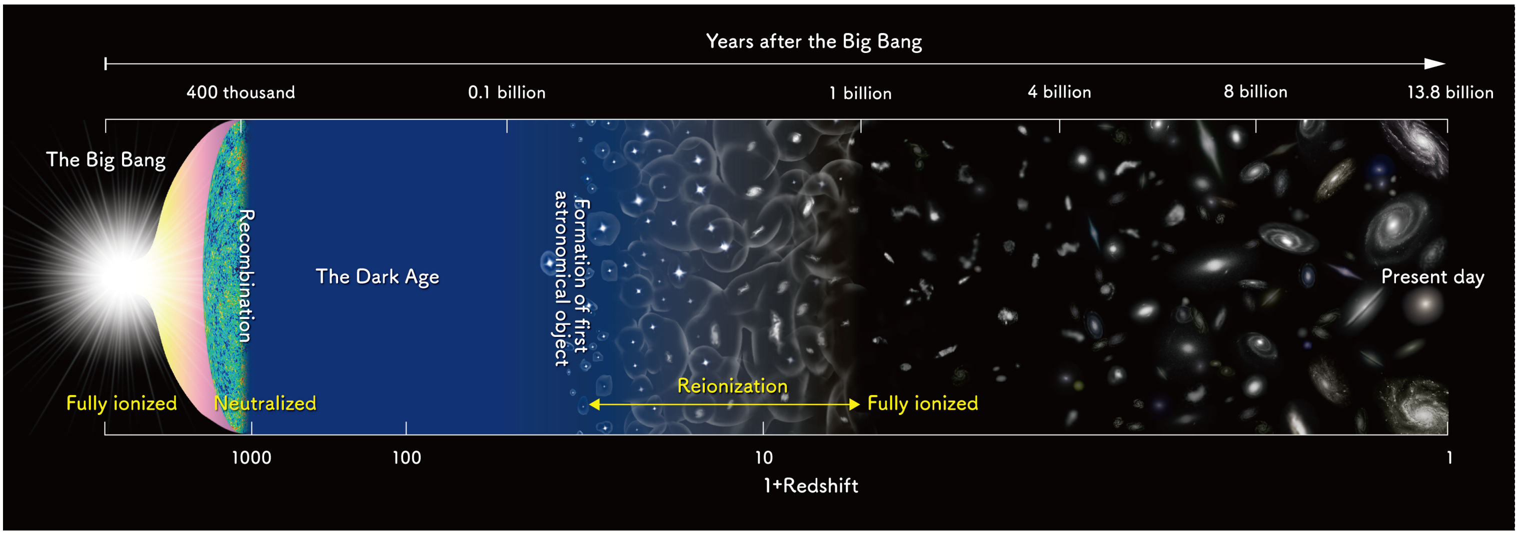

Figure 1.3: Our understanding of the timeline of the Universe. After the cosmic recombination the Universe became neutral, leading to the dark ages. First stars lit up the Universe, triggering the Epoch of Reionization and returning the Universe to an ionized state. The precise timing of these events remains uncertain and is the subject of ongoing research. Credits: NAOJ

describe galaxies too faint to ever be observed. EoR also affects the 21cm signal—another key probe of the high-redshift Universe. Improved understanding of reionization enhances the power of the 21cm line to address cosmological questions (Andrei Mesinger, 2016).

1.4 Observing the Epoch of Reionization

Understanding EoR from first principles is impossible given the dynamic range of the problem, starting from the star-formation cloud scales to cosmological distance of ionized bubbles. However, as of today we have a wealth of information from many large-scale probes. This allows us to help understand this important epoch. In this section we outline the most important ones, highlighting the ones that will be important for this thesis.

1.4.1 *QSO spectra: Lyman-α Forest and Damping Wing* We have already mentioned one observational set that was historically the most important: high-redshift QSOs. With their bright emission, they could be observed up to very large redshifts with a high signal-to-noise ratio, even for wavelengths bluer than the Lyman-α transition ($\lambda = 1215.67$ Å). The observed spectra of QSOs at these wavelengths were interlaced with a pattern called the Lyman-α forest, caused by absorption from neutral patches along the line of sight at the redshift of resonance. Given the resonant nature of the Lyman-α line, even a modest neutral fraction, $x_{\rm HI} \sim 10^{-5}$, is enough to saturate the absorption. Despite the Universe being ionized, there are many self-shielded regions connected to overdensities – and ultimately galaxies – giving rise to troughs. These troughs depend on the column density of the absorber, but also on the temperature (F. B. Davies and Steven R. Furlanetto, 2016). However, the presence of patches with little to no absorption indicates a Universe that is highly ionized (i.e. neutral fraction $< 10^{-5}$). This picture changes when looking at QSOs at $z > 6$, where the troughs become larger and larger. The presence of such large patches that are not fully ionized (sometimes larger than 100 cMpc) historically provided evidence that we were probing the epoch of reionization (R. H. Becker et al., 2001; G. D. Becker, Bolton, et al., 2015).

Today, we have observations of a large number of QSOs. Early attempts at quantifying EoR progression relied on relatively simple but robust statistics, such as quantifying the number of dark pixels in spectra of high-redshift quasars (Andrei Mesinger, 2010; McGreer, Andrei Mesinger, and Fan, 2011; McGreer, Andrei Mesinger, and D'Odorico, 2015; Jin et al., 2023). With the increasing number of spectroscopic observations, we can map the transmission along many lines of sight (using, for example, the XQR-30 survey; (D'Odorico et al., 2023; Sarah E. I. Bosman, F. B. Davies, et al., 2022)). This gives rise to a large statistical sample of Lyman-α forest measurements, allowing for more sophisticated methods.

Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. (2021) used 21cmFAST semi-numerical simulations to forward-

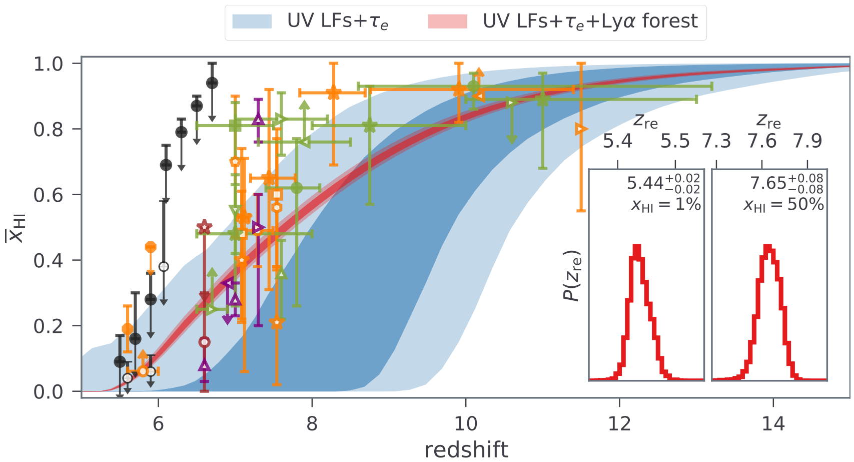

Figure 1.4: Posterior of the neutral hydrogen fraction evolution from Qin, Andrei Mesinger, Prelogović, et al. (2024). Dark (light) blue contours correspond to the 68% (95%) confidence intervals with the constraints from UV LFs and CMB optical depth, while analogously red contours correspond to the posterior with the inclusion of observations of Lyman-α forest optical depth distribution. Constraints are compared with other probes: dark gaps upper limits (McGreer, Andrei Mesinger, and D'Odorico, 2015; Jin et al., 2023), damping wing absorption in QSO (Andrei Mesinger and Zoltán Haiman, 2004; Bañados et al., 2018; Greig and Andrei Mesinger, 2017; Greig and Andrei Mesinger, 2018; Greig, Trott, et al., 2021; F. B. Davies, Hennawi, et al., 2018; F. Wang et al., 2020; Spina et al., 2024; Zhu et al., 2023) and galaxies (Emma Curtis-Lake et al., 2023; Hsiao et al., 2024; Umeda et al., 2024), Lyα EWs (Andrei Mesinger, Aykutalp, et al., 2015; C. A. Mason, Adriano Fontana, et al., 2019; Jung, Steven L. Finkelstein, Dickinson, et al., 2020; L. R. Whitler et al., 2020; Bolan, Lemaux, et al., 2022; Bruton et al., 2023; Jones, Bunker, Saxena, Witstok, et al., 2024; Nakane et al., 2024; Tang, Stark, Topping, et al., 2024) and LF and Lyα clustering (Emanuele Sobacchi and Andrei Mesinger, 2015; Inoue et al., 2018). Inset show the posterior of the end of EoR (redshift at which $\overline{x}_{\mathrm{HI}} = 0.01$) (left panel) and midpoint (redshift at which $\overline{x}_{\mathrm{HI}} = 0.5$) (right panel). Taken from Qin, Andrei Mesinger, Prelogović, et al. (2024)

model distributions of optical depths in the Lyman-α forest and compared them with observations to constrain the evolution of the mean ionized fraction (see also Sarah E. I. Bosman, F. B. Davies, et al. (2022)). With the increased sample, Qin, Andrei Mesinger, Prelogović, et al. (2024) were able to place very narrow constraints, especially on the end stages of EoR. The result was that the end of reionization occurred at $z \approx 5.5$, with an extended reionization history and a midpoint at $z \approx 7.5$. We show the posterior on the EoR history from Qin, Andrei Mesinger, Prelogović, et al. (2024) in Fig. 1.4. This tight posterior makes the Lyman-α forest the strongest current probe of the end of EoR. However, the model still allows for different ionization histories at higher redshifts.

Neutral hydrogen is also able to influence the SEDs of QSOs redward of the Lyman-α line through damping wing absorption. The amount of damping can be used to infer the neutral fraction of the region around a high-redshift QSO. The utility of this method depends on accurately reconstructing the intrinsic QSO spectrum. Many works have aimed to do this, for example using various analytic and machine learning techniques (Greig and Andrei Mesinger, 2017; F. B. Davies, Hennawi, et al., 2018; Fathivavsari, 2020; Sarah E. I. Bosman, Ďurovčíková, et al., 2021; B. Liu and Bordoloi, 2021; Z. Sun, Ting, and Cai, 2023; Ďurovčíková et al., 2024) (see also Greig, S. E. I. Bosman, et al. (2024) for a review and comparison of different frameworks). These methods have been applied to QSOs at $z \gtrsim 6$, eight of which are at $z > 7$, constraining the progression of EoR near its midpoint.

1.4.2 *CMB Observations* Another historically important probe targeting EoR is the CMB. As CMB photons travel the Universe, they are scattered by the electrons liberated during the epoch of reionization. The scattering along the line of sight creates the damping of the signal, quantified by the Thomson optical depth τ_{CMB}. CMB optical depth can be inferred from the total temperature signal, but there it is completely degenerate with the amplitude of the primordial fluctuations. On the other hand, polarization of the local CMB quadrupole on the scale of the horizon during EoR gives an estimate of τ_{CMB} independently of cosmology (see the review in Christian L. Reichardt, 2016). This allowed D. N. Spergel et al. (2007) to publish the first detection of the optical depth using three years worth of data with WMAP, obtaining $\tau_{\mathrm{CMB}} = 0.09 \pm 0.03$. With more observing time, and later with the Planck satellite, measurements improved and an interesting trend appeared, where new constraints became lower and lower (see Fig.41 in Planck Collaboration, Aghanim, Ashdown, et al., 2016). The current best estimate is $\tau_{\rm CMB} = 0.0561 \pm 0.0071$ (1σ) from Planck Collaboration, Aghanim, Akrami, et al. (2020), which implies a midpoint of reionization at $z \sim$ 7–9.

Despite the historical significance and relative model independence (see for example Qin, Poulin, et al. (2020)), CMB optical depth gives limited information. Given its integrated nature, it is mostly sensitive to the midpoint of the EoR, leaving freedom for variations in duration and early contributions to it. Furthermore, the optical depth is inferred using large scales ($l \lesssim 20$), which are affected by cosmic variance. Current measurements are close to the value expected from cosmic variance, further limiting the informativeness of the probe in the future.

However, EoR further distorts the CMB signals at smaller scales. The scattering of CMB photons during EoR produces a signal at $l \gtrsim 1000$, the kinetic Sunyaev–Zel'dovich effect (R. A. Sunyaev and Zeldovich, 1970). This effect is sourced by the Doppler shifting of CMB photons by the bulk motion of ionized gas within ionized bubbles. We leave the discussion of this effect for Chapter 2.

1.4.3 *Galaxy Observations* Galaxies are believed to be the dominant supplier of ionizing photons that drove the EoR, with only a marginal contribution from active galactic nuclei of high-redshift quasars (Trebitsch et al., 2021; Pratika Dayal, Volonteri, et al., 2025). Therefore, it is impossible to disentangle EoR science from galaxy evolution at high redshifts. Galaxies are more abundant than QSOs, so they have a higher constraining power; however, they are complex and intrinsically fainter. This makes the inference more difficult, requiring the modeling of many different processes described above.

There are two ways to use galaxies for EoR science. First, using population statistics about galaxies, we can deduce how galaxies evolve, linking them to the progress of EoR. Second, similar to QSOs, individual lines of sight probe the IGM, leading to additional constraints from an even larger population of objects. We discuss both methods in turn.

Luminosity Functions The most important summary for any galaxy observation is the one-point statistics, the luminosity function. Here we write it in the integral form:

$$\frac{\mathrm{d}n}{\mathrm{d}L_{\rm i}}(L_i, z) = \int \mathrm{d}M_h \frac{\mathrm{d}n}{\mathrm{d}M_h} p(M_*|M_h) \int \mathrm{d}M_*\ p(\mathrm{SFR}|M_*) \int \mathrm{dSFR}\ p(L_i|\mathrm{SFR}, M_h, M_*, ...), \tag{1.39}$$

where we link the galaxy abundance to the underlying halo distribution by a series of well-established relations. These are in turn the stellar-to-halo mass relation, the star-formation main sequence, and luminosity scaling relations. We leave the luminosity scaling general, as luminosities at different wavelengths follow different distributions. However, we show the equation in the general form, which integrates over the conditional distribution, emphasizing that the link between the halo mass and luminosity in any given band is not deterministic, and that the luminosity function is obtained by marginalizing over distributions. With such a formulation, we can consider any galaxy-to-galaxy scatter, which we consider as stochastic. In Chapter 4, we will discuss this stochasticity in more detail.

Given that the most important telescopes probe the high-redshift sky in rest-frame ultraviolet, the UV luminosity function has become the golden standard for describing galaxy populations at high redshifts. It is usually expressed in terms of magnitude at 1500 or 1600 Å. The Hubble Space Telescope was able to probe the UV LF deep into the EoR ($z \sim 10$; (R. J. Bouwens, Illingworth, Oesch, et al., 2015; R. J. Bouwens, Oesch, I. Labbé, et al., 2016; Oesch, R. J. Bouwens, et al., 2018a; Mauro Stefanon, Ivo Labbé, et al., 2019; R. J. Bouwens, Oesch, M. Stefanon, et al., 2021)), giving the picture of a Universe in which galaxies' star-formation efficiencies were constant (defined as the ratio of star formation rate to gas accretion rate; e.g., Tacchella, Bose, et al., 2018). These ideas were challenged by early observations taken with JWST. With its coverage into the infrared, JWST is better positioned to tackle the UV LF at $z > 10$ (see Fig. 1.5). There are now UV LF estimates close to what we believe is the cosmic dawn ($z > 15$), with some photometric candidates reaching $z > 20$ (M. Castellano, A. Fontana, et al., 2025; Pérez-González, Östlin, et al., 2025). Spectroscopic confirmation of all of these candidates will elucidate the evolution of the massive galaxies. As mentioned earlier, initial photometric estimates of the UV LF from JWST showed an increased abundance of massive halos compared to what was expected from extrapolation of HST data (G. Roberts-Borsani, Morishita, et al., 2022; Rychard J. Bouwens et al., 2023a). Different mechanisms were proposed, such as higher star-formation efficiency (Qin, Balu, and Wyithe, 2023; Harikane, Nakajima, et al., 2024), lack of dust (Andrea Ferrara, Andrea Pallottini, and Pratika Dayal, 2023), presence of AGN (Pacucci, Pratika Dayal, et al., 2022), change in the IMF (Yung, Somerville, Steven L. Finkelstein, et al., 2024; Cueto et al., 2024), among which stochasticity was also mentioned (Guochao Sun et al., 2023;

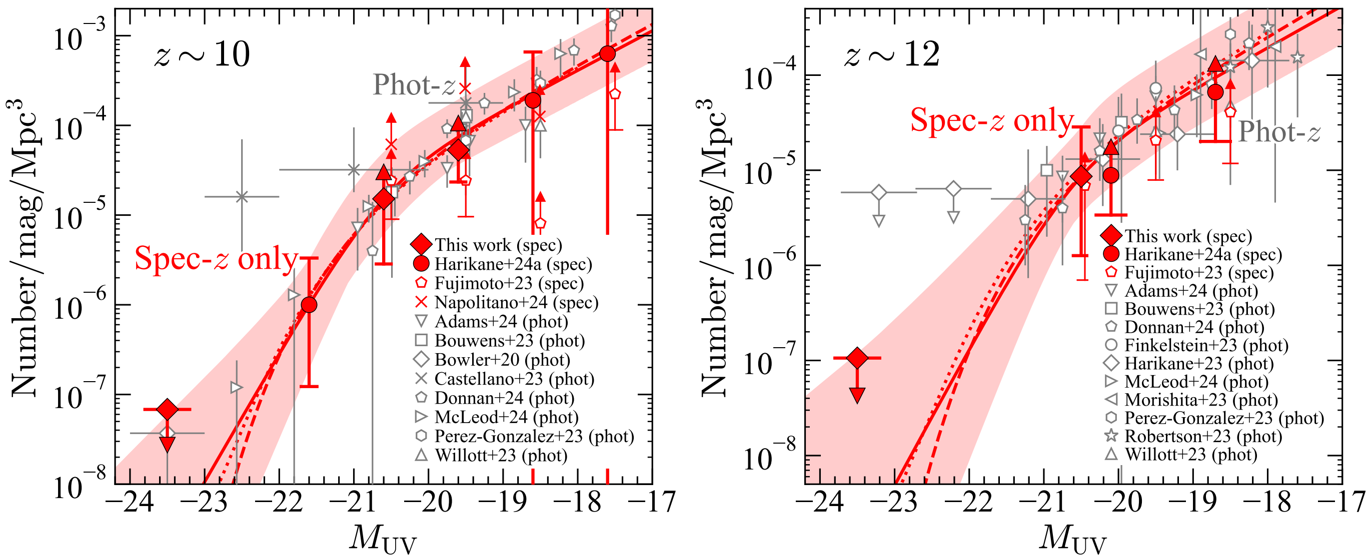

Figure 1.5: Spectroscopic UV LF estimate from Harikane, Inoue, et al. (2025) for $z \sim 10$ (left panel) and $z \sim 12$ (right panel) is shown in red, together with estimates from Harikane, Nakajima, et al. (2024). Shown are also photometric estimates with JWST, HST, and VISTA (Adams et al., 2024; Rychard J. Bouwens et al., 2023a; Bowler et al., 2020; Marco Castellano et al., 2023; Donnan, McLure, et al., 2024; McLeod et al., 2024; Pérez-González, Costantin, et al., 2023; Willott et al., 2024; Steven L. Finkelstein, Bagley, et al., 2023a; Harikane, Ouchi, Oguri, et al., 2023; Morishita et al., 2024; B. Robertson et al., 2024), as well as spectroscopic constraints from Napolitano, M. Castellano, et al. (2025) and Fujimoto, B. Wang, et al. (2024).

C. A. Mason, Trenti, and Tommaso Treu, 2023). Spectroscopic confirmation of the photometric candidates will help find the correct solution among them (or a combination thereof).

Current telescopes, including JWST, target the massive galaxies bright enough to be observed. However, most of the galaxies that reionized the Universe are too faint to be observed (Atek et al., 2024, see also Fig. 1.5). Despite that, any galaxy evolution model needs to be able to match UV LF constraints from JWST and HST. We will demonstrate how this can be done in the following chapters.

Galactic Lyman-alpha Damping Wing Attenuation Another important probe of the EoR is the attenuation of the galactic emission redward of the Lyman-α. Galaxies are intrinsically fainter than QSOs, making it harder to infer the neutral fraction on an object-to-object basis. However, galaxies are more abundant. The population statistics of attenuation allow us to use a large number of sources and connect them to EoR models. Already with HST, Lyman-α emitter luminosity functions were constructed and compared to the underlying Lyman-break galaxies, indicating a Universe that is reionizing (Tilvi, Papovich, et al., 2014; Schenker et al., 2014; Pentericci et al., 2014; Z.-Y. Zheng et al., 2017; Konno et al., 2018; Ota et al., 2017). On top of that, Bayesian analysis in C. A. Mason, Tommaso Treu, et al. (2018) and C. A. Mason, Adriano Fontana, et al. (2019) allowed individual galaxies to be put into a framework to infer the neutral fraction, obtaining constraints at $z \sim 7$.

The situation changed dramatically with JWST. With its increased sensitivity and infrared coverage, Lyman-α emission was mapped even more easily and to higher redshifts (even at $z \gtrsim 10$ Bunker et al., 2023; Witstok, Peter Jakobsen, et al., 2025). The increased number of Lyman-α emitters led to tighter and tighter constraints on the neutral fraction (Tang, Stark, Topping, et al., 2024; Umeda et al., 2024; Jones, Bunker, Saxena, Arribas, et al., 2025; Kageura et al., 2025; C. A. Mason, Z. Chen, et al., 2025), though statistical and systematic uncertainties are still a problem.

Lyman-α emission can also be used to track the EoR locally. Lyman-α photons are predominantly absorbed by the nearest hydrogen atoms, thus making the attenuation sensitive to the distance to the nearest neutral patch. This makes the Lyman-α line a great probe of individual ionized bubbles. First attempts at estimating ionized bubbles were done with HST (Tilvi, Malhotra, et al., 2020; Endsley and Stark, 2022) by observing galaxy overdensities around multiple Lyman-α emitters located close by. With JWST, more and more Lyman-α emitters were observed, allowing more and more fields to be identified as locations of ionized bubbles (Saxena et al., 2023; Tang, Stark, Topping, et al., 2024; Witstok, Smit, et al., 2024; Napolitano, Pentericci, et al., 2024; Z. Chen, Stark, C. Mason, et al., 2024; Z. Chen, Stark,

C. A. Mason, et al., 2025). Detection of ionized bubbles would be a powerful way of inferring the connection between galaxies and the EoR. However, statistical uncertainties need to be addressed. We propose a framework to do so in Chapter 3.

1.4.4 *21cm Signal* The 21cm signal that arises from the spin-flip transition of neutral hydrogen promises to be the strongest probe of the Universe during the dark ages, cosmic dawn, and EoR. The full science potential of the 21cm signal is beyond the scope of this introduction; we encourage readers to read the state of the field in (Andrei Mesinger, 2019).

The 21cm power spectrum is the prime target of several radio interferometers that are still only obtaining upper limits on the signal (Abdurashidova et al., 2022; Mertens et al., 2020; Nunhokee et al., 2025) (though see The HERA collaboration et al. (2023) and Greig, Trott, et al. (2021) for constraints on x-ray heating from upper limits). On the other hand, single-dish radio experiments provide information about the globally-averaged signal. A tentative detection of the global signal was reported by the EDGES telescope (Bowman et al., 2018), but its confirmation is still under question (see results from the SARAS experiment, Singh et al. (2022); also Cang et al. (2025)). In the near future, the SKA will take first light, revolutionizing the field and providing definitive answers to many questions still present (Braun et al., 2015).

1.5 Open Questions About EoR

Despite a number of observations targeting very high redshifts, increased theoretical knowledge, and improvements in hydrodynamical simulations, there are a number of open questions. In this section, we outline the most important ones, highlighting those I tried to answer during my PhD.

1.5.1 *Timing and Morphology of Reionization* The most important question about the epoch of reionization is: ***When did it happen?*** Throughout the previous decades, we have gathered evidence of the mean evolution of the process (see a summary of results in Fig. 1.4 taken from Qin, Andrei Mesinger, Prelogović, et al. (2024)). However, there are many caveats to essentially any EoR observations, as discussed above. While Lyman-α forest observations give tight constraints on the end stages of the evolution of the mean hydrogen fraction, they allow freedom in the mid- and early stages of EoR. On the other hand, CMB observations provide limited information on the evolution, targeting only the midpoint and/or duration of the process. Galaxy observations are a promising tool to study the EoR, but their faintness and systematics introduce large uncertainties. The 21cm signal is poised to revolutionize the field, but current experiments are still dominated by foregrounds. Combining all of the datasets mentioned above is necessary to map the full evolution, but large uncertainties still leave room for different scenarios.

EoR is a complex process that cannot be described solely by its mean evolution. Another important question is: ***How did EoR proceed locally?*** This is connected to the broader question: ***What is the morphology of EoR?*** Inferring the curve of mean evolution doesn't allow us to distinguish between what kinds of galaxies dominated the ionizing budget. This leads to another important question about EoR: ***What were the sources driving the reionization?*** We are only observing ”the tip of the iceberg” of the galaxy population responsible for reionization, so this question cannot be answered by direct observations. Furthermore, this question is linked to the complex problem of estimating the escape fractions of galaxies. Escape fraction cannot be measured directly even for the brightest galaxies, and extrapolating scaling relations from lower redshifts is still not a viable option. Therefore, we don't know how much even the observed galaxies contributed to reionization. The question of the sources of reionization can be approached from the other side, by systematically inferring the EoR morphology. In Chapter 3, we explore an exciting way of detecting ionized bubbles using Lyman-α observations. Detecting individual EoR bubbles could help answer how the galaxy population within them reionized their surroundings. We discuss how that can be done in Chapter 3.

Deciphering EoR is necessarily connected to answering questions about the evolution of the first galaxies. Galaxies are the main sources of ionizing photons, but they are also subject to feedback from ionization shocks. Answering questions about EoR is necessarily linked to another important problem: ***What are the properties of the galaxies during EoR?***

1.6 Inferring Knowledge from Data.

High-redshift astrophysics is a data-starved field of physics where controlled experiments are impossible and data is expensive to obtain. Additionally, astrophysical processes are impossible to model from first principles. The questions asked in the previous section need to be answered on scales as small as stars (and smaller) and as large as the Universe itself. Therefore, we rely on simulations and models of galaxy evolution to describe the processes of the first stars, first galaxies, and EoR. These models are constructed with constraints from the limited data we have. In this section, I describe how these models are built and how they are inferred from data. I emphasize points that are tackled in this thesis.

1.6.1 *Improvement of Modeling and Simulations* The Epoch of Reionization is difficult to model given the large range of scales involved. Circumventing this problem can be done in a couple of ways. On one end of the spectrum, we have large hydrodynamic simulations that often include radiative transfer to model the ionization of the IGM (Lewis, Ocvirk, Sorce, et al., 2022; Kannan et al., 2022). These simulations are very computationally expensive and cannot be varied in a Bayesian way (though see J. E. Davies, Bird, et al. (2023)). Furthermore, despite the large range of scales they cover, they cannot resolve the smallest scales where star formation happens. For example, cosmological simulations cannot capture the modes of escape of ionizing photons at the ISM level, relying on subgrid prescriptions (Yeh et al., 2022; Kostyuk, Nelson, et al., 2023). However, the usefulness of cosmological simulations in understanding the complexity of scales and providing tools for forecasting is of great value.

On the other end of the spectrum, we have analytic and semi-analytic models (Steven R. Furlanetto, Zaldarriaga, and Hernquist, 2004; McQuinn, Oliver Zahn, et al., 2006; Somerville, Hopkins, et al., 2008). They are flexible enough to be computed quickly in a physically motivated way, but capturing all of the relevant physics and scales for EoR is challenging. Usually, these approaches include a level of simplification on galaxy scales which may or may not be valid. One of the most important is that galaxy properties scale deterministically with halo mass or some other property. This simplification allows the integrals over the halo mass to be performed easily. However, ***every*** scaling relation has scatter around it. An example of stochasticity important for galaxy evolution is burstiness, discussed above. The impact of stochasticity is uncertain and needs to be studied. It is also worthwhile to note that even numerical simulations include simplifications at some scales that are often deterministic. For that reason, in Chapter 3 we tackle the question: ***What is the impact of stochasticity in EoR models?***

In between these two ends are semi-numerical simulations. These are fast simulations (see Chapter 2 as well) that do not employ radiative transfer, relying instead on the physical assumption of the excursion set formalism for reionization (Steven R. Furlanetto, Zaldarriaga, and Hernquist, 2004; Andrei Mesinger, S. Furlanetto, and Cen, 2011a). These approaches also treat galaxies as an analytic function of the underlying density field, i.e., integrating the effect of galaxies over some spatial scale. This allows semi-numerical simulations to be computed in a short amount of time, often in minutes, enabling the use of simulations in a Bayesian way. It is this approach that allows us, in Chapter 2, to simulate multiple large-scale probes and efficiently explore parameter space, drawing conclusions about EoR. Semi-numerical simulations also have shortcomings related to stochasticity (though in J. E. Davies, Andrei Mesinger, and S. Murray (2025), stochasticity is explicitly included).

It is also important to note that this stochasticity can be learned directly from data. In Chapter 5, we discuss a current project where stochasticity, along with scaling relations, is inferred from JWST UVLF data, using clustering to break many degeneracies. In the future, stochasticity in semi-numerical simulations will allow for exploration of parameter space using various other probes.

1.6.2 *Improvements in Inference Techniques* In previous sections, we outlined a large number of observations targeting EoR. However, given the large uncertainties and systematics, an important question is: ***How can we extract information from high-redshift data?*** The most appropriate way to infer knowledge from data is through the domain of Bayesian statistics. Bayes' theorem is a key theorem in statistics, describing the probability of some parameters of the model (θ) given observed data (D):

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}, \tag{1.40}$$

where $p(D|\theta)$ is the "likelihood" of the data given some parameters of the model, while $p(\theta)$ is the prior probability of the parameters. $p(D)$ is the evidence of the model and is a very useful quantity to evaluate different models that try to explain the same data. Bayes' theorem is a key way to place robust constraints on parameters, and it naturally extends to multiple datasets (in the likelihood).

Bayesian statistics has been extensively used in EoR research. The use of various datasets and quick simulations has allowed researchers to infer the EoR history from the Lyman-α forest, CMB optical depth, and galaxy observations (e.g., Greig and Andrei Mesinger, 2015; C. A. Mason, Tommaso Treu, et al., 2018; Qin, Andrei Mesinger, Sarah E. I. Bosman, et al., 2021; Qin, Andrei Mesinger, Prelogović, et al., 2024; C. A. Mason, Z. Chen, et al., 2025). In Chapter 2, I will describe one such approach, where we show how constraints change when adding the observation of the patchy kinetic Sunyaev-Zel'dovich signal (taken from Nikolić, Andrei Mesinger, Qin, et al., 2023). Additionally, Bayesian inference is an indispensable tool for inferring the properties of galaxies deep into the EoR, where cosmological simulations disagree on scaling relations. We outline a work in progress on that topic in Chapter 5. In the previous sections, it was demonstrated that we need to infer the EoR morphology together with the mean evolution of the neutral hydrogen fraction. Precisely measuring the properties of ionized bubbles at the mid- and early stages of EoR would allow us to do that. In Chapter 3, I show an approach that robustly infers the properties of ionized bubbles around galaxy observations. Carefully accounting for sources of stochasticity allows robust inference at the 10% level.

Together with improvements in observations and models, inference techniques are also modernizing. The rise of machine learning algorithms has allowed us to improve our inference through simulation-based inference (SBI; Prelogović and Andrei Mesinger, 2023; B. Semelin et al., 2025), field-level inference, and emulators (Breitman et al., 2024), to name a few. In Chapter 5, we will discuss potential implementations of these methods in current and future projects, further enhancing our knowledge of the EoR.

1.7 Thesis Overview

In Chapter 2, I present the work published in Nikolić, Andrei Mesinger, Qin, et al. (2023). There, we used patchy kSZ signal observations from C. L. Reichardt, Patil, et al. (2021) to put constraints on the EoR and galaxy properties. We used 21cmFAST simulations in a Bayesian framework, incorporating other data from CMB optical depth from Planck Collaboration, Aghanim, Akrami, et al. (2020) and the Lyman-α forest (Sarah E. I. Bosman, Fan, et al., 2018). We express the results in terms of astrophysical scaling laws and EoR histories, finding consistency between the patchy kSZ signal and the Lyman-α forest. We also provide forecasts on the constraining power of future patchy kSZ measurements and compare the scaling of the signal with EoR properties.

In Chapter 3, I present the work published in Nikolić, Andrei Mesinger, C. A. Mason, et al. (2025) (accepted for publication in A&A, at the time of writing in the production stage). There, we developed a pipeline that statistically infers the properties of ionized bubbles around observations using Lyman-α data from JWST/NIRSpec. We take into account stochasticity from galaxy scaling relations and the IGM, as well as complementary information from multiple galaxies in the field. We test the pipeline with different Lyman-α properties and by performing inference on 21cmFAST simulations. We also developed a survey strategy that can be implemented with future JWST observations.

Stochasticity is further discussed in Chapter 4. There, I present the work published in Nikolić, Andrei Mesinger, J. E. Davies, et al. (2024), where we developed a model that predicts emissivities, EoR histories, and UV LFs while taking into account scatter around all galaxy scaling relations. We found that scatter around the star-formation main sequence, representing the burstiness of high-redshift galaxies, is important at all redshifts ($z > 5$), for wavelengths from X-rays to near-UV, and at all scales. A wide distribution of escape fractions affects the ionizing emissivity, while other sources of scatter are important at the $\sim 10\%$ level. We find that the current levels of scatter in our model, rooted in simulations and galaxy observations ($z < 10$), cannot explain the photometric estimates of the UV LF from JWST at $z > 12$.

I conclude the thesis in Chapter 5, presenting also future work. First, I detail a project where we infer galaxy relations and scatter around them using the JWST UV LF, together with information about the angular clustering. The work is nearing completion at the time of writing this thesis. Finally, I present some future projects that are closely connected to the work presented in other chapters.

2

Inferring Reionization and Galaxy Properties from the Patchy Kinetic Sunyaev-Zel'dovich Signal

2.1 Introduction

The epoch of reionization (EoR) is a major milestone in the Universe's evolution. Although many questions remain, recent years have seen a dramatic increase in the volume of data available to probe the cosmological frontier. These include: (i) high-redshift QSO spectra (e.g. Bolton, G. D. Becker, et al., 2010; G. D. Becker, Bolton, et al., 2015; Iršič et al., 2017; Sarah E. I. Bosman, Fan, et al., 2018; Gaikwad, Rauch, et al., 2020); (ii) Lyman alpha emitting galaxies (e.g. Stark, Ellis, Chiu, et al., 2010; Stark, Ellis, Charlot, et al., 2017a; Konno et al., 2018; Hoag et al., 2019; C. A. Mason, Adriano Fontana, et al., 2019; Leonova et al., 2022a; Endsley, Stark, Rychard J. Bouwens, et al., 2022); (iii) the optical depth to the CMB (e.g. Planck Collaboration, Adam, et al., 2016; Planck Collaboration, Aghanim, Akrami, et al., 2020; Heinrich and Wayne Hu, 2021); (iv) UV luminosity functions (LF, e.g. Steven L. Finkelstein, Ryan, et al., 2015; R. J. Bouwens, Illingworth, Oesch, et al., 2015; R. J. Bouwens, Oesch, I. Labbé, et al., 2016; Oesch, R. J. Bouwens, et al., 2018a; Bhatawdekar et al., 2019; Rychard J. Bouwens et al., 2023b; Harikane, Ouchi, Oguri, et al., 2023); (v) preliminary upper limits on the 21 cm power spectrum (Mertens et al., 2020; Trott et al., 2020; Z. The HERA collaboration A. et al., 2022; The HERA collaboration et al., 2023). This trend is set to culminate in the coming decade with 21-cm maps of the first billion years from the Square Kilometre Array (SKA)[1].

A complementary probe that has arguably seen less attention is provided by the patchy kinetic Sunyaev-Zel'dovich (kSZ) signal. The kSZ is sourced by the Doppler shifting of CMB photons that scatter off of free electrons, resulting in secondary temperature anisotropies. It is typically separated into post-EoR (or homogeneous) and EoR (or patchy) contributions. The patchy kSZ is determined by the timing, duration and morphology of the EoR. Thus, measuring its shape and amplitude could inform us about the evolution of this cosmic milestone as well as the galaxies that sourced it (e.g. Iliev, Pen, et al., 2007; Andrei Mesinger, McQuinn, and David N. Spergel, 2012; H. Park et al., 2013; Tirthankar Roy Choudhury, Mukherjee, and Paul, 2021; Bégin, A. Liu, and Adélie Gorce, 2022).

Measurements of the patchy kSZ have historically focused on the angular multipole $l = 3000$ (roughly corresponding to 4 arcmin, or a comoving scale of 20 Mpc during the EoR). At lower multipoles the primary CMB anisotropies are increasingly dominant, while at higher multipoles systematics such as the cross-correlation between the thermal Sunyaev-Zel'dovich (tSZ) and dusty galaxies become even more challenging. The two telescopes actively targeting the kSZ, the Atacama Cosmology Telescope (ACT) and the South Pole Telescope (SPT), have until recently only published upper limits (Lueker et al., 2010; S. Das, Marriage, et al., 2011; Shirokoff et al., 2011; C. L. Reichardt, L. Shaw, et al., 2012; Dunkley et al., 2013; S. Das, Louis, et al., 2014; George et al., 2015). Strong foregrounds, including bright extragalactic sources, as well as modelling uncertainties remain very challenging. However, the SPT collaboration recently claimed a low signal to noise (S/N) measurement of the patchy kSZ signal: $\mathcal{D}_{3000}^{\rm pkSZ} = 1.1^{+1.1}_{-0.7}\,\mu{\rm K}^2$ (68% C.I., C. L. Reichardt, Patil, et al., 2021). These relatively low values qualitatively point to a much later and more rapid EoR compared to original estimates (e.g. McQuinn, Steven R. Furlanetto, et al., 2005; Oliver Zahn, Zaldarriaga, et al., 2005; Iliev, Pen, et al., 2007; O. Zahn et al., 2012; Andrei Mesinger, McQuinn, and David N. Spergel, 2012; H. Park et al., 2013; Calabrese

---

[1] https://www.skatelescope.org/

et al., 2014; Alvarez, 2016; Paul, Mukherjee, and Tirthankar Roy Choudhury, 2021). Future telescopes, such as the Simons Observatory[2] (P. Ade et al., 2019), CMB-Stage 4[3] (Abitbol et al., 2017) and CMB-HD[4] (Sehgal et al., 2019), should help better characterize the CMB foregrounds and related systematics to narrow down error bars.

However, interpreting a tentative detection of the kSZ is difficult. Firstly, one needs to statistically separate the homogeneous and patchy contributions from the total kSZ power. Secondly, the patchy kSZ power is an integral measurement of the EoR, and as such is prone to strong astrophysical parameter degeneracies. Robust interpretation therefore must rely on additional, complementary observations of the EoR and high-redshift galaxies.

Here we quantify what we can learn from the recent kSZ measurement using a fully Bayesian framework. Unlike previous works, we directly sample empirical properties of galaxies that drive the EoR, creating 3D lightcones on-the-fly. This allows us to: (i) self-consistently sample different EoR morphologies when comparing against kSZ observations (instead of the common approach of fixing the morphology and empirically varying the midpoint and duration of the EoR); (ii) combine independent high-z galaxy and EoR observations when computing the posterior; and (iii) set physically-meaningful priors.

This paper is organized as follows. In Sec. 2.2 we discuss how we compute the patchy kSZ signal. Our Bayesian framework, combining the kSZ with complementary observations, is summarized in Sec. 2.3. We present and discuss our results in Sec. 2.4. In Sec. 2.5 we quantify how accurately the midpoint and duration of reionization can predict the patchy kSZ at $l = 3000$. Finally, we conclude in Sec. 2.6. Throughout this work, we assume standard ΛCDM cosmological parameters $(\Omega_\mathrm{m}, \Omega_\mathrm{b}, \Omega_\Lambda, h, \sigma_8, n_\mathrm{s} = 0.321, 0.049, 0.679, 0.67, 0.81, 0.963)$, consistent with the latest estimates from Planck Collaboration, Aghanim, Akrami, et al. (2020). Unless stated otherwise, we quote all quantities in comoving units.

2.2 The Patchy Kinetic Sunyaev-Zel'dovich Signal

The secondary temperature anisotropy of the CMB due to the kinetic Sunyaev-Zel'dovich effect in the line of sight (LoS) direction $\hat{\mathbf{u}}$ can be written as:

$$\begin{aligned}\delta T_\mathrm{kSZ} &= \frac{\Delta T}{T}(\hat{\mathbf{u}}) \\ &= \sigma_\mathrm{T} \int \mathrm{d}z \, \left(\frac{\mathrm{d}t}{\mathrm{d}z}\right) \, e^{-\tau_e(z)} n_e \, \hat{\mathbf{u}} \cdot \mathbf{v} \\ &= \sigma_\mathrm{T} \int \mathrm{d}z \, \left(\frac{\mathrm{d}t}{\mathrm{d}z}\right) \, e^{-\tau_e(z)} \, x_e \, n_b \, \hat{\mathbf{u}} \cdot \mathbf{v}.\end{aligned} \tag{2.1}$$

Here, σ_T is the Thomson scattering cross-section, n_e is the number density of electrons[5] which can be expanded as the product of the ionized fraction (x_e) and baryon density (n_b), $\mathbf{v}$ is the velocity of electrons, and τ_e is the optical depth of CMB photons up to redshift z:

$$\tau_e(z) = \sigma_\mathrm{T} \int_0^z \mathrm{d}z' \, c \frac{\mathrm{d}t}{\mathrm{d}z'} n_e. \tag{2.2}$$

The redshift integral in equation (2.1) is generally separated into a post-reionization (or homogeneous) component and one due to patchy reionization. Observations measure the total kSZ power spectrum[6], which is the sum coming from these two respective components: $\mathcal{D}_l^\mathrm{kSZ} = \mathcal{D}_l^\mathrm{hkSZ} + \mathcal{D}_l^\mathrm{pkSZ}$. The post-reionization kSZ power spectrum, $\mathcal{D}_l^\mathrm{hkSZ}$, is dominated by fluctuations in n_b during the era of cluster formation at $z \lesssim 1$ (e.g. L. D. Shaw, Rudd, and Nagai 2012), while the patchy kSZ power spectrum, $\mathcal{D}_l^\mathrm{pkSZ}$, is dominated by order unity fluctuations in x_e during the EoR at $z \gtrsim 5$ (e.g. Alvarez 2016). Constraining the patchy kSZ thus requires statistically accounting for the post-EoR (homogeneous) kSZ signal; we summarize how this was done for recent observations in Sec. 2.2.2. Because we do not know *a priori* the reionization redshift, here we *define* the patchy kSZ component as the contribution to equation (2.1) of redshifts above $z \geq 5$. We note that this is a lower value compared to some previous choices in the literature. It is

---

[2] https://simonsobservatory.org/

[3] https://cmb-s4.org/

[4] https://cmb-hd.org/

[5] Here we assume helium is doubly ionized at $z < 3$, and singly ionized at the same fraction as hydrogen during the EoR.

[6] The power spectrum is defined as $\mathcal{D}_l^\mathrm{kSZ} = \frac{l(l+1)}{2\pi} C_l^\mathrm{kSZ}$, where $C_l^\mathrm{kSZ} = T_\mathrm{CMB}^2 |\overline{\delta T}_\mathrm{kSZ}(k)|^2$, T_CMB is the mean CMB temperature and $\overline{\delta T}_\mathrm{kSZ}(k)$ is the Fourier transform of δT_kSZ.

motivated by recent Lyman alpha forest data whose interpretation requires a late reionization, ending at $5.3 \lesssim z \lesssim 5.6$ (Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. 2021; T. Roy Choudhury, Paranjape, and Sarah E. I. Bosman 2021; Sarah E. I. Bosman, F. B. Davies, et al. 2022, Qin et al. in prep).

The kSZ power is typically measured at $l = 3000$; smaller multipoles are increasingly dominated by primary CMB anisotropies, while larger multipoles become swamped by other foregrounds such as dusty galaxies (e.g. O. Zahn et al., 2012; Alvarez, 2016; C. L. Reichardt, L. Shaw, et al., 2012). During the EoR, $l = 3000$ roughly corresponds to physical scales of $\sim 20\,\mathrm{cMpc}$. Therefore, measurements of the patchy kSZ at this multipole are sensitive to the EoR morphology on these scales, as well as the timing and duration of the corresponding epochs. Simulation box sizes larger than about $300\,\mathrm{cMpc}$ are sufficient to capture the ionization power spectra on those scales (e.g. Iliev, Mellema, et al., 2014; Kaur, N. Gillet, and Andrei Mesinger, 2020). Unfortunately, the kSZ is determined to leading order by the velocity-ionization cross power, and much larger scales (above 1 cGpc) are required to capture the fluctuations in the velocity field and corresponding velocity-ionization cross-power at $l \sim 3000$ (e.g. L. D. Shaw, Rudd, and Nagai, 2012; Alvarez, 2016). Given that radiative transfer simulations on such large scales are computationaly prohibitive, more approximate schemes are required to calculate the patchy kSZ signal.

The patchy kSZ power is sometimes computed analytically (with some terms calibrated to smaller numerical simulations; e.g. H. Park et al. 2013; A. Gorce, Ilić, et al. 2020) but at the price of neglecting the contribution of higher order correlations (above two points) which can represent up to 10% of the total patchy power (Alvarez, 2016). More importantly, it is difficult to associate prior probabilities on the "effective" parameters of such models; priors are important for inference from a low S/N detection whose likelihood is not strongly constraining. Instead, in this work, we choose to compute the patchy kSZ signal by ray-tracing through large 3D lightcone simulations with approximate radiative transfer (so-called semi-numerical simulations; e.g. O. Zahn et al. 2012; Andrei Mesinger, McQuinn, and David N. Spergel 2012; Battaglia et al. 2013; Seiler et al. 2019; Tirthankar Roy Choudhury, Mukherjee, and Paul 2021; N. Chen et al. 2022). Our self-consistent approach allows us to incorporate multi-frequency observations of the EoR and high-z galaxies in the likelihood. We discuss how this is done in the following section.

2.2.1 *Computing the Patchy kSZ from Galaxy-Driven EoR Simulations* In this work we extend the public simulation package `21cmFAST`[7] (e.g. Andrei Mesinger and S. Furlanetto 2007a; Andrei Mesinger, S. Furlanetto, and Cen 2011b; S. Murray et al. 2020a) to forward-model the patchy kSZ signal together with other observables. `21cmFAST` is a semi-numerical code used for generating cosmological simulations of the early Universe. It computes the evolved density and velocity fields using second-order Lagrangian perturbation theory (e.g. Scoccimarro 1998). The ionization field is generated from the density field by comparing the cumulative number of ionizing photons produced by galaxies to the number of hydrogen atoms plus cumulative number of IGM recombinations, in spherical regions with decreasing radii, R (e.g. Steven R. Furlanetto, Zaldarriaga, and Hernquist, 2004). Specifically, a cell is marked as ionized if at any radius:

$$n_{\mathrm{ion}} \geq (1 + n_{\mathrm{rec}})(1 - x_e), \tag{2.3}$$

where n_{rec} is the cumulative number of recombinations per baryon computed according to the sub-grid scheme of Emanuele Sobacchi and Andrei Mesinger (2014), x_e accounts for pre-ionization by X-ray photons, and n_{ion} is the cumulative number of ionizing photons per baryon, with quantities averaged over the sphere of radius R:

$$n_{\mathrm{ion}} = \frac{1}{\rho_b} \int_0^{\infty} \mathrm{d}M_{\mathrm{h}} \, \frac{\mathrm{d}n(M_{\mathrm{h}}, z|R, \delta_R)}{\mathrm{d}M_{\mathrm{h}}} \, f_{\mathrm{duty}} \, M_* \, f_{\mathrm{esc}} \, N_{\gamma/b}. \tag{2.4}$$

Here dn/dM_{h} is the conditional halo mass function, $N_{\gamma/b}$ is the number of ionizing photons per stellar baryon, f_{esc} is the escape fraction of ionizing photons, and f_{duty} corresponds to the fraction of halos that host star forming galaxies.

Here we adopt the flexible parameterization from J. Park, Andrei Mesinger, et al. (2019). Specifically, f_{duty} decreases exponentially below a characteristic mass scale, M_{turn}, due to inefficient gas cooling and/or feedback (e.g. E. Sobacchi and A. Mesinger 2013; H. Xu, Wise, et al. 2016; Mutch et al. 2016):

$$f_{\mathrm{duty}}(M_{\mathrm{h}}) = \exp\left(-\frac{M_{\mathrm{h}}}{M_{\mathrm{turn}}}\right). \tag{2.5}$$

---

[7] `https://github.com/21cmfast/21cmFAST`

The ionizing escape fraction $f_{\rm esc}$ and stellar mass M_* are taken to be power law functions of halo mass:

$$f_{\rm esc}(M_{\rm h}) = f_{\rm esc,10}\left(\frac{M_{\rm h}}{10^{10}M_\odot}\right)^{\alpha_{\rm esc}}, \tag{2.6}$$

$$M_*(M_{\rm h}) = f_{*,10}\left(\frac{M_{\rm h}}{10^{10}M_\odot}\right)^{\alpha_*}\left(\frac{\Omega_b}{\Omega_m}\right) M_{\rm h}. \tag{2.7}$$

Here, $f_{\rm esc,10}$ is the ionizing photon escape fraction normalized to the value in halos of mass $10^{10}M_\odot$, $f_{*,10}$ is the fraction of galactic gas in stars also normalized to the value in halos of mass $10^{10}M_\odot$, and $\alpha_{\rm esc}$ and α_* are the corresponding power law indices. Both $f_{\rm esc}$ and $f_* \equiv f_{*,10}\left(\frac{M_{\rm h}}{10^{10}M_\odot}\right)^{\alpha_*}$ have a physical upper limit of 1. This model also assumes that the star formation rate can be expressed on average as the stellar mass divided by some characteristic time scale:

$$\dot{M}_*(M_{\rm h}, z) = \frac{M_*}{H^{-1}(z)t_*}, \tag{2.8}$$

where $H(z)$ is the Hubble parameter and t_* is the characteristic time-scale for star formation (with this definition, its value varies from zero to unity).

This six-parameter galaxy model ($f_{*,10}$, α_*, $f_{\rm esc,10}$, $\alpha_{\rm esc}$, $M_{\rm turn}$, t_*) is able to capture the average properties of the faint galaxies that dominate the ionizing photon budget, both from theoretical models and observations (e.g. P. Behroozi et al. 2019; Ishigaki et al. 2018; X. Ma, Quataert, et al. 2020; J. Park, N. Gillet, et al. 2020; R. J. Bouwens, Illingworth, P. G. van Dokkum, et al. 2022). Further details about the code and the parametrization can be found in Andrei Mesinger, S. Furlanetto, and Cen (2011b) and J. Park, Andrei Mesinger, et al. (2019) and S. Murray et al. (2020a).

For a given combination of astrophysical parameters, `21cmFAST` outputs 3D lightcones of the relevant cosmological fields. We thus compute the patchy kSZ signal by ray-tracing through the ionization, density and LoS velocity lightcones, directly calculating the integral in equation (2.1), accounting also for the angular evolution of $\hat{\mathbf{u}}$ (Andrei Mesinger, McQuinn, and David N. Spergel, 2012).

In Fig. 2.1 we show an example of this procedure using a simulation that is 1.5 Gpc on a side. The astrophysical parameters of this simulation are taken from the posterior distribution of Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. 2021 (discussed further below), specifically: $[\log_{10}(f_{*,10}), \alpha_*, \log_{10}(f_{\rm esc,10}), \alpha_{\rm esc}, \log_{10}(M_{\rm turn}), t_*] = (-1.42, 0.614, -1.78, 0.474, 8.62, 0.392\ )$. The midpoint of EoR is at $z_r = 6.1$, while the neutral fraction drops to zero at $z_{\rm end} = 4.9$. The duration of the EoR, defined throughout as $\Delta_z \equiv z(\overline{x}_{\rm HI} = 0.75) - z(\overline{x}_{\rm HI} = 0.25)$, is $\Delta_z = 0.76$ and the CMB optical depth for the simulation is $\tau_e = 0.042$. In the top panel we show a 2D slice (with a thickness of 1.4 Mpc) through the neutral fraction lightcone. In the bottom panels, we show the map of the patchy kSZ signal and the corresponding angular power spectrum. While this model was chosen to have patchy kSZ power that agrees with the median estimate reported by C. L. Reichardt, Patil, et al. 2021, complementary EoR and galaxy observations pull the posterior towards larger values of the $l = 3000$ kSZ power, as we quantify further below.

2.2.2 *Observations of the Patchy kSZ* Observing the kSZ power spectrum is very challenging due to the presence of strong foregrounds as well as the primary CMB anisotropies. Deep integration over multiple frequencies is essential in separating these different components of the power spectra. Over the past decade, ACT and SPT have published increasingly tighter upper limits on the cosmic kSZ signal (Dunkley et al., 2013; S. Das, Louis, et al., 2014; C. L. Reichardt, L. Shaw, et al., 2012; George et al., 2015). Using SPT-SZ and SPTpol measurements at 95, 150 and 220 GHz, combined with a prior on the CIB-tSZ foregound from Crawford et al. (2014), C. L. Reichardt, Patil, et al. (2021) recently claimed a 3σ measurement of the total kSZ power: $\mathcal{D}_{3000}^{\rm kSZ} = 3.0 \pm 1.0\ \mu{\rm K}^2$ (68% C.I.).

To isolate the patchy contribution to this total kSZ power, the authors subtracted an estimate of the $z < 5.5$ homogeneous component based on the simulations of L. D. Shaw, Rudd, and Nagai (2012): $\mathcal{D}_{3000}^{\rm hkSZ} = 1.65\mu{\rm K}^2$. The uncertainty around this value is bracketed by rescaling the best guess by a factor of 0.75 and 1.25. Doing so and using the bispectrum prior on tSZ, C. L. Reichardt, Patil, et al. (2021) estimate the patchy kSZ power at $l = 3000$ to be $\mathcal{D}_{3000}^{\rm pkSZ} = 1.1^{+1.0}_{-0.7}\ \mu{\rm K}^2$ (68% C.I.). Since our choice of lower bound in this work is $z = 5.0$ instead of $z = 5.5$, we add to the patchy kSZ estimate from C. L. Reichardt, Patil, et al. (2021) the contribution of the homogeneous component over the redshift interval $5 < z < 5.5$. We estimate this be approximately $0.1\mu{\rm K}^2$ (e.g. fig. 6 in L. D. Shaw, Rudd, and Nagai 2012; fig. 5 in Andrei Mesinger, McQuinn, and David N. Spergel 2012). Therefore we use the following observational constraint when performing inference in Section 2.4.1: $\mathcal{D}_{3000}^{\rm pkSZ} = 1.2^{+1.0}_{-0.7}\ \mu{\rm K}^2$ (68% C.I.).

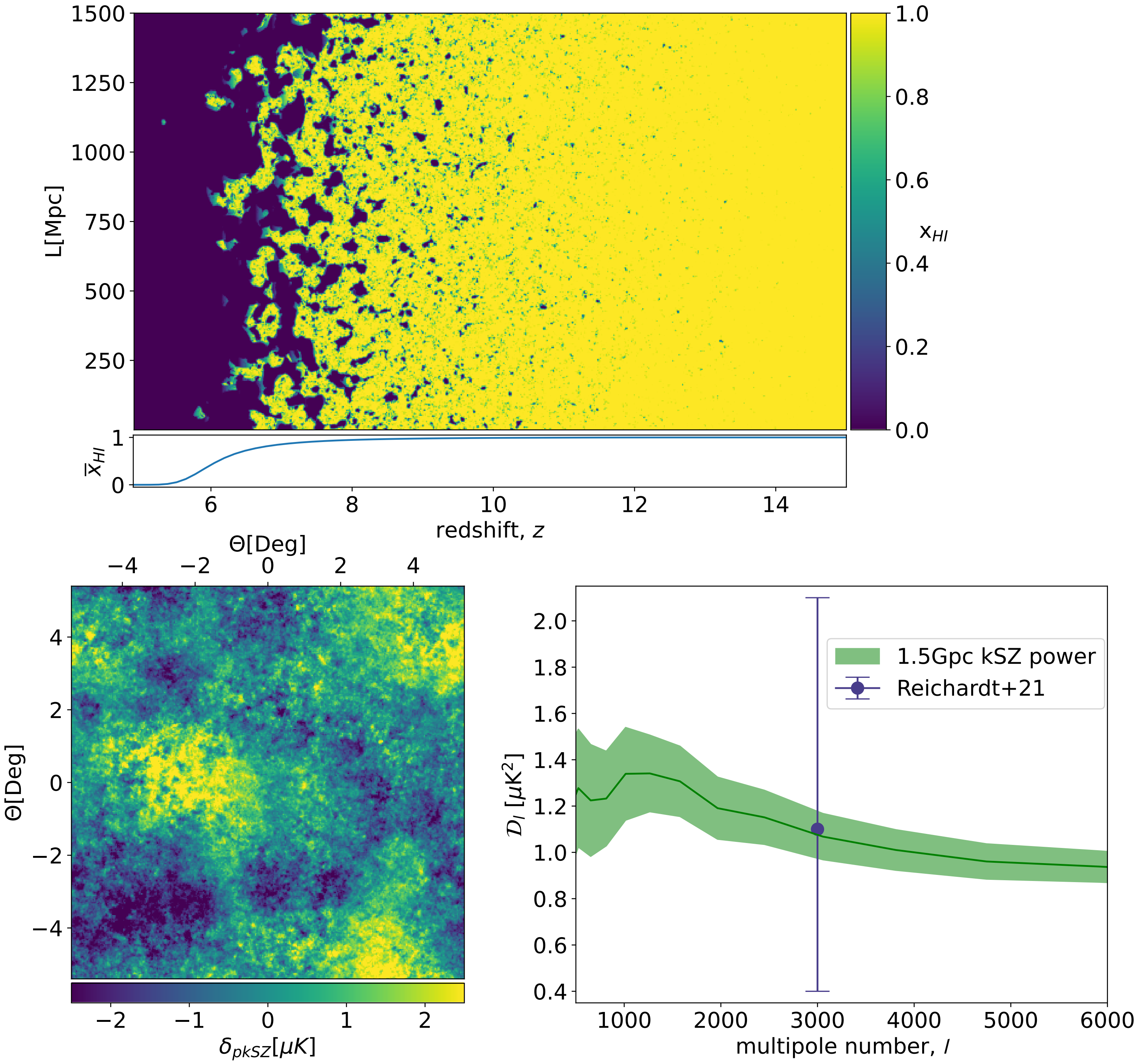

Figure 2.1: *Upper panel:* 2D slice through the neutral hydrogen fraction lightcone together with its mean evolution on the bottom. The lightcone slice is 1.5 Gpc in height and 1.4 Mpc thick. *Lower left panel:* map of the patchy kinetic Sunayev-Zel'dovich signal, defined as being sourced by redshifts greater than five. *Lower right panel:* Corresponding angular power spectrum of the patchy kSZ (*solid line*). The green shaded area highlights the 1σ Poisson sample variance. Also shown is the recent measurement by C. L. Reichardt, Patil, et al. (2021) at $l = 3000$. This simulation used the following astrophysical parameters: $\log_{10}(f_{*,10}) = -1.42$, $\alpha_* = 0.614$, $\log_{10}(f_{\rm esc,10}) = -1.78$, $\alpha_{\rm esc} = 0.474$, $\log_{10}(M_{\rm turn}) = 8.62$ and $t_* = 0.392$.

A more robust foreground model and a consistent analysis across scales can improve constraints, as demonstrated in Adélie Gorce, Marian Douspis, and Salvati (2022) where the authors give an upper limit of $\mathcal{D}_{3000}^{\rm pkSZ} < 1.58\,\mu{\rm K}^2$ (95% C.L.) using the same data as C. L. Reichardt, Patil, et al. (2021).[8] Reducing the uncertainties on the total kSZ require

---

[8] As this project was started before the publication of Adélie Gorce, Marian Douspis, and Salvati (2022), here we use the original patchy kSZ estimate by (C. L. Reichardt, Patil, et al., 2021). The estimate in Adélie Gorce, Marian Douspis, and Salvati (2022) would imply an even

deeper integration, lower noise levels, and more frequency channels to better characterize foregrounds and systematics, which future telescopes such as CMB-S4 and the Simons Observatory (e.g. Abitbol et al. 2017; P. Ade et al. 2019) are expected to achieve. Furthermore, robustly isolating the patchy component of the total kSZ signal requires exhaustively sampling models of galaxy clusters in order to better characterize the post-reionization (homogeneous) component. Motivated by upcoming data and improved analysis, we also perform a forecast run from a mock measurement with error bars corresponding to the uncertainty expected from future experiments. This is presented in subsection 2.4.3.

2.3 Complementary EoR and Galaxy Observations

We now have several, independent observational probes of the EoR which can help constrain astrophysical parameters (e.g. T. Roy Choudhury and A. Ferrara, 2006; Greig and Andrei Mesinger, 2015; A. Gorce, M. Douspis, et al., 2018; J. Park, Andrei Mesinger, et al., 2019). Here we follow Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. (2021), who used the same galaxy parametrization as we do, and use the following observational data:

1. *Lyman α forest opacity distributions* – the $5.4 \leq z \leq 6.0$ probability density functions (PDFs) of the forest effective optical depth, $\tau_{\rm eff} \equiv -\ln\langle f\rangle_{50\,{\rm Mpc}}$, computed from the mean normalized flux, f, of the QSO sample in Sarah E. I. Bosman, Fan, et al. (2018). Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. (2021) showed that this data require reionization to end late, $z \leq 5.6$ (see also T. Roy Choudhury, Paranjape, and Sarah E. I. Bosman 2021).

2. *Dark fraction in the Lyα and Lyβ forests* – the fraction of QSO spectral pixels that are dark (zero transmission) in both Lyman alpha and Lyman beta from the sample in McGreer, Andrei Mesinger, and D'Odorico (2015). This so-called dark fraction provides a model-independent upper limit on the neutral hydrogen fraction, with the value at $z \sim 5.9$ corresponding to $\bar{x}_{\rm HI} < 0.06 + 0.05$ (1σ). This dataset favors earlier reionization models.

3. *High-redshift galaxy UV luminosity functions (UV LFs)* – the 1500 Å restframe UV LFs at $z = 6 - 10$, estimated by R. J. Bouwens, Illingworth, Oesch, et al. (2015), R. J. Bouwens, Oesch, I. Labbé, et al. (2016), and Oesch, R. J. Bouwens, et al. (2018a). To constrain our models, we assume a conversion factor between the star formation rate (SFR) and UV luminosity, $\dot{M}_* = \mathcal{K}_{\rm UV} L_{\rm UV}$, and take $\mathcal{K}_{\rm UV} = 1.15 \cdot 10^{-28}\,{\rm M}_\odot{\rm yr}^{-1}{\rm erg}^{-1}$ s Hz, following G. Sun and S. R. Furlanetto (2016)[9]. UV luminosities are then related to magnitudes using the AB magnitude relation (Oke and J. E. Gunn, 1983): $\log_{10}\left(\frac{L_{\rm UV}}{{\rm ergs}^{-1}{\rm Hz}^{-1}}\right) = 0.4 \times (51.63 - M_{1500})$. UV LFs are very useful in anchoring our SFR relations (i.e. the ratio f_*/t_*), using the more massive reionization-era galaxies bright enough to be observed directly with the *Hubble* (and eventually *JWST*) telescope.

4. *The CMB optical depth* – the Thomson scattering optical depth of CMB photons as computed by Planck Collaboration, Aghanim, Akrami, et al. (2020), $\tau_e = 0.0561 \pm 0.071$ (1σ). Although it is more accurate to directly forward model the CMB EE power spectra, Qin, Poulin, et al. (2020) show that computing the likelihood from τ_e (a compressed summary statistic of the CMB power spectra) does not notably impact the resulting posterior for the astrophysical model used here.

These four complementary datasets are used in all of our inferences, each contributing a factor in the final likelihood. We write out explicitly all likelihood terms in Appendix 2.8. For further details, we refer the interested reader to Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. (2021).

2.4 What Do We Learn from the Patchy kSZ Signal?

We now explore what astrophysical constraints can be obtained from reionization observations, including the recent kSZ measurement (C. L. Reichardt, Patil, et al., 2021). We first discuss our Bayesian sampler and the set up of our forward models, before showing results using current and future kSZ measurements.

---

later reionization than shown here, also consistent with the newest analysis of the Lyman alpha forest spectra (Qin et al. in prep) as well as the forest dark fraction (Jin et al. 2023; Campo et al. in prep). We aim to revisit this in future work when more of these new constraints become public.

[9] This value was obtained assuming a stellar metallicity of $Z_* = 10^{-0.15z} Z_\odot$ and a Salpeter initial mass function (see also Madau and Dickinson (2014)).

2.4.1 *Inference Set-Up* To perform Bayesian inference, we use `21cmMC`[10] (Greig and Andrei Mesinger, 2015; Greig and Andrei Mesinger, 2018), a public Monte Carlo sampler of `21cmFAST`. For each set of model parameters (see section 2.2.1), `21cmMC` computes a 3D lightcone realization of cosmological fields, comparing the model to the observations (see sections 2.2.2 and 2.3). Here we use the MultiNest (Feroz, Hobson, and Bridges, 2009; Qin, Andrei Mesinger, Greig, et al., 2021) sampler, which is fully implemented in `21cmMC` and scales well to high-dimensional inference (e.g. Z. The HERA collaboration A. et al., 2022). We use 1000 live points, an evidence tolerance of 0.5 and a sampling efficiency of 0.8. We checked for convergence by launching a run with 2000 live points and found no significant difference in the inferred posterior distributions. Our fiducial posterior converges after $\sim$ 45k samples, taking $\sim$ 260k core hours.

Unfortunately, due to computational limitations, we cannot use ultra-large simulations (e.g. Fig. 2.1) when forward modeling. Instead we use smaller boxes, calibrating their output to account for the missing large-scale modes in the kSZ signal (see also Iliev, Pen, et al. 2007; L. D. Shaw, Rudd, and Nagai 2012; H. Park et al. 2013; Alvarez 2016). Specifically, we use simulations of $(500\,\mathrm{Mpc})^3$ on a 256^3 grid. When constructing the lightcones, we rotate the coeval boxes to minimize duplication of structures due to periodic boundary conditions (e.g. Andrei Mesinger, McQuinn, and David N. Spergel 2012). We account for the missing large-scale power by sampling several realizations (different cosmic seeds) of 500 Mpc boxes, and comparing their power spectra to those from 1.5 Gpc boxes, constructed using the same astrophysical parameters. We compute the mean ratio of the missing power, $f_{0.5\mathrm{Gpc}}^{1.5\mathrm{Gpc}} \equiv \mathcal{D}_{3000}^{\mathrm{pkSZ},1.5\mathrm{Gpc}}/\mathcal{D}_{3000}^{\mathrm{pkSZ},500\mathrm{Mpc}}$, adjusting our forward models by this factor and including the corresponding variance in the denominator of the likelihood. We obtain $f_{0.5\mathrm{Gpc}}^{1.5\mathrm{Gpc}} = 1.27 \pm 0.19$. Further details on this calibration procedure can be found in Appendix 2.7.

2.4.2 *Inference Results Using the Recent SPT Measurement* We compute two posteriors:

- *without kSZ* – this corresponds to the posterior based on the observational data (i)–(iv) from the previous section, i.e. large-scale Lyα forest opacity PDFs, the forest dark fraction, UV LFs, and the CMB optical depth.[11]
- *with kSZ* – this is the same as *without kSZ*, but including an additional factor in the likelihood, $\mathcal{L}_{\mathrm{kSZ}}$ (see Appendix 2.7 for details) corresponding to the patchy kSZ measurement by C. L. Reichardt, Patil, et al. (2021), adjusted for the slightly different lower redshift bound as discussed above: $\mathcal{D}_{3000}^{\mathrm{pkSZ}} = 1.2^{+1.0}_{-0.7}\,\mu\mathrm{K}^2$ (68% C.I.).

Comparing the *without kSZ* and *with kSZ* posteriors, we quantify the additional constraining power provided by the patchy kSZ. We begin by showing the constrains on the fundamental galaxy parameters, before discussing the corresponding derived quantities such as the EoR history and the halo-galaxy connection.

Galaxy Parameters and EoR History In the bottom left of Fig. 2.2, we show the resulting two- and one-dimensional posteriors *without kSZ* (blue) and *with kSZ* (red). We also show the model posteriors together with two of the observational data used in the likelihood: the $l = 3000$ patchy kSZ power (*top center*), and the UV LFs at $z = 6$, 8, 10 (*top right*).

From the $\mathcal{D}_{3000}^{\mathrm{pkSZ}}$ PDFs shown in the top center panel, we see that the recent measurement by C. L. Reichardt, Patil, et al. (2021) is in mild tension with the *without kSZ* posterior: the kSZ data favor the low amplitude tail of the *without kSZ* posterior, corresponding to late reionization models. Indeed, by including the kSZ measurement, the distribution is shifted in favor of smaller kSZ power. Most of the *with kSZ* posterior is still above the mean estimate of the kSZ power by C. L. Reichardt, Patil, et al. (2021), though perfectly consistent given the large observational uncertainty.

The biggest difference between the two galaxy parameter posteriors is in the recovered ionizing escape fraction, parametrized in our model with $f_{\mathrm{esc},10}$ and α_{esc} (see Eq. 2.6). The SPT measurement favors slightly lower values of $f_{\mathrm{esc},\,10}$ and higher values of α_{esc}. As a result, the inferred ionizing efficiency slightly increases in more massive, late-appearing galaxies (discussed further in the following section), so that the EoR occurs later and more rapidly, as can be seen from the EoR histories shown in Fig. 2.3. Including the relatively low patchy kSZ amplitude claimed by C. L.

---

[10] Available at `https://github.com/21cmfast/21CMMC`.

[11] Even though we used the same parametrization and observational data as Qin, Andrei Mesinger, Sarah E. I. Bosman, et al., 2021, our *without kSZ* posterior distribution is slightly different. This is because here we use the ionizing photon conservation correction from J. Park, Greig, and Andrei Mesinger (2022), which results in roughly a shift of 0.2 in the recovered α_{esc} (as also shown in J. Park, Greig, and Andrei Mesinger 2022). When computing the Lyman alpha forest we use a harder UV background (with energy index $\beta_{\mathrm{uv}} = -2$ instead of -5) and a higher post-ionization front temperature ($T_{\mathrm{re}} = 2.0 \times 10^4$K instead of 1.0×10^4K), motivated by recent estimates from hydrodynamic simulations (e.g. D'Aloisio et al. 2019b). The harder UV background shifts the end of reionization to slightly earlier times, compared with Qin, Andrei Mesinger, Sarah E. I. Bosman, et al., 2021.

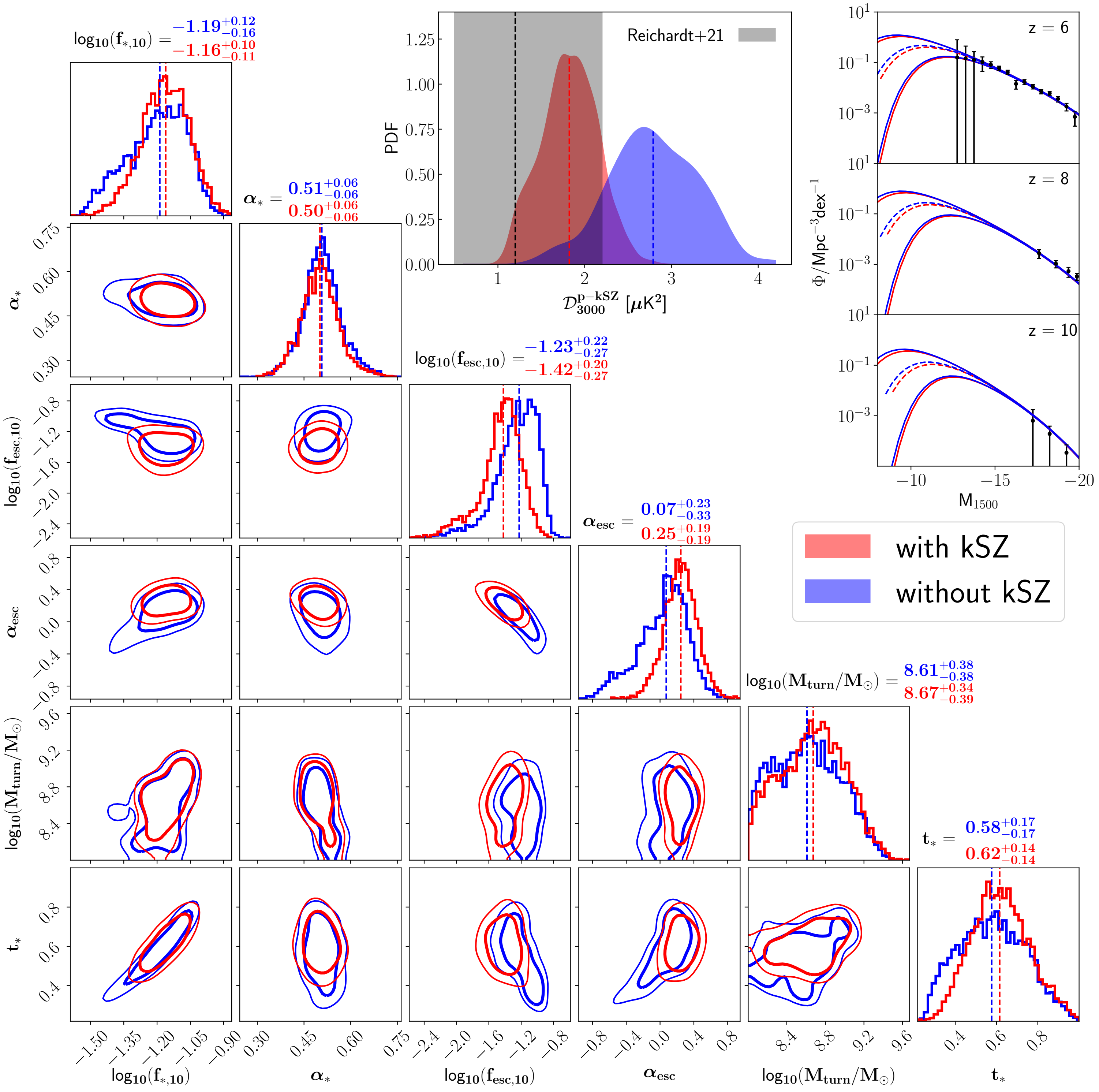

Figure 2.2: Marginalized posteriors of *without kSZ* (*blue*) and *with kSZ* (*red*). As discussed in the text, *without kSZ* is constrained using large-scale Lyα forest opacity PDFs, the forest dark fraction, UV LFs, and the CMB optical depth, while *with kSZ* additionally includes the recent measurement of the patchy kSZ power at $l = 3000$. The 1D and 2D posterior distributions of the model parameters are shown in a corner plot on the left, with thin and thick lines representing 95% and 68% credible intervals (C.I.) respectively. The marginalized median values (shown also as dashed lines) with the 68% central C.I. are given over the 1D distribution functions for the two runs. In the upper middle panel we show the PDFs of the patchy kSZ signal power spectrum at $l = 3000$, together with the C. L. Reichardt, Patil, et al. (2021) observational estimate in grey. Also shown are the median and [14, 86]% C.I. (dashed and solid lines respectively) of the inferred UV luminosity functions at $z = 6$, 8 and 10. Black points with error bars are UV LF observations used for the inference from R. J. Bouwens, Illingworth, Oesch, et al. (2015) and R. J. Bouwens, Oesch, I. Labbé, et al. (2016) and Oesch, R. J. Bouwens, et al. (2018a).

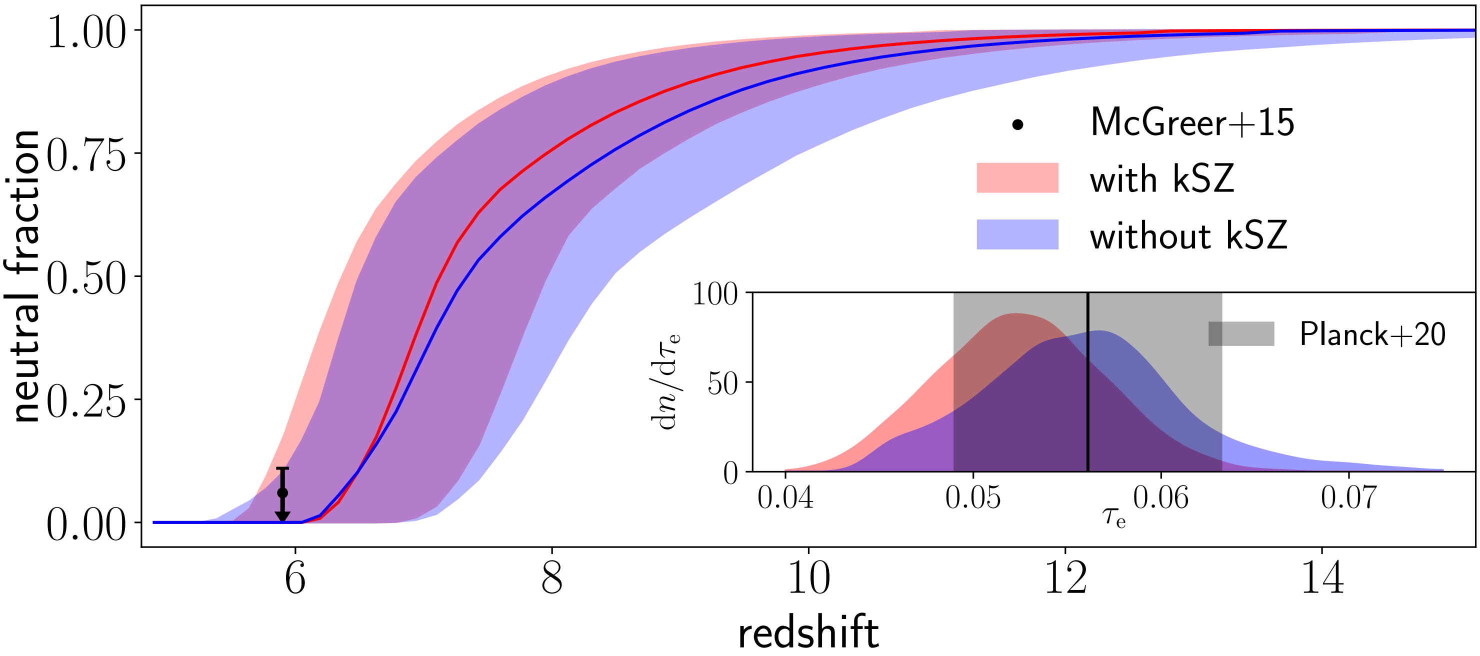

Figure 2.3: Median EoR histories with 95% C.I. for the *without kSZ* (*blue*) and *with kSZ* (*red*) posteriors. Also shown is the upper limit from McGreer, Andrei Mesinger, and D'Odorico (2015) at $z \sim 5.9$ that is used in the likelihood (Section 2.3) for both posteriors. The insert shows probability density distributions of τ_e for the *with kSZ* and *without kSZ* posteriors. The vertical black line and gray shaded region correspond to the Planck Collaboration, Aghanim, Akrami, et al. (2020) measurement of τ_e, also used in the likelihood for inferences. Including kSZ data increasing the preference for a later and more rapid reionization.

Reichardt, Patil, et al. (2021) disfavors the more extended EoR histories present in the *without kSZ* posterior. While the end of the EoR remains fairly unchanged, constrained by Lyman alpha forest observations (e.g. Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. 2021; T. Roy Choudhury, Paranjape, and Sarah E. I. Bosman 2021), the middle and early stages are shifted to later times with the addition of kSZ data. This translates into lower CMB optical depths as seen in the inset of Fig. 2.3: $\tau_e = 0.052^{+0.009}_{-0.008}$ for the *with kSZ* posterior compared to $\tau_e = 0.055^{+0.012}_{-0.009}$ for the *without kSZ* ome. The EoR histories implied by the *with kSZ* posterior are consistent with the Lyman α forest data, but in slight ($\lesssim 1\sigma$) tension with the CMB optical depth τ_e as well as the QSO dark pixel fraction. We note that an updated estimate of the QSO dark pixel fraction using more recent, much larger QSO samples from D'Odorico et al. (2023) results in weaker upper limits on the neutral fraction at $z \sim 6$, making them perfectly consistent with later EoR models (Campo et al. in prep). This would leave the CMB τ_e as the only dataset preferring a slightly earlier EoR. Such a mild tension between the two CMB datasets could come from calibration or analysis inconsistencies between large- and small-scale data, that is between the SPT and Planck data (e.g. Adélie Gorce, Marian Douspis, and Salvati, 2022).

On the other hand, constraints on parameters governing the star formation rates and stellar-to-halo mass relations (i.e. $f_{*,10}$, α_*, t_*, $M_{\rm turn}$) are fairly unchanged when including kSZ data. As already shown in J. Park, Andrei Mesinger, et al. 2019; Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. 2021, observed high-z UV LFs constrain the stellar-to-halo mass relation ($f_{*,10}$, α_*) and place an upper limit on a faint end turnover ($M_{\rm turn}$). Therefore, these parameters have only limited freedom to impact the timing of reionization while still being consistent with the UV LFs data.

Scaling Relations of EoR Galaxies To gain further insight into the implications of our results, we show the corresponding galaxy scaling relations in Fig. 2.4. In the top panel, we plot the inferred stellar-to-halo mass relation (SHMR), defined as the average stellar mass inside a halo of mass $M_{\rm h}$ (including the $f_{\rm duty}$ occupation fraction term from Eq. 2.5). The redshift-independent median relation from the *with kSZ* posterior is denoted with the red solid line, corresponding to:

$$\frac{\bar{M}_*}{M_h} = 0.011^{+0.003}_{-0.002} \left(\frac{M_{\rm h}}{10^{10}{\rm M}_\odot}\right)^{0.50^{+0.06}_{-0.06}} \qquad (68\%{\rm C.I.}), \tag{2.9}$$

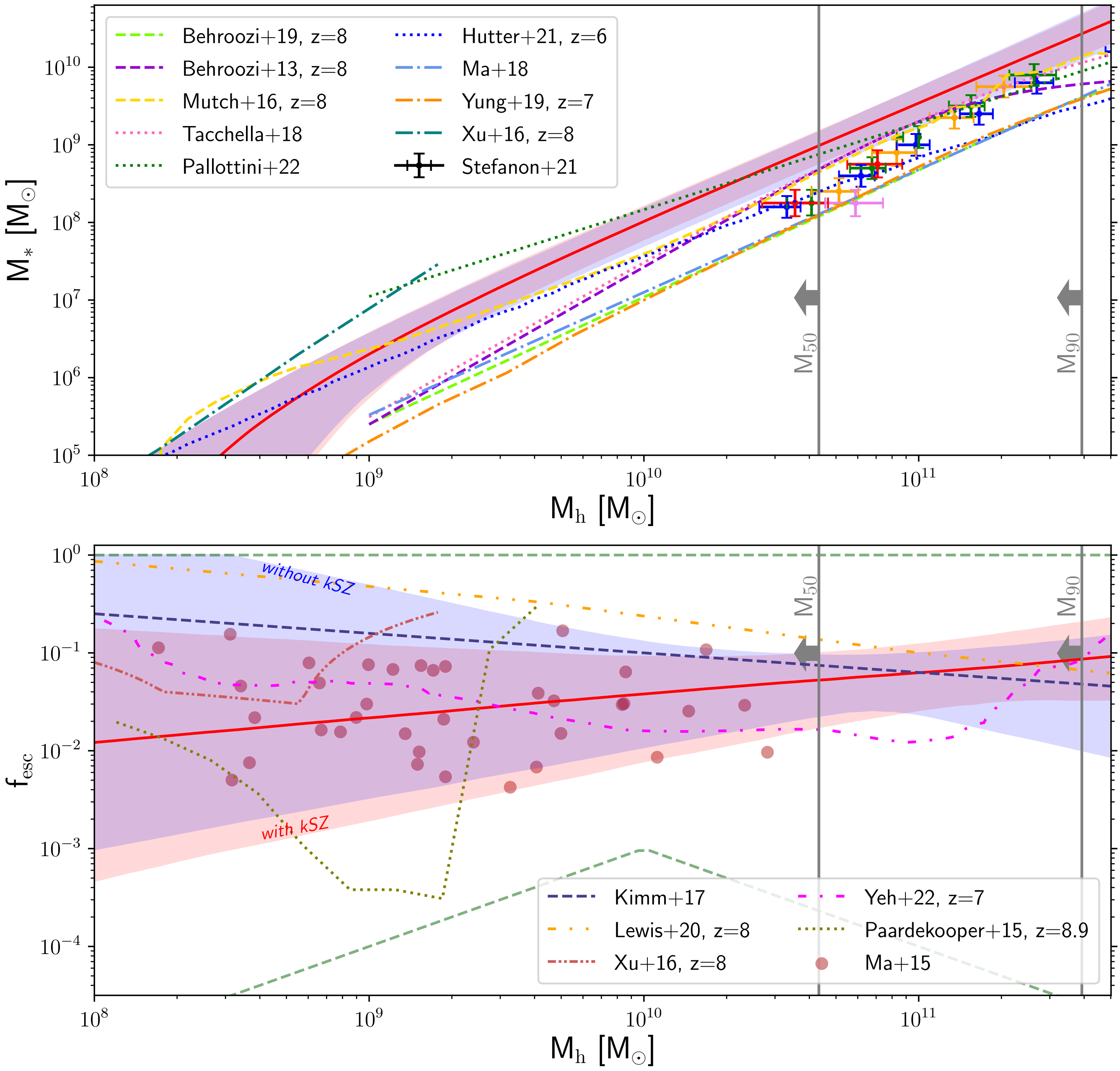

Figure 2.4: The dependence of average galaxy properties with halo mass. Shaded regions represent 95% C.I. of *without kSZ* (*blue*) and *with kSZ* (*red*). Vertical lines demarcate the posterior-averaged mean of M_{50} (M_{90}), defined as halo mass upper limit below which galaxies source 50% (90%) of the ionizing emissivity at $z = 7$. *Upper panel:* stellar to halo mass relation (SHMR). For illustrative purposes, we also show a selection of independent results from empirical and semi-analytic models (P. S. Behroozi, Wechsler, and Conroy, 2013; Mutch et al., 2016; Tacchella, Bose, et al., 2018; P. Behroozi et al., 2019; Yung, Somerville, Popping, et al., 2019; Hutter, Pratika Dayal, et al., 2021), as well as from cosmological, hydro simulations (H. Xu, Wise, et al., 2016; X. Ma, Hopkins, et al., 2018; A. Pallottini, A. Ferrara, et al., 2022). For Tacchella, Bose, et al. (2018) and X. Ma, Hopkins, et al. (2018), we use the redshift independent fits from their Eq. 12 and Eq. 1, respectively. The H. Xu, Wise, et al. (2016) result is taken for their Void region at z= 8, though the SHMR is similar for other environments and redshifts (see their fig. 16). The A. Pallottini, A. Ferrara, et al. (2022) curve is a linear fit to their data points. Colored points with error bars correspond to abundance matching estimates assuming a constant duty cycle (Mauro Stefanon, Rychard J. Bouwens, et al., 2021), where blue/green/orange/red/pink points are for z= 6/7/8/9/10 . *Lower panel:* the ionizing escape fraction. The green dashed lines demarcate the prior range. Again for illustrative purposes, we show estimates from some cosmological, hydrodynamic simulations (Paardekooper, Khochfar, and Dalla Vecchia, 2015; X. Ma, Kasen, et al., 2015; H. Xu, Wise, et al., 2016; Kimm, Katz, et al., 2017; Lewis, Ocvirk, Aubert, et al., 2020; Yeh et al., 2022). For Kimm, Katz, et al. (2017) we show their "fiducial" model. The X. Ma, Kasen, et al. (2015) points represents time-averaged escape fractions obtained using their SHMR relation (see their figures 3 and 9.)

while the 95% C.I. is shown with the purple shading.[12]. We also show the corresponding 95% C.I. of the *without kSZ* posterior in blue. The two posteriors overlap in this space, again illustrating that the SHMR for our model is determined by the UV LFs, and is unaffected by kSZ data. We also include some other estimates from the literature, which show sizable scatter for the high-redshift, small-mass regime that is relevant for the EoR. Our inferred relation is roughly consistent with current estimates, given their large scatter. It is unsurprising that, despite the fairly large scatter, the slopes of the SHMRs shown in this panel are roughly similar. This is because in most cases the observed UV LFs are used either directly or indirectly to calibrate the models. The slope of the UV LFs combined with the slope of the HMF, both power-laws in this range, sets the slope of the SHMR, with the normalization being more sensitive to the star formation – L_{1500} conversion[13].

Similarly, in the bottom panel of Fig. 2.4, we show the ionizing escape fraction to halo mass relation. Our redshift-independent median relation for *with kSZ* is denoted with a solid red line:

$$f_{\rm esc} = 0.038^{+0.021}_{-0.017} \left(\frac{M_{\rm h}}{10^{10}{\rm M}_\odot}\right)^{0.25^{+0.19}_{-0.19}} \qquad (68\%{\rm C.I.};\ \textit{with kSZ}), \tag{2.10}$$

Also shown is the result for *without kSZ* in blue:

$$f_{\rm esc} = 0.060^{+0.038}_{-0.028} \left(\frac{M_{\rm h}}{10^{10}{\rm M}_\odot}\right)^{0.07^{+0.23}_{-0.33}} \qquad (68\%{\rm C.I.};\ \textit{without kSZ}). \tag{2.11}$$

As in the panel above, the corresponding shaded areas demarcate the 95% C.I. The green dashed lines denote the range of our prior in this space, uniform over $\log_{10} f_{\rm esc,10} \in [-3, 0]$, $\alpha_{\rm esc} \in [-1, 0.5]$; the fact that our posterior is tighter than the prior illustrates the constraining power of current observations and that our results are not sensitive to our choice of prior.

Again for illustrative purposes, we show some theoretical estimates from the literature. Compared to the SHMR in the top panel, there is far less consensus on the ionizing escape fraction. This is because the relevant small scales are impossible to resolve in cosmological simulations; therefore results are sensitive to the resolution/sub-grid prescriptions. Indeed, some simulations suggest an increasing trend with halo mass while others suggest a decreasing trend.

In contrast, Bayesian inference allows the observations to inform us about the (mean) $f_{\rm esc}(M_{\rm h})$ relation. By comparing the blue and red shaded regions we see that the addition of kSZ data favors a slight increase in the mean escape fraction towards more massive halos. While the uncertainties are still large at the small mass end, the ionizing escape fraction for galaxies hosted by $\sim 10^{10} - 10^{11}\ M_\odot$ halos is reasonably well constrained to be a few percent. Interestingly, strong evolution with halo mass is disfavored.

In Fig. 2.4 we also demarcate the posterior-averaged mean of M_{50} (M_{90}), defined as halo mass upper limit below which galaxies source 50% (90%) of the ionizing emissivity at $z = 7$, i.e. M_{50} is calculated by solving the equation: $\int_0^{\rm M_{50}} {\rm dM_h}\ {\rm dn_{ion}/dM_h} = 1/2 \int_0^\infty {\rm dM_h}\ {\rm dn_{ion} dM_h}$. We see that over half of the ionizing photons are sourced by galaxies that are below current detection limits.

2.4.3 *Forecast Assuming Future kSZ Measurments*

The improved precision of future experiments, as well as their larger sky coverage, will allow for lower noise levels and decreased sample variance. With CMB-S4, we expect the errors on the measurement of the amplitude of the CMB temperature power spectrum at $l = 3000$ to decrease by a factor of 5 to 10, depending on the bandpower (Abazajian et al., 2016). Improved foreground modelling should also help reduce the uncertainty on the kSZ amplitude by roughly 30% (Adélie Gorce, Marian Douspis, and Salvati, 2022). On the theoretical side, suites of simulations could better characterize the contribution of the homogeneous kSZ to the total power.

To quantify the corresponding improvement in parameter constraints, we repeat the *with kSZ* inference in Sec. 2.4.2, but using a mock future kSZ measurement instead of C. L. Reichardt, Patil, et al. (2021). We assume $\mathcal{D}_{3000}^{\rm pkSZ} = 2.0 \pm 0.10\,\mu{\rm K}^2$. The mean value corresponds to the maximum a posteriori (MAP) model from the *with kSZ* posterior in the previous section, while the choice of uncertainty is (very roughly) motivated by the arguments above.

---

[12] We note that the inferred median scaling of the SHMR $\propto M_h^{0.5}$, is close to the value expected by simply assuming SNe feedback scales with the gravitational potential of the host halo, SHMR $\propto M_h^{0.67}$ (e.g. Wyithe and Abraham Loeb, 2013). As discussed further below, any mass dependence of $\mathcal{K}_{\rm UV}$ (here assumed to be constant) would also impact the inferred scaling.

[13] We caution that our SHMRs are likely overconstrained, because we do not include any uncertainty in the SFR – L_{1500} conversion (i.e. we fix $\mathcal{K}_{\rm UV}$ from Sec. 2.3 to be a constant). This conversion depends on the IMF and the duration of recent star formation, with different assumptions changing $\mathcal{K}_{\rm UV}$ by factors of ~ 2 (e.g. Wilkins, Lovell, and Elizabeth R. Stanway 2019; E. R. Stanway, Chrimes, et al. 2020).

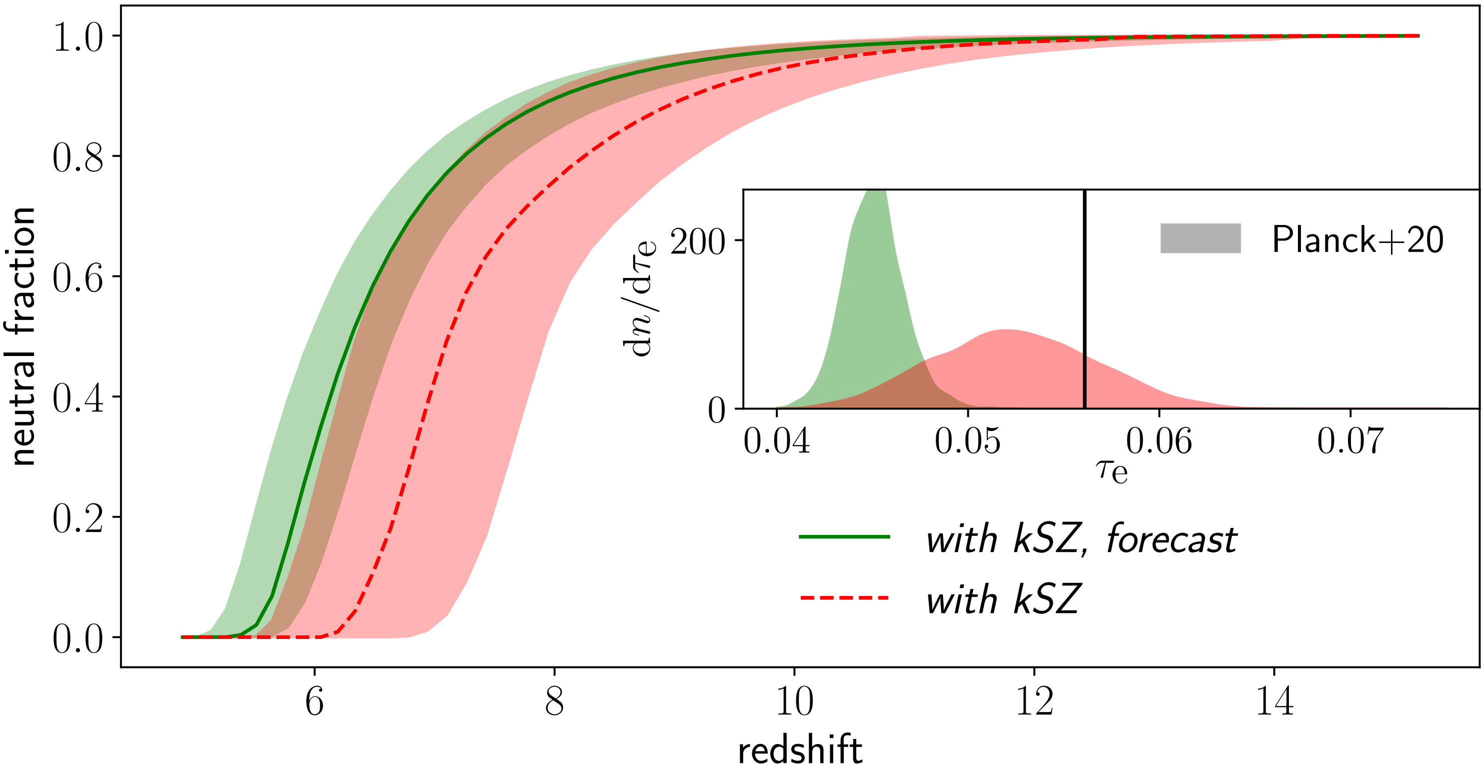

Figure 2.5: Median EoR histories with 95% C.I. for the *with kSZ* (*red*) posterior and new posterior with the forecast value of patchy kSZ amplitude: $\mathcal{D}_{3000}^{\rm pkSZ} = 2.0 \pm 0.10\,\mu{\rm K}^2$ (*green*). In the inset we also show the corresponding PDFs of $\tau_{\rm e}$.

In Figure 2.5 we show constraints on the EoR history using the mock kSZ observation (*green shaded region*), together with the current constraints (*red shaded region*). We see that if the kSZ error bars could be reduced by a factor of ~ 10, it would result in a dramatic improvement on the recovered EoR history, with the midpoint of reionization being constrained to an r.m.s. uncertainty of $\sigma_{z_r} = 0.16$, compared to 0.4 for the current *with kSZ* posterior in red. A similar improvement is also obtained for the duration of EoR: $\Delta_z \equiv z(\overline{x}_{\rm H} = 0.75) - z(\overline{x}_{\rm H} = 0.25)$. Using the mock kSZ observation we recover $\Delta_z = 1.09^{+0.12}_{-0.09}$, which compared to the current *with kSZ* constraints of $\Delta_z = 1.16^{+0.24}_{-0.19}$, reduces the uncertainty by a factor of ~ 2.

It is interesting to note that the change in the recovered history is primarily in delaying reionization; the duration decreases only marginally. Because galaxies sit inside halos, the duration of reionization cannot be arbitrarily short; it will be limited by the growth of the HMF. The most rapid EoR models are those dominated by the rare, bright galaxies hosted by massive halos in the exponential tail of the HMF. Their fractional abundance increases more rapidly compared to that of the more common, smaller halos. However, the observed UV LFs set a lower limit on Δ_z because we actually see galaxies down to $M_{1500} \sim -13$, and the rare bright galaxies cannot have $f_{\rm esc} > 1$. Since we cannot physically decrease Δ_z to values below unity, the only physically plausible way of decreasing the kSZ amplitude to agree with the mock observation is to lower the redshift of reionization [14]. This is seen in the figure, and it causes the CMB τ_e PDF to pile up in the lower $\sim 30\%$ C.I. inferred from *Planck* alone. Interestingly this later EoR history is in very good agreement with the latest, independent estimates coming from the Lyman alpha forest opacity fluctuations which imply that EoR finishes at $z \approx 5.3$ (Qin et al., in prep).

---

[14] This picture can be changed somewhat if additional ionizing sources are present, such as AGN. A rapid increase in the number density of AGN could make reionization end somewhat more abruptly. However recent estimates imply AGN cannot contribute more than $\sim 10\%$ to the EoR (e.g. Qin, Mutch, et al. 2017; Harikane, Zhang, et al. 2023), so we expect their eventual impact to be modest. We also note that our escape fraction parametrization effectively captures EoR histories in which AGN contribute significantly through high values of the $\alpha_{\rm esc}$ parameter. In future work we will further increase the flexibility of our model, allowing for an explicit contribution from AGN.

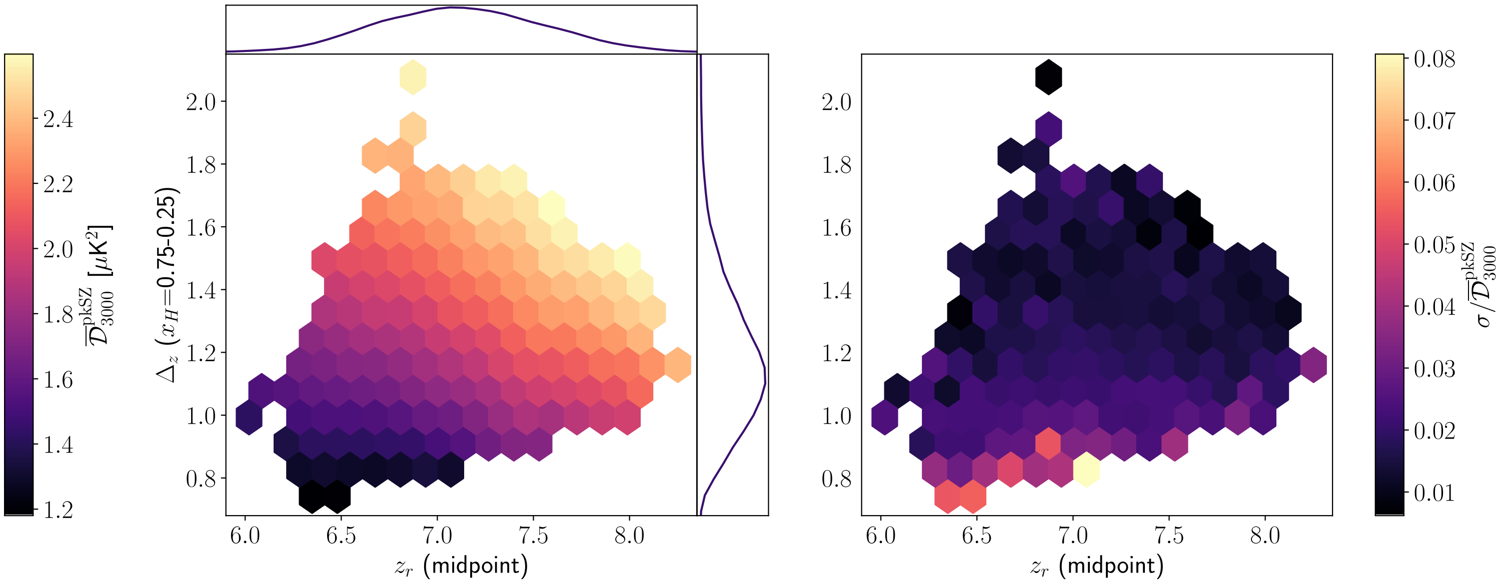

Figure 2.6: *Left panel:* Mean value of the patchy kSZ power spectrum amplitude at $l = 3000$ binned as a function of z_r (midpoint of reionization) and Δ_z (duration of reionization; $\Delta_z \equiv z(\overline{x}_{\rm H} = 0.75) - z(\overline{x}_{\rm H} = 0.25)$). The samples are taken from the *with kSZ* run. Plotted are the bins for which the scatter in the bin is larger than the uncertainty of the mean. 1D posterior distribution of Δ_z and z_r are shown on the sides. *Right panel:* Standard deviation of the patchy kSZ power spectrum at $l = 3000$ within each bin divided by the mean for the same bin.

2.5 Do We Need Self Consistent Forward Models of the kSZ?

When interpreting kSZ observations, it is common to vary the amplitude of the patchy kSZ power but with a fixed power spectrum shape (e.g. O. Zahn et al. 2012; Battaglia et al. 2013; C. L. Reichardt, Patil, et al. 2021). Generally the power at $l = 3000$, $\mathcal{D}_{3000}^{\rm pkSZ}$, is related to empirical parameters characterizing the EoR history, such as its midpoint and duration (with several definitions found in the literature). This is in contrast to our approach in which the patchy kSZ power spectra are self-consistently forward-modeled directly from galaxy properties.

Using only empirical parameters for the EoR history has two important drawbacks: (i) for a given EoR history, the patchy kSZ power can also vary due to the EoR *morphology* (Andrei Mesinger, McQuinn, and David N. Spergel, 2012; Battaglia et al., 2013; A. Gorce, Ilić, et al., 2020; Paul, Mukherjee, and Tirthankar Roy Choudhury, 2021; Tirthankar Roy Choudhury, Mukherjee, and Paul, 2021; N. Chen et al., 2022); and (ii) it is more difficult to physically-motivate priors for *derived* EoR history parameters, than it is for the fundamental galaxy parameters (e.g. Qin, Poulin, et al., 2020). The choice of priors is especially important when the likelihood is not overly constraining (e.g. Trotta 2017; Efstathiou 2021).[15] Here we briefly explore the impact of (i). We sample our *with kSZ* posterior from Section 2.4.2, computing for each sample the midpoint of the EoR, z_r, and its duration, Δ_z. In Fig. 2.6 we plot the mean (*left panel*) $\bar{\mathcal{D}}_{3000}^{\rm pkSZ}$ and normalized r.m.s. (*right panel*) of the $l = 3000$ patchy kSZ, as a function of z_r and Δ_z. We leave blank under-sampled bins of (z_r, Δ_z), defined as those for which the variance of the mean is larger than the mean of the variance.

We note that our estimates of $\bar{\mathcal{D}}_{3000}^{\rm pkSZ}$ are tens of percent higher compared to some recent estimates, for a given combination of z_r and Δ_z (Adélie Gorce, Marian Douspis, and Salvati, 2022; N. Chen et al., 2022). This might be in part due to different sampling of EoR morphologies, or to differences in how the patchy kSZ power is defined. Indeed, we estimate a difference of $\Delta z = 1$ on the lower redshift bound of the integral in equation 2.1 to result in a $\sim 0.2\,\mu{\rm K}^2$ difference in the patchy kSZ amplitude. Another potential source of disagreement could stem from using the Limber approximation to compute the patchy kSZ spectrum from the power spectrum of the density-weighted peculiar velocity field of the free electrons (e.g. C.-P. Ma and Fry 2002; A. Gorce, Ilić, et al. 2020; Paul, Mukherjee, and Tirthankar Roy

---

[15] In the previous section, we showed that our likelihood was indeed quite constraining, and therefore our posterior was not sensitive to our choice of priors. This is because we use several complementary EoR and galaxy observations to construct the likelihood. However, when only using the kSZ observation and ignoring for example the UV LFs, the likelihood is not overly constraining and the posterior can strongly depend on the choice of priors over the EoR history parameters (e.g. Greig and Andrei Mesinger, 2017; J. Park, Andrei Mesinger, et al., 2019; Z. The HERA collaboration A. et al., 2022).

Choudhury 2021; Tirthankar Roy Choudhury, Mukherjee, and Paul 2021), rather than ray-tracing the signal. However, we find a good agreement between the two approaches on the scales of interest ($l \gtrsim 1000$).We also compute the slope of the $\bar{\mathcal{D}}_{3000}^{\rm pkSZ}$ - Δ_z relation using the full posterior sample, finding values that are roughly 15% larger than Battaglia et al. (2013) and N. Chen et al. (2022). A more detailed comparison with other analysis is not possible given the differences in modelling, definitions, and EoR parameters, and would require a dedicated study.

In the left panel we also show the marginalized 1D PDFs of z_r (*top*) and Δ_z (*right*). Our *with kSZ* posterior corresponds to the following constraints on the EoR history parameters: $z_r = 7.12^{+0.44}_{-0.41}$ and $\Delta_z = 1.16^{+0.24}_{-0.19}$ (68% C.I.). As noted in previous studies, there is a strong degeneracy between z_r and Δ_z, as either a later or a shorter EoR decreases the patchy kSZ power. Our median recovered values of Δ_z are consistent with those from other recent analyses of the SPT observation, including C. L. Reichardt, Patil, et al. (2021) who found $\Delta_z = 1.1^{+1.6}_{-0.7}$ (68% C.I.) and Tirthankar Roy Choudhury, Mukherjee, and Paul (2021) who found $\Delta_z = 1.30^{+0.19}_{-0.60}$. Our limits are however ~ 3 times tighter compared to C. L. Reichardt, Patil, et al. (2021) since we use additional, complementary observations in the likelihood.

In the right panel of Fig. 2.6 we quantify the scatter in the $l = 3000$ patchy kSZ power, *at fixed values of* z_r *and* Δ_z. We see that the r.m.s. scatter in the power is generally at the level of a few percent. Thus varying the $l = 3000$ kSZ power amplitude at a function of only z_r and Δ_z, without considering the EoR morphology, cannot yield an accuracy on $\bar{\mathcal{D}}_{3000}^{\rm pkSZ}$ better than $\sim$ few percent. We stress also that this is a conservative estimate, since we only compute the scatter in the kSZ power for our relatively narrow *with kSZ* posterior. Studies that do not consider complementary EoR and galaxy observations in the likelihood would result in broader posteriors with correspondingly larger scatter in the mean power. Indeed, by sampling a broader range of models, Paul, Mukherjee, and Tirthankar Roy Choudhury (2021) find a larger r.m.s. scatter, of order $\sim 0.4\mu\mathrm{K}^2$ for the kSZ power at a fixed z_r and Δ_z.

It is important to note that without using complimentary observations in the likelihood, both the distributions of (z_r, Δ_z) seen in the left panel and the scatter in the kSZ power at a fixed EoR history seen in the right panel would be considerably broader. As noted earlier, the current SPT detection is low S/N and by itself is very not constraining. Only in combination with complimentary observations can we obtain tight constraints on the EoR history and not be sensitive to our choice of priors.

2.6 Conclusions

The patchy kSZ signal is an integral probe of the timing and morphology of the EoR. Recently, C. L. Reichardt, Patil, et al. (2021) have claimed a detection of the patchy kSZ signal ($\mathcal{D}_{3000}^{\rm pkSZ} = 1.1^{+1.0}_{-0.7}\,\mu\mathrm{K}^2$). In the future, we expect a dramatic increase in S/N from telescopes such as CMB-S4 and Simons Observatory enhancing the potential of using kSZ measurements for EoR science.

In this work we quantify what we can learn about the EoR from the patchy kSZ signal. We modify the public `21cmFAST` code to produce forward-models of the patchy kSZ signal. We then perform Bayesian inference by sampling galaxy properties and using the recent kSZ measurement together with other observations in the likelihood. These include: (i) high-z UV LFs; (ii) Lyα forest opacity distributions; (iii) the Lyman forest pixel dark fraction and (iv) CMB optical depth.

In order to quantify the additional constraining power of the patchy kSZ we computed two posteriors: one based on Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. (2021) (using datasets (i)–(iv); *without kSZ*) and one with an additional likelihood term for the recent measurement of the patchy kSZ cited above (*with kSZ*). We found that the addition of the kSZ measurement shifts the posterior distribution in favor of faster and later reionization models (Fig. 2.2). This results in a lower optical depth to the CMB: $\tau_e = 0.052^{+0.009}_{-0.008}$ (68% C.I.).

The shift to later and more rapid EoR implies a lower ionizing escape fraction with a very weak positive scaling with halo mass. The average $f_{\rm esc}$ of typical galaxies driving the EoR is a few percent. We disfavor a strong evolution of $f_{\rm esc}$ with galaxy mass.

We also present constraints on common empirical parameters characterizing the midpoint and duration of reionization, respectively $z_r = 7.10^{+0.44}_{-0.41}$ and $\Delta_z = 1.16^{+0.24}_{-0.19}$ (68% C.I.), consistent with other recent results (C. L. Reichardt, Patil, et al., 2021; Tirthankar Roy Choudhury, Mukherjee, and Paul, 2021). We show that the scatter in patchy kSZ power at $l = 3000$, *at a fixed* z_r *and* Δ_z, is of order $\sim$ few percent. Thus the interpretation of current kSZ data can be done using only these two summary statistics. However, without a physical model it would be difficult to assign prior probabilities or use complimentary observations in the likelihood.

Future observations should further improve the measurement of the patchy kSZ signal (Abazajian et al., 2016). To forecast the resulting improvement in parameter constraints, we also create a mock observation with the measurement

Table 2.1: Astrophysical parameters used for the scaling test. $f_{0.5\text{Gpc}}^{1.5\text{Gpc}}$ is the power spectrum scaling factor (see text). Note that the parameter combination 2 is used in Fig. 2.1.

	$\log_{10}(f_*)$	α_*	$\log_{10}(f_{\text{esc}})$	α_{esc}	$\log_{10}(M_{\text{turn}}/M_\odot)$	t_*	$\mathcal{D}_{3000}^{\text{pkSZ}}\,[\mu\text{K}^2]$	$f_{0.5\text{Gpc}}^{1.5\text{Gpc}}$
1	-1.437	0.559	-1.239	0.093	8.515	0.332	1.522	1.176
2	-1.416	0.614	-1.780	0.474	8.622	0.392	1.076	1.132
3	-1.498	0.493	-1.201	0.175	8.668	0.282	1.469	1.132
4	-1.144	0.477	-1.577	0.209	8.787	0.591	1.442	1.136

error reduced to $0.1\mu K^2$, centered on the MAP model from our inference. Such a futuristic observation can reduce the uncertainties on the recovered EoR history by factors of ∼2, 3. However, if the patchy kSZ power is confirmed to be low ($\mathcal{D}_{3000}^{\text{pkSZ}} \lesssim 2\,\mu\text{K}^2$), it would result in a mild tension with the CMB τ_e inferred from primary CMB anisotropies.

2.7 Calibrating Simulations to Account for Missing Large-Scale kSZ Power

As discussed in Section 2.2.1, large boxes are required to accurately simulate the patchy kSZ signal. L. D. Shaw, Rudd, and Nagai (2012) find that a simulation box of side length $100\,h^{-1}$Mpc would miss about 60% of the kSZ power. However, using large simulations which also resolve small-scale physics in forward modeling is computationally impractical. Although one could account for missing large-scale power analytically (H. Park et al., 2013; A. Gorce, Ilić, et al., 2020), such perturbative approaches are approximate and have only been tested with a few models. Instead, here we compute the kSZ signal directly from multiple, smaller-box realizations of the signal and statistically characterize the missing power comparing to a large-box realization (see also Iliev, Pen, et al., 2007).

We pick a random sample from the posterior distribution of Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. (2021), corresponding to the *without kSZ* posterior. For this set of astrophysical parameters, we compute the patchy kSZ power using a 1.5 Gpc simulation, run on a 1050^3 grid. We then generate 20 realizations of the smaller-box simulations used in our inference (500 Mpc on a 256^3 grid), using the same astrophysical parameters but varying the initial random seed. When constructing the lightcones using the 500 Mpc simulations, we rotate the coeval boxes to minimize duplication of structures due to periodic boundary conditions (e.g. Andrei Mesinger, McQuinn, and David N. Spergel 2012). We also performed a resolution check and found a negligible difference in the kSZ power with respect to the resolution. The resulting histogram of $\mathcal{D}_{3000}^{\text{pkSZ}}$ from the 500 Mpc simulations is shown in Fig. 2.7, together with the value from the 1.5 Gpc simulation (blue vertical line). We compute the ratio of the missing power as $f_{0.5\text{Gpc}}^{1.5\text{Gpc}} = \mathcal{D}_{3000}^{\text{pkSZ}-1.5\text{Gpc}}/\mathcal{D}_{3000}^{\text{pkSZ}-500\text{Mpc}}$. Using these 20 realizations, we find $f_{0.5\text{Gpc}}^{1.5\text{Gpc}} = 1.27 \pm 0.19$. We include this scaling factor and associated uncertainty in the likelihood when performing inference:

$$\ln \mathcal{L}_{\text{kSZ}} = -\frac{1}{2}\left(\frac{\mathcal{D}_{3000}^{\text{kSZ,SPT}} - \mathcal{D}_{3000}^{\text{kSZ,model}}}{\sigma_a + \sigma_b\left(\mathcal{D}_{3000}^{\text{kSZ,SPT}} - \mathcal{D}_{3000}^{\text{kSZ,mock}}\right)}\right)^2, \tag{2.12}$$

with $\sigma_a = 2\frac{\sigma_u\sigma_l}{\sigma_u+\sigma_l}$, $\sigma_b = \frac{\sigma_u-\sigma_l}{\sigma_u+\sigma_l}$. Here, σ_u and σ_l are upper and lower 68% C.I. limits of the measurement. Since we are adding the scaling factor uncertainty in the quadrature, the final expressions for σ_u and σ_l are $\sigma_u = \sqrt{\sigma_{u,SPT}^2 + \sigma_{f_{0.5\text{Gpc}}^{1.5\text{Gpc}}}^2}$ and $\sigma_l = \sqrt{\sigma_{l,SPT}^2 + \sigma_{f_{0.5\text{Gpc}}^{1.5\text{Gpc}}}^2}$, where the measurement is expressed as $(\mathcal{D}_{3000}^{\text{kSZ,SPT}})_{-\sigma_{l,SPT}}^{+\sigma_{u,SPT}} = 1.1_{-0.7}^{+1.0}\mu\text{K}^2$. Log-likelihood written in equation (2.12) is a Gaussian whose width depends on the parameter value and it's used for the asymmetric statistical errors of the measurement (see e.g. Barlow, 2004).

Note that since we are varying the seed, the variance in $f_{0.5\text{Gpc}}^{1.5\text{Gpc}}$ also includes the Poisson uncertainty on the mean power, stemming from the fact that the power is estimated from a finite number of wavemodes. The later is illustrated as a solid black segment in Fig. 2.7, and has a subdominant contribution to the scatter in $\mathcal{D}_{3000}^{\text{pkSZ}}$ from the 500 Mpc simulations.

How much does the scaling factor, $f_{0.5\text{Gpc}}^{1.5\text{Gpc}}$, depend on the choice of astrophysical parameters? Unfortunately, it would be computationally impractical to repeat the above calibration procedure over our entire 6D astrophysical parameter space. Instead we sample four different astrophysical parameters from the *without kSZ* posterior, and compute $f_{0.5\text{Gpc}}^{1.5\text{Gpc}}$

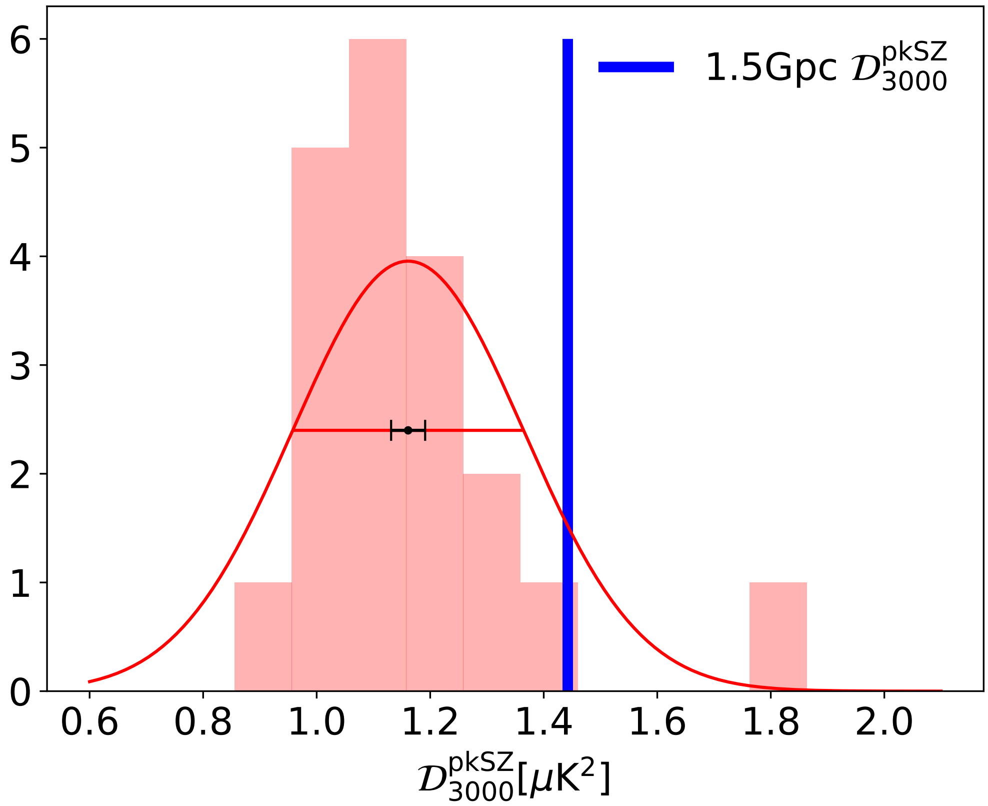

Figure 2.7: Histogram of the $l = 3000$ patchy kSZ amplitudes generated from 500 Mpc boxes by varying the cosmic initial seed (see text for details). The solid red line corresponds to a Gaussian fit to the histogram, with the standard error σ denoted as a solid line. The vertical blue line denotes the value for the 1.5 Gpc box. We see that the small boxes on average underestimate the kSZ power at $l = 3000$ by $\sim 20\%$, though with sizable scatter. For illustrative purposes, we also demarcate with the black segment the size of the 1 σ Poisson uncertainty on the mean power, arising from sampling the power spectrum with a limited number of modes for a 500 Mpc box. We see that the cosmic variance from varying the seed is much larger than the Poison sample variance.

using a single 1.5 Gpc and 500 Mpc simulation for each parameter set. The parameters and corresponding scale factors are listed in the Table 2.1. Reassuringly the scaling factors vary by only $\sim 3\%$ between the four parameter combinations. This is much smaller than the $\sim 0.7\mu\text{K}^2$ measurement uncertainty, justifying our assumption of a constant $f_{0.5\text{Gpc}}^{1.5\text{Gpc}}$.

2.8 Likelihoods Used for Inference

In Section 2.3 we introduce various EoR observations that we used to perform inference of astrophysical parameters (θ). Here we write out likelihoods for those observations, used in Section 2.4.2:

- Lyman α forest opacity distributions - The log likelihood for one redshift, z, and effective optical depth, τ_{eff}, bin is assumed to be Gaussian:

  $$\ln \mathcal{L}_{\alpha,z,\tau_{\text{eff}}}(\theta) = -0.5 X^{\text{T}} \Sigma^{-1} X, \tag{2.13}$$

  where X is the difference between the model and the observed effective optical depth PDF in that bin and Σ is the total error covariance matrix. More details about the covariance matrix can be found in appendices of Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. (2021). The total log-likelihood for the Lyman α forest is the sum over redshift and optical depth bins:

  $$\ln \mathcal{L}_{\alpha}(\theta) = \sum_{z} \sum_{\tau_{\text{eff}}} \ln \mathcal{L}_{\alpha,z,\tau_{\text{eff}}}(\theta) \tag{2.14}$$

for $z \in \{5.4, 5.6, 5.8, 6.0\}$. and $\tau_{\mathrm{eff}} < 8$.

- Dark fraction in the Ly α and Lyβ forests - the log likelihood is given as:

$$\ln \mathcal{L}_{\mathrm{DF}}(\theta) = \begin{cases} 0 & \text{if } \overline{x}_{\mathrm{HI,z}}(\theta) \leq 0.06 \\ -\frac{1}{2}\frac{\left(\overline{x}_{\mathrm{HI,z}}(\theta)-0.06\right)^2}{\sigma_{\mathrm{DF}}^2} & \text{otherwise} \end{cases} \tag{2.15}$$

  where $\overline{x}_{\mathrm{HI,z}}$ is the modelled neutral fraction at $z = 5.9$ and $\sigma_{\mathrm{DF}} = 0.05$.

- High-redshift galaxy UV luminosity functions (UV LFs) - The log likelihood is the sum over redshifts and magnitudes given by R. J. Bouwens, Illingworth, Oesch, et al. (2015) and R. J. Bouwens, Oesch, I. Labbé, et al. (2016) and Oesch, R. J. Bouwens, et al. (2018a):

$$\ln \mathcal{L}_{\mathrm{LF}}(\theta) = -0.5 \sum_{z} \sum_{M_{\mathrm{UV}}} \left(\frac{\phi_{\mathrm{LF,model}}(\theta, z, M_{\mathrm{UV}}) - \phi_{\mathrm{LF,obs}}(z, M_{\mathrm{UV}})}{\sigma_{\mathrm{LF}}(z, M_{\mathrm{UV}})} \right)^2 . \tag{2.16}$$

  Here $\phi_{\mathrm{LF,model}}$ is the modelled luminosity function at a given redshift and magnitude and $\phi_{\mathrm{LF,obs}}$ is the observed one, with the corresponding uncertainties, σ_{LF}. Summation over redshifts is done for $z \in \{6, 7, 8, 10\}$.

- The CMB optical depth - The log likelihood for the CMB optical depth is given as:

$$\ln \mathcal{L}_{\tau_{\mathrm{e}}}(\theta) = -\frac{1}{2} \left(\frac{\tau_{\mathrm{e,model}}(\theta) - \tau_{\mathrm{e,obs}}}{\sigma_{\tau_{\mathrm{e}}}} \right)^2 , \tag{2.17}$$

  where $\tau_{\mathrm{e,model}}$ is the modelled CMB optical depth and $\tau_{\mathrm{e,obs}} = 0.0561$ is the observed one from Planck Collaboration, Aghanim, Akrami, et al. (2020) with $\sigma_{\tau_{\mathrm{e}}} = 0.071$ the corresponding uncertainty.

3

Mapping Reionization Bubbles in JWST Era II: Inferring the Position and Characteristic Size of Individual Bubbles

3.1 Introduction

The Epoch of Reionization (EoR) marks an important milestone in the Universe's evolution. UV radiation from the first, clustered galaxies reionized their surrounding intergalactic medium (IGM). These HII "bubbles" expanded, percolated, and eventually permeated all of space, completing the final phase change of our Universe. The timing and morphology of the EoR tell us which galaxies were responsible as well the role of IGM clumps that regulated the end stages (e.g., McQuinn, Lidz, et al., 2007; Emanuele Sobacchi and Andrei Mesinger, 2014).

Lyman α emission from galaxies is an especially useful tool for studying the early stages of the EoR, when HII regions are relatively small such that the neutral IGM leaves a strong imprint via damping wing absorption (e.g., see review in Dijkstra, 2014). A common approach is to estimate the mean neutral fraction of the IGM using a statistically large sample of galaxies (e.g., Andrei Mesinger and Steven R. Furlanetto 2008b; Stark, Ellis, Chiu, et al. 2010; Andrei Mesinger, Aykutalp, et al. 2015; C. A. Mason, Tommaso Treu, et al. 2018; Jung, Steven L. Finkelstein, Dickinson, et al. 2020; Bolan, Lemaux, et al. 2022; Jones, Bunker, Saxena, Witstok, et al. 2024; Nakane et al. 2024). However, galaxies reside in biased regions of the IGM, and connecting the corresponding damping wing signature to the mean neutral fraction is very model dependent (e.g., Andrei Mesinger and Steven R. Furlanetto, 2008a; Lu, C. A. Mason, Hutter, et al., 2024). In contrast, unbiased probes such as the Lyman alpha forest are sourced from representatively-large volumes of the IGM, and can already tightly constrain the mean neutral fraction during the latter half of the EoR (Qin, Andrei Mesinger, Sarah E. I. Bosman, et al., 2021; Sarah E. I. Bosman, F. B. Davies, et al., 2022; Qin, Andrei Mesinger, Prelogović, et al., 2024).

In addition to estimating the *global* neutral fraction from the IGM Lyα damping wings, one could instead infer the presence (or lack thereof) of an *individual* HII region surrounding an observed group of galaxies (Tilvi, Malhotra, et al., 2020; Endsley and Stark, 2022; Jung, Steven L. Finkelstein, R. L. Larson, et al., 2022; Hayes and Scarlata, 2023). This could potentially allow us to connect the growth of the local HII region to the properties of the galaxies inside it. Having several such estimates of HII bubble sizes will allow us to understand which kind of galaxies drove reionization (e.g., faint/bright; McQuinn, Lidz, et al., 2007; Andrei Mesinger, Greig, and Emanuele Sobacchi, 2016), well before the advent of tomographic 21cm maps with the Square Kilometer Array (SKA). Fortunately, the James Webb Space Telescope (JWST) is providing spectra from an ever-increasing number of galaxy groups at high redshifts which can be used for this purpose (e.g., Saxena et al., 2023; Witstok, Maiolino, et al., 2024; Tang, Stark, Z. Chen, et al., 2023; Tang, Stark, Topping, et al., 2024; Z. Chen, Stark, C. Mason, et al., 2024; Umeda et al., 2024; Napolitano, Pentericci, et al., 2024)

However, the interpretation of these observations has so far been fairly approximate. The presence of an IGM damping wing in each galaxy is estimated *independently of its neighbors*. This wastes invaluable information, as neighboring galaxies provide complimentary sightlines into the local EoR morphology. The result is that the studies focusing on individual galaxies generally only predict lower limits for the radii of local HII regions. Furthermore, studies tend to ignore one or more important sources of stochasticity in the EoR morphology, intrinsic galaxy emission and/or telescope noise.

In this work we develop a new framework to *infer the local HII region size and location* from Lyα observations of a galaxy group. Our formalism accounts for the *relative position* of each galaxy with respect to the host and nearby HII bubbles by creating self-consistent forward models of JWST/NIRSpec spectra for each galaxy. We account for the relevant sources of uncertainty/stochasticity, including: (i) the IGM mean neutral fraction, $\bar{x}_{\mathrm{HI}}$; (ii) the EoR morphology, given $\bar{x}_{\mathrm{HI}}$; (iii) the emergent Lyα emission, given the observed UV magnitudes; and (iv) NIRSpec instrument noise. Unlike many previous studies, we do not make any assumptions about the unknown relative contribution of observed versus unobserved galaxies to the growth of the local HII region. We quantify how many galaxies are required to robustly detect individual ionized regions with a $\lesssim 10\%$ accuracy in their inferred location and characteristic size, during the early stages of EoR. This work is a companion to Lu, C. A. Mason, Andrei Mesinger, et al. (2025), in which we presented a complementary formalism to detect edges of ionized regions, using empirically-calibrated relations.

This paper is structured as follows. In Section 3.2 we present our forward modeling pipeline for Lyα galaxy spectra during the EoR. We introduce our procedure to infer the size and location of the surrounding HII region in Section 3.3. We apply our framework to mock data and show our main results in Section 4.3. We build further confidence by performing out-of-distribution tests in Section 3.5. In Section 3.6 we quantify observational requirements for detecting individual HII regions and we conclude in Section 3.7. All quantities are presented in comoving units unless stated otherwise. We assume a standard ΛCDM cosmology $(\Omega_{\mathrm{m}}, \Omega_{\mathrm{b}}, \Omega_{\Lambda}, h, \sigma_8, n_{\mathrm{s}} = 0.310, 0.049, 0.689, 0.677, 0.81, 0.963)$, with parameters consistent with the latest estimates from Planck Collaboration, Aghanim, Akrami, et al. (2020). All quantities are quoted in comoving units and evaluated in the rest-frame, unless stated otherwise.

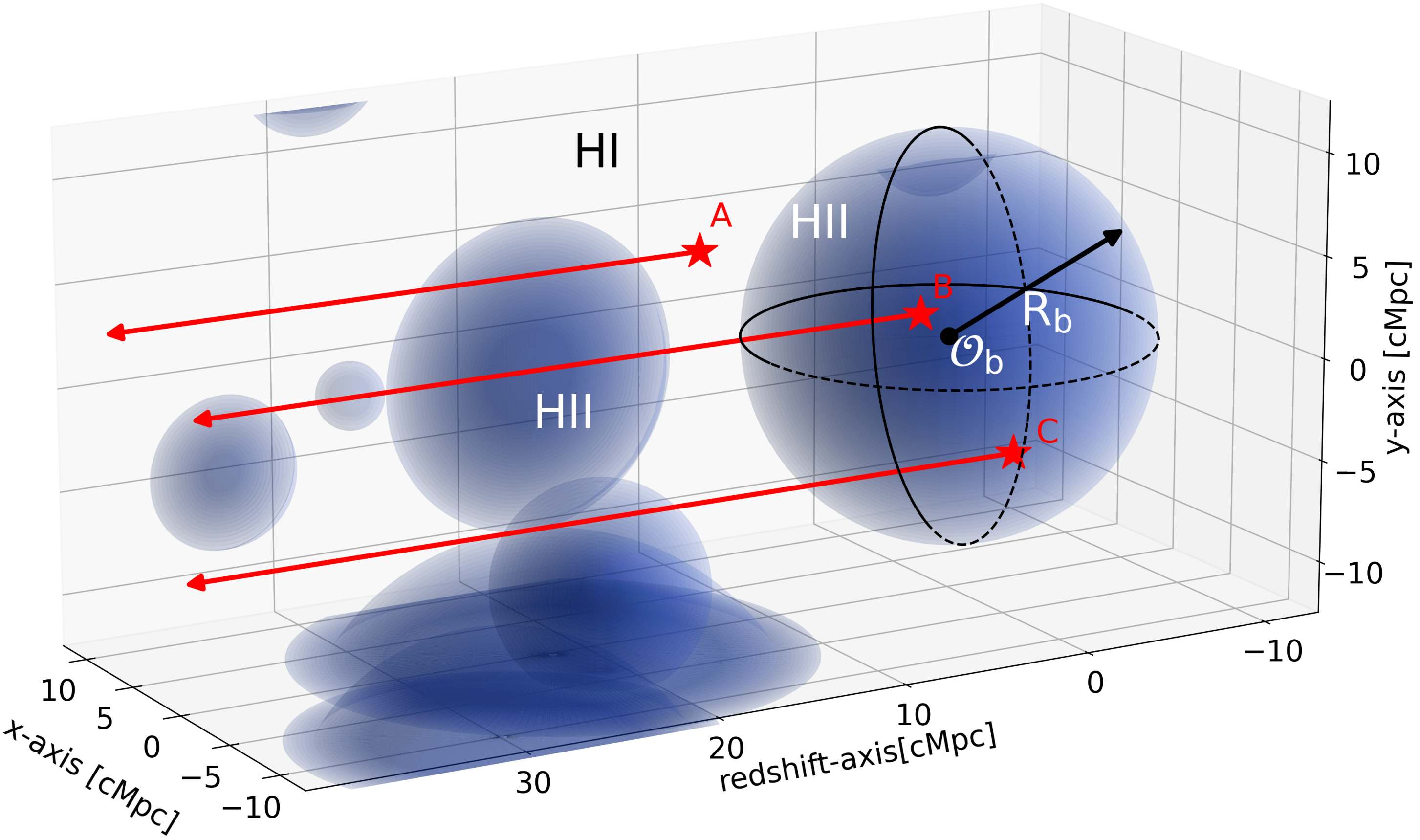

Figure 3.1: Schematic of our framework. Cosmic HII regions are shown in blue, embedded in an otherwise neutral IGM. We observe Lyman alpha spectra from a group of galaxies (here three example galaxies are denoted with 'A', 'B', and 'C'), and we wish to infer the central HII bubble (characterized by a sphere with radius 'R_{b}' and center '$\mathcal{O}_{\mathrm{b}}$'). Our framework combines the complimentary information provided by neighboring sight-lines towards the galaxies (red arrows), accounting for the main sources of stochasticity.

3.2 Observing Lyman-Alpha Spectra from Galaxies During the EoR

Our fiducial set-up is shown in Fig. 3.1. An HII region in an observed galaxy field is characterized as a sphere, with a center location ($\mathcal{O}_\mathrm{b}$) and characteristic radius (R_b). This is the "local" or "target" HII region whose properties we aim to infer. Nearby ionized regions are also shown in blue in the diagram. Observed galaxies can be located both inside and outside HII regions; here we denote three such galaxies with 'A', 'B', 'C', highlighting their sightlines towards the observer with red arrows.

Specifically, we wish to determine the conditional probability of the HII bubble center, $\mathcal{O}_\mathrm{b}$, and radius, R_b, given observed Lyα spectra of N_gal galaxies in a field with a central redshift z,

$$\mathcal{P}(\mathcal{O}_\mathrm{b}, R_\mathrm{b} | \boldsymbol{x}^i, f_\alpha^i(\lambda), M_\mathrm{UV}^i, z) \ , \ i \in [1, N_\mathrm{gal}] \tag{3.1}$$

Here $\boldsymbol{x}^i$, M_UV^i, and $f_\alpha^i(\lambda)$ are vectors of the galaxies' Cartesian coordinates, UV magnitudes, and observed Lyα spectra. For each galaxy, the observed spectrum in the rest-frame can be written as:

$$f_\alpha(\lambda) = L_\alpha J(\lambda) e^{-\tau_\mathrm{EoR}(\lambda)} + \mathcal{N}(\lambda) \tag{3.2}$$

where L_α is the emergent[1] Lyman-alpha luminosity of a galaxy, $J(\lambda)$ is the normalized, emergent Lyman-α profile, $\exp[-\tau_\mathrm{EoR}(\lambda)]$ accounts for IGM attenuation, and $\mathcal{N}(\lambda)$ is the spectrograph noise.

In the schematic shown in Fig. 3.1, galaxy 'A' is outside of an ionized bubble, and its Lyman α flux will therefore be strongly attenuated by the neutral IGM (i.e. having a large $\tau_\mathrm{EoR}(\lambda)$). We should only detect Lyman alpha flux from galaxy 'A' if it has a high emergent luminosity, L_α, *and* its Lyα profile, $J(\lambda)$, is strongly redshifted from the systemic z (e.g., Dijkstra 2014). Galaxy 'B' is close to the center of the local HII bubble, and will have (*on average*; c.f. right panel of Fig. 3.5), the lowest Lyα damping wing attenuation from the patchy EoR. However, the observed flux depends on all of the terms in eq. (3.2), each of which can have sizable stochasticity. Below we detail how each of these terms.

3.2.1 *Emergent Lyman-Alpha Profile* We start with the Lyman-alpha profile emerging into the IGM, $J(\lambda)$, normalized to integrate to unity. In order to escape the ISM of the galaxy, Lyman-α photons must diffuse spectrally which leads to a double-peaked Lyman-alpha shape (Neufeld, 1990; Weida Hu et al., 2023; Hutter, Trebitsch, et al., 2023; Almada Monter and Max Gronke, 2024). Due to the resonant nature of the line, the blue peak generally gets absorbed even by the ionized IGM at $z \gtrsim 5$ (though see Romain A. Meyer et al. 2021 for some putative, rare counter examples). Following C. A. Mason, Tommaso Treu, et al. (2018), we model the remaining red peak as a Gaussian:

$$J(\lambda) = \frac{2}{\nu_\alpha} \sqrt{\frac{\ln 2}{\pi}} \exp\left(-\frac{(\Delta \mathrm{v} - v_\alpha)^2}{v_\alpha^2 / (2 \ln 2)} \right), \tag{3.3}$$

where ν_α represents the velocity offset from systemic of the center of the line, and $\Delta \mathrm{v} = ((\lambda - \lambda_\alpha) c / \lambda_\alpha)$ is the velocity difference from the resonant wavelength of the line, $\lambda_\alpha = 1215.16 \mathring{A}$. As in C. A. Mason, Tommaso Treu, et al. (2018), for simplicity we assume that the FWHM of the line is equal to the velocity of the offset (e.g., Verhamme et al., 2018). Although these profiles are motivated by lower redshift observations (e.g. Yamada et al. 2012; Orlitová et al. 2018; Weida Hu et al. 2023), we note that our framework can easily accommodate any distribution for $J(\lambda)$, once we have better models for the emergent spectra. We also assume that all Lyman-alpha photons with a velocity offset below the circular velocity, v_circ, of the host halo are absorbed by the CGM (Dijkstra, Andrei Mesinger, and Wyithe, 2011; Laursen, Sommer-Larsen, and Razoumov, 2011). In Fig. 3.2 we show an example of the emergent profile in blue, for a galaxy with UV magnitude $M_\mathrm{UV} = -20.0$ and velocity offset $v_\alpha = 270$km/s, with a $v_\mathrm{circ} = 160$km/s.

The PDF of the emergent velocity offset is well described by a log-normal distribution (Steidel, Rudie, et al., 2014; De Barros et al., 2017; Stark, Ellis, Charlot, et al., 2017b; C. A. Mason, Tommaso Treu, et al., 2018):

$$\mathcal{P}(v_\alpha | M_\mathrm{UV}) = \frac{1}{\sqrt{2\pi} \ln 10 \ v_\alpha \ \sigma_\mathrm{v}} \exp\left(-\frac{(\log_{10} v_\alpha - \overline{v_\alpha}(M_\mathrm{UV}))^2}{2\sigma_\mathrm{v}^2} \right) \tag{3.4}$$

---

[1] Throughout we use the term "emergent" to refer to values escaping from the galaxy into the IGM. Therefore the emergent amplitude, L_α, and profile, $J(\lambda)$, are determined by Lyman alpha radiative transfer through the interstellar medium (ISM) and the circumgalactic medium (CGM; e.g. Neufeld 1990; Laursen, Sommer-Larsen, and Razoumov 2011). We do not model the details of this radiative transfer in this work, but instead rely on empirical relations based on post-EoR observations to determine the conditional distributions of L_α and $J(\lambda)$.

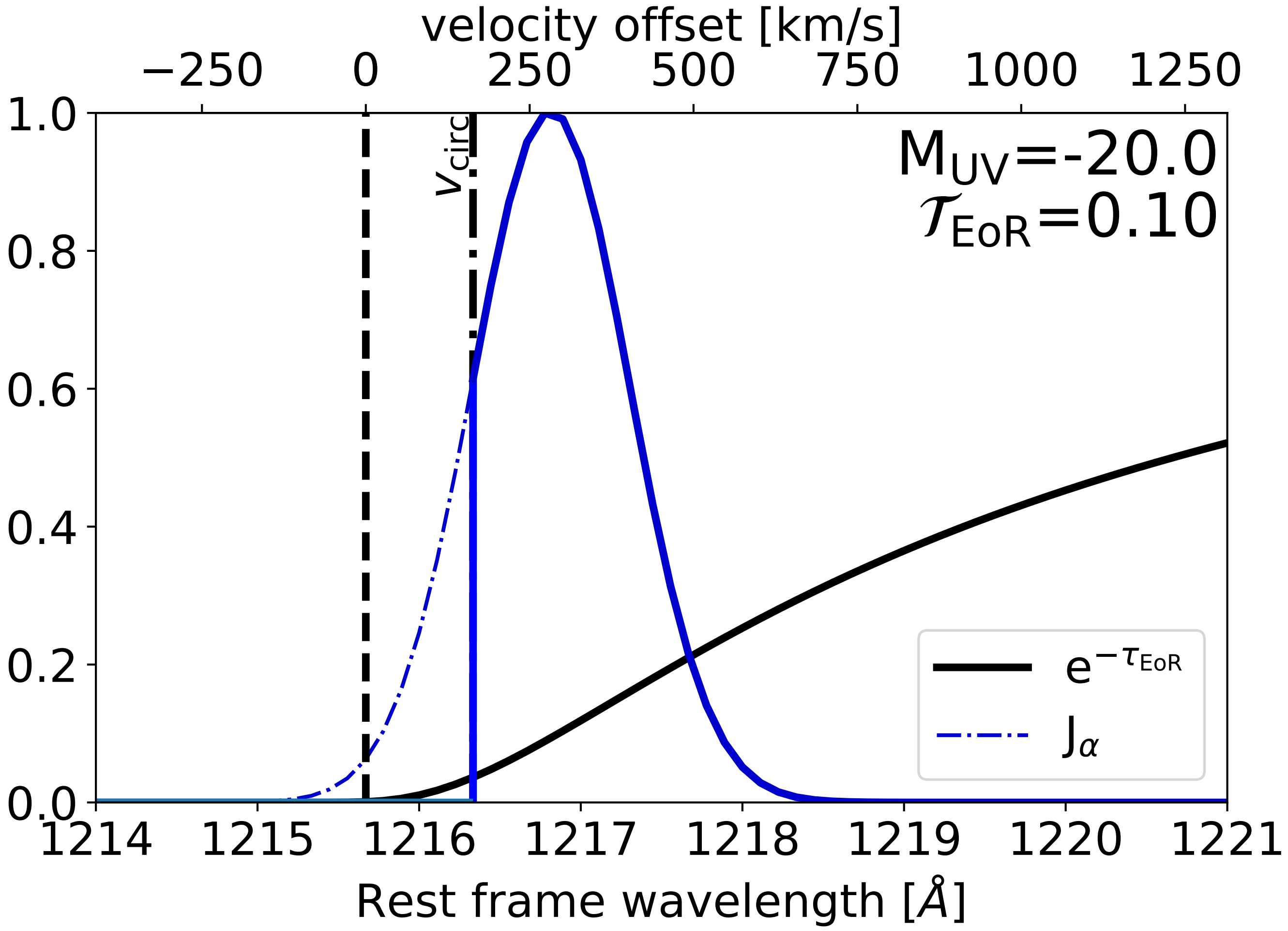

Figure 3.2: An example of an emergent Lyman alpha line and IGM opacity as a function of wavelength. The blue solid line represents the normalized Lyman-α emergent profile, while the blue dot-dashed line demarcates the flux absorbed by the CGM, blueward of the circular velocity of the galaxy. The solid black line illustrates an IGM damping wing attenuation profile, taken from a random sightline at $\bar{x}_{\rm HI} = 0.65$ and $R_b = 10$cMpc (see text for details). The Lyman alpha transmission integrated over all wavelengths for this example would be $\mathcal{T}_{\rm EoR} \equiv \int {\rm d}\lambda J(\lambda) e^{-\tau_{\rm EoR}(\lambda)} = 0.10$.

where the mean velocity offset is correlated with the UV magnitude (Stark, Ellis, Charlot, et al. 2017b, though see Bolan, Bradăc, et al. 2024):

$$\log_{10} \overline{v_\alpha}(M_{\rm UV}, z) = 0.32\gamma(M_{\rm UV} + 20.0 + 0.26z) + 2.34 \tag{3.5}$$

and $\sigma_v = 0.24$, $\gamma = -0.3$ for $M_{\rm UV} \geq -20.0 - 0.26z$, and $\gamma = -0.7$ otherwise. We show this distribution in the upper panel of Fig. 3.3. Although there is some indication of a mild redshift evolution in this distribution (e.g., Tang, Stark, Topping, et al. 2024; Witstok, Smit, et al. 2024), we show in Section 3.5 that our results are insensitive to such changes.

3.2.2 *Emergent Lyman-alpha Luminosity* The absolute normalization of the profile discussed above (i.e., the emergent Lyman alpha luminosity L_α) is generally defined via the so-called rest-frame equivalent width (e.g., Dijkstra and Wyithe, 2012):

$$W = \frac{L_\alpha}{L_{1500,\nu}} \frac{\lambda_\alpha}{\nu_\lambda} \left(\frac{\lambda_{\rm UV}}{\lambda_\alpha}\right)^{\beta+2} \tag{3.6}$$

where $L_{1500,\nu}$[erg/s/Hz] is the specific UV luminosity evaluated at 1500$\mathring{A}$ obtained from the continuum flux $f_{1500,\nu}$: $L_{1500,\nu} = 4\pi d_{\rm L}^2 f_{1500}$ where f_{1500} is given in units of [erg/s/cm^2/Hz], where $d_{\rm L}$ is the luminosity distance to the source, $\nu_\lambda = 2.47 \times 10^{15}$Hz, $\lambda_\alpha = 1215.67\mathring{A}$, $\lambda_{\rm UV} = 1500\mathring{A}$ is the rest-frame wavelength at which the UV magnitude is measured and β is the UV slope (which we assume to be $\beta = -2.0$ for simplicity).

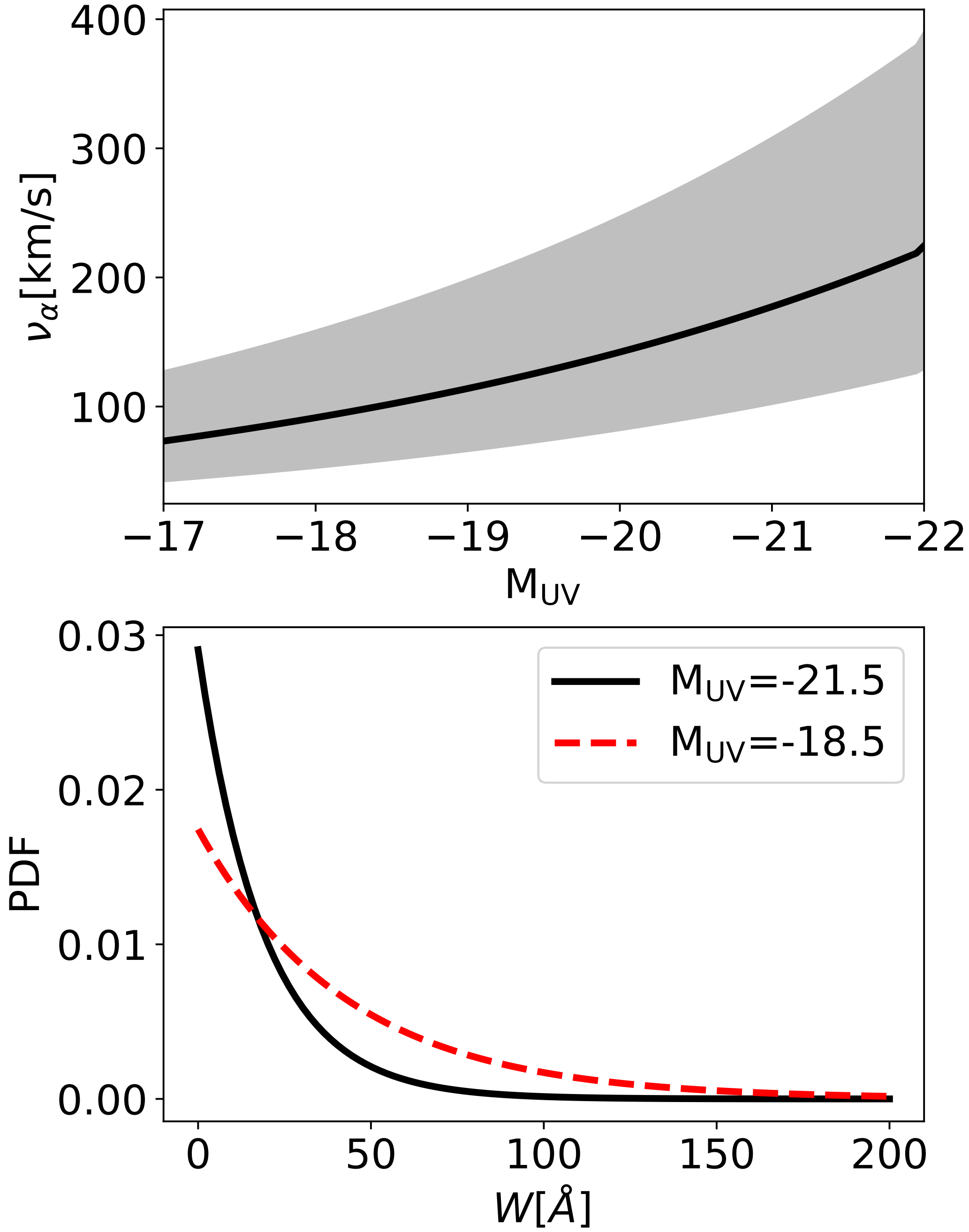

Figure 3.3: *Upper panel:* Mean (black line) and 1-sigma (shaded region) of the velocity offset distribution as a function of UV magnitude (Eq. 3.4). *Lower panel:* PDF of the Lyman-alpha rest-frame equivalent widths. The black and red dashed curves represent the distribution of equivalent widths (W) for Lyman-α emitters (Eq. 3.7). The non-emitters have equivalent widths described by a delta-function at $W = 0\mathring{A}$ (not shown) normalized such that the integral of the PDF is 1. The distribution is shown for $\mathrm{M_{UV}} = -18.5$ and -21.5

For galaxies at $z < 6$, where we expect τ_{EoR} to be negligible, C. A. Mason, Tommaso Treu, et al. (2018) found the following fit based on data from De Barros et al. (2017):

$$p_6(W|M_{\mathrm{UV}}) = \frac{A(M_{\mathrm{UV}})}{W_{\mathrm{c}}(M_{\mathrm{UV}})} e^{-\frac{W}{W_{\mathrm{c}}(M_{\mathrm{UV}})}} H(W) + [1 - A(M_{\mathrm{UV}})]\delta(W), \tag{3.7}$$

where $A(M_{\mathrm{UV}})$ is the fraction of intrinsic emitters for a given M_{UV} and $W_{\mathrm{c}}(M_{\mathrm{UV}})$ is the characteristic scale of the distribution, which is anti-correlated with M_{UV}. $H(W)$ is the Heaviside step function and $\delta(W)$ is a Dirac delta function. We use the following fit, as in C. A. Mason, Tommaso Treu, et al. (2018): $A = 0.65 + 0.1 \tanh\left[3(M_{\mathrm{UV}} + 20.75)\right]$ and $W_{\mathrm{c}} = 31 + 12 \tanh\left[4(M_{\mathrm{UV}} + 20.25)\right]\mathring{A}$. The distribution of emergent equivalent widths is shown in the lower panel of

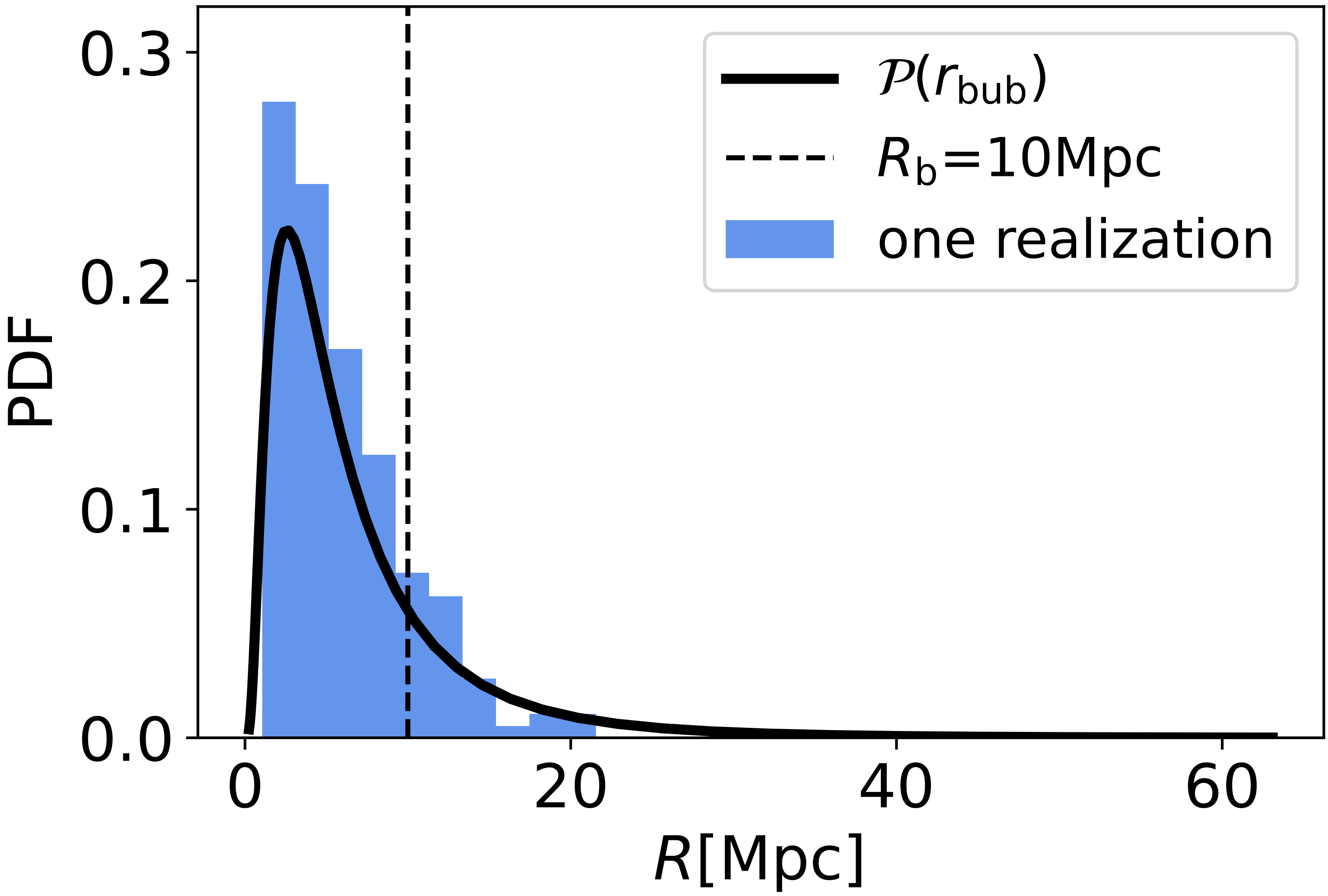

Figure 3.4: Distribution of HII region radii used in our "overlapping spheres" algorithm for the large-scale EoR morphology (see text for details). The vertical dashed line marks $R_{\rm b} = 10$cMpc which is our fiducial size for the central ionized region whose properties we aim to infer (c.f. Fig. 1).

Figure 3.3 for two UV magnitudes. We note that our framework can easily accommodate different EW distributions (e.g., Tommaso Treu et al. 2012; Lu, C. A. Mason, Andrei Mesinger, et al. 2025; Tang, Stark, Topping, et al. 2024). However, we show in Section 3.5 that our results are not sensitive to the choice of distribution shape.

3.2.3 *IGM Damping Wing Absorption* During the EoR, the damping wing absorption from the residual HI patches along the line of sight can strongly attenuate the Lyα line (c.f. the $\exp[-\tau_{\rm IGM}]$ curve in Fig. 3.2). The damping wing optical depth is mostly sensitive to the distance to the nearest neutral HI patch (e.g. Jordi Miralda-Escudé 1998). Indeed, this is why we will be able to infer the size of the local HII bubble in this work (see also the complementary empirical approach in Lu, C. A. Mason, Andrei Mesinger, et al. 2025 based on empirical $\tau_{\rm IGM} - R_b$ relations from an EoR simulation). Nevertheless, the surrounding EoR morphology beyond the local HII region does contribute to the total $\tau_{\rm IGM}$ as an additional source of scatter (e.g. Andrei Mesinger and Steven R. Furlanetto 2008a).

Here we generate an EoR morphology *at a given* $\bar{x}_{\rm HI}$, by placing overlapping spherical HII regions, with a radius distribution given by (cf. Fig. 3.1):

$$\mathcal{P}(r_{\rm bub}|\bar{x}_{\rm HI}) = \frac{1}{\sqrt{2\pi}\sigma_{\rm bub}} \cdot \exp\left(-\frac{(\log r_{\rm bub} - \mu_{\rm bub})^2}{\sigma_{\rm bub}^2}\right). \tag{3.8}$$

We sample from the above distribution of radii, randomly choosing the center location, and stopping when the volume filling factor of ionized regions reaches the desired value, $\bar{x}_{\rm HI}$. Although simplistic, overlapping ionized spheres does result in a similar EoR morphology as is seen in cosmological radiative transfer simulations (e.g. Oliver Zahn, Andrei

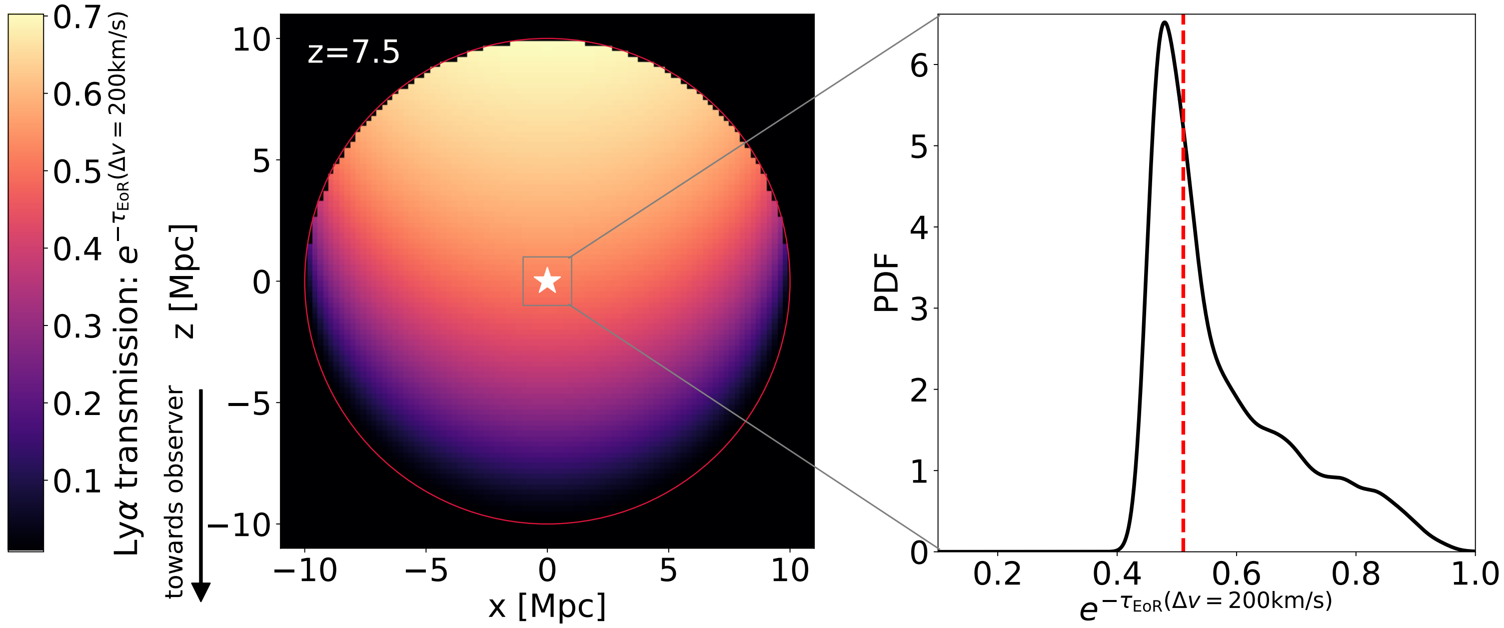

Figure 3.5: *Left panel*: The *mean* Lyα transmission evaluated at Δv = +200km/s ($e^{-\tau_{\rm EoR}}(\Delta v = +200)$km/s)as a function of the position inside an ionized bubble. The observer is located towards the bottom of the figure. We compute the mean transmission by averaging over 10000 realizations of EoR morphologies, given an assumed neutral fraction of $\bar{x}_{\rm HI} = 0.65$ at $z = 7.5$ (see text for details). *Right panel*: The *distribution* of Lyα transmission at Δv = +200km/s from these realizations, evaluated at a fixed position inside the bubble denoted by the white star (see text for details). The mean is marked by the vertical dashed line.

Mesinger, et al., 2011; Andrei Mesinger, S. Furlanetto, and Cen, 2011a; Ghara et al., 2018; Doussot and Benoît Semelin, 2022). We do not assume we know the true value $\bar{x}_{\rm HI}^{\rm true}$ a priori; instead, we sample a prior distribution of $\bar{x}_{\rm HI}$ from complimentary observations while performing inference (see Section 3.3 for more details).

For simplicity, in this proof-of-concept work we ignore the $\bar{x}_{\rm HI}$ dependence of the bubble size distribution in eq. (3.8), taking constant values of $\mu_{\rm bub} = \log(5{\rm cMpc})$ and $\sigma_{\rm bub} = 0.5$. These choices in our algorithm roughly reproduce the ionized bubble scales seen in simulations during the early stages of the EoR (e.g., Andrei Mesinger and S. Furlanetto 2007b; Giri et al. 2018; Lu, C. A. Mason, Andrei Mesinger, et al. 2025; Doussot and Benoît Semelin 2022; Neyer et al. 2024). Note that the $r_{\rm bub}$ from Eq. 3.8 does not directly translate to any of the metrics commonly used to characterize EoR morphology (e.g. Y. Lin et al., 2016; Giri et al., 2018) and so comparisons must be done a-posteriori. In future work, we will calibrate Eq. (3.8) to EoR simulations, conditioning also on the matter field to account for the (modest) bias of HII regions at early times (e.g. Fig. 12 in Emanuele Sobacchi and Andrei Mesinger, 2014). We show our assumed bubble size distribution in Fig. 3.4, as well as one realization in a 200 × 80 × 80 cMpc^3 volume.

For a given realization of EoR morphology and galaxy field (c.f. Fig. 3.1), we compute the damping wing optical depth by casting rays from the galaxy locations and summing the optical depth contributions from all HI patches along the LOS (e.g. Jordi Miralda-Escudé 1998):

$$\tau_{\rm EoR}(\lambda_{\rm em}) = \frac{\tau_{\rm GP} R_\alpha}{\pi} \sum_i x_{\rm H,i} \left(\frac{1+z_{b,i}}{1+z_\lambda}\right)^{3/2} \left[I\left(\frac{1+z_{b,i}}{1+z_\lambda}\right) - I\left(\frac{1+z_{e,i}}{1+z_\lambda}\right)\right] \tag{3.9}$$

where $z_\lambda = \frac{\lambda_{\rm em}}{\lambda_\alpha}(1+z_{\rm obs}) - 1$, $\lambda_{\rm em}$ is the wavelength at which we evaluate the optical depth, $\lambda_\alpha = 1215.67\mathring{A}$ is the Lyman-alpha resonant wavelength and $z_{\rm obs}$ is the redshift of the emitting galaxy. $\tau_{\rm GP} \approx 7.16 \cdot 10^5\left((1+z_{\rm obs})/10\right)^{3/2}$ is the Gunn-Peterson optical depth, $R_\alpha = \frac{\Lambda}{4\pi\nu_\alpha}$, $\Lambda = 6.25 \cdot 10^8{\rm s}^{-1}$ is the decay constant and $\nu_\lambda = 2.47 \cdot 10^{15}$Hz is the Lyman-α resonant frequency. In the above equation, $I(x)$ is given by:

$$I(x) \equiv \frac{x^{9/2}}{1+x} + \frac{9}{7}x^{7/2} + \frac{9}{5}x^{5/2} + 3x^{3/2} + 9x^{1/2} - \ln\left|\frac{1+x^{1/2}}{1-x^{1/2}}\right| \tag{3.10}$$

The summation accounts for every neutral patch encountered along the LOS, with a given patch, i, extending from $z_{b,i}$ to $z_{e,i}$. We assume that ionized patches have no neutral hydrogen atoms so they do not contribute to the attenuation.

In the left panel of Fig 3.5 we show the *mean* IGM transmission, evaluated at $\Delta\mathrm{v} = +200\mathrm{km/s}$ redward of the systemic redshift, as a function of position inside the central HII bubble of $R_\mathrm{b} = 10$ cMpc. The observer is located towards the bottom of the figure. This mean transmission was computed by averaging over 10000 realizations of EoR morphologies outside the central bubble, assuming $\bar{x}_\mathrm{HI} = 0.65$ (e.g. one such realization is shown in Fig. 3.1). As expected, there is a clear trend of increased transmission for galaxies located at the far end of the central HII bubble. The mean transmission is a function of the distance of the galaxy to the bubble edge in the direction towards the observer. In Lu, C. A. Mason, Andrei Mesinger, et al. 2025 we used these trends to define empirical "edge detection" algorithms.

However, at every location in the bubble, there is sizable sightline-to-sightline scatter in the IGM transmission. We quantify this in the right panel of Fig. 3.5, showing the transmission PDF constructed from the 10000 realizations of EoR morphology external to the central bubble. The sightlines used to compute this PDF originated from the location marked by the white star in the left panel. The PDF is quite broad and asymmetric (see also Fig. 2 in Andrei Mesinger and Steven R. Furlanetto 2008b as well as Lu, C. A. Mason, Hutter, et al. 2024). While it is difficult for the IGM to completely attenuate Lyα for a galaxy located at the center of this bubble, some morphologies can result in large stretches of ionized IGM in the direction of the observer, driving a high-transmision tail in the PDF. The width of this PDF highlights the importance of accounting for stochasticity in the EoR morphology when interpreting galaxy Lyman alpha observations (e.g. Andrei Mesinger and Steven R. Furlanetto 2008a; Andrei Mesinger, Aykutalp, et al. 2015; C. A. Mason, Tommaso Treu, et al. 2018; Bruton et al. 2023; Keating et al. 2024; Lu, C. A. Mason, Andrei Mesinger, et al. 2025).

3.2.4 *Including NIRSpec Noise* We take NIRSpec on JWST for our fiducial spectrograph in this work, which is already measuring Lyα spectra from galaxy groups during the EoR (e.g. Tang, Stark, Ellis, et al. 2024; Witstok, Maiolino, et al. 2024). To forward-model NIRSpec observations, we bin the observed flux $f_\alpha(\lambda) = L_\alpha J(\lambda) e^{-\tau_\mathrm{EoR}(\lambda)}$ to $R \sim 2700$ (high-resolution of NIRSpec), then add Gaussian noise $\mathcal{N}$ to each spectral bin, with a standard deviation of $\sigma(\mathcal{N}) = 2 \times 10^{-20}\mathrm{ergs}^{-1}\mathrm{cm}^{-2}\mathring{A}^{-1}$. This level of noise is obtainable with roughly a few hours of integration on NIRSpec (Bunker et al., 2023; Saxena et al., 2023; Tang, Stark, Ellis, et al., 2024), and corresponds to an uncertainty on the integrated flux ($\mathcal{F}_\mathrm{int} \equiv \int f_\alpha(\lambda)\mathrm{d}\lambda$) of $\sigma(\mathcal{N}_\mathrm{int}) = 1 \times 10^{-19}\mathrm{ergs}^{-1}\mathrm{cm}^{-2}$ (estimated assuming the emission line is spectrally unresolved). This can be further translated to 5σ limiting equivalent widths of $W = 25\mathring{A}$ ($W = 60\mathring{A}$) for $M_\mathrm{UV} = -18$ ($M_\mathrm{UV} = -17$). We also find worse recovery assuming shallower observations, while our results do not improve significantly assuming deeper integrations than this fiducial value.

We re-bin the spectra to lower values of R and test the inference with a coarser resolution. By performing additional binning, we lose some information on the observed line profile (e.g. Byrohl and M. Gronke, 2020), but lower the dimensionality of our likelihood (see Section 3.3). In future work, we will explore more sophisticated inference approaches that can scale to high-dimensional likelihoods (Cranmer, Brehmer, and Louppe, 2020; Anau Montel, Alvey, and Weniger, 2024). Here we empirically settle on rest-frame $\Delta\lambda \sim 1\mathring{A}$ as our bin width (i.e. $R \sim 1000$), corresponding to the medium resolution NIRSpec grating (P. Jakobsen et al., 2022, see Section 3.3).

3.3 Inferring the Local HII Bubble

With the above framework, we can create mock observations and corresponding forward models by sampling each of the terms in Eq. 3.2. We detail this procedure below.

3.3.1 *Mock Observations* We first construct a mock observation of N_gal galaxies in a survey volume of $V_\mathrm{survey} =$ 20x20x20 cMpc^3 ($\sim$ 7' ×7' × $\Delta z = 0.07$) at $z = 7.5$. This FoV is roughly motivated by *JWST* (corresponding to roughly 4 NIRSpec pointings), though in practice the forward-modeled volume should be tailored to the specific observation that is being interpreted. We place a bubble with radius $R_b = 10$ cMpc at the (arbitrarily-chosen) center of the volume, and construct the surrounding EoR morphology out to distances of 200 cMpc[2], using the prescription from Sect. 3.2.3 and assuming $\bar{x}_\mathrm{HI} = 0.65$. Below we demonstrate that our results are insensitive to these fiducial choices.

We assign random locations to the galaxies, $\boldsymbol{x}^i$ with $i \in [1, N_\mathrm{gal}]$, using rejection sampling to ensure that on average 75% of the galaxies are inside HII regions. This number is motivated by analysis in (Lu, C. A. Mason, Hutter, et al.,

---

[2] Neutral IGM at larger distances contributes a negligible amount to the total attenuation, due to the steepness of the damping wing profile (Andrei Mesinger and Steven R. Furlanetto, 2008a).

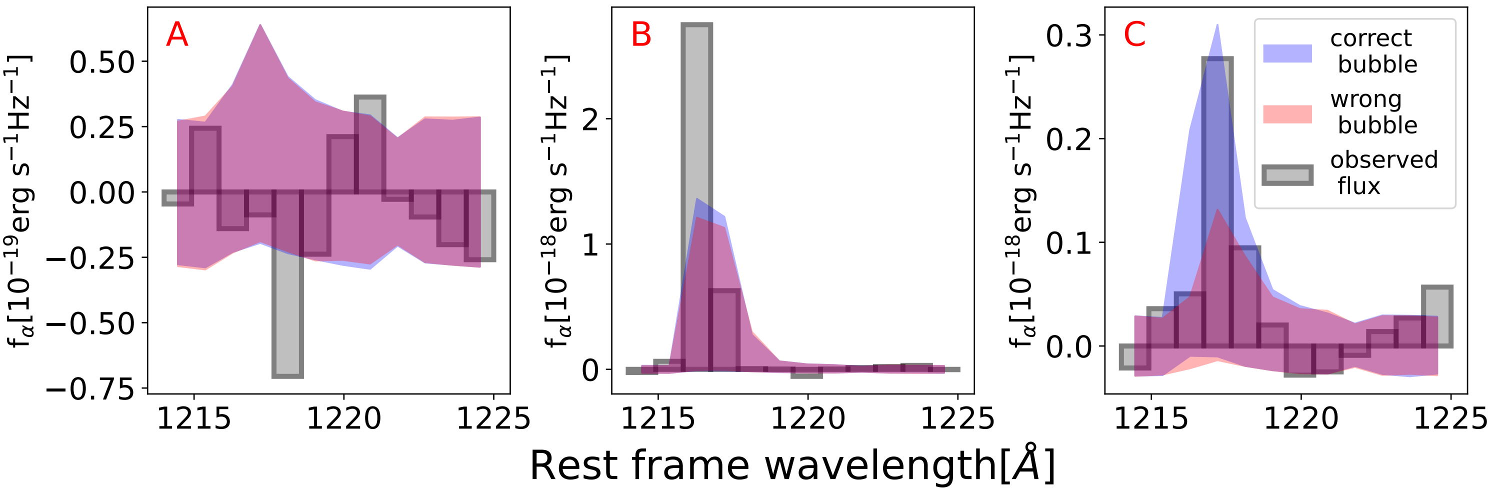

Figure 3.6: The observed flux (*gray*) corresponding to the galaxies 'A', 'B', and 'C' from the mock observation shown in Fig. 3.1. In blue we show the 68% C.L. of the likelihood assuming the correct HII bubble location and radius, $(R_\mathrm{b}, \mathcal{O}_\mathrm{b}) = (R_\mathrm{b}^\mathrm{true}, \mathcal{O}_\mathrm{b}^\mathrm{true})$. In red we show the 68% C.L. of the flux likelihood assuming the correct HII bubble location but a slightly smaller radius, $R_\mathrm{b} = 0.8R_\mathrm{b}^\mathrm{true}$. We see that the observed flux is more consistent with the blue likelihood than the red likelihood, mostly due to galaxy 'C' that is located close to the edge of the bubble (note the different y-axis ranges in each panel, with the observed flux from galaxy 'A' being noise dominated).

2024) which shows that $\sim 25\%$ galaxies are found outside of ionized bubbles for $M_\mathrm{UV} > -18$ at $\overline{x}_\mathrm{H} = 0.7$ (their Fig.3). This is a very approximate way of accounting for galaxy bias, as both galaxies and HII regions are correlated to the large-scale matter field.[3]

We generate UV magnitudes for each galaxy by sampling the UV luminosity function (LF) from J. Park, Andrei Mesinger, et al. (2019) down to a magnitude limit of $M_\mathrm{UV} = -18.0$. Each galaxy is then assigned an emergent emission profile according to the procedure in the previous section, which is attenuated by its sightline through the realization of the EoR morphology. Finally, a noise realization is added to the binned flux to create a mock spectrum for each galaxy (c.f. Eq. 3.2).

3.3.2 *Maximum likelihood Estimate of Bubble Size and Location* We then interpret this mock observation by forward modeling the observed flux for each galaxy, varying: (i) the position and radius of the central HII bubble; (ii) the surrounding EoR morphology; (iii) the neutral fraction of the Universe (within ±0.1 of the "truth", conservatively wider than current limits Qin, Andrei Mesinger, Prelogović, et al. 2024); (iv) the emergent Lyman alpha flux given the galaxy's observed M_UV (i.e. W and v_α); (v) NIRSpec noise realizations.

For each forward model, we compute the likelihood of the mock observation, given the location and radius of the central HII region. Our sampling procedure effectively marginalizes over the unknowns (ii) – (v) from above. Because mapping out the joint likelihood over all of the observed galaxies would be numerically challenging, we make the simplifying assumption that the likelihood of the observed flux from each galaxy, $f_\alpha^i(\lambda)$, is independent from the other galaxies. This allows us to write the total likelihood of the observation as a product of the likelihoods of the individual galaxies:

$$\mathcal{L} = \prod_{i}^{N_\mathrm{gal}} \mathcal{L}^i(f_\alpha^i(\lambda) \mid \mathcal{O}_\mathrm{b}, R_\mathrm{b}, \boldsymbol{x}^i, M_\mathrm{UV}^i, z) \tag{3.11}$$

While this assumption is clearly incorrect, here we present results only in terms of the maximum likelihood, $\hat{\mathcal{L}}$. We demonstrate below that Eq. (3.11) provides an unbiased estimate of $\hat{\mathcal{L}}$.[4]

---

[3] This is a reasonable approximation, as evidenced by our results in Section 3.5.2, where we apply our framework to simulations that self-consistently account for galaxy and HII region bias. Indeed, observations of galaxies demonstrate that the bias dominates clustering at larger scales ($\gtrsim$ tens of cMpc), while the smaller scales relevant for this work are dominated by Poisson noise (Bhowmick et al., 2018; Kragh Jespersen et al., 2024, Davies et al. in prep).

[4] Ideally, we would want to map out the full posterior:

$$\mathcal{P}(\mathcal{O}_\mathrm{b}, R_\mathrm{b}|\boldsymbol{x}, f_\alpha(\lambda), M_\mathrm{UV}, z) \propto \mathcal{L}(f_\alpha(\lambda)|\mathcal{O}_\mathrm{b}, R_\mathrm{b}, \boldsymbol{x}, M_\mathrm{UV}, z)\ \pi(\mathcal{O}_\mathrm{b}, R_\mathrm{b}),$$

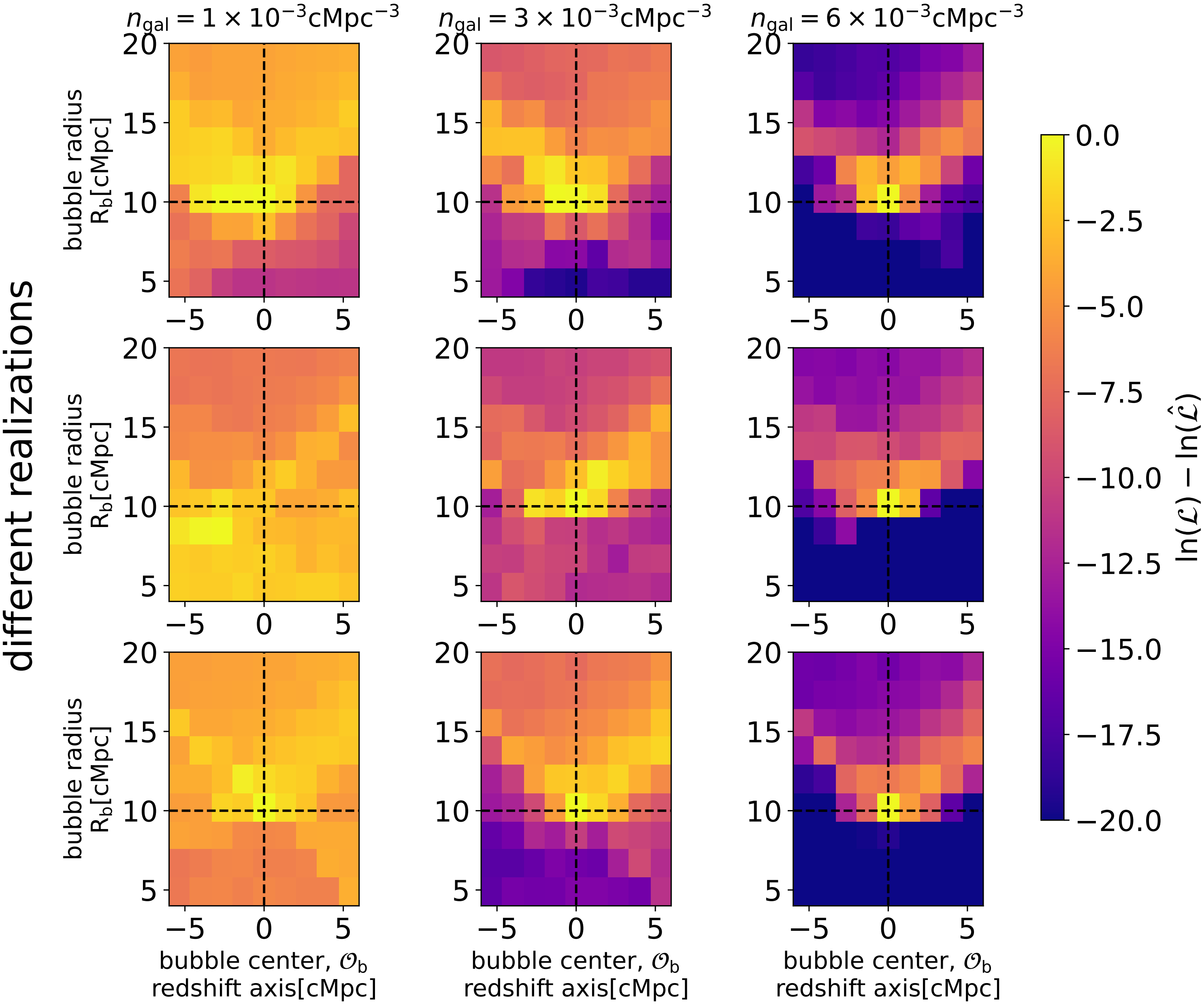

Figure 3.7: 2D slices through the log likelihood, normalized to the maximum value in each panel. The vertical and horizontal axis in each panel correspond to the bubble radius, $R_{\rm b}$, and redshift axis of the center, $\mathcal{O}_{\rm b} = (x_b = x_b^{\rm true}, y_b = y_b^{\rm true}, z_b)$. The true values, $(R_{\rm b}^{\rm true}, z_b^{\rm true}) = (10\ {\rm cMpc}, 0\ {\rm cMpc})$, are demarcated with the dashed lines. The mock observation in this example was made at $z = 7.5$, assuming $\overline{x}_{\rm H} = 0.65$. Different columns represent different numbers of galaxies used for inference (8 on the left, 24 in the middle and 48 on the right). Different rows correspond to different realizations of EoR morphology and galaxy samples.

It is important to note that *we do not assume a Gaussian likelihood for the observed flux at each wavelength*, as is commonly done. With high S/N spectra, correlations between flux bins can be significant. Instead we directly map out the *joint PDF of flux over all wavelength bins*, $\mathcal{L}^i(f_\alpha^i(\lambda_1), f_\alpha^i(\lambda_2), f_\alpha^i(\lambda_3), ... \mid \mathcal{O}_{\rm b}, R_{\rm b}, \boldsymbol{x}^i, M_{\rm UV}^i, z)$, using kernel density estimation over the forward-modeled spectra (see Appendix 3.9 for details). This preserves the covariances between the wavelength bins, and is commonly known as implicit likelihood or simulation based inference.

We demonstrate this procedure for the three galaxies shown in Fig. 3.1. The observed flux from galaxies 'A', 'B', and 'C', is denoted in gray in the three panels of Fig. 3.6, left to right respectively. In blue we show the 68% C.L.

---

where $\pi(\mathcal{O}_{\rm b}, R_{\rm b})$ is a prior for the center location and radius of the HII bubble. However, including the correlations of the (non-Gaussian) likelihoods at the location of every galaxy is analytically not tractable, and would require high dimensional simulation based inference (e.g., de Santi et al., 2023; Lemos et al., 2023). We save this for future work.

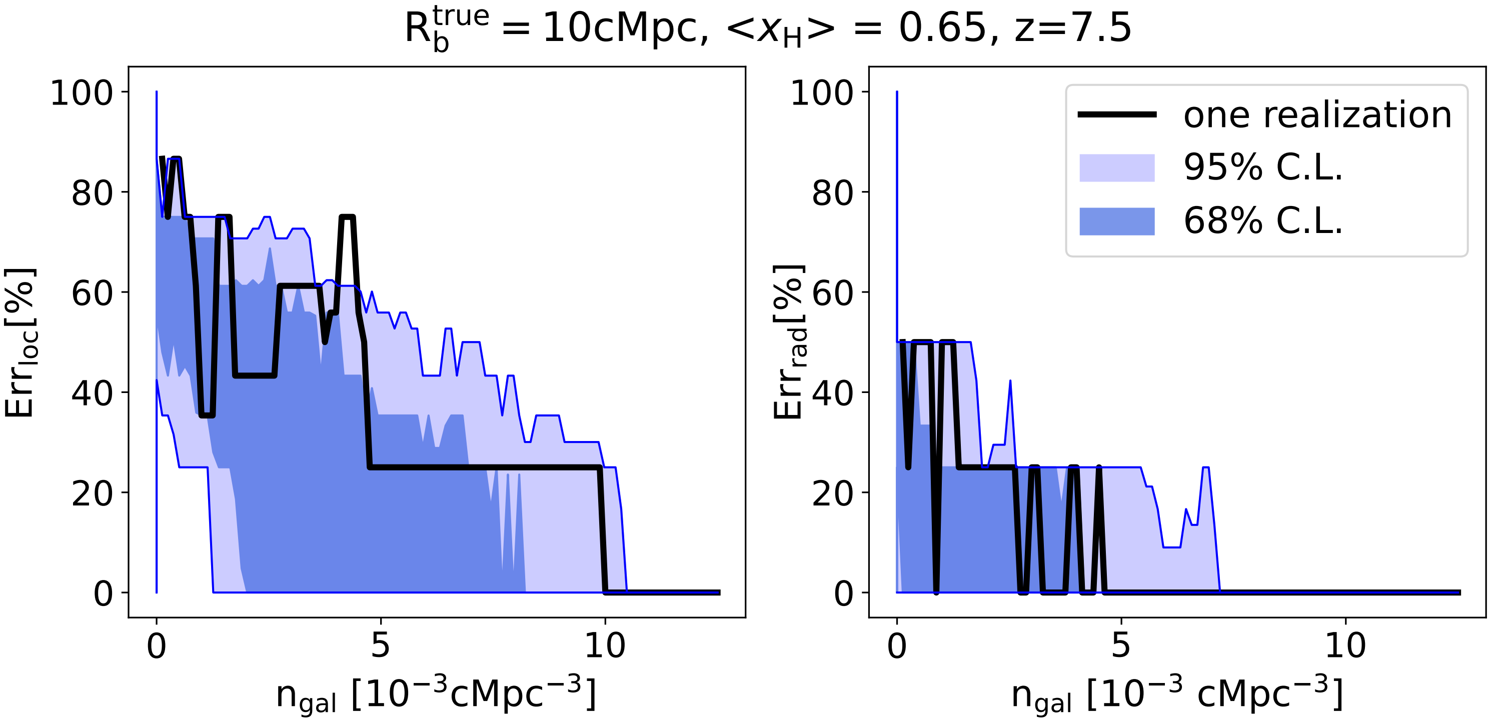

Figure 3.8: Fractional errors of the maximum likelihood HII bubble center (*left*) and radius (*right*) as a function of the number density of observed galaxies. The black curve corresponds to a single realization of the EoR morphology and galaxy samples. The light (dark) blue region corresponds to the 95th (68th) percentile of fractional errors obtained from 100 realizations. With $n_{\rm gal} \gtrsim 0.01$ galaxies cMpc^{-3}, the fractional errors drop below the grid resolution we use to calculate the likelihoods (1.5cMpc).

of the likelihood assuming the correct HII bubble location and radius, $(R_{\rm b}, \mathcal{O}_{\rm b}) = (R_{\rm b}^{\rm true}, \mathcal{O}_{\rm b}^{\rm true})$. In red we show the 68% C.L. of the flux likelihood assuming the correct HII bubble location but a slightly smaller radius, $R_{\rm b} = 0.8R_{\rm b}^{\rm true}$. Galaxy 'A' in this mock observation is outside of the central HII bubble, and therefore the true and slightly smaller values of $R_{\rm b}$ result in the same likelihood. Galaxy 'B' is located close to the center of the bubble, so both $R_{\rm b} = R_{\rm b}^{\rm true}$ and $R_{\rm b} = 0.8R_{\rm b}^{\rm true}$ give similar values of transmission $\exp[-\tau_{\rm EoR}]$ (see Fig. 3.5). This results in only a slight preference for $R_{\rm b} = R_{\rm b}^{\rm true}$. On the other hand, galaxy 'C' is located close to the edge of the bubble in Fig. 3.1. For that galaxy, we see that the observed spectrum in gray is more consistent with the correct likelihood in blue than with the incorrect one in red. Having a smaller bubble would imply more IGM attenuation on average at the location of this galaxy, making the observed strong Lyα emission less likely. In this specific realization, the joint likelihood (Eq. 3.11) of the observed fluxes of 'A', 'B', and 'C' is two times larger for the correct value of bubble radius than for the incorrect one. As one includes more and more galaxies, the maximum likelihood becomes increasingly peaked around the true values for the bubble size and location.

We illustrate this explicitly in Figure 3.7, in which we plot 2D slices through the log likelihood. The log likelihood normalized to the maximum value is indicated with the color bar. The vertical and horizontal axis in each panel correspond to the sampled bubble radius, $R_{\rm b}$, and redshift axis of the center, $\mathcal{O}_{\rm b} = (x_b = x_b^{\rm true}, y_b = y_b^{\rm true}, z_b)$. The true values, $(R_{\rm b}^{\rm true}, z_b^{\rm true}) = (10\ {\rm cMpc}, 0\ {\rm cMpc})$, are demarcated with the dashed lines. The columns correspond to increasing number density of observed galaxies (*left to right*), while rows correspond to different realizations of forward models.

There are several interesting trends evident in Figure 3.7. Firstly, we see that increasing the number of observed galaxies (going from left to right in each row) results in an increasingly peaked likelihood, with the maximum settling on the true values for the bubble radius and location. Different realizations of the EoR morphology give different likelihood distributions when a small number of galaxies is observed. However, as the number density of observed galaxies is increased, the likelihood distributions between different realizations converge (i.e. all sources of stochasticity are "averaged out").

We also see a degeneracy between the redshift of the bubble center and its radius. Because the transmission is mostly determined by the distance to the bubble edge along the line of sight, moving the bubble center further away from the observer at a fixed radius is roughly degenerate with decreasing the radius at a fixed center location. This degeneracy is mitigated by having a larger number of sightlines to observed galaxies, allowing us to constrain the radius of curvature of the bubble.

3.4 How Many Galaxies Do We Need to Confidently Infer the Local HII Bubble?

In the previous section, we demonstrated that our framework gives an unbiased estimate of the HII region size and location when applied to a large galaxy sample. Here we quantify just how "large" does this galaxy sample need to be, in order for us to be confident in our results. For this purpose we define two figures of merit:

$$\begin{aligned}\mathrm{Err}_{\mathrm{loc}} &= \left| \hat{\vec{\mathcal{O}}}_{\mathrm{b}}(n_{\mathrm{gal}}) - \vec{\mathcal{O}}_{\mathrm{b}}^{\mathrm{true}} \right| / R_{\mathrm{b}}^{\mathrm{true}} \qquad (3.12)\\ &= \sqrt{(\hat{x}_b - x_b^{\mathrm{true}})^2 + (\hat{y}_b - y_b^{\mathrm{true}})^2 + (\hat{z}_b - z_b^{\mathrm{true}})^2} / R_{\mathrm{b}}^{\mathrm{true}}\end{aligned}$$

$$\mathrm{Err}_{\mathrm{rad}} = \left| \hat{R}_{\mathrm{b}}(n_{\mathrm{gal}}) - R_{\mathrm{b}}^{\mathrm{true}} \right| / R_{\mathrm{b}}^{\mathrm{true}}. \qquad (3.13)$$

Here, $\hat{\vec{\mathcal{O}}}_{\mathrm{b}}(n_{\mathrm{gal}})$ and $\hat{R}_{\mathrm{b}}(n_{\mathrm{gal}})$ are the maximum likelihood estimates of the HII bubble center and radius computed from a galaxy field with number density $n_{\mathrm{gal}} = N_{\mathrm{gal}}/V_{\mathrm{survey}}$, and $\mathrm{Err}_{\mathrm{loc}}$ and $\mathrm{Err}_{\mathrm{rad}}$ are the corresponding fractional errors (normalized to the true bubble radius).

Since we are calculating the likelihood on the grid, we can say that our framework "has found" the optimal bubble when the fiducial and inferred locations and radii coincide, i.e. $\mathrm{Err}_{\mathrm{loc}} = \mathrm{Err}_{\mathrm{rad}} = 0$. In that case, the error of location and radius is below the grid on which we calculate the likelihood (1.5cMpc in our fiducial case, corresponding to a $\lesssim 15\%$ fractional error).

The solid black curve in Fig. 3.8 shows how these fractional errors change with galaxy number density for a single realization. Here the realization of EoR morphology and "observed" galaxies are held constant, with the maximum likelihood computed each time a new "observed" galaxy is added. The more galaxies we observe, the smaller the error on our inferred HII bubble location and radius. The sizable stochasticity in galaxy properties and sightline-to-sightline scatter in the IGM opacity manifest as "noise" in this evolution, making it non-monotonic. Nevertheless, even a single galaxy is able to shrink the fractional error by a factor of two from our prior range, ruling out extreme values.

We repeat this calculation with 100 different realizations of the mock observation (EoR morphology and observed galaxy samples). In Fig. 3.8 we show the resulting 68% and 95% C.L. on the fractional errors as more and more galaxies are added to the field. We see that in 68% (95%) of cases, number densities of 7.7× (10.5×) 10^{-3}cMpc^{-3} are sufficient to obtain a $\lesssim$ 15% error on the center position. The corresponding requirements are 4.2× (7.2×) 10^{-3}cMpc^{-3} for $\lesssim$ 15% error on the bubble radius (comparable to Lu, C. A. Mason, Andrei Mesinger, et al. 2025). In other words, Lyα spectra from ~ 0.01 galaxies per cMpc^3 are required to be $\gtrsim$95% confident that the HII bubble location and size recovered by our method is accurate at $\lesssim$ 1 cMpc. This corresponds roughly to 80 galaxies in 2x2 tiled pointings with *JWST*/NIRSpec.

3.4.1 *Including a Prior on the Emergent Lyman-Alpha*

In this work we assume the emergent Lyman alpha emission follows post-EoR empirical relations, as described in Section 3.2.1. Beyond this, we assume no prior knowledge on each galaxy's intrinsic emission. However, lower opacity nebular lines such as the Balmer lines (Hα, Hβ), can provide complimentary estimates of the intrinsic production of Lyman alpha photons (e.g. Hayes, Runnholm, et al. 2023; Saxena et al. 2023; Z. Chen, Stark, C. Mason, et al. 2024). In this section we repeat the analysis from above, but including a simple prior on the emergent Lyα emission.

Specifically, we apply rejection sampling to keep only those forward-models for which the transmission is within 20% of the true one (or putting an upper bound if $\mathcal{T}_{\mathrm{EoR}} \leq 0.2$)[5] Then we calculate the likelihood the same way as we did in Section 3.3. The width of the prior is motivated by current observations using JWST (c.f. Saxena et al. 2023 for

---

[5] A detection of Balmer lines puts constraints on the production rate of Lyman alpha photons, who then have to pass through the ISM, CGM and IGM. Our distribution of Lyman-α equivalent widths and profile shapes described in Section. 3.2.2 effectively already includes radiative transfer and escape through the ISM + CGM, which we assume does not evolve significantly with redshift at $z \gtrsim 6$. Thus our prior is effectively a prior on the IGM transmission. We leave a self-consistent treatment of the redshift evolution of ISM+CGM escape fraction and IGM transmission for future work.

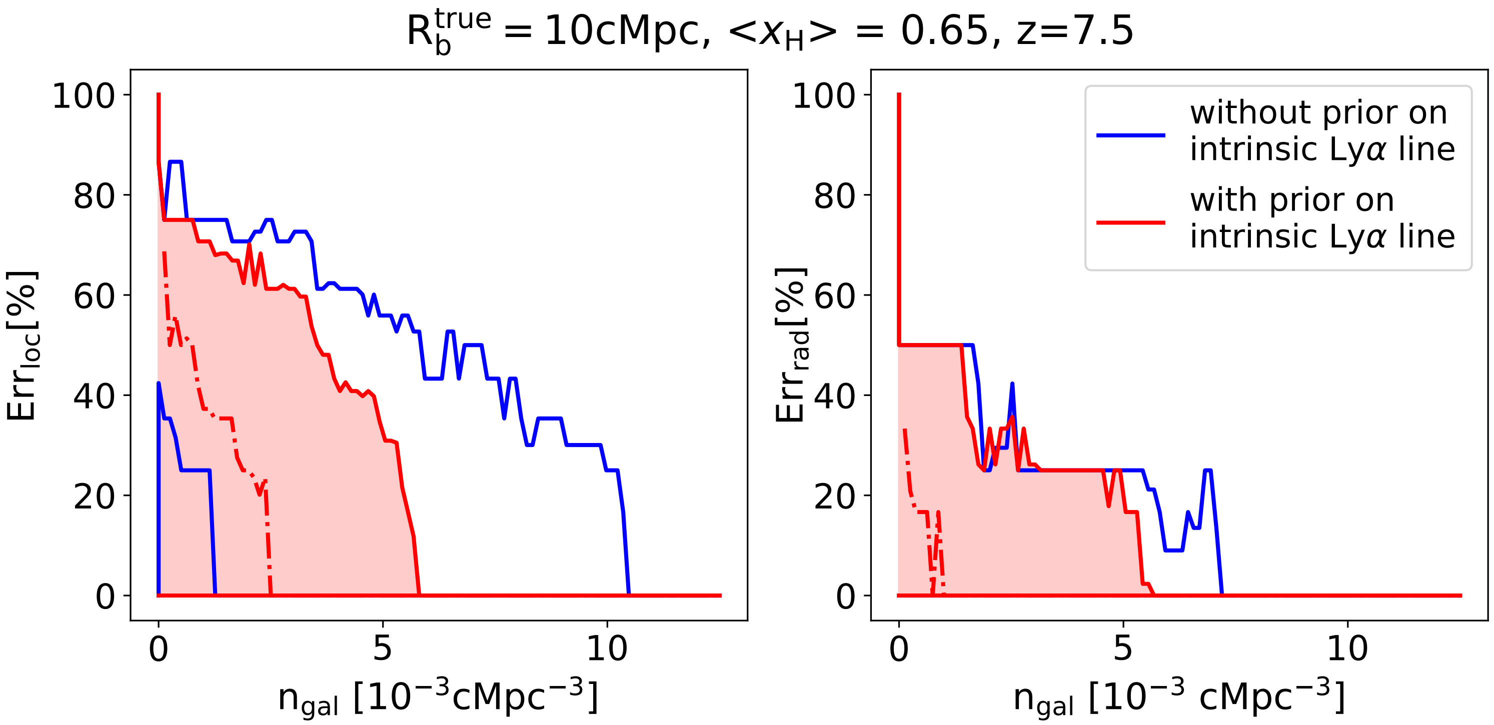

Figure 3.9: Same as Fig. 3.8, but we include in red the 95% C.L. on the fractional errors assuming a prior on the emergent Lyman alpha emission, motivated by observed Balmer lines (see text for details). Compared to the analogous results without including a prior (*blue*), we see that the required number of galaxies to achieve the same accuracy is reduced by roughly a factor of two.

Hβ, and X. Lin et al. 2024 for Hα from ground instruments). Note that Hα observations above $z > 7$ are not possible with NIRSpec. For higher redshifts, we therefore have to rely on a fainter Hβ. Despite this, Hβ is regularly observed in high$-z$ galaxies, so using the prior for all galaxies is not unreasonable (e.g., R. A. Meyer et al., 2024).

In Fig. 3.9 we show the 95% C.L. on the fractional errors with (*red*) and without (*blue*; same as Fig. 3.8) the prior information on the emergent Lyman alpha emission. We see that using additional nebular lines to constrain the emergent Lyman alpha emission can reduce by a factor of $\sim$ two the required number of galaxies to obtain the same constraints.

3.4.2 *Results for Different Bubble Sizes* Our fiducial choice for the radius of the central HII bubble is $R_\mathrm{b} = 10$Mpc, motivated by HII bubble sizes in the early stages of the EoR. Here we test the performance of our pipeline for other radii, $R_\mathrm{b} = 5$ and 15Mpc.

In the top (bottom) panels of Fig. 3.10 we show the analysis with and without the 20% prior on the emergent emission, assuming $R_\mathrm{b}^\mathrm{true} = 5$Mpc ($R_\mathrm{b}^\mathrm{true} = 15$Mpc). Comparing to the fiducial results in Fig. 3.9, we see that fewer galaxies are required to infer the size and location for larger HII bubbles. For $R_\mathrm{b}^\mathrm{true} = 15$ cMpc roughly 2-3 times fewer galaxies are required to constrain the HII bubble with the same accuracy, compared to $R_\mathrm{b}^\mathrm{true} = 5$ cMpc. This is most likely due to the fact that the larger bubbles allow for a broader range of IGM opacities. Galaxies inside small bubbles have roughly the same attenuation regardless of their relative location inside the bubble. In contrast, for larger bubbles there is a noticeable difference between the (average) attenuation on the near side and far of the bubble, allowing us to more easily constrain their geometry (c.f. the left panel of Fig. 3.5).

3.5 Building Confidence in our Framework

In this section, we explore how do our results depend on our model assumptions. To do this, we apply our framework on a different reionization morphology and on different emergent EW distributions. Even though our framework is a proof-of-concept, having it perform well on such "out-of-distribution" tests would help build confidence that it is not sensitive to uncertainties in the details of our model.

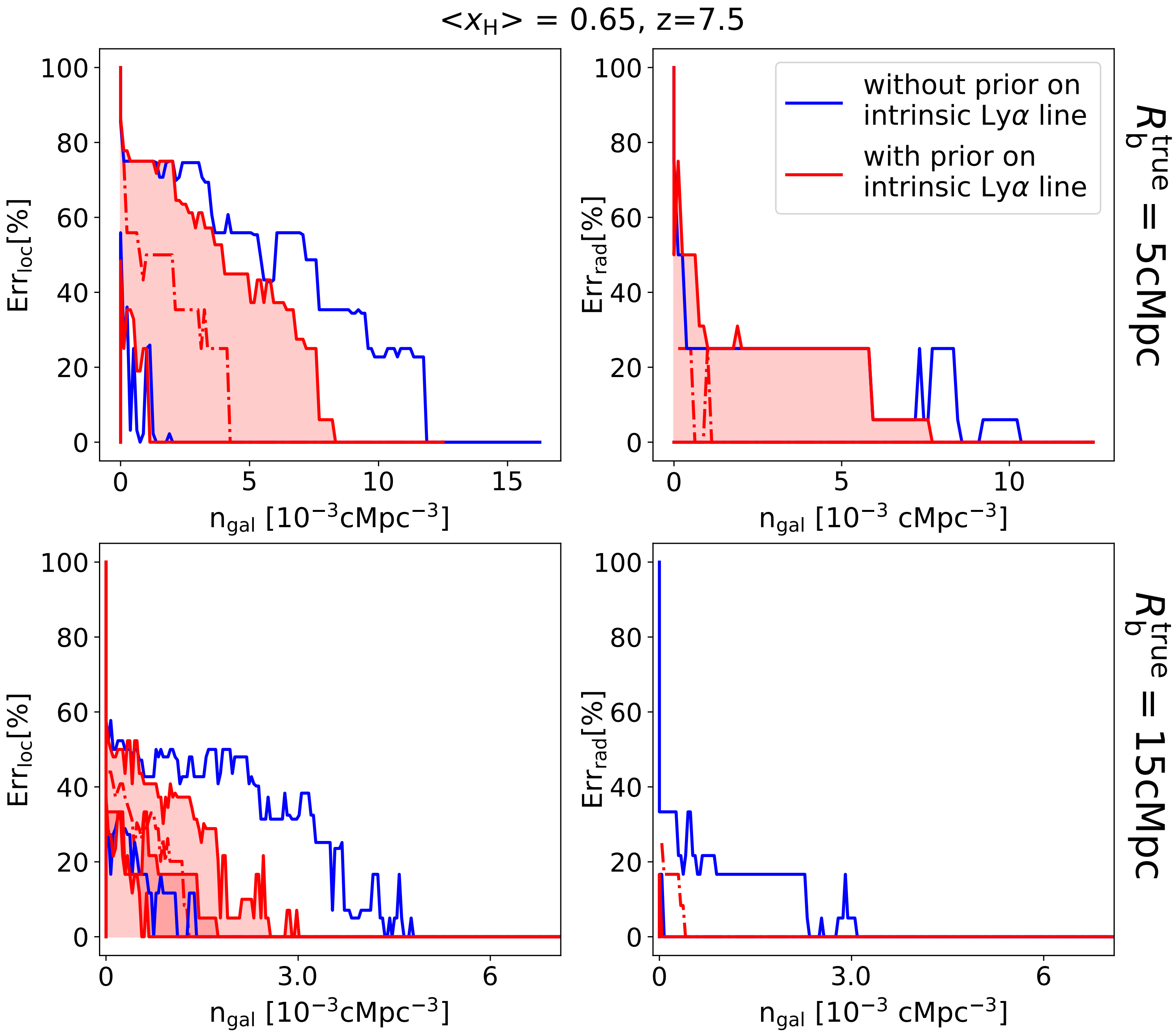

Figure 3.10: Same as Fig. 3.9, but assuming $R_b^{\mathrm{true}} = 5$ (15) cMpc in the top (bottom) row. Larger bubbles can be inferred with the same accuracy using a smaller number density of galaxies.

3.5.1 *Different EW Distribution* First, we test the performance of our pipeline assuming the emergent Lyman-α luminosities follow a different EW distribution than the we use in our forward-models (Sec. 3.2.2). Specifically, we generate mock observations assuming a Gaussian distribution from Tommaso Treu et al. (2012) (based on Stark, Ellis, and Ouchi (2011), but with the same parameters as in Section 3.2.2):

$$p_{6;\mathrm{G}}(W|M_{\mathrm{UV}}) = \frac{A(M_{\mathrm{UV}})}{W_{\mathrm{c}}(M_{\mathrm{UV}})}\sqrt{\frac{2}{\pi}}e^{-\frac{1}{2}\left(\frac{W}{W_{\mathrm{c}}(M_{\mathrm{UV}})}\right)^2} H(W) \\ + [1 - A(M_{\mathrm{UV}})]\delta(W) \tag{3.14}$$

We then interpret these mock observations with our fiducial pipeline, which uses the exponential EW distribution from Eq. 3.7.

The resulting 95% C.L. on the fractional errors are highlighted in gray in Fig. 3.11. In blue, we show the same 95% C.L. from Fig. 3.8, which use same EW distribution for the forward models and the mock observation. The fact that

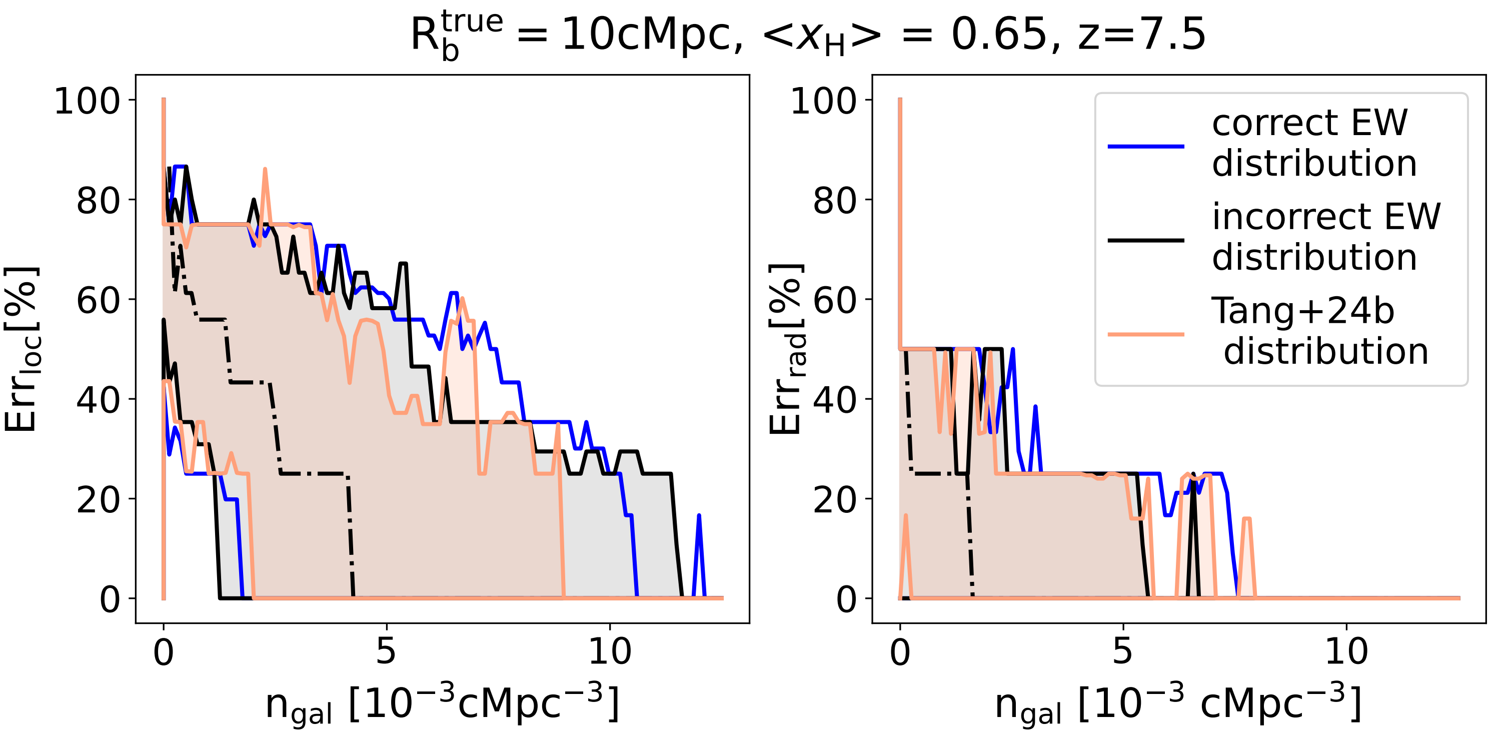

Figure 3.11: Same as Fig. 3.8, but including in gray (orange) the 95% C.L. on the fractional errors assuming that the EW distribution we use to interpret the mock observation is a Gaussian constructed from the same data (log-normal distribution from Tang, Stark, Topping, et al., 2024). In other words, the mock observations and forward models are generated using different EW distributions (see text for details). The fact that the C.L. demarcated in gray, orange and blue roughly overlap indicates that our results are not sensitive to our choice of the emergent EW distribution.

the gray and blue regions demarcate roughly the same fractional errors suggests that our analysis is not sensitive to the choice of EW distribution.

As a further test we repeat the analysis, but using a log-normal distribution from Tang, Stark, Topping, et al., 2024 which was constructed using completely separate data. The distribution is given with:

$$p(W|M_{\rm UV}) = \frac{1}{\sqrt{2\pi}\cdot\sigma_W\cdot W}\exp\left(-\frac{\log(W)-\mu_W)^2}{2\cdot\sigma_W^2}\right) \tag{3.15}$$

with $\mu_W = \log(5.0)$ and $\sigma_W = 1.74$. The results are shown with the orange color in Fig. 3.11. Again the blue and orange contours follow each other closely indicating that our pipeline is robust to the changes in the EW distribution.

3.5.2 *Demonstration on a 3D Reionization Simulation* Because this work is intended as a proof-of-concept, throughout we made several simplifying assumptions. For example, our reionization morphology is generated by overlapping ionized spheres and we have only a simplified treatment for galaxy – HII bubble bias. Including a more realistic bias for both galaxies and HII regions would be straightforward to do analytically, but it is necessary? Here we test how our simple model performs on self-consistent 3D simulations of reionization.

We apply our framework on the galaxy catalogs and ionization maps from the simulations of Lu, C. A. Mason, Andrei Mesinger, et al. (2025), generated with the public `21cmFAST` code (Andrei Mesinger and S. Furlanetto, 2007b; Andrei Mesinger, S. Furlanetto, and Cen, 2011a; S. Murray et al., 2020b). These simulations capture the complex morphology of reionization, which is self-consistently generated from the underlying galaxy fields. We process the galaxies with the procedure outlined in Section 3.2 to create mocks on which we can perform the inference.

We use two ionization boxes, one at $\overline{x}_{\rm HI} = 0.69$ and $\overline{x}_{\rm HI} = 0.90$. We select ionized bubbles and associated galaxies that match the volumes we use in our fiducial set-up, specifically 20x20x20 cMpc. When applying the inference framework to real data, one should of course tailor the forward models to match the specific details of the survey.

We illustrate the results in Figures 3.12 and 3.13, for a couple of example surveys at $\overline{x}_{\rm HI} = 0.69$ and $\overline{x}_{\rm HI} = 0.9$, respectively. In the left panels, we show 1.5 cMpc thick slices through the simulation boxes. Ionized/neutral regions are

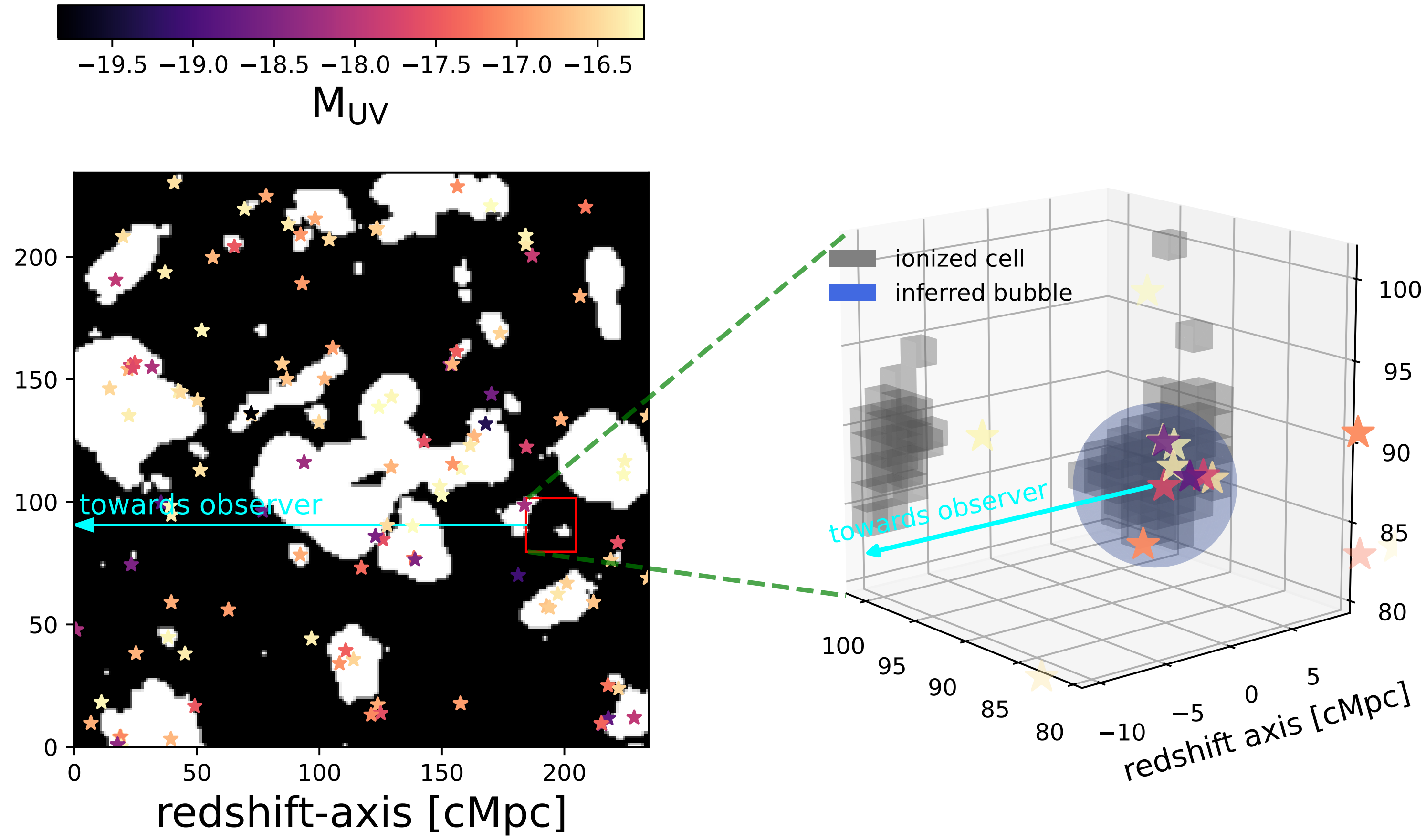

Figure 3.12: Application of our bubble-finding procedure to a 3D simulation. *Left:* 2D slices through the ionization field. Ionized/neutral regions are shown as white/black. The brightest galaxies in this 1.5 cMpc slice are denoted with stars, colored according to their UV magnitudes. The line of sight direction used to compute the IGM opacity for each observed galaxy is indicated with the arrow. The 20x20x20 cMpc sub-volume used for the mock survey is denoted with the red square. *Right:* Zoom-in of the mock survey. A random sub-sample (to avoid overcrowding) of the observed galaxies are shown with stars. Ionized voxels are shown in gray. Our maximum likelihood solution for the central HII bubble is overlaid with the blue sphere. The reasonable overlap of the central gray blob with the blue sphere lends confidence in our framework.

shown as white/black. The brightest galaxies are shown with stars, whose colors correspond to their UV magnitudes. The line of sight direction used to compute the IGM opacity for each observed galaxy is indicated with the arrow.

The volumes of the mock surveys are illustrated with the zoom-ins on the right. Here, gray cells show the ionized voxels of the simulation, and the blue sphere is the maximum likelihood solution for the central HII region. We assumed all galaxies inside this 20x20x20 cMpc volume brighter than $M_{\rm UV} < -17.0$ are observed with NIRSpec. This corresponds to $\sim 50(30)$ galaxies at $\overline{x}_{\rm HI} = 0.69$ ($\overline{x}_{\rm HI} = 0.9$).

We see from the zoom-ins in these figures that the inferred HII regions in blue provide a reasonable characterization of the "true" HII region in gray. This gives us confidence in our simplified treatment of the EoR morphology and the associated spatial correlation with galaxies. As mentioned above, we will tailor future applications to actual NIRSpec surveys of galaxy groups.

3.6 JWST Observational Requirements

We currently have several spectroscopically-targeted fields containing groups of galaxies with detection in Lyman-α, such as COSMOS (Endsley and Stark, 2022; Witten et al., 2024), EGS (Oesch, P. G. van Dokkum, et al., 2015; Zitrin et al., 2015; G. W. Roberts-Borsani et al., 2016; Tilvi, Malhotra, et al., 2020; Leonova et al., 2022b; R. L. Larson et al., 2022; Jung, Steven L. Finkelstein, R. L. Larson, et al., 2022; Tang, Stark, Z. Chen, et al., 2023; Tang, Stark, Topping, et al., 2024; Z. Chen, Stark, C. Mason, et al., 2024; Napolitano, Pentericci, et al., 2024), BDF (M. Castellano,

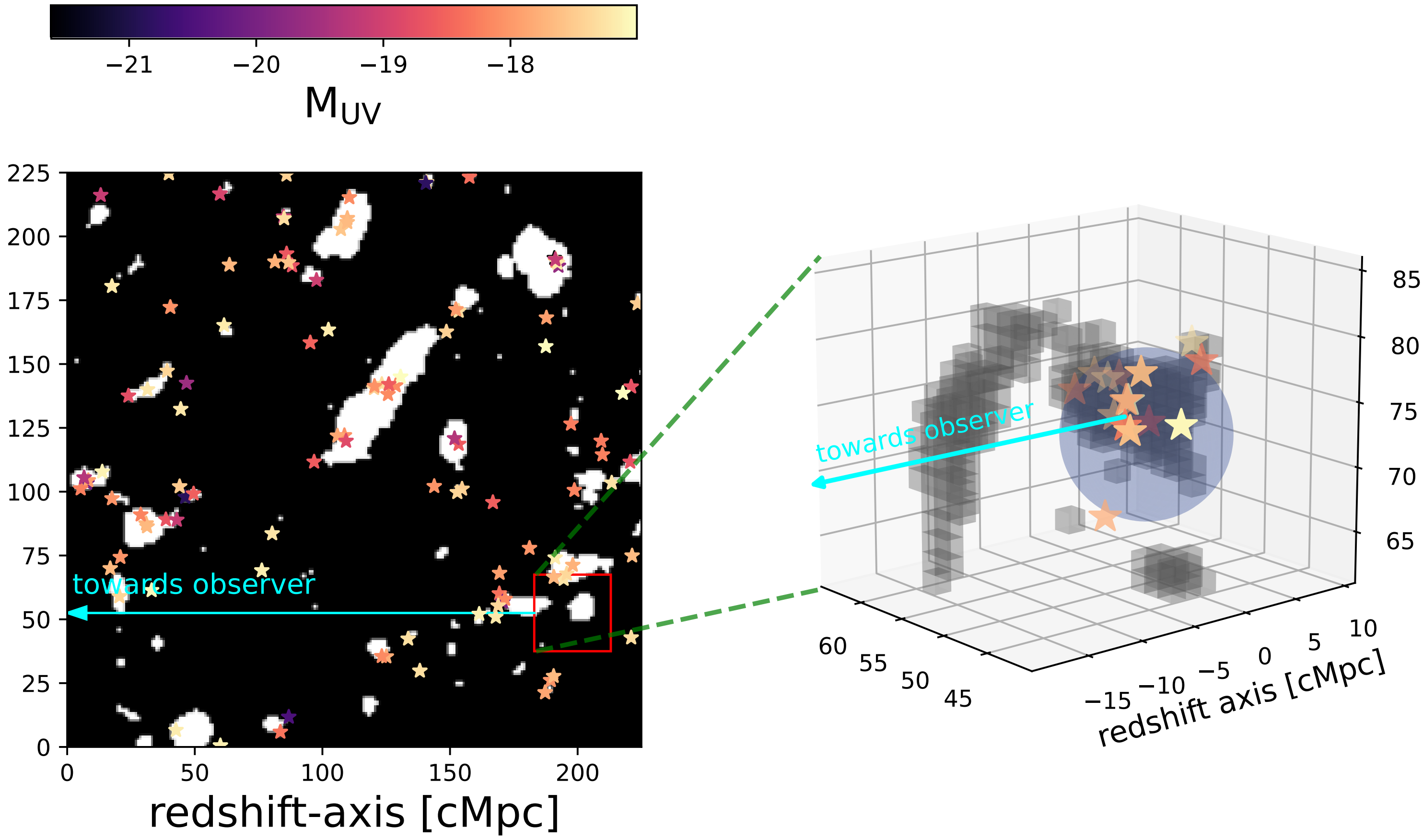

Figure 3.13: The same as Fig. 3.12, but for $\overline{x}_{\rm HI} = 0.9$.

P. Dayal, et al., 2016; M. Castellano, Pentericci, et al., 2018), GOODS-N (Oesch, G. Brammer, et al., 2016; Eisenstein et al., 2023; Bunker et al., 2023; Tacchella, Eisenstein, et al., 2023) and GOODS-S (Witstok, Smit, et al., 2024; Tang, Stark, Topping, et al., 2024). Although the number densities are a factor of few lower ($\sim 1.0 \times 10^{-3}$cMpc^{-3}) than the required values in Section. 4.3, the data sets are expanding rapidly. On-going and proposed programs are extending fields and going deeper, resulting in larger areas and number densities. Here we quantify in more detail the *JWST* survey requirements in order to be able to robustly constrain HII bubbles with our procedure.

In Fig. 3.14 we show the number density of galaxies as a function of limiting UV magnitude, $M_{\rm UV}^{\rm lim}$ and galaxy overdensity, $n(< M_{\rm UV}^{\rm lim}/\bar{n}(< M_{\rm UV}^{\rm lim})$.We plot three curves corresponding to $M_{\rm UV}^{\rm lim}$= -17, -18, -19. We demarcate in gray the required number density for accurate HII bubble recovery we found in the previous section; the lower/upper range correspond to including/not including a prior on the emergent Lyα emission from Balmer lines (c.f. Section 4.3).

At the mean number density, we would require the photometric survey to detect galaxies down to $M_{\rm UV}^{\rm lim} \lesssim$ -17.0 – -17.5. The quoted range spans what can be achieved with or without a prior on the Lyman alpha production rate from Balmer lines. In an unlensed field, obtaining such number densities would require roughly 600 hours of integration with NIRCam on JWST (estimated based on exposure times in Morishita et al., 2024, for 4 pointings): achievable, but ambitions.

However, several current fields are known to contain overdensities. For example, COSMOS contains a 140 pMpc^3 volume that is estimated to be three times overdense at $z \approx 6.8$ (Endsley and Stark, 2022), while EGS is estimated to contain a 12pMpc^3 volume that is also 3× overdense at $z \approx 8.7$ (Zitrin et al., 2015; Leonova et al., 2022b; R. L. Larson et al., 2022; Tang, Stark, Z. Chen, et al., 2023; Z. Chen, Stark, C. Mason, et al., 2024; Lu, C. A. Mason, Hutter, et al., 2024). Also, GOODS-S and GOODS-N contain 4x and 8x overdensities over volumes larger than the fiducial one used in this work (62 arcmin^{2}x$\Delta z \approx 0.2$ Tang, Stark, Topping, et al., 2024). Furthermore, one of the most distant observed LAE is also believed to be located in an overdensity (up to 24× overdense at $z = 10.6$ in 2.6pMpc^3 volume;Oesch, G. Brammer, et al. 2016; Bunker et al. 2023; Tacchella, Eisenstein, et al. 2023; see also Lu, C. A. Mason, Hutter, et al. 2024 for other examples). Observing a field with an overdensity of three (eight) would require photometric completeness down to $M_{\rm UV}^{\rm lim} \lesssim$ -18.5 – -18 (-19 – -18.5). This would require ”only” 50 (4.2) hours of integration with NIRCam. Note

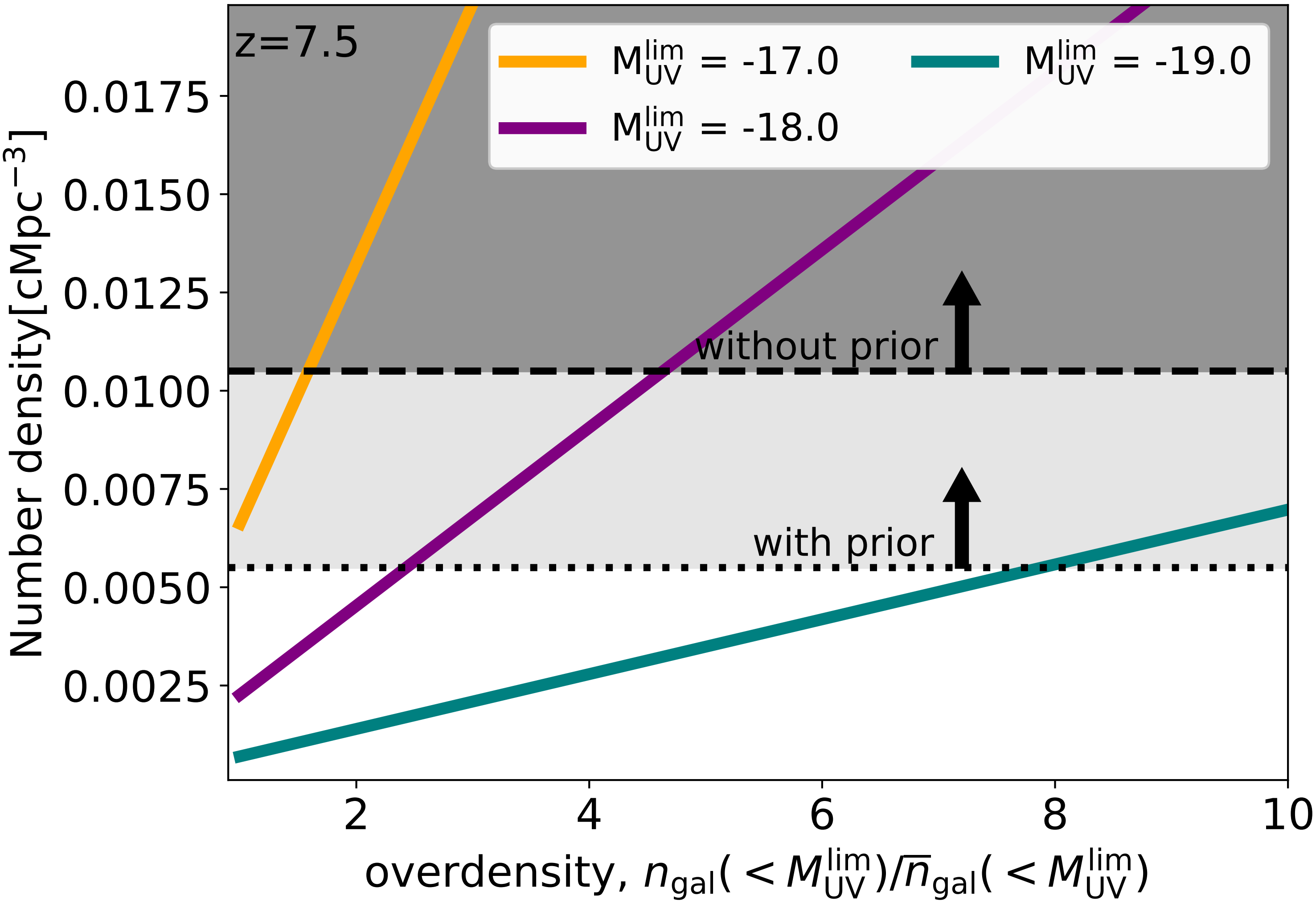

Figure 3.14: Cumulative number density of galaxies down to various magnitude limits, as a function of galaxy overdensity. Dotted (dashed) lines demarcate the minimum number densities found in this work for accurate HII bubble recovery, including (not including) a prior on the emergent emission of Lyman alpha.

that here we assume that emergent Ly-α properties remain the same depending on the overdensity (i.e. there is no environment dependencies in Sections 3.2.1 and 3.2.2).

These candidates then require spectroscopic follow-up. For spectroscopic confirmation, the best candidate is [OIII]5007$\AA$ emission line. The [OIII] emission line is routinely observed with JWST (e.g., Endsley, Stark, L. Whitler, Topping, Z. Chen, et al., 2023; Endsley, Stark, L. Whitler, Topping, Johnson, et al., 2024; R. A. Meyer et al., 2024). We can estimate the required exposure time for detecting [OIII] line by assuming the $M_{\rm UV}$ dependent distribution of equivalent widths from Endsley, Stark, L. Whitler, Topping, Johnson, et al. (2024). For that distribution, the equivalent width limit for detecting 90% galaxies at $M_{\rm UV} = -18.0$ is 120$\AA$. Obtaining such an equivalent width (at 5σ) at with the G395M filter requires 18 hours per pointing. The number density requirement (and thus exposure time) would increase if we do not detect Hβ as required to put a prior on the intrinsic Lyman-α emission. It would also increase if the [OIII] distribution is lower at lower magnitudes.

Following that, to detect Lyman α at a noise level of 5×10^{-19}erg s^{-1}cm^{-2} (corresponding to a detection of a line of W$\approx$ 25$\AA$) would require 23 hours with G140M per pointing. These numbers are comparable to existing large JWST surveys (Eisenstein et al., 2023).

We also note that the volume of the potential *JWST* surveys used for this analysis has a direct impact on the sizes of the HII regions that can be inferred (see also Lu, C. A. Mason, Andrei Mesinger, et al. 2025). Tiled surveys of the same depth but larger volumes, could hope to detect correspondingly-larger HII regions. The fact that larger HII regions occur later in the EoR, with a correspondingly higher galaxy number densities, might mitigate the integration time required. We postpone more systematic survey design to subsequent work in which we apply our framework to simulated 3D EoR lightcones.

Another interesting approach could be coupling the framework from Lu, C. A. Mason, Andrei Mesinger, et al. (2025) with the one outlined in this work. The large-scale empirical method from Lu, C. A. Mason, Andrei Mesinger, et al. (2025) could isolate "interesting" sub-volumes that could then be analyzed with the quantitative inference approach presented here. Finally, the pipeline we developed here could be used to target multiple fields and can be used to infer the sizes of multiple ionized bubbles which would help determine the bubble size distribution. Constraining it, in turn, would allow us answer which sources ionized the Universe.

3.7 Conclusions

The morphology of reionization tells us which galaxies dominated the epoch of cosmic reionization. Individual bubbles surrounding groups of galaxies encode information on the contribution of unseen, faint sources and allow us to correlate the properties of bubbles and the galaxies they host.

In this work we build a framework to use Lyman-α observations from *JWST* NIRspec to constrain ionization morphology around a group of galaxies. Our framework for the first time uses complementary information from sightlines to neighboring galaxies, and samples all important sources of stochasticity to robustly place constraints on the size and position of local HII bubbles.

We find that Lyα spectra from ~ 0.01 galaxies per cMpc^3 are required to be $\gtrsim$95% confident that the HII bubble location and size recovered by our method is accurate to better than $\lesssim 1.5$ cMpc. This corresponds roughly to 80 galaxies in 2x2 tiled pointings with *JWST*/NIRSpec. These requirements can be reduced by using additional nebular lines (for example Hβ) to constrain the intrinsic Lyman alpha emission. We find that a simple prior on the emergent Lyman-α emission reduces by a factor of $\sim$ two the required number of galaxies to obtain the same constraints. Such number densities are achievable with a targeted survey with completeness down to $M_{\rm UV} = -17 -$ -19, depending on the over-density of the field.

We demonstrate that our framework is not sensitive to the assumed distribution for the emergent Lyman alpha emission. We also find accurate recovery of ionized bubbles when applied to 3D EoR simulations.

Our pipeline can be applied to existing observations of Lyman-α spectra from galaxy groups. Additionally, observational requirements for a statistical detection of local HII regions presented here can be used to design complimentary, new *JWST* surveys. Applying our framework to multiple, independent fields would allow us to constrain the *distribution* of bubble sizes, which helps us understand which galaxies reionized the Universe.

3.8 Data Availability

The code related to the work is publicly available at IvanNikolic21/Lyman alpha bubbles.

3.9 Mapping the Joint Distribution Over All Flux Bins with Kernel Density Estimation

In Section 3.3 we showed how we generate forward models of JWST observations. Here we detail how we compute flux PDFs, i.e. likelihoods, in the space of local HII region parameters. In Fig. 3.15 in black we show a corner plot of forward models of a galaxy located in the center of its local HII region with size $R_{\rm b} = 10$cMpc. Black contours display 68% and 95% C.L. of the distribution respectively. We only show bins in our fiducial resolution (R 1000) where the Lyman-α emission is located. In order to fit a density estimator to obtain a smooth PDF, we scale fluxes by a common function:

$$s(f_\alpha(\lambda)) = 93.5 + 5\log\left(10^{-18} + 2f_\alpha(\lambda)\right). \tag{3.16}$$

Each column and row represents one wavelength bin of the flux scaled by the function in Eq. 3.16. We fit a Kernel Density Estimator to the forward models displayed in the corner plot with an exponential kernel of bandwidth $h = 0.12$. The scaling function and bandwidth were selected by optimizing the results of Section 4.3 (i.e., kernel density bandwidth and normalizing function that allows the inference of bubble properties for the least number of galaxies). We show contours of the fitted KDE in red in Fig. 3.15.

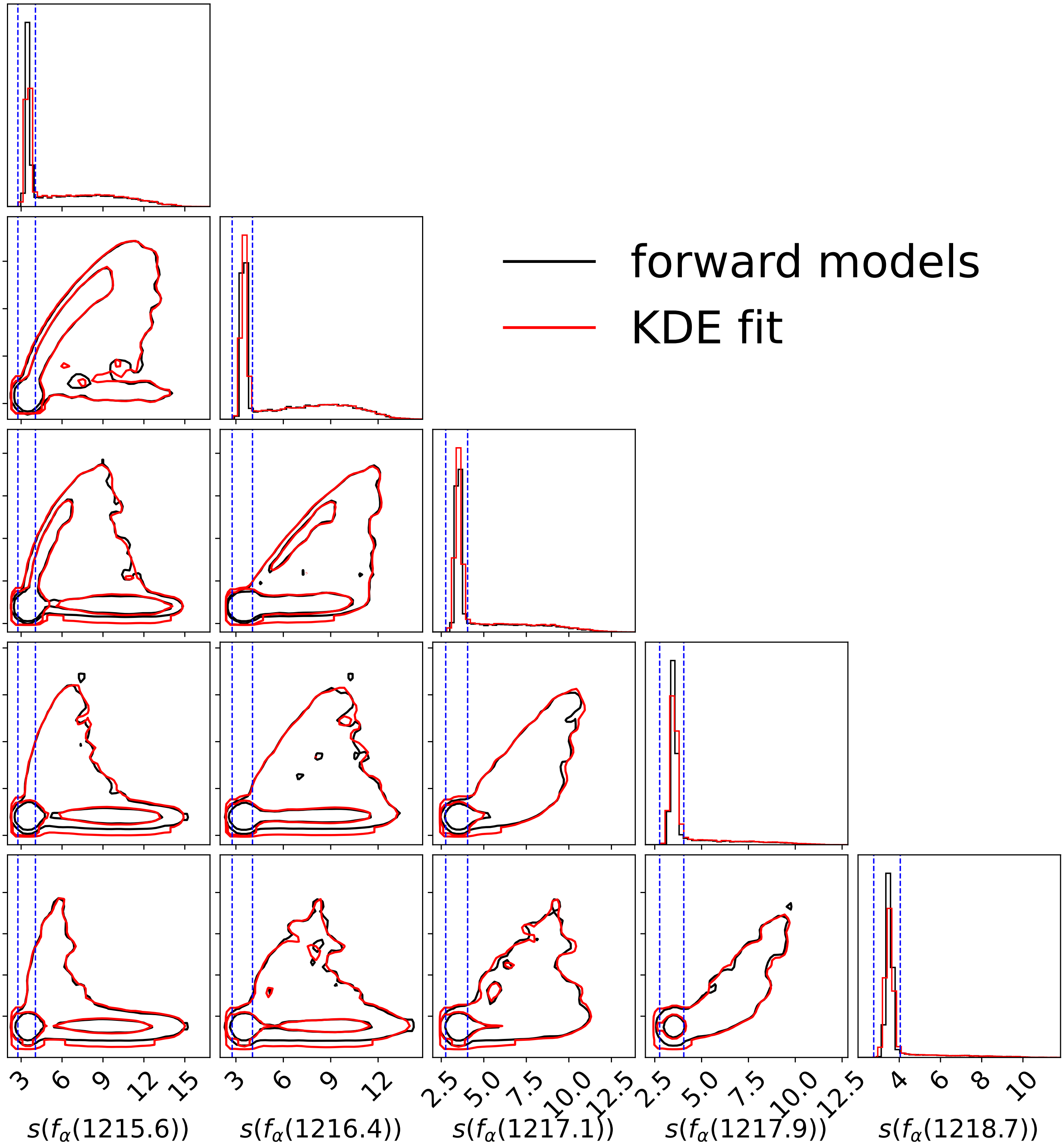

Figure 3.15: Corner plot of scaled flux spectra and the KDE fit to the forward models. Black contours represent 68% and 95% C.L. of the distribution of forward models for a galaxy located in the center of a $R_\mathrm{b} = 10$cMpc ionized bubble. Red contours represent same distributions for the KDE fit to the forward models. Blue dashed lines mark regions with $\pm 2\sigma$ of the noise of the spectrum. Contours outside of these dashed lines represent Lyman-α emission, with non-negligible correlation between bins.

Several features can be noted in Fig.3.15. Firstly, there is a strong peak for lower values of $s(f_\alpha(\lambda))$. This peak corresponds to the noise which is Gaussian by construction. This is further demonstrated with the blue dashed lines that show 5σ noise levels from Section. 3.2.4. On the other hand, there are additional features for larger values of $s(f_\alpha(\lambda))$. These features represent the Lyman-α signal that is higher than the noise. It is important to note that there is non-negligible correlation between different bins and the distribution is highly non-trivial. This shows that a simple likelihood that uses independent Gaussians for each bin would potentially bias the inference and an approach that fits the whole distribution is necessary.

KDE fits the data for all of the bins, in the noise and signal regimes. There is a small overestimation of the width of the distribution for bins that are noise-dominated. Since these bins do not give a lot of information for Lyman-α inference, our KDE is not biasing results.

4

The Importance of Stochasticity in Determining Galaxy Emissivities and UV LFs during Cosmic Dawn and Reionization

4.1 Introduction

The Universe underwent dramatic changes during the first billion years. Following cosmic recombination, the Universe was cold, dark and fairly empty. During the cosmic dawn (CD) when the first galaxies formed, their ultraviolet (UV) and X-ray radiation spread out, heating and ionizing the intergalactic medium (IGM). This culminated in the final major phase change of our Universe: the epoch of reionization (EoR; see for example reviews in Zaroubi 2013, Andrei Mesinger, Greig, and Emanuele Sobacchi 2016 and Pratika Dayal and Andrea Ferrara 2018).

Understanding how the first galaxies heated and ionized the Universe requires modeling their UV and X-ray emission, and constraining these models with data (e.g., Qin, Andrei Mesinger, Sarah E. I. Bosman, et al., 2021; Abdurashidova et al., 2022). The emission of any single galaxy is highly variable, depending on the time evolution of star formation, feedback, and geometry of interstellar absorption (e.g., Tacchella, Dekel, et al., 2016; Barrow, Wise, et al., 2017; Lovell et al., 2021; A. Pallottini and A. Ferrara, 2023). These processes are not known from first principles, and they are extremely challenging to simulate for a single galaxy, let alone for a cosmological sample of galaxies.

Luckily, the relevant cosmic radiation fields are sourced by the combined contribution from many galaxies, which allowed us to take advantage of the central limit theorem and use only average scaling relations to connect galaxy properties to their host dark matter halos (whose abundances and evolution are reasonably well known). This is the general approach taken by many analytic, semi-numerical and numerical models of the EoR and CD (e.g., Zoltán Haiman, Abel, and Martin J. Rees, 2000; B. Ciardi, A. Ferrara, and White, 2003; Steven R. Furlanetto, Zaldarriaga, and Hernquist, 2004; Andrei Mesinger, S. Furlanetto, and Cen, 2011a; Holzbauer and Steven R. Furlanetto, 2012; T. Fragos et al., 2013; Ross et al., 2017; Mirocha, La Plante, and A. Liu, 2021; Schaeffer, Giri, and Schneider, 2023).

However, it is not clear when is it safe to ignore galaxy-to-galaxy scatter (that is, stochasticity). Stochasticity can be important even when estimating globally averaged quantities such as the mean EoR history. Assuming population-averaged quantities (e.g., ionizing escape fraction and stellar to halo mass relation) can give biased results for correlated distributions (e.g., the average of a product is not the same as the product of the averages; c.f. Appendix 4.9 for simple examples). Moreover, some measurements (e.g., 21-cm interferometry, Lyα forest and kinetic Synaev-Zel'dovich signal) are sensitive to the spatial fluctuations in the galaxy emissivity, on some range of spatial scales. As that scale is reduced, there are fewer galaxies over which to average, and stochasticity becomes more important (e.g., F. B. Davies and Steven R. Furlanetto, 2016). The importance of stochasticity also increases at high redshifts, where sources are rarer and more biased. It has been evoked to explain controversial claims at $z > 10$, such as a rapid redshift evolution of the global 21cm signal during the CD (e.g., Kaurov et al., 2018), and an overabundance of massive galaxy candidates from JWST photometry at $z > 10$ (e.g., Mirocha and Steven R. Furlanetto, 2023b; C. A. Mason, Trenti, and Tommaso Treu, 2023; Shen et al., 2023; Guochao Sun et al., 2023; Kravtsov and Belokurov, 2024).

For this work we constructed a model of galaxy emissivity in the bands that are relevant for interpreting current and upcoming observations of the EoR and the CD: (i) ionizing UV (which drives the EoR and determines the residual HI fraction in the ionized IGM); (ii) soft X-ray (which heats and partially ionizes the IGM during the CD); and (iii) Lyman

Werner (which determines when H_2 cooling stops being efficient in the first galaxies). We computed the distribution of these multi-frequency emissivities as a function of scale and redshift. Our model sampled the largest expected sources of stochasticity, including the following: the abundance of dark-matter halos, the stellar-to-halo mass relation, the galaxy main sequence, the fundamental metallicity relation, luminosity, and escape fraction scalings. We quantified the relative importance of each term to the total emissivity in each of the considered bands. We also evaluated the importance of these stochastic terms for simple estimates of the EoR history, as well as the high redshift UV luminosity functions (LFs). Our results can be used to improve estimates of cosmic radiation fields and guide models of the EoR and CD by highlighting the most important sources of scatter.

The structure of this paper is as follows. In Section 4.2 we introduce our model for calculating galaxy emissivities. In Section 4.3 we present the resulting UV, X-ray, and LW emissivity distributions, quantifying the relative importance of each source of stochasticity. In Section 4.4 we show two analytic estimates of the EoR history, quantifying the relative impact of ignoring galaxy-to-galaxy scatter. In Section 4.5 we show the UV LFs implied by our fiducial model, comparing them to observational estimates from photometric candidates. Finally, we conclude in Section 4.6. All quantities are presented in comoving units unless stated otherwise. Throughout this work, we assume standard ΛCDM cosmological parameters $(\Omega_{\rm m}, \Omega_{\rm b}, \Omega_\Lambda, h, \sigma_8, n_{\rm s} = 0.310, 0.049, 0.689, 0.677, 0.81, 0.963)$, consistent with the latest estimates from Planck Collaboration, Aghanim, Akrami, et al. (2020).

4.2 Computing Emissivities at High Redshifts

If galaxy properties could be written as deterministic functions of the mass of their host halos and/or redshift, we could write the emissivity (e.g., erg s^{-1} cMpc^{-3}) in some spectral band, i, at a redshift z as:

$$\varepsilon_i(z) = \int \mathrm{d}M_h \frac{\mathrm{d}n(M_h, z)}{\mathrm{d}M_h} L_i(M_h, z)\ f_{\rm esc,i}(M_h, z)\ . \tag{4.1}$$

Here $\frac{\mathrm{d}n}{\mathrm{d}M_{\rm h}}$ is the number density of halos per unit mass – that is, the halo mass function (HMF) – L_i is the intrinsic luminosity of a galaxy hosted in a halo of mass M_h at redshift z, and $f_{{\rm esc},i}$ is the fraction of photons that escape the galaxy to make it into the IGM.

However, we know that the above relations are not deterministic functions of halo mass and redshift. Complex physics of galaxy evolution and radiative transfer induces a spread around relations linking different galaxy properties. Nevertheless, there are empirically well-established relations that characterize some of the main correlations of galaxy properties. Therefore, a more general form for the mean of the emissivity would marginalize over these relations. Specifically, we can write the emissivity in a spectral band, i, at redshift z, averaged over comoving volumes $(4/3)\pi R_{\rm nl}^3$, as follows:[1]

$$\begin{aligned}
\bar{\varepsilon}_i(R_{\rm nl}, z) = &\int \mathrm{d}M_h \int \mathrm{d}\delta_0 \frac{\mathrm{d}n(M_h, z \mid R_0, \delta_0)}{\mathrm{d}M_h} p_z(\delta_0 \mid R_{\rm nl}) && \text{(HMF)}\\
&\times \int \mathrm{d}M_*\ p(M_* \mid M_{\rm h}) && \text{(SHMR)}\\
&\times \int \mathrm{dSFR}\ p_z(\mathrm{SFR} \mid M_*) && \text{(SFMS)}\\
&\times \int \mathrm{d}Z\ p_z(Z \mid \mathrm{SFR}, M_*) && \text{(FMR)}\\
&\times \int \mathrm{d}L_i\ L_i\ p(L_i \mid \mathrm{SFR}, Z) && \text{(L)}\\
&\times \int \mathrm{d}f_{\rm esc,i}\ f_{\rm esc,i}\ p(f_{\rm esc,i}). && \text{(EF)}
\end{aligned} \tag{4.2}$$

Here $\delta_0(R_0)$ is the linear matter overdensity of a spherical volume of Lagrangian radius R_0 corresponding to the final Eulerian radius $R_{\rm nl}$, $p(M_* \mid M_{\rm h})$ is the conditional probability of stellar mass M_* for a given $M_{\rm h}$, $p_z(\mathrm{SFR} \mid M_*)$ is

---

[1] Throughout we use the subscript "nl" to indicate nonlinear (Eularian) quantities and the subscript "0" to indicate Lagrangian quantities linearly evolved to $z = 0$ (following convention). We recall that all length scales are in comoving units, unless otherwise specified.

the conditional probability of a star formation rate (SFR) for a given stellar mass[2], $p_z(\mathrm{Z} \mid \mathrm{SFR}, M_*)$ is the conditional probability of a stellar metallicity Z for a given SFR and M_*, $p(L_i \mid \mathrm{SFR}, Z)$ is the conditional probability of a luminosity L_i in a given wavelength band i for a given SFR and Z, and $p(f_{\mathrm{esc,i}})$ is the probability of an escape fraction $f_{\mathrm{esc,i}}$ in band i[3]. Loosely speaking, the running averages of the conditional probabilities in the first four rows are commonly referred to as the halo mass function (HMF)[4], stellar-to-halo mass relation (SHMR), star forming main sequence (SFMS) of galaxies, and fundamental metallicity relation (FMR); while (L) and (EF) represent running averages of conditional distributions of the luminosity and escape fraction at wavelength band i , respectively. We label the corresponding rows in the equation above with these acronyms, and go through each probability distribution in more detail below. In principle, the PDFs above could be conditioned on additional galaxy properties, which could further increase the importance of stochasticity. Note that higher order moments of the emissivity, such as its variance $\langle(\varepsilon_i - \bar{\varepsilon}_i)^2\rangle$, can be similarly expressed in terms of the above conditional probability distributions.

For general distributions, Eqs. 4.1 and 4.2 do not give the same mean. This means that even interpreting average quantities like the EoR history could be biased if not accounting for stochasticity. More fundamentally, the various sources of scatter in Eq. 4.2 result in spatial fluctuations in the emissivity which can be important for many EoR/CD observations. To date, the impact of this scatter on EoR/CD observables has only been explored in a limited fashion. For example, Hassan et al. (2022) found that scatter in the intrinsic production rate of ionizing photons predicted by the Simba simulation (Davé et al., 2019) has only a modest impact on the EoR morphology. On the other hand, Reis, Rennan Barkana, and Fialkov (2022) used a toy model to characterize the effective scatter in star formation efficiency, finding in some cases a large impact on the EoR and CD morphology; though see Murmu et al. (2024) for the opposite conclusions using a different astrophysical model. Indeed, a sizable scatter in the ionizing emissivity is needed to explain the latest Lyman alpha forest data at $z = 5 - 6.3$ (Qin, Andrei Mesinger, Sarah E. I. Bosman, et al. 2021; Gaikwad, Haehnelt, et al. 2023; F. B. Davies, Sarah E. I. Bosman, Gaikwad, et al. 2024, Qin et al. in prep.).

We could analytically derive the emissivity distribution, $p(\varepsilon_i \mid R_{\mathrm{nl}}, z)$, if the conditional distributions in Eqs. HMF-EF followed simple Gaussian forms. In the more general case, we can solve for $p(\varepsilon_i \mid R_{\mathrm{nl}}, z)$ by numerically sampling the above relations. Specifically, to compute a single realization (denoted below with a "tilde") of the emissivity, we perform the following Monte Carlo (MC) procedure and describe each step in turn below:

Algorithm 1 Algorithm 1: Computing a single realization of the emissivity in spectral band i of a region of radius R_{nl} at redshift z: $\tilde{\varepsilon}_i(R_{\mathrm{nl}}, z)$

```
sample the linear matter overdensity $\tilde{\delta}_0 \sim p_z(\delta_0 \mid R_{\mathrm{nl}})$
obtain a realization of the halo field by sampling the CHMF: $\widetilde{\{M_h^{\mathrm{j}}\}} \sim \mathrm{d}n(M_h, z \mid \tilde{\delta}_0, R_0)/\mathrm{d}M_h$
for all halos $j$ with mass $\tilde{M}_h^{\mathrm{j}}$ do
    sample probability that the halo hosts an actively star-forming galaxy, $p(t_{\mathrm{duty}} \mid \tilde{M}_h^{\mathrm{j}})$
    if halo does not host a star-forming galaxy then
        CONTINUE
    end if
    sample stellar mass $\tilde{M}_*^j \sim p(M_* \mid \tilde{M}_{\mathrm{h}}^j)$
    sample star formation rate $\tilde{\mathrm{SFR}}^j \sim p_z(\mathrm{SFR} \mid \tilde{M}_*^j)$
    sample metallicity $\tilde{Z}^j \sim p_z(Z \mid \tilde{\mathrm{SFR}}^j, \tilde{M}_*^j)$
    sample intrinsic luminosity $\tilde{L}_i^j \sim p(L_i \mid \tilde{\mathrm{SFR}}^j, \tilde{Z}^j)$
    sample escape fraction $\tilde{f}_{\mathrm{esc,i}}^j \sim p(f_{\mathrm{esc,i}})$
end for
$\tilde{\varepsilon}_i = \sum_{\mathrm{halo}=j} \tilde{L}_i^j \; \tilde{f}_{\mathrm{esc,i}}^j$
```

4.2.1 *Halo Mass Function (HMF)*

In this work we wish to compute the distribution of galaxy emissivities of regions of a given Eulerian scale at a given redshift, $p(\varepsilon_i | R_{\mathrm{nl}}, z)$. Our model is anchored by the fact that galaxies are hosted by

---

[2] We use "z" subscripts to indicate probability distributions that are also functions of redshift (see below for more details).

[3] Note that our choice of conditional distributions is motivated by well-known relations, but this choice is not unique. Furthermore, we do not assume any direct correlation of the ionizing escape fraction to other properties, as there is currently no consensus on what galaxy property would constitute an appropriate basis to characterize the distribution.

[4] Strictly speaking, the average HMF is not the mean of the conditional halo mass function (CHMF), but its limit as $R_0 \to \infty$ and $\delta_0 \to 0$.

dark matter halos, whose relative abundances are described by conditional halo mass functions.

Here we use the hybrid CHMF proposed by Rennan Barkana and Abraham Loeb (2004), in which the analytically-tractable Press-Schecher CHMF (Press and Schechter, 1974) is normalized to have the same mean as the (non conditional) Sheth-Tormen HMF (ST, Sheth and Tormen, 1999):

$$\begin{aligned}\frac{\mathrm{d}n}{\mathrm{d}M_\mathrm{h}}(M_\mathrm{h}, z \mid \delta_0, R_0) = \frac{\overline{f}_\mathrm{ST}}{\overline{f}_\mathrm{PS}}\sqrt{\frac{2}{\pi}}\frac{\overline{\rho}}{M_\mathrm{h}}\times \\ \times \frac{\delta_\mathrm{c}-\delta_0}{\sigma^2(M_h)-\sigma^2(R_0)}\left|\frac{\mathrm{d}\sigma}{\mathrm{d}M_\mathrm{h}}\right|\exp\left(-\frac{(\delta_\mathrm{c}-\delta_0)^2}{\sigma^2(M_h)-\sigma^2(R_0)}\right)\end{aligned} \tag{4.3}$$

In the above, $\overline{f}_\mathrm{ST}$ and $\overline{f}_\mathrm{PS}$ correspond to the mean Sheth-Tormen and Press-Schechter collapsed fractions above the atomic cooling threshold of $T_\mathrm{vir} \geq 10^4$ K, respectively, $\delta_\mathrm{c}(z)$ is the critical linear density from the spherical collapse model, and $\sigma^2(M)$ is the mass variance of the Lagrangian (linear) density field on scales $M = (4/3)\pi R_0^3$.

In order to sample from Eq. 4.3, we need to connect Lagrangian and Eulerian quantities (see also, e.g., Trapp and Steven R. Furlanetto 2020). In Lagrangian space, $p(\delta_0)$ follows a zero-mean Gaussian distribution whose width is determined by $\sigma^2(R_0)$. We transform this distribution to Eulerian space using the spherical collapse model (e.g., Mo and White 1996):

$$p(\delta_\mathrm{nl} \mid R_\mathrm{nl}, z)\mathrm{d}\delta_\mathrm{nl} = \frac{1}{1+\delta_0} f_\mathrm{R}(\sigma^2 \mid R_0, z)\mathrm{d}\sigma^2 \ , \tag{4.4}$$

$$\delta_0 = -1.35(1+\delta_\mathrm{nl})^{-2/3} + 0.78785(1+\delta_\mathrm{nl})^{-0.58661} - 1.12431(1+\delta_\mathrm{nl})^{-1/2} + 1.68647 \ . \tag{4.5}$$

Here f_R is the first-crossing distribution from Sheth (1998), and $\delta_\mathrm{nl} = \rho/\bar{\rho} - 1$ is the Eulearian (nonlinear) overdensity.

With the above relations, we generate a Lagrangian overdensity sample, $\tilde{\delta}_0 \sim p(\delta_0 \mid R_\mathrm{nl}, z)$ (step 1 of the MC procedure in the previous subsection). We then compute a corresponding realization of the halo field according to the following procedure. We first sample the total number of halos with masses above some arbitrary minimum value, obtaining $\tilde{N}(> M_\mathrm{min} \mid \tilde{\delta}_0, R_0)$, by assuming a Poisson distribution whose mean is given by the integral of Eq. 4.3 from M_min to infinity. We then assign each halo a mass by sampling the normalized cumulative mass function, using rejection to ensure the total mass is within ±10% of the target mean.[5]

We show some example CHMFs in Fig.4.1. The dashed black curves correspond to the target mean CHMFs in regions of Eularian scale R_nl, at mean density for infinite realizations, at redshifts $z = 5, 10$ and 15 (top to bottom panels). The red solid curves show a single realization, computed according to the above procedure, while the red shaded region corresponds to the 95% C.L. of the halo field conditioned on a region of scale $R_\mathrm{nl} = 5$Mpc. The impact of stochasticity in the red curves is very evident as the mean number decreases, that is, toward high redshifts and high masses. In the top panel we also show the mean (non conditional) HMF (that is, the limit as $R_0 \to \infty$ at $\delta_0 = 0$).

Fig.4.1 also highlights that there are effectively two sources of scatter when determining the halo abundances in a given volume: (i) the scatter in the mean value of the CHMF, driven by its dependence on the underlying matter overdensity (that is, the difference between the blue and black curves); and (ii) the scatter due to discrete sampling around the target mean CHMF (that is, the difference between the black and red curves). The former determines cosmological signals like 21cm since it is correlated to the underlying matter field. The latter on the other hand is effectively a sample noise term. Both sources of scatter are naturally accounted for in N-body simulations, although periodic boundary conditions mean that (i) is underestimated due to limited box sizes (e.g., Rennan Barkana and Abraham Loeb 2004). On the other hand, analytic and semi-numerical models of inhomogeneous radiation fields account for (i), but often assume (ii) is negligible in order to reduce computation costs. Below we confirm the validity of this approximation.

4.2.2 *Stellar-to-Halo Mass Relation (SHMR)* Both observations and theory established a strong relation between the stellar and halo masses of galaxies (e.g., Harikane, Ouchi, Ono, et al., 2016; Ceverino, Klessen, and Glover, 2018;

---

[5] Although approximate, our approach has a couple of notable advantages over other simple MC implementations of stochasticity in which halo numbers are sampled from independent Poisson distributions in fixed mass bins (e.g., Reis, Rennan Barkana, and Fialkov, 2022). Firstly, by sampling a continuous CDF, we avoid binning halo masses and forcing them to have discrete values. Moreover, the mean total number of halos is much larger than the mean number in any given mass bin, validating the assumption of a Poisson distribution. Furthermore, having a mass error threshold ensures approximate mass conservation in each realization. The alternative of not correlating halo samples in neighboring mass bins and not ensuring mass conservation can significantly overestimate the importance of stochasticity when halos become rare, which can explain why our results are different from those in Reis, Rennan Barkana, and Fialkov (2022).

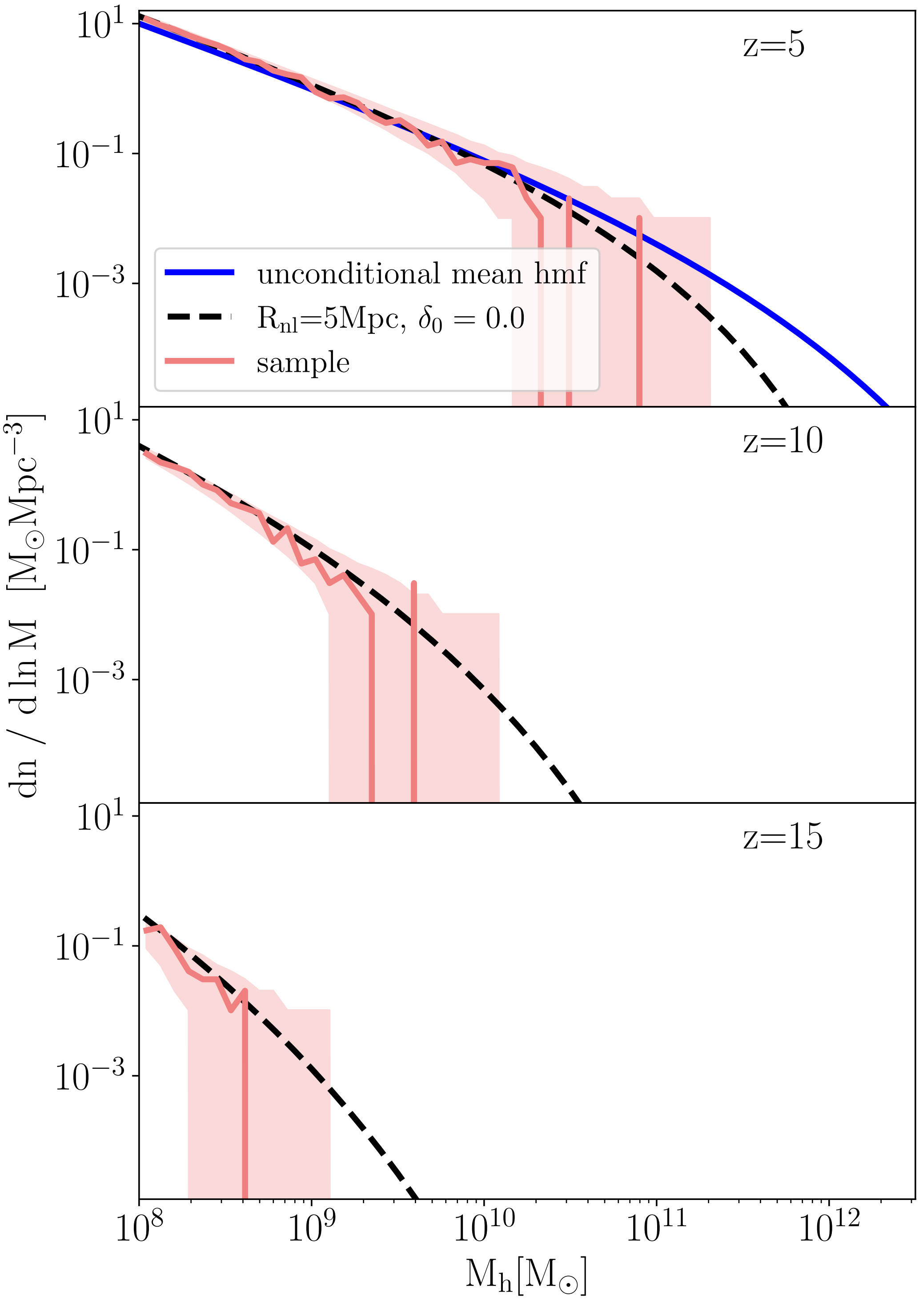

Figure 4.1: Example halo mass functions used in this work at three different redshifts. The dashed black curves and surrounding red regions correspond to the theoretical mean (Eq. 4.3) and 95% C.L. of the halo field conditioned on a region of scale $R_{\rm nl} = 5$ Mpc having a density equal to the cosmic average. The solid red curve corresponds to a single realization sampled from these distributions. The sample variance scatter in the red curve is seen to increase toward large masses and high redshifts, as the target mean values become smaller. In the top panel we also show in blue the non-conditional HMF (that is, the limit as $R_0 \to \infty$ at $\delta_0 = 0$).

Mauro Stefanon, Rychard J. Bouwens, et al., 2021; Lovell et al., 2021; Kannan et al., 2022; A. Pallottini, A. Ferrara, et al., 2022; Di Cesare et al., 2023). Here we assume a log-normal conditional probability of a galaxy having a stellar mass, M_*, given a host halo mass, M_h: $p(\log M_*|\log M_h) = \mathcal{N}(\log M_* \mid \mu_{M_*}(M_h), \sigma_{M_*})$. We assume a mass-independent σ_{M_*} of 0.25 dex (e.g., Ceverino, Klessen, and Glover, 2018; Lovell et al., 2021; A. Pallottini, A. Ferrara, et al., 2022) and a

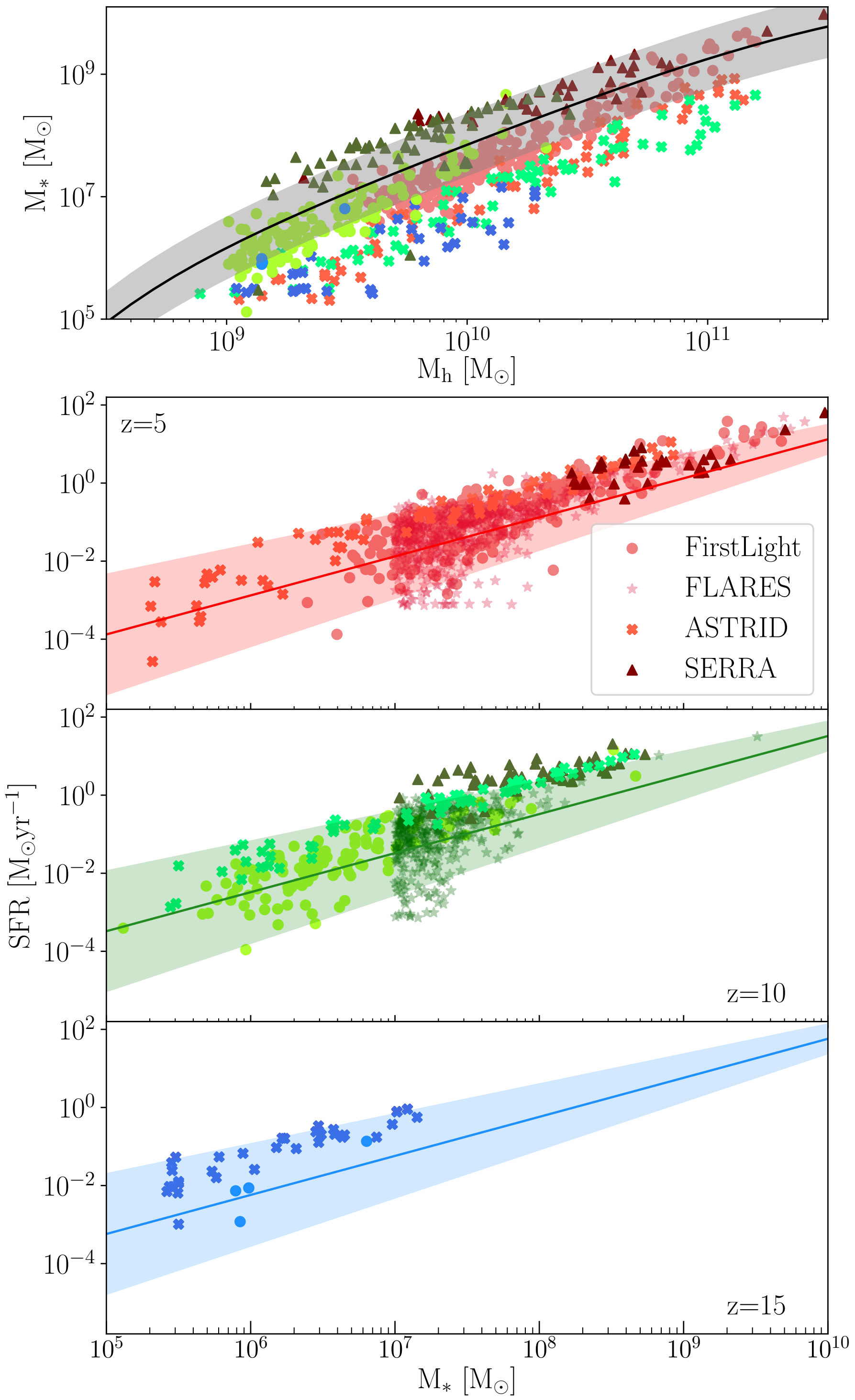

Figure 4.2: Stellar-to-halo mass mass relation and star formation main sequence used in this work. *Uppermost panel:* Our (redshift independent) stellar to halo mass relation (*solid curve*) and 2σ scatter (*shaded region*). *Lower panels:* Galaxy star-forming main-sequence (*solid curves*) and 2σ scatter (*shaded regions*) at $z = 5$, 10, 15 (*top to bottom*).) Coloured symbols represent galaxies from cosmological simulations, circles for FirstLight (Ceverino, Klessen, and Glover, 2018), stars for FLARES (Lovell et al., 2021), crosses for ASTRID (Bird et al., 2022; J. E. Davies, Bird, et al., 2023) and triangles for SERRA (A. Pallottini, A. Ferrara, et al., 2022). For ASTRID we randomly select galaxies in fixed mass bins, to avoid over crowding the plot, and for SERRA we use their $z = 6$ and $z = 12$ snapshots for $z = 5$ and 10, respectively.

mean given by the following double power law SHMR:

$$\mu_{M_*}(M_h) = -1.412 + \log M_{\rm h} - \log\left[\left(\frac{M_{\rm h}}{2.6\times10^{11}M_\odot}\right)^{-0.5} + \left(\frac{M_{\rm h}}{2.6\times10^{11}M_\odot}\right)^{0.6}\right]. \tag{4.6}$$

A standard physical interpretation of the double power law form is that the low-mass scaling is determined by stellar feedback while the high-mass scaling is determined by AGN feedback (e.g., Wechsler and Tinker 2018; P. Behroozi et al. 2019 and references therein). In Eq. (4.6) the normalization and the low-mass power-law index correspond to the maximum a posteriori (MAP) values inferred from a combination of CMB, QSO and high-redshift UV LF observations in Nikolić, Andrei Mesinger, Qin, et al. (2023), while the high-mass power-law index is taken from the bright-end UV LF empirical fits in Mirocha, Steven R. Furlanetto, and Guochao Sun (2017). Our results mostly depend on the former, as the steepness of the HMF at high redshifts means that early radiation fields are dominated by the faint (low mass) galaxies (e.g., R. J. Bouwens, Illingworth, Oesch, et al., 2015; Rychard J. Bouwens et al., 2023a; N. J. F. Gillet, Andrei Mesinger, and J. Park, 2020, see also below).

Gas accreting from the IGM onto halos is gravitationally heated, and can also be photo-heated by the ionizing UV background (UVB) during the EoR. In order to condense onto the galaxy and form stars, this gas needs to cool. Cooling can be inefficient in halos with small virial temperatures, with an exponentially decreasing fraction of halos capable of sustaining star formation (e.g., E. Sobacchi and A. Mesinger, 2013; H. Xu, Wise, et al., 2016). Here we account for this effect by assuming only a fraction $f_{\rm duty}(M_{\rm h}) = \exp[-M_{\rm turn}/M_{\rm h}]$ of halos host star-forming galaxies, taking $M_{\rm turn} = 5\times10^8 M_\odot$ based on the inference result in Nikolić, Andrei Mesinger, Qin, et al. (2023). Specifically, for each halo we sample a random variable uniformly between 0 and 1, and only populate the halo with a star forming galaxy if the value of the random variable is less than $f_{\rm duty}$.

In the top panel of Fig. 4.2 we show our mean SHMR and 2σ scatter (solid black line and gray shaded region, respectively). The mean SHMR is a power law over most of the mass range shown. At the high (low) mass end we see a flattening due to our parametrization of AGN feedback (inefficient accretion), as discussed above. For comparison, we also show galaxies from several hydrodynamic simulations: FirstLight (Ceverino, Klessen, and Glover, 2018), ASTRID (Bird et al., 2022; J. E. Davies, Bird, et al., 2023), and SERRA (A. Pallottini, A. Ferrara, et al., 2022). The simulated galaxies are colored according to their redshift, with red for $z = 5$, green for $z = 10$ and blue for $z = 15$.

We see significant differences in the (mean) SHMR between different simulations. The cosmological zoom-in SERRA simulations imply a mean SHMR that is roughly two orders of magnitude higher at the smallest halo masses compared with the ASTRID simulations. FLARES and FirstLight are somewhere in between these two extremes, as is our fiducial model. We remind our reader that our mean relation was inferred from data, as discussed in Nikolić, Andrei Mesinger, Qin, et al. (2023), and not based on these simulations.

Conversely, our choice of 0.25 dex scatter around the mean relation is roughly motivated by the galaxy-to-galaxy scatter found in any given hydrodynamic simulation. This scatter is driven primarily by stellar/AGN feedback and mergers. As our fiducial choice of scatter is motivated by the simulations, we are implicitly including these effects; however, our parametric approach can be used to infer the mean and scatter in these relations from data in a simulation-agnostic manner. Interestingly, despite the fact that different simulations predict different means, the scatter around the mean is roughly comparable. Furthermore, we see that the simulations do not show strong evidence of a redshift evolution of the SHMR, justifying our fiducial model (see also, e.g., Mutch et al., 2016; Harikane, Ouchi, Ono, et al., 2016; Tacchella, Dekel, et al., 2016; X. Ma, Hopkins, et al., 2018; Yung, Somerville, Popping, et al., 2019).

4.2.3 *Galaxy Star Formation Main Sequence (SFMS)* The SFRs of galaxies are known to be strongly correlated with their stellar mass content. The mean of this SFR – M_* relation is loosely referred to as the galaxy SFMS; galaxies with SFRs significantly above (below) the SFMS are referred to as bursty (quenched). The SFMS is well established observationally at low redshifts and (comparably) large masses (e.g., Brinchmann et al., 2004; Santini et al., 2017; E. Curtis-Lake et al., 2021; Popesso et al., 2023). The observed mean relation at small masses follows a power law, whose index is fairly constant but whose normalization decreases with redshift. This decrease with redshift is naturally reproduced if one assumes that the star formation timescale is related to the free-fall time at the mean viral density of host halos, $t_{\rm ff}$, which during matter domination scales as the Hubble time: $t_{\rm ff} \propto H^{-1}(z)$ (e.g., see J. Park, Andrei Mesinger, et al., 2019, and references therein).

At a given z, we again assume a log-normal conditional probability $p_z(\log{\rm SFR} \mid \log M_*) = \mathcal{N}[\log{\rm SFR} \mid \mu_{\rm SFR}(M_*, z), \sigma_{\rm SF}$ For the mean SFMS we use the model of J. Park, Andrei Mesinger, et al. (2019), with the normalization set by the

MAP values in Nikolić, Andrei Mesinger, Qin, et al. (2023):

$$\mu_{\mathrm{SFR}}(M_*, z) = \log M_* - \log\,[0.43 H^{-1}(z)]. \tag{4.7}$$

We assume a mass-dependent scatter that increases toward smaller masses, as these galaxies are expected to be more bursty[6]:

$$\sigma_{\mathrm{SFR}}(M_*) = \begin{cases} -0.12\log M_* + 1.35 & \text{if } \log M_* < 10 \\ 0.19. & \text{otherwise} \end{cases} \tag{4.8}$$

The normalization and scaling of the scatter was fit to the hydrodynamic simulations of Ceverino, Klessen, and Glover (2018), but the mean (that is, Eq. 4.7) is the same as the one in Nikolić, Andrei Mesinger, Qin, et al. (2023). Just like with the SHMR, the scatter is mostly driven by galactic feedback and mergers.

We plot our assumed SFMS and 2σ scatter in the bottom three panels of Figure 4.2. Panels correspond to $z = 5$, 10, 15 (top to bottom), with different symbols indicating values taken from hydrodynamic simulations: FirstLight (Ceverino, Klessen, and Glover, 2018), FLARES (Lovell et al., 2021), ASTRID (Bird et al., 2022), and SERRA (A. Pallottini, A. Ferrara, et al., 2022). The figure illustrates that our fiducial model is in general agreement with results from these hydrodynamic simulations.[7]

4.2.4 *Fundamental Metallicity Relation (FMR)* The galaxy emissivity also depends on the metallicity of the stellar population. Here we relate the metallicity to the SFR and stellar mass of a galaxy, taking advantage of the well-studied fundamental metallicity relation (FMR; F. Mannucci et al. 2010; Curti, Filippo Mannucci, et al. 2020b). Specifically, we assume a log-normal conditional probability of a galaxy having a stellar metallicity Z, given its SFR and stellar mass, $p_z(\log Z \mid \log \mathrm{SFR}, \log M_*) = \mathcal{N}[\log Z \mid \mu_Z(M_*, \mathrm{SFR}, z), \sigma_Z]$. We assume a constant scatter of $\sigma_Z = 0.1$ dex, and a mean given by the following (c.f. Curti, Filippo Mannucci, et al. 2020b):

$$\mu_Z(M_*, \mathrm{SFR}, z) = 0.296 \left(1 + \left(\frac{M_*}{M_0}\right)^{-2.1}\right)^{-0.148} 10^{\Delta_z}\ Z_\odot\ , \tag{4.9}$$

where $M_0(\mathrm{SFR}) \equiv 10^{10.11} \times (\mathrm{SFR}/\mathrm{M_\odot yr^{-1}})^{0.56}\mathrm{M_\odot}$, and $\Delta_z = -0.056z + 0.064$ accounting for putative redshift evolution (Curti, Maiolino, et al., 2024). In the above, we converted from gas phase to stellar metalicities using $Z/Z_\odot = 10^{(12+\log\,(\mathrm{O/H})-8.69)}$ with solar metallicity $Z_\odot = 0.02$ (Asplund et al., 2004), and adjusting for gas phase metallicities being higher by a factor of ≈ 2.63 on average (Strom et al., 2018).

In Fig. 4.3 we show galaxies from a single realization of a comoving volume with radius $R_{\mathrm{nl}} = 5$ cMpc, at mean density at redshift 6. Each point denotes a single galaxy with the color corresponding to its typical stellar metallicity. It is important to note that the apparent scatter in the metallicity at a fixed M_* is considerably larger than the 0.1 dex scatter we set around the mean FMR at a given M_* and SFR. This is because the intrinsic scatter in the combination of SHMR and SFMS (that is, the width of $p_z(\mathrm{SFR}, M_*)$ from the previous sections) dominates over our choice of scatter in the metallicity given these properties (that is, the width of $p_z(Z \mid \mathrm{SFR}, M_*)$; see also e.g., Garcia et al. 2024). In the figure we also show the binned values for the metallicity of $z > 6$ galaxies from Curti, Maiolino, et al. (2024) as well as the metallicity estimate of GN-z11 from Bunker et al. (2023). Although these observations span a range of redshifts, they are generally consistent with our samples.

4.2.5 *Luminosity Scalings* The intrinsic luminosity of a galaxy depends primarily on the SFR and its history, as well as the metallicity of the stellar population (e.g., G. B. Brammer, Pieter G. van Dokkum, and Coppi, 2008; Allende Prieto et al., 2018; E. R. Stanway and Eldridge, 2018; Bret D. Lehmer et al., 2021; Tassos Fragos et al., 2023). Here we describe how we compute the intrinsic luminosities for each of the wavelength bands of interest: X-ray, ionizing and Lyman Werner.

---

[6] Throughout this work we use "burstiness" to indicate a wide scatter around the mean SFMS. We do not investigate what such distributions imply for the star formation histories of individual galaxies.

[7] Detailed comparisons to other works would require standardizing definitions. For example, here we define the SFR as an instantaneous quantity, while elsewhere it could be averaged over ~ 100 Myr to allow for a more direct comparison to photometric observations. Here we are just interested in confirming that our fiducial choices are reasonable.

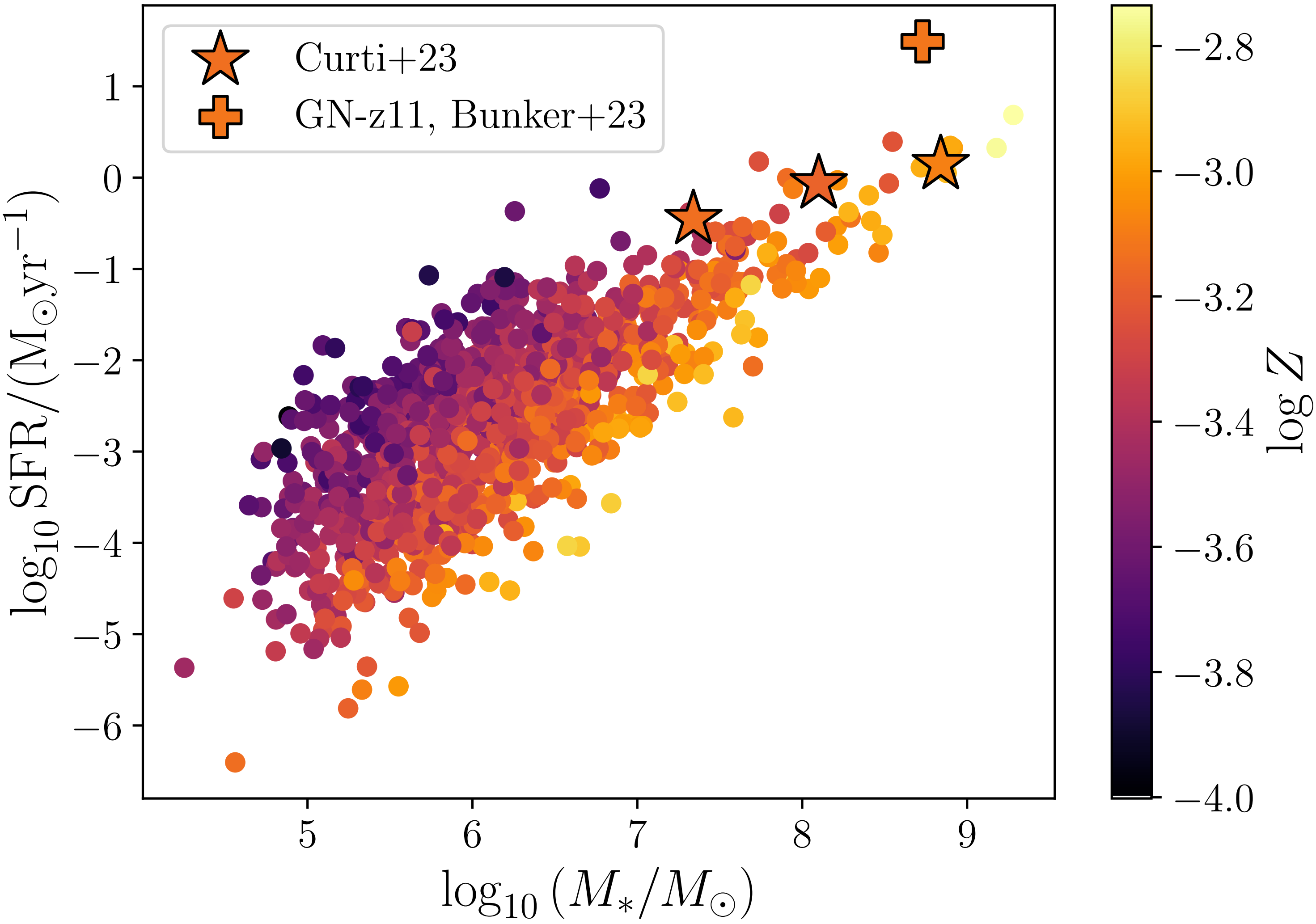

Figure 4.3: Fundamental metallicity relation used in this work. Each point corresponds to a stellar mass and star formation rate of a galaxy from a single realization of a $R_{\rm nl} = 5$ cMpc volume at mean density at $z = 6$, with the color denoting its metallicity: $\tilde{Z}^j \sim p_z(Z \mid \tilde{\mathrm{SFR}}^j, \tilde{M}_*^j)$ (see text for details). Stars denote the observationally estimated means in three bins for $z > 6$ galaxies in Curti, Maiolino, et al. (2024) (converted from gas metalicities, see text for details), as well as the metallicity estimate of GN-z11 from Bunker et al. (2023).

Soft-Band X-ray Luminosity Soft-band[8] X-rays emerging from the first galaxies are responsible for heating and partially ionizing the IGM during the cosmic dawn (e.g., McQuinn, 2016), which can have a dramatic imprint in the cosmic 21cm signal (e.g., Andrei Mesinger, Andrea Ferrara, and Spiegel, 2013; Abdurashidova et al., 2022). It is likely that the X-ray emissivity of $z > 6$ galaxies is dominated by high-mass X-ray binaries (HMXBs; e.g., Steven R. Furlanetto, McQuinn, and Hernquist, 2006; T. Fragos et al., 2013; Pacucci, Andrei Mesinger, et al., 2014; Eide et al., 2018). HMXBs are massive stars accreting onto a compact companion. The total X-ray output of a galaxy from HMXBs should therefore scale with the SFR of the galaxy (due to the rapid stellar evolution timescales of massive stars) and its metallicity (which determines the efficiency of radiative-driven winds and the resulting mass loss of the massive companion). Indeed we observe a strong dependence of the X-ray luminosity on the galaxy's SFR and metallicity in local galaxies and in stacks out to $z \sim 2.5$ (e.g., Brorby et al., 2016; B. D. Lehmer et al., 2016; Fornasini et al., 2019; Bret D. Lehmer et al., 2021).

Here we assume a log-normal conditional probability of a galaxy having an intrinsic soft-band X-ray luminosity, $L_{\rm X}$ (in units of erg s^{-1}), given a SFR and metallicity: $p(\log L_{\rm X} \mid \log \mathrm{SFR}, \log Z) = \mathcal{N}(\log L_{\rm X} \mid \mu_{\rm X}(\mathrm{SFR}, Z), \sigma_{\rm X})$. We

---

[8] Here we define the soft band to be $0.5 - 2$ keV. Roughly speaking, photons with higher energies do not interact with the high-z IGM (e.g., Oh, 2001; H. Xu, Ahn, et al., 2014; Madau and Tassos Fragos, 2017), while photons with lower energies get absorbed inside the host galaxies (e.g., A. Das et al., 2017, see also Section 4.2.6).

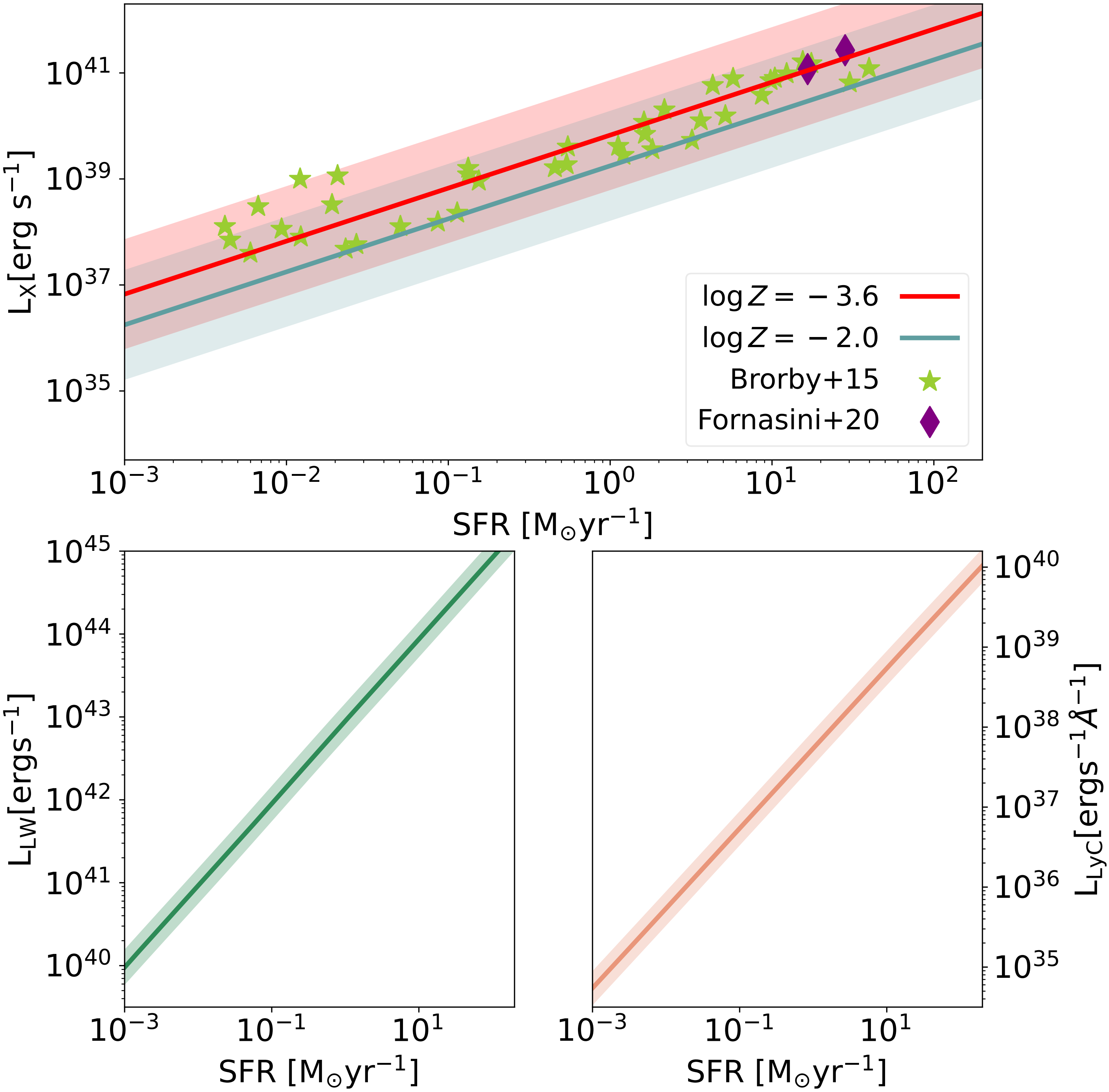

Figure 4.4: Luminosity scalings used in this work. *Upper panel:* scaling of soft-band, high-mass X-ray binary luminosity with SFR. Red and blue lines with the corresponding shaded regions represent the mean and 2σ range for metalicities $Z = -3.6, -2.0$, respectively. Green stars correspond to values from local star forming galaxies, discussed in Brorby et al. (2016). Purple diamonds are redshift-binned stacks from Fornasini et al. (2019) for $z = 1.5$ and 2.3. *Lower panels:* SFR scaling of the integrated Lyman-Werner (11.2 – 13.6 eV) luminosity (*left panel*) and specific ionizing luminosity at the Lyman-limit (*right panel*). Shaded regions represent the 2σ scatter around the mean relation. Both panels assume that metallicity follows the mean FMR.

assume a constant σ_X of 0.5 dex and a mean given by:

$$\mu_X(\mathrm{SFR}, Z) = \log \mathrm{SFR} + 40.5 + \log\left[\left(\frac{Z/Z_\odot}{0.05}\right)^{0.64} + 1\right]. \tag{4.10}$$

These fiducial choices are obtained by assuming a double power-law function for the X-ray luminosity function that results in a flattening at lower metallicities and fits the data at the high SFR/metallicity end (e.g., T. Fragos et al., 2013; Bret D. Lehmer et al., 2021; Kaur, Qin, et al., 2022; Geda et al., 2024). They are roughly consistent with empirical fits to local galaxies, which could however suffer from incompleteness at the lowest SFR bins (e.g., Brorby et al. 2016). We have converted the hard band X-ray luminosity in Bret D. Lehmer et al. (2021) (0.5-8 keV) to the soft band one by multiplying the luminosities by a factor of 0.3, consistent with observational estimates (e.g., Basu-Zych et al., 2013) and corresponding to an intrinsic SED with a power-law index of $\Gamma = 2.0$ (e.g., Mineo, Gilfanov, and R. Sunyaev, 2012).

We show this dependence of the X-ray luminosity with SFR, for two different metallicity values, in the top panel of Figure 4.4. For comparison, we include redshift-binned stacked observations of star-forming galaxies at $z \sim 2$ from Fornasini et al. (2019) and Lyman-Break analogs from Brorby et al. (2016). We see that our fiducial model is consistent with current data; however, it is highly uncertain how these relations scale to the first galaxies whose metallicity ranges are not sampled by current observations (e.g., Magg et al., 2022; Kaur, Qin, et al., 2022).

Ionizing and Lyman Werner Luminosities We use the Binary Population and Spectral Synthesis (BPASS) code to compute intrinsic ionizing and Lyman Wener UV luminosities (E. R. Stanway and Eldridge, 2018; Byrne, E. R. Stanway, et al., 2022). BPASS provides a deterministic prediction for the UV luminosity, $L_{\rm UV}$, as a function of SFR, Z, and SFR history. For the latter we assume that our sampled SFR is exponentially declining toward higher redshifts, as implied by Equation 4.7. Therefore $L_{\rm UV}$ is sampled assuming a log-normal conditional probability $p(\log L_{\rm UV} \mid \log {\rm SFR}, \log Z) = \mathcal{N}(\log L_{\rm UV} \mid \mu_{\rm UV,BPASS}({\rm SFR}, Z), \sigma_{\rm L})$ where $\mu_{\rm UV,BPASS}$ is the predicted luminosity from BPASS. We add an additional scatter of $\sigma_{\rm L} = 0.1$ dex around the mean to compensate for unaccounted sources of stochasticity, for example, the mean IMF, alpha-element distribution, etc. (Byrne and E. R. Stanway, 2023). However, this level of scatter is negligible compared to the scatter of the bulk galaxy properties like SFR and stellar mass. We show the scaling relation of $L_{\rm UV}$ with SFR in the bottom panels of Fig. 4.4 for the 11.2–13.6 eV Lyman-Werner band (left panel; in units of erg s^{-1}) and Lyman-limit (right panel; in units of erg s$^{-1}\mathring{A}^{-1}$ evaluated at 13.6 eV).

4.2.6 *Escape Fractions* Our final step in computing the emissivity is determining what fraction of the produced photons manage to escape the host galaxy into the IGM. This is referred to as the escape fraction. We use different prescriptions for the escape fraction in our three bands of interest. We describe each in turn below.

Both hydrodynamic simulations (e.g., Cen and Kimm, 2015; H. Xu, Wise, et al., 2016; Barrow, B. E. Robertson, et al., 2020; Yeh et al., 2022; Kostyuk, Nelson, et al., 2023) and direct observations of low redshift galaxies (e.g., Izotov et al., 2016; Grazian et al., 2017; Steidel, Bogosavljević, et al., 2018; Pahl et al., 2023) show sizable stochasticity in the ionizing escape fraction, though there is no consensus on what is an appropriate distribution. Here we take two scenarios. Our fiducial model assumes a log-normal distribution for the ionizing escape fraction with a width of 0.3 dex (c.f. Mascia, Pentericci, Calabrò, et al. 2023), while we also show a bimodal distribution in which galaxies have an ionizing escape fraction of either 0 or 1 (resulting in maximum scatter). In both cases we take the inference result of Nikolić, Andrei Mesinger, Qin, et al. (2023): $\bar{f}_{\rm esc} = 0.053$ for the mean. We do not assume that the mean or scatter of the escape fraction depend on galaxy properties, as such dependencies are not yet well established for unbiased galaxy samples, especially at high redshifts. For example, assuming that scatter depends on mass could accelerate or decelerate reionization, by effectively shifting the population-averaged mean as a function of redshift. Our framework can easily be extended to include putative dependencies on galaxy properties (e.g., Mascia, Pentericci, Calabrò, et al., 2023), as well as accommodating different functional distributions (e.g., Kreilgaard et al. 2024). We defer such studies to future work.

For the X-ray escape fraction, we adopt the results of A. Das et al. (2017), where they computed the X-ray opacities of simulated high-z galaxies, finding that most photons with energies above 0.5 keV manage to escape. Following that work, we assume an escape fraction of unity above 0.5 keV and zero below that value. Similarly, we assume values of unity for the Lyman Werner escape fraction, given the typical low opacities of such photons through the host ISM (e.g., Zoltán Haiman, Abel, and Martin J. Rees, 2000; J. Wolcott-Green, Z. Haiman, and G. L. Bryan, 2011).

4.3 Results: Emissivities

Here we present our distributions for the emissivity in each of the three bands in turn. We show the full distributions as a function of redshift, before quantifying the relative importance of each source of scatter. For the latter, we compute the mean and standard deviations of the emissivity PDF when one source of scatter is removed (that is, using only the corresponding mean relation with zero scatter), normalized to the values of the full distribution containing all sources of scatter: $\mu_\epsilon/\mu_\epsilon^{\rm full}$ and $\sigma_\epsilon/\sigma_\epsilon^{\rm full}$. As mentioned above, we consider the following sources of scatter:

(i) spatial dependence of the mean CHMF on the large scale matter density

(ii) Poisson sample variance in halo number around the target mean CHMF

(iii) scatter around the SHMR

(iv) scatter around the SFMS

(v) scatter around the FMR

(vi) scatter in the mapping of the intrinsic luminosity to SFR, M_* and Z

(vii) scatter in the escape fraction

We need to define a comoving volume over which to sum up the contributions of galaxies, in order to compute the emissivity PDFs. Here we chose a fiducial scale of $R_{\rm nl} = 5$ cMpc. This is roughly comparable to several relevant scales during the EoR and CD: (i) the typical HII bubble sizes during the early-middle stages of reionization (e.g., McQuinn, Lidz, et al., 2007; Y. Lin et al., 2016); (ii) the resolution of 21cm maps achievable after a 1000h observation with SKA1-low (Koopmans et al., 2015; Prelogović, Andrei Mesinger, et al., 2022); (iii) the Lyman limit mean free path at $z \sim 6$ (e.g., G. D. Becker, D'Aloisio, et al., 2021); and (iv) the field of view of *JWST* (e.g., T. Treu et al., 2022; Steven L. Finkelstein, Bagley, et al., 2023b; Bunker et al., 2023). Our emissivity PDFs are generated from 10000 realizations of such volumes. In Appendix 4.8 we vary this scale and demonstrate that the estimated mean emissivities have converged to within a few percent.

4.3.1 *Ionizing UV Emissivity*

In Figure 4.5 we show the distributions of the ionizing emissivity in our fiducial model, sampling all of the above-mentioned sources of stochasticity. On the left axis we report the specific emissivity in units of erg s^{-1} Hz^{-1} evaluated at the Lyman limit, while on the right axis we show the total number of ionizing photons above the Lyman limit per baryon per Gyr. Red violins show the ionizing emissivity PDF, with crosses (horizontal bars) demarcating the mean (99% C.L.) of the distributions. The fraction of our $R_{\rm nl} = 5$ cMpc realizations that have a zero emissivity is denoted at the bottom of each violin.

As galaxies become rarer toward higher redshifts, the mean ionizing emissivity decreases and the region to region scatter increases. The emissivity PDF becomes bimodal, with some regions having an emissivity of zero while those that have a non-zero emissivity show an approximately log-normal distribution. This is shown in Fig. 4.6 with a rectangle at the bottom of the figure representing the probability that a region of $R_{\rm nl} = 5$ cMpc has zero emissivity. At $z \gtrsim 18$ the majority of $R_{\rm nl} = 5$ cMpc volumes are expected not to have any galaxies that are actively emitting ionizing photons, in this fiducial scenario with a log-normal $p(f_{\rm esc})$. If instead we assume a binomial $p(f_{\rm esc})$ distribution, the majority of $R_{\rm nl} = 5$ cMpc have zero ionizing emissivity already by $z \gtrsim 15$.

We now quantify the main sources of scatter driving the variance in Fig. 4.5. As discussed above, we do this by repeating our emissivity calculation but omitting one source of stochasticity (that is, only using the corresponding mean relation with no scatter). In Figure 4.6 we plot the corresponding mean (*top panel*) and standard deviation[9] (*bottom panel*), normalized to the corresponding values from the full calculation shown in Fig. 4.5. Note that since most of our sources of scatter are log-normal, assuming a mean relation instead of the full distribution would underestimate the mean of the emissivity shown in the top panel (see Appendix 4.9).

The most important source of scatter is the escape fraction, if the escape fraction is binomial. In this scenario, assuming only the mean escape fraction for all galaxies would underestimate the standard deviation (std) by 60 – 70% throughout the EoR and CD (*gray dash-dotted curves*). However, the mean emissivity is unchanged (since the bimodal distribution has the same median and mean; see Appendix 4.9). If we instead assume that the escape fraction is log-normally distributed (*red circled curves*), not including scatter in this quantity only underestimates the mean and std of the emissivity by $\sim 10\%$.

Another important source of stochasticity is the burstiness of star formation. Assuming all galaxies follow the mean SFMS without scatter (*blue dashed curves*) would underpredict the mean (std) of the ionizing emissivity by 40% (15%) at $z \sim 5$. This underprediction in the std rises to $\sim 50\%$ toward $z \sim 20$, as the typical galaxies have smaller stellar

---

[9] For numerical stability, we calculate the standard deviation by fitting a log-normal to the nonzero distribution of emissivities.

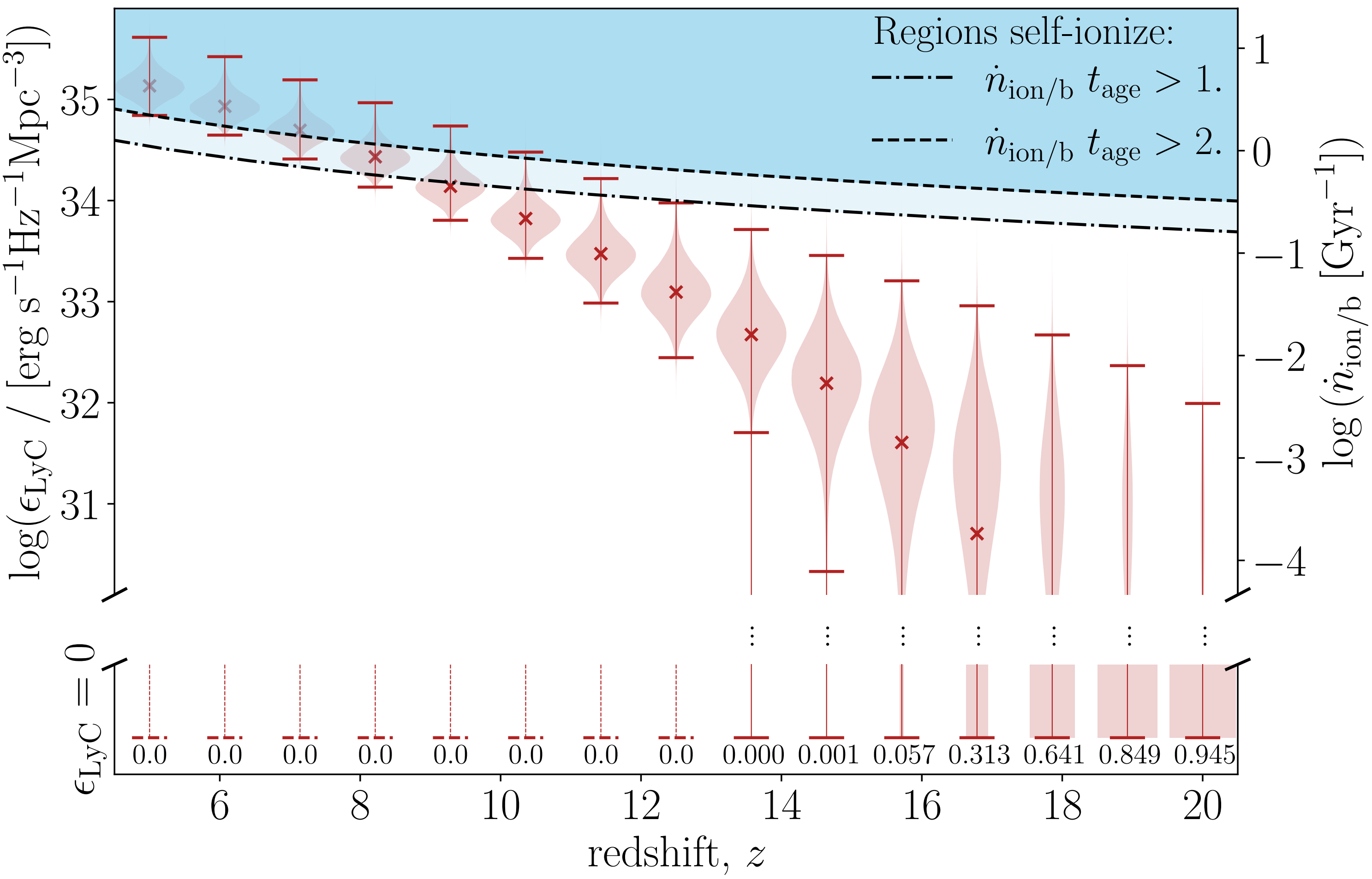

Figure 4.5: Distribution of Lyman limit emissivities for regions with a radius of 5 cMpc. Violin plots correspond to the full emissivity PDFs, while the crosses and horizontal bars demarcate the mean and 99th percentiles, respectively. The rectangle on the bottom with the matching number represents the fraction of 5 cMpc regions with zero emissivity. On the left axis we show the specific emissivity at the Lyman limit, while on the right axis we show the corresponding number of ionizing photons (¿13.6 eV) per baryon per Gyr. The blue shaded region at the top demarcates the approximate criteria for a 5 cMpc to ionize: having an emissivity greater than one (dot dashed line) or two (dashed line) ionizing photons per baryon in the age of the Universe. Assuming a threshold value of two ionizing photons per stellar baryon (e.g., Bolton and Haehnelt 2007; Emanuele Sobacchi and Andrei Mesinger 2014).) we see that roughly half of 5 cMpc regions can self-ionize by $z \sim 7$, consistent with the latest estimates of the EoR history e.g., Qin et al. in prep

.

masses and therefore a broader $p(\mathrm{SFR} \mid M_*)$ (c.f. the bottom panels in Fig. 4.2). In other words, the increased "bursty" nature of star formation at higher redshifts (at which the emissivity is dominated by galaxies with smaller masses) drives a correspondingly larger spatial variance in the ionizing emissivity.

On the other hand, ignoring scatter around the SHMR results in an underprediction of the mean and std of the emissivity by only 10%. Other sources of scatter have a negligible impact on the mean and variance of the ionizing emissivity. In particular, we note that only $\sim 5\%$ of our realizations of 5 Mpc regions at $z = 20$ contain fewer than 10 actively star forming galaxies. Therefore, it is not surprising that Poisson scatter in the halo number is unimportant in determining emissivities. We note that here we only consider galaxies above the atomic cooling threshold; had we considered an additional population of molecular-cooling galaxies, Poisson scatter would have been even less important since their expected mean number density is much larger.

4.3.2 *X-ray Emissivity* In Figure 4.7 we show the distributions of soft-band X-ray emissivities for our fiducial model, accounting for all sources of stochasticity. Red violins again represent PDFs averaged over comoving volumes with

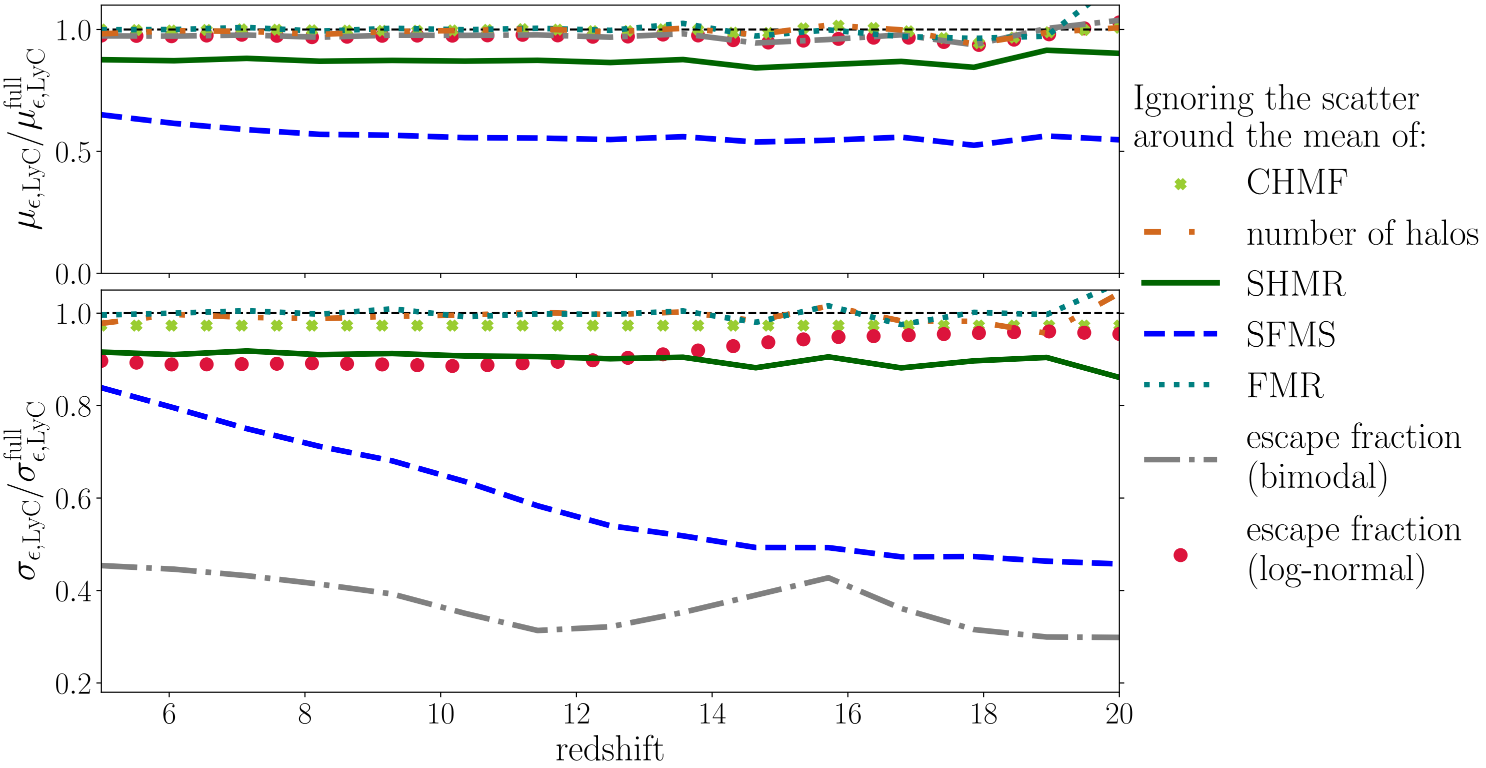

Figure 4.6: Fractional contribution of different sources of scatter (see legend) to the mean and standard deviation of the ionizing emissivity, computed over $R_{\mathrm{nl}} = 5$ Mpc volumes. The top (bottom) panel shows the mean (standard deviation) when removing one source of stochasticity, normalized to the fiducial value that includes all scatter.

$R_{\mathrm{nl}} = 5$ cMpc, while the rectangles represent probabilities that a region of size $R_{\mathrm{nl}} = 5$cMpc has zero X-ray emissivity. As expected the means and widths of the distributions decrease toward higher redshifts.

We see that the X-ray emissivities have broader distributions compared with the ionizing emissivities in Fig. 4.5. For example, at $z \sim 10$ the region-to-region std of X-ray emissivities is 300% of the mean, while for ionizing emissivities it is only 50% of the mean. This is primarily due to the fact that the HMXB LFs that source the X-ray emission in our model are fairly shallow (see Bret D. Lehmer et al., 2021). Thus the galaxy-averaged X-ray luminosity is sensitive to sample variance as it can be determined by a small number of HMXBs. This is evident by comparing the widths of the conditional $p(L \mid \mathrm{SFR}, M_*, Z)$ distributions for X-rays and ionizing photons in Fig. 4.4. The additional stochasticity in the ionizing emissivity due to the ionizing escape fraction (assuming it is log-normaly distributed) is subdominant compared with the wider X-ray intrinsic luminosity distribution.

We isolate the relative importance of each source of scatter to the X-ray emissivity in Figure 4.8. As in the previous section, we show $\mu_i/\mu_{\mathrm{full}}$ and $\sigma_i/\sigma_{\mathrm{full}}$ in the upper and lower panels, respectively.

The biggest impact on the mean and std comes from the scatter in the SFR$-M_*$ relation (blue dashed curves). Ignoring the scatter around the SFMS results in an underprediction of the mean (standard deviation) of the X-ray emissivity by 20% at $z = 5$ rising to a factor of 60% at $z = 20$. As in the previous section, this is driven by the mass-dependence of scatter in SFMS. The physical interpretation is the same: increased burstiness of star formation in small mass galaxies (that dominate at higher redshifts) boosts the variance of the X-ray emissivity. The scatter around the SFMS is even more important for X-ray emissivity, compared with ionizing emissivity, due to the strong dependence of the intrinsic X-ray luminosity on the SFR (see Fig. 4.4 and associated discussion).

Another important source of scatter is the $L_{\mathrm{X}}-$SFR relation (violet dash-dotted curves). The relative difference in standard deviations is roughly 30% at $z = 20$, rising to 45% at $z = 5$. At $z = 5$ it is more important than the scatter in SFR$-M_*$ relation. Note that complex physics of the formation of binary stars could induce an additional redshift dependence in this scatter, with different IMF's giving different populations of binary stars. This would go in the direction of increasing the importance of modeling $L_{\mathrm{X}}-$SFR scatter at earlier times.

Scatter in the SHMR has a $\sim 10\%$ effect, again without redshift dependence since we chose a constant width for $p(M_* \mid M_h)$. Scatter in the other terms has a negligible impact on the X-ray emissivity.

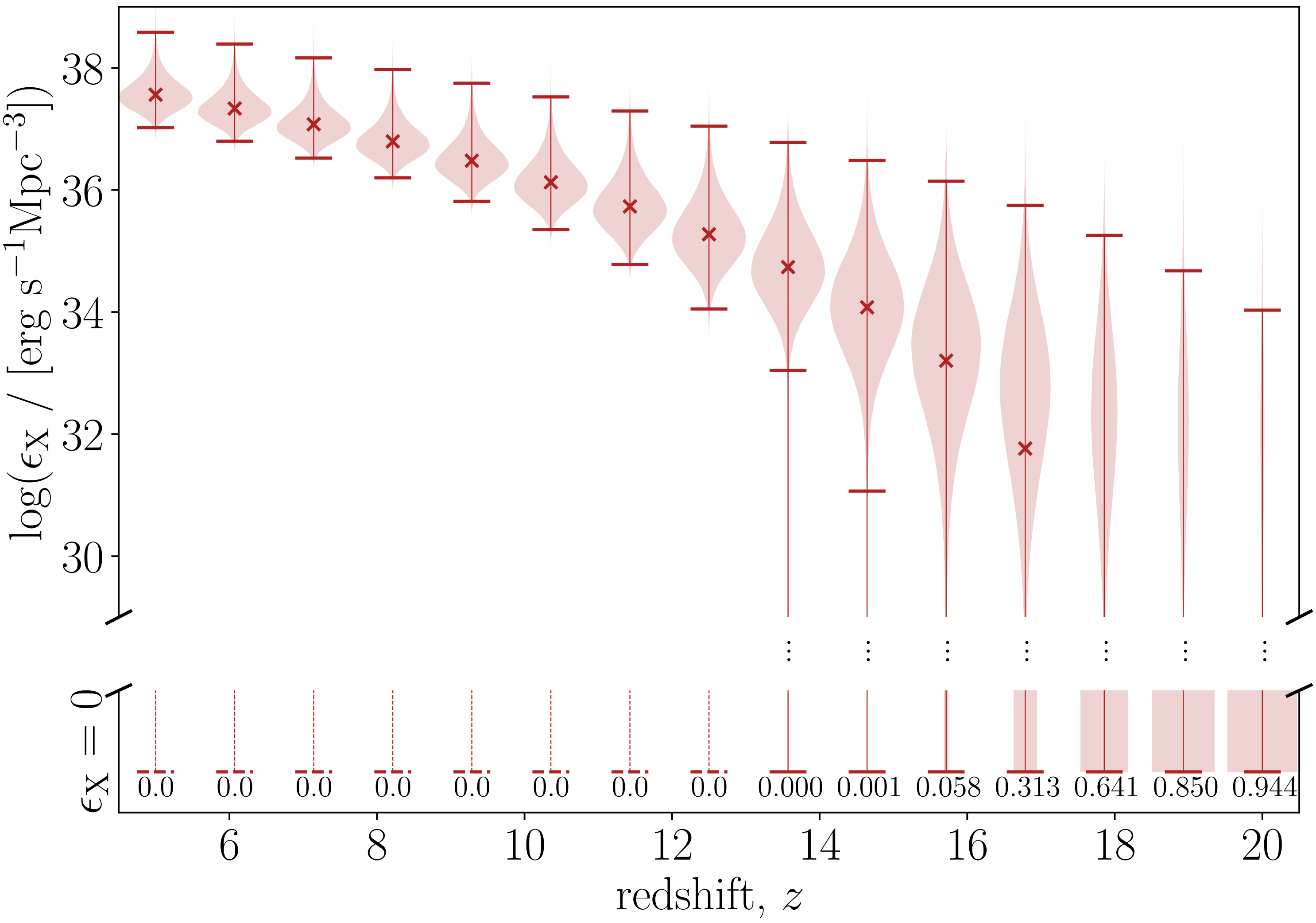

Figure 4.7: Same as Fig. 4.5, but for soft band (0.5–2 keV) X-ray emissivity.

4.3.3 *Lyman Werner Emissivity* Soft UV photons are important during the cosmic dawn as they regulate the abundances of H_2 (which provides an important cooling channel for the first galaxies) and the excited spin state of HI (which determines the cosmic 21cm signal). For concreteness, here we evaluate the emissivity in the Lyman-Werner band (11.2-13.6eV) noting that our conclusions would remain the same regardless of the specific soft UV range of interest.

In Fig. 4.9 we show the distribution of LW emissivities in 5 cMpc regions. We see that LW emissivities are more uniform (that is, with narrower PDFs) than both ionizing or X-ray emissivities from the previous subsections. This is to be expected, as the latter bands are sensitive to stochasticity in the ionizing escape fraction and HMXB LFs, neither of which contribute to the LW emissivity.

In Fig. 4.10 we show the fractional contribution of different sources of scatter to the mean and std of LW emissivity. Again, the SFMS (blue dashed lines) is the most important contributing source to the variance of emissivity, but less so compared to X-rays (as could be expected from Fig. 4.4). At $z \sim 20$ the burstiness of SFR contributes at a $\sim 50\%$ level to the std of the distribution, but this drops to $\sim 10\%$ at $z \sim 5$.

Also important is the scatter in the SHMR (green full curves) which contributes at a $\sim 10\%$ level to the mean and std for all redshifts. Other sources of scatter are negligible.

4.4 Results: EoR History

The ionizing emissivities shown in the previous section can be used to estimate the redshift evolution of the volume filling factor of ionized regions, $Q_{\rm HII}(z)$ – the EoR history. Even though it is an average quantity, computing the EoR history accurately requires accounting for the spatial and temporal coevolution of sources and sinks of ionizing photons,

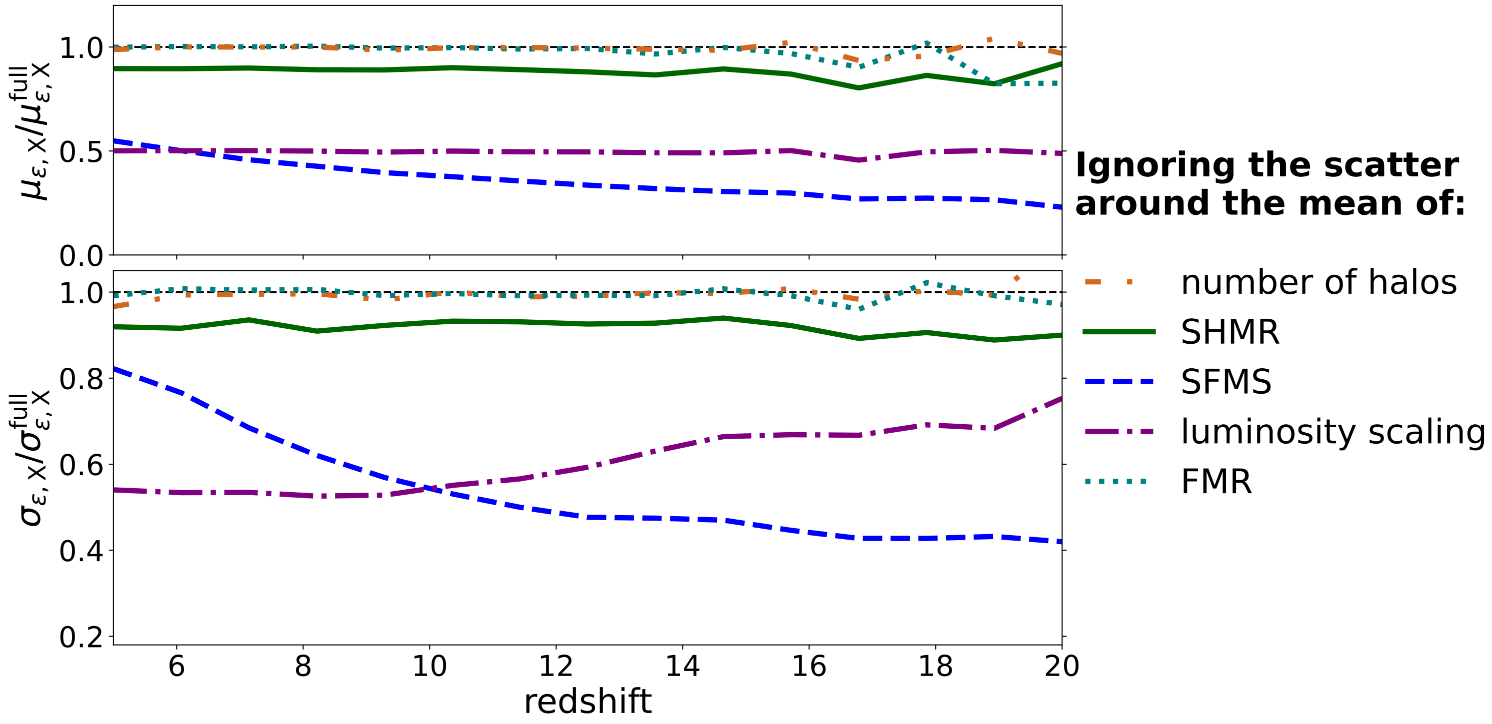

Figure 4.8: Like Fig. 4.6, but for the soft band (0.5–2 keV) X-ray emissivity.

and is therefore best done numerically. However popular analytic approximations exist and can provide insight into the relative impact of scatter in galaxy properties.

Here we compute two proxies for the EoR history. The first is the most common approximation in the literature (e.g., Madau, Haardt, and Martin J. Rees, 1999), obtained by:

$$\frac{\mathrm{d}Q_{\mathrm{HII}}}{\mathrm{d}t} = \dot{n}_{\mathrm{ion/b}} - \alpha_{\mathrm{A}}\ C\ \langle n_H \rangle\ Q_{\mathrm{HII}}. \tag{4.11}$$

Here $\dot{n}_{\mathrm{ion/b}}$ is the ionizing emissivity per baryon predicted by our model[10], α_{A} is the case-A recombination coefficient, $\langle n_H \rangle$ is the mean hydrogen density, and $C \equiv \langle n_H^2 \rangle / \langle n_H \rangle^2$ is the so-called "clumping factor" computed only over the ionized (not self-shielded) gas. By assuming a constant clumping factor, this equation ignores the correlation between sources and sinks of ionizing photons.[11] Estimates of the EoR history obtained with eq. (4.11) underpredict the duration of the EoR by $\Delta z \sim 1$–2, with the error increasing toward the end stages (see, e.g., Figure 6 in Emanuele Sobacchi and Andrei Mesinger 2014). Here we take $C = 2$, noting that we are only interested in the relative impact of galaxy stochasticity on the EoR history.

We show the resulting estimates in the left panel of Fig. 4.11. The black curve corresponds to our fiducial model, in which we account for all of the aforementioned sources of scatter. The green (blue) curve is computed ignoring scatter around the mean SHMR (SFMS). The orange curve does not account for any scatter, taking only the mean values for each relation. We see that scatter in the SHMR only delays the EoR history by $\Delta z \sim 0.1$. Ignoring scatter around the SFMS has a bigger impact, delaying the EoR history by $\Delta z \sim 0.5 - 1$. Ignoring scatter in all galaxy properties underestimates the duration of the EoR and delays the end stages by up to $\Delta z \sim 2$.

Our second proxy for the EoR history is obtained directly from Fig. 4.5. Specifically, we compute the fraction of 5 cMpc regions whose emissivities are larger than two ionizing photons per baryon per age of the Universe at that redshift, that is, $\dot{n}_{\mathrm{ion/b}} t_{age} > 2$ (shown by the dashed line in Fig 4.5; c.f. Bolton and Haehnelt 2007; Emanuele Sobacchi

---

[10] We make the standard assumption that helium is singly ionized by stellar sources together with hydrogen, due to their comparable ionization thresholds.

[11] In reality, most recombinations will come from the earliest patches of the IGM to ionize, which are those with the highest densities of galaxies. As a result, the growth of HII regions surrounding the highest galaxy densities begins to stall as reionization progresses, with an increasing fraction of ionizing photons required to balance recombinations. This process naturally results in a "soft landing," with the recombinations starting to balance ionizations in the late EoR stages, smoothly transitioning to the post-EoR regime (e.g., Emanuele Sobacchi and Andrei Mesinger 2014).

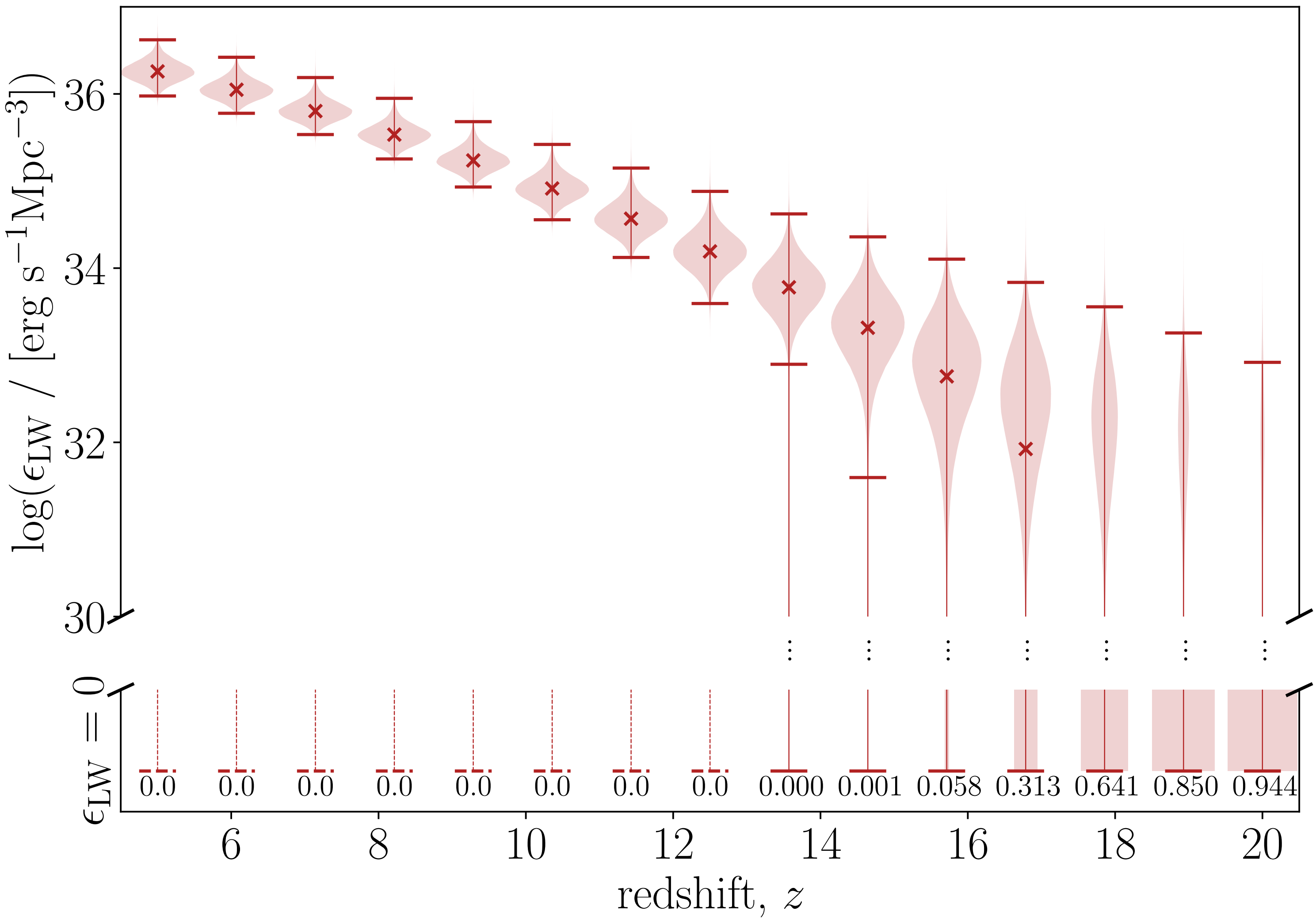

Figure 4.9: Same as Fig. 4.5, but for the Lyman-Werner (11.2-13.6 eV) emissivity.

and Andrei Mesinger 2014). This approximation of the EoR history assumes that each 5 cMpc region of the Universe is instantaneously ionized when this criterion is reached, and that each such region is independent. However, it does correctly compute the spatial variation in the emissivity, allowing us to account for local recombinations by increasing the required ionizing photon threshold. Again, we stress that here we are only interested in the relative impact of galaxy stochasticity on the EoR history.

We show the resulting estimate in the right panel of Fig. 4.11, for the same models as shown in the left panel. The qualitative evolution of this quantity is different from the one in the left panel. By its definition, taking only mean values (orange curve) would result in a step function at the end of the EoR. Importantly however, the relative impact of ignoring scatter in the SFMS and SHMR is the similar in both panels.

Regardless of the proxy used in estimating the EoR history, neglecting scatter around the SFMS results in a delayed EoR history by $\Delta z \sim 0.5$–1. Neglecting all galaxy-to-galaxy scatter by assuming mean values (c.f. Eq. 4.1) results in an overextended EoR history and delays its completion by $\Delta z \sim 1$–2.

Our results suggest that inferring galaxy properties from EoR history data without accounting for stochasticity could bias recovery toward brighter galaxies or higher escape fractions. We will investigate this further in future work, using 3D simulations that can more accurately capture the evolution of photon sinks and thus better predict the EoR history.

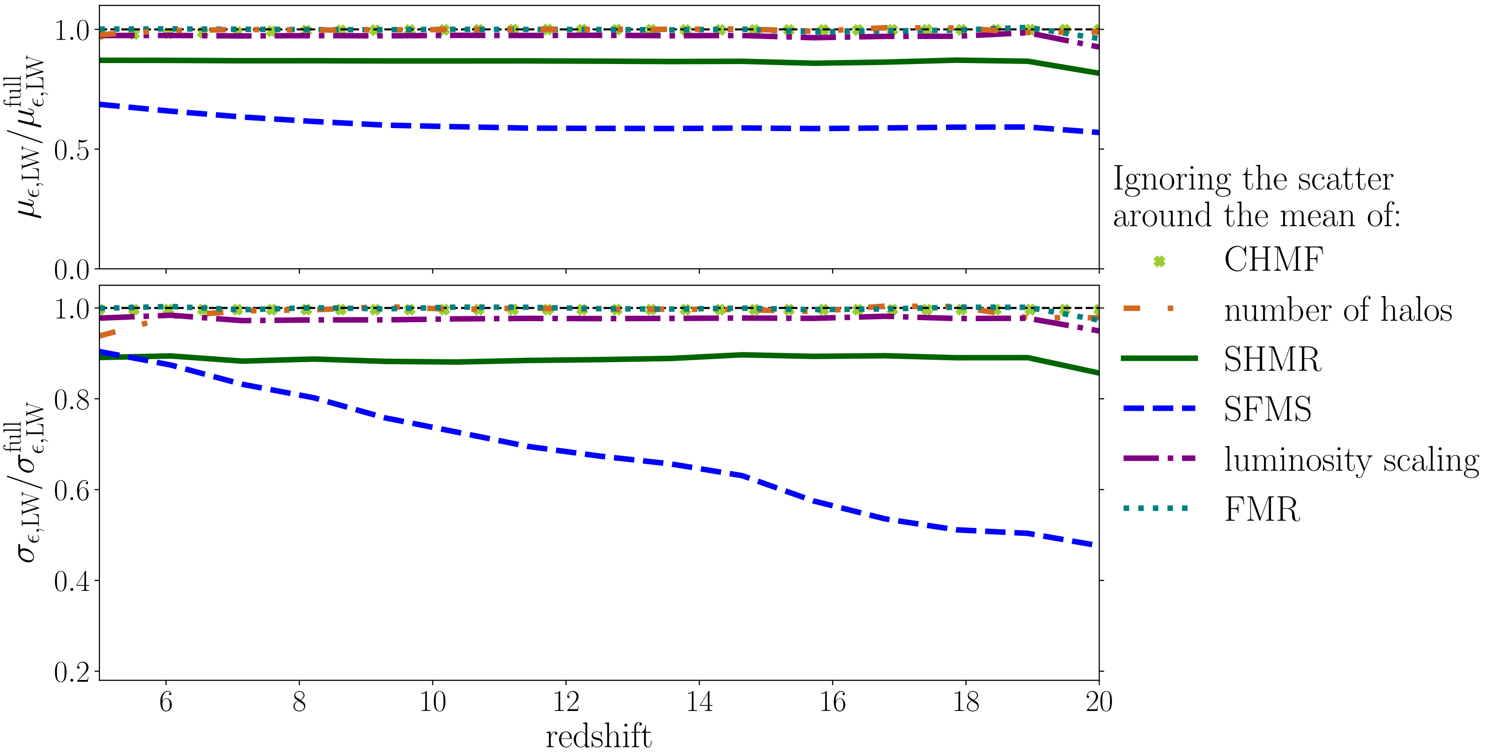

Figure 4.10: Same as Fig. 4.6, but for the Lyman Werner (11.2-13.6 eV) emissivity.

4.5 Results: UV Luminosity Functions

The framework developed in Section 4.2 also allows us to compute the corresponding galaxy UV LFs. Specifically, for each galaxy realization, we measure its rest frame magnitude using the luminosity from 1450$\AA$ to 1550$\AA$ directly from BPASS (see Section 4.2.5). For simplicity, we do not account for nebular emission, nor dust attenuation (e.g., Andrea Ferrara, Andrea Pallottini, and Pratika Dayal, 2023). We will include these in future work focused on interpreting UV LFs.

In Fig. 4.12 we plot the mean UV LF at each redshift (red line) along with the 68% C.L. (red shaded region). In green, we show the UV LFs calculated assuming only mean relations without any scatter. Also shown are various observational estimates from *HST* (R. J. Bouwens, Illingworth, Oesch, et al., 2015 (B15); R. J. Bouwens, Oesch, I. Labbé, et al., 2016 (B16); R. J. Bouwens, Oesch, M. Stefanon, et al., 2021 (B21); Livermore, S. L. Finkelstein, and Lotz, 2017 (L17); Ishigaki et al., 2018 (I18); Oesch, G. Brammer, et al., 2016 (O16); Oesch, R. J. Bouwens, et al., 2018b (O18); Leethochawalit et al., 2023 (L22); Kauffmann et al., 2022 (K22)) and *JWST* (Naidu et al., 2022 (N22); Steven L. Finkelstein, Bagley, et al., 2023b (F22); Donnan, McLeod, et al., 2023 (D23); Donnan, McLure, et al., 2024 (D24); Pérez-González, Costantin, et al., 2023 (P23); B. Robertson et al., 2024 (R24); Harikane, Nakajima, et al., 2024 (H23); McLeod et al., 2024 (M24); Willott et al., 2024 (W24)).

Comparing the green and red curves, we see that including scatter shifts the mean to brighter magnitudes and flattens the UV LFs. This is a well-known effect of upscattering some fraction of the more abundant faint galaxies to brighter magnitudes. For our fiducial model, the shift is roughly 1-2 magnitudes at $M_{\mathrm{UV}} \sim$ -18, consistent with other estimates of the impact of stochasticity on UV LFs (e.g., C. A. Mason, Trenti, and Tommaso Treu, 2023; Shen et al., 2023; Gelli, C. Mason, and Hayward, 2024).

Comparing the red curve to the observational estimates, we see that our fiducial model is consistent with UV LFs at $z \lesssim 10$. However, the mean underpredicts the recent estimates of UV LFs at $z \gtrsim 11$ from broad-band *JWST* photometry. Although the observational data points are mostly within the 68% C.L. of our model, they are systematically higher. Assuming there is no observational bias and that there are no correlations between the magnitude bins, this systematic underprediction would imply that our fiducial model is strongly disfavored by the data at $z \gtrsim 11$. This is qualitatively consistent with previous conclusions from the literature that larger than expected levels of scatter would be required to explain *JWST* results, provided the $z > 10$ photometric estimates are accurate (e.g., Mirocha and Steven R. Furlanetto 2023b; C. A. Mason, Trenti, and Tommaso Treu 2023; Shen et al. 2023; A. Pallottini and A. Ferrara 2023; Gelli, C.

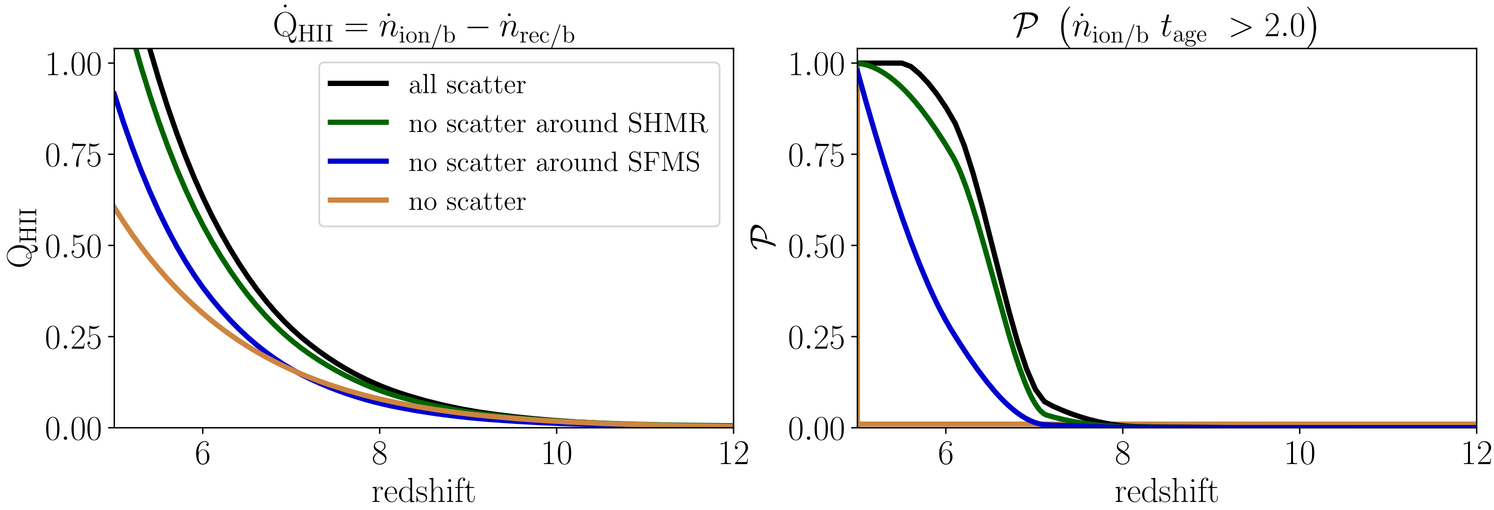

Figure 4.11: Relative contribution of galaxy stochasticity to the EoR history. In the left panel we show the common approximation of an EoR history calculated assuming a constant clumping factor (c.f. Eq. 4.11), while in the right we show the fraction of 5 cMpc regions that exceed the threshold of two ionizing photons per baryon per age of the Universe (c.f. Fig. 4.5). Black curves correspond to our fiducial model, including all sources of scatter. Green / blue curves ignore scatter around the SHMR / SFMS, while the orange curves ignore all sources of scatter. We see that not accounting for the burstiness of star formation (that is, scatter around the SFMS) can result in EoR histories that are delayed by $\Delta z \sim 0.5$–1. Not accounting for any scatter and assuming only mean galaxy properties delays the completion of the EoR by $\Delta z \sim 2$.

Mason, and Hayward 2024). Alternately, a correlation between observational estimates in different magnitude bins (for example through cosmic variance; e.g., Willott et al., 2024) could alleviate this apparent tension.

We also calculate the star formation density evolution from galaxies down to $M_{\rm UV} < -17$. In order to provide a more like-to-like comparison with observations, we do not use the sampled SFRs directly but instead convert from the UV luminosity via a constant conversion factor: $\mathrm{SFR}(M_\odot \mathrm{yr}^{-1}) = K_{\rm UV} L_{\rm UV}(\mathrm{erg\ s^{-1} Hz^{-1}})$, where $K_{\rm UV}$ is the conversion factor which depends on the IMF and star formation history. We take $K_{\rm UV} = 1.15 \times 10^{-28} M_\odot \mathrm{yr}^{-1}/\mathrm{ergs}^{-1}\mathrm{Hz}^{-1}$ (G. Sun and S. R. Furlanetto, 2016), consistent with other works. The result is shown in Fig. 4.13. The solid curve represents our fiducial model which includes all sources of stochasticity, while the dashed curve represents the model where only the mean relations are considered. We see that our model reproduces the data very well at $z \lesssim 12$ (R. Bouwens et al., 2020), though is somewhat lower than some estimates at higher redshifts. As this statistic is fundamentally an integral over the UV LFs, we reach the same qualitative conclusions. Quantitatively, the discrepancy with the data seems less than for the UV LFs. Due to the steepness of the UV LFs, the integrated SFRD is dominated by the faint end limit used to compute it, while the $z > 12$ JWST observations are more discrepant on the bright end.

The empirical framework we developed here is very flexible, and allows us to explicitly define the mean and scatter in every fundamental relation that leads to the 1500 Å UV magnitude. In future work we will use our model combined with physically motivated priors to infer these conditional distributions from *JWST* UV LFs and other observational data.

4.6 Conclusions

For this work we quantified how the galaxy-to-galaxy scatter in their properties impacts estimates of their emissivities and related observables. We used a semi-empirical model that explicitly defines scatter around well-studied mean relations: (i) the conditional halo mass function (CHMF); (ii) the SHMR; (iii) th galaxy SFMS; (iv) the FMR; (v) the conditional intrinsic luminosity; and (vi) the photon escape fraction. We computed the corresponding multi-frequency (ionizing UV, X-rays, and LW) emissivities, EoR histories, and UV LFs, quantifying the relative importance of the above sources of scatter.

We find that the burstiness of star formation (that is, scatter around the mean SFMS) is important for all emissivities. Because we assume burstiness increases toward smaller mass galaxies, the scatter around the SFMS becomes increasingly

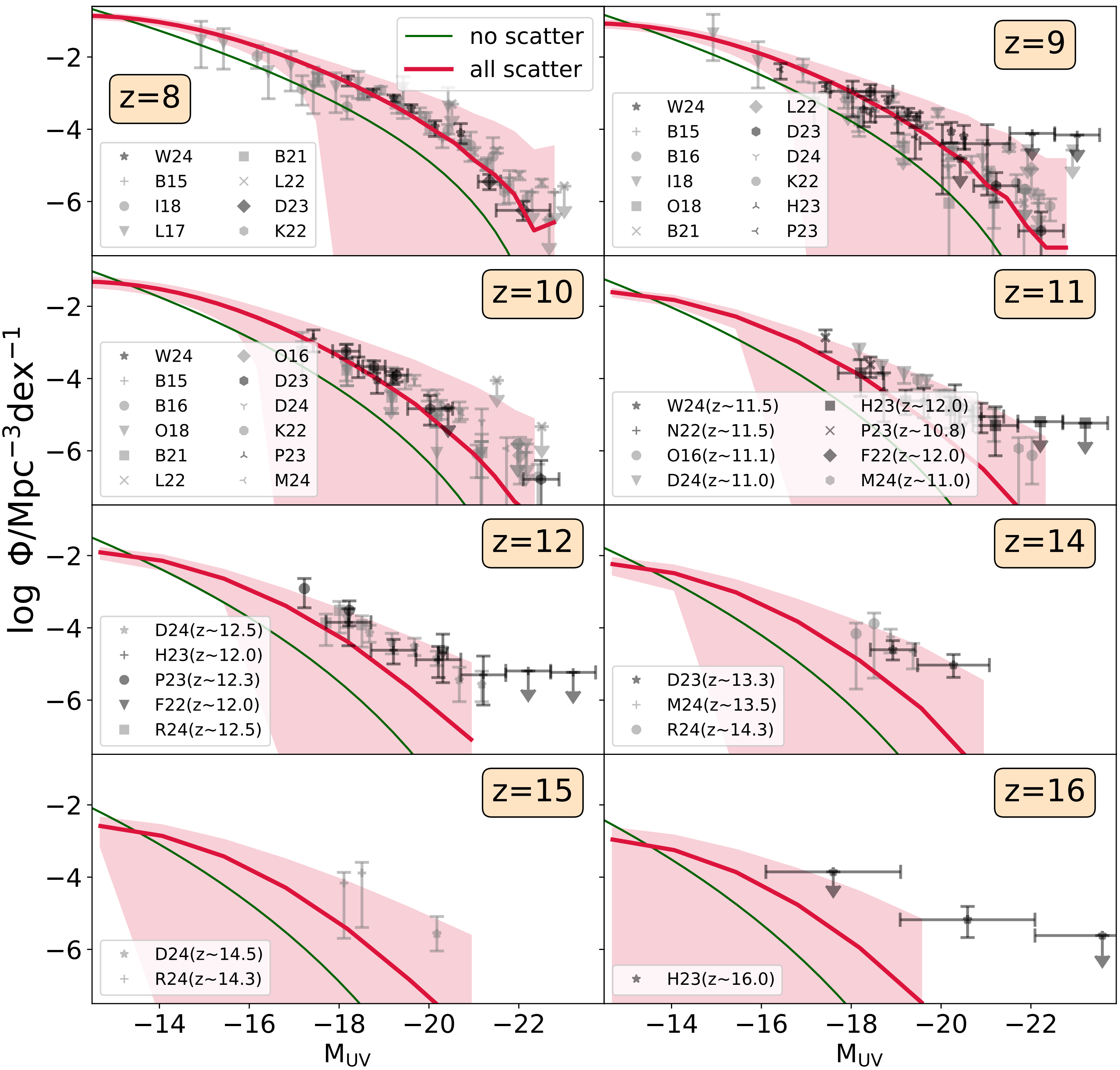

Figure 4.12: High redshift UV luminosity functions. Our fiducial model including all of the aforementioned sources of scatter is shown with the red solid lines (*mean values*) and surrounding shaded regions (*68% C.L.s*). The solid green curves correspond to UV LFs calculated using only mean relations without any scatter. Also shown in each panel are various observational estimates from *HST* and *JWST* (see text for details).

important at higher redshifts. Neglecting this source of stochasticity could underpredict the mean and std of emissivities by factors of up to a few tens during the EoR and CD. Stochasticity in the ionizing escape fraction can dominate the spatial scatter in the ionizing emissivity if its distribution is binomial. If instead the escape fraction is log-normally distributed, its contribution to the total emissivity scatter is only on the order of $\sim$ 10%. For the X-ray emissivity, one must account for scatter in the intrinsic luminosity, which in our fiducial model is driven by HMXB luminosity functions.

We find that neglecting stochasticity overestimates the duration of reionization, delaying its completion by $\Delta z \sim$

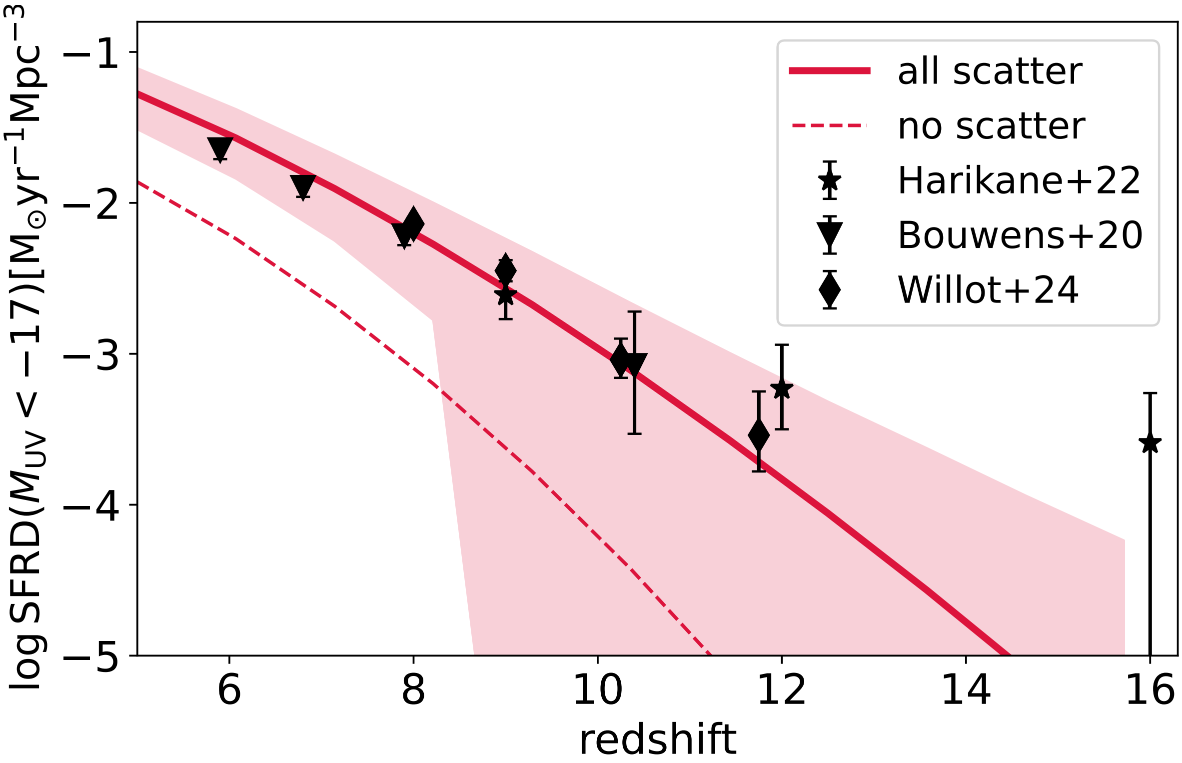

Figure 4.13: Star formation rate density from bright galaxies ($M_{\mathrm{UV}} < -17$) as a function of density. Our fiducial model including all of the aforementioned sources of scatter is shown with the solid line and surrounding shaded regions (68% C.L.). The dashed line corresponds to the SFRD calculated using only mean relations without any scatter. Also shown are high-redshift estimates from R. Bouwens et al. (2020), Harikane, Ouchi, Oguri, et al. (2023), and Willott et al. (2024).

1–2. Neglecting only scatter around the mean SFMS results in a delay of the EoR history by $\Delta z \sim 0.5 - 1$. This suggests that inferring galaxy properties from EoR history data without accounting for stochasticity could bias recovery toward brighter galaxies or higher escape fractions.

We recovered the well-known effect of stochasticity flattening the UV LFs. In our fiducial model, this results in a shift of 1–2 UV magnitudes at $M_{\mathrm{UV}} \sim$ -18. Our UV LFs are consistent with observational data at $z \leq 10$ but consistently under-predict recent estimates at higher redshifts. This is qualitatively in line with other studies, and implies that larger scatter is required in order for it to be the sole explanation for photometric estimates at $z > 10$.

We conclude that models of the EoR and CD should at least account for scatter around the SFMS. Simulating the X-ray background during these epochs (for example when computing the 21cm signal) additionally requires accounting for scatter in the intrinsic X-ray luminosities of galaxies.

The semi-empirical framework we used for this work is flexible and transparent. It can easily be extended to accommodate for additional observables, different functional distributions, and/or dependencies on additional galaxy properties.

4.7 Data Availability

The code related to the work is publicly available at IvanNikolic21/Stochasticity Sampler.

4.8 Convergence of the Mean Emissivity with Scale

Here we confirm that our fiducial choice of $R_{\mathrm{nl}} = 5$ cMpc when computing emissivity distributions converges to the correct mean. In principle, the averaging over the scale-dependent overdensity distribution, $p_z(\delta_0|R_{\mathrm{nl}})$, in the top row of Eq. 4.2 should self-consistently ensure that the correct mean is recovered, regardless of the choice of scale. In

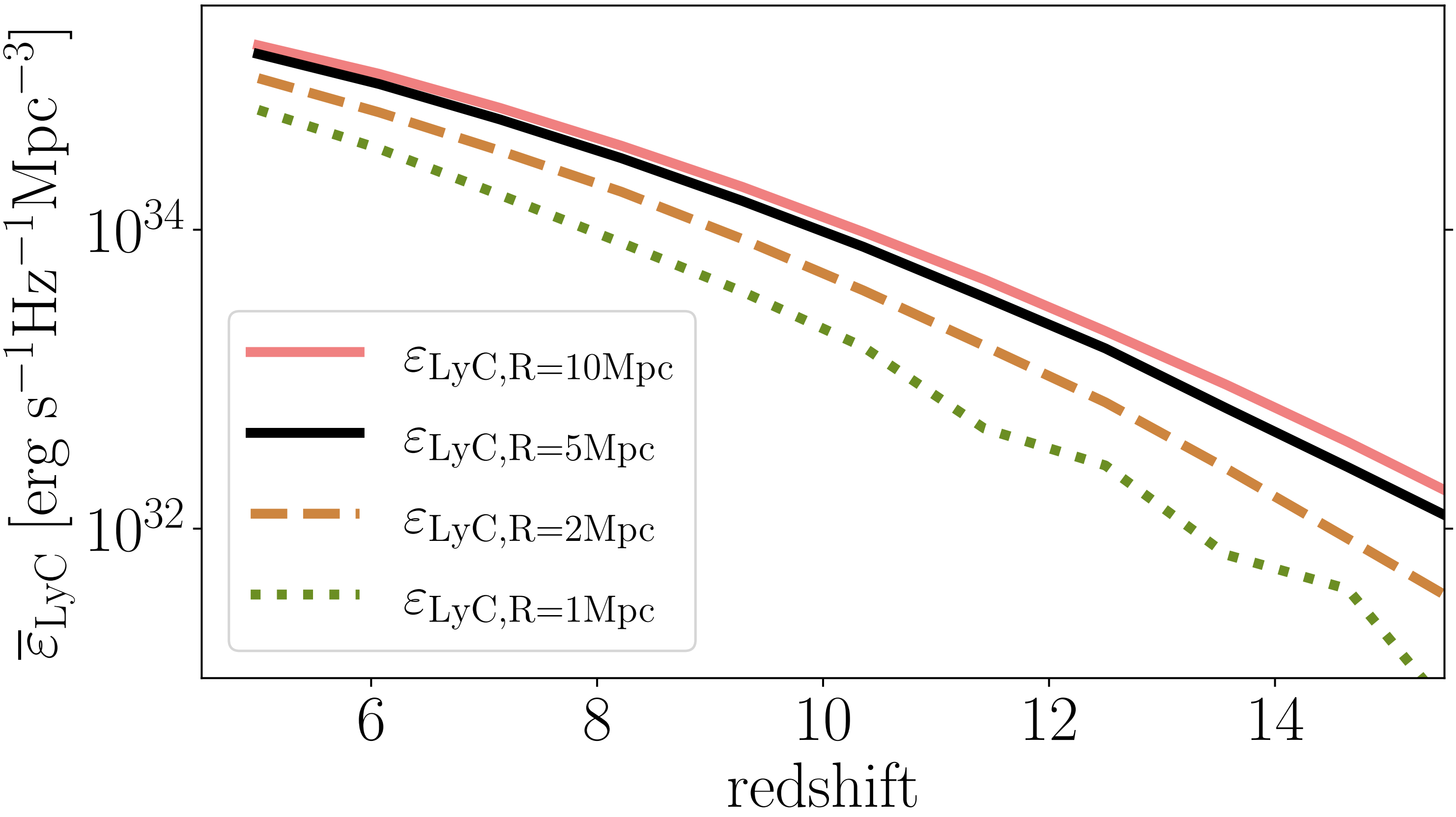

Figure 4.14: The mean of the ionizing emissivity computed over regions of varying scales, $R_{\rm nl}$ = 1, 2, 5, 10 cMpc. Our fiducial choice of 5 cMpc has converged in the mean to within a few percent. The analytical excursion set model misses massive halos when conditioned smaller scales, resulting in an underprediction of the mean for $R_{\rm nl}$ = 1 cMpc by factors of few – 10.

practice however, conditional halo mass functions underpredict the numbers of relatively massive halos whose Lagrangian volumes are close to the conditioning scale.

In Figure 4.14 we illustrate how the mean emissivity changes with scale. We use the ionizing emissivity to illustrate the trend; however, the result is the same for the other two bands of interest. We chose three additional scales: 1, 2, 10 Mpc. The first two choices roughly span the cell sizes used in semi-numerical (e.g., Andrei Mesinger, S. Furlanetto, and Cen, 2011a; Muñoz, Qin, et al., 2022; Schaeffer, Giri, and Schneider, 2023), or low-resolution (e.g., Dixon et al., 2016; Meriot and B. Semelin, 2024) radiative transfer simulations of the EoR/CD, while the latter roughly corresponds to the HII bubble size late in the EoR.

Comparing the red and black curves, we see that the mean for our fiducial choice of 5 cMpc has converged to within a few percent. We are thus reassured that our fiducial choice can be used to predict global quantities such as the EoR history and UV LFs.

As the scale is further reduced, we see that the mean emissivity can be significantly underestimated when using conditional excursion set formalism. For example, the mean using $R_{\rm nl}$ = 1 cMpc is underestimated by factors of few - ten. This serves as a caution against computing halo fields only at the cell level for low-resolution EoR/CD simulations (e.g., Appendix A in F. B. Davies and Steven R. Furlanetto 2022; Reis, Rennan Barkana, and Fialkov 2022). Instead, N-body (e.g., Dixon et al. 2016; Schaeffer, Giri, and Schneider 2023; Meriot and B. Semelin 2024), excursion-set that accounts for larger scales (e.g., Steven R. Furlanetto, Zaldarriaga, and Hernquist, 2004; Andrei Mesinger, S. Furlanetto, and Cen, 2011a; Cen and Kimm, 2015), or a mixture of the two (e.g., McQuinn, Lidz, et al., 2007, Davies et al. in prep) should be used.

4.9 Shift in the Mean Emissivity for Correlated Log-Normal Distributions

In Section 4.2 we wrote the emissivity in the form:

$$\bar{\varepsilon} = \int \mathrm{d}M_h \frac{\mathrm{d}n(M_h, z)}{\mathrm{d}M_h} L(M_h), \tag{4.12}$$

where we ignore the escape fraction for the moment. As we mentioned in that section, this formula holds if galaxy properties are deterministic functions of halo mass. However, we know that is not the case and for that reason we wrote the general formula for the mean in Eq. 4.2. In our case, we assume log-normal distribution for most of the scaling relations (except for the ones relating to the halo abundances) so we can analytically integrate over some of the distributions in Eq. 4.2. For simplicity, in this section we reduce the dependencies of luminosity to only $L(M_h)$ with an appropriate log-normal PDF $p(\log L|M_h) = \mathcal{N}(\mu_L, \sigma_L)$ without loss of generality. The Eq. 4.2 becomes:

$$\begin{aligned}\bar{\varepsilon} &= \int \mathrm{d}M_h \frac{\mathrm{d}n(M_h)}{\mathrm{d}M_h} \int \mathrm{d}L \; L \; p(\log_{10} L | \log_{10} M_h) \\ &= \int \mathrm{d}M_h \frac{\mathrm{d}n(M_h)}{\mathrm{d}M_h} \int \mathrm{d}L \; L \; \frac{1}{\sqrt{2\pi} \log_e 10 \; \sigma_L L} \exp\left(-\frac{(\log_{10} L - \mu_L)^2}{2\sigma_L^2}\right).\end{aligned} \tag{4.13}$$

The integral on the right can be analytically computed:

$$\bar{\varepsilon} = \int \mathrm{d}M_h \frac{\mathrm{d}n(M_h)}{\mathrm{d}M_h} 10^{\left(\mu_L + \frac{\log_e 10 \;\; \sigma_L^2}{2}\right)}. \tag{4.14}$$

Since the halo mass function in general is not an analytic function of halo mass and mean of the luminosity scaling depends on the halo mass, this integral cannot be computed analytically. However we can already gain intuition about the mean looking at the last term of Eq. 4.14. The factor 10^{μ_L} corresponds to the mean of the $L(M_h)$ scaling relation, that is, what one would obtain if one did not consider scatter around the mean. The second part, $10^{\frac{\log_e 10 \;\; \sigma_L^2}{2}}$ represents the shift of the mean when integrating over the whole PDF of the distribution. This is proportional to the width of the distribution, indicating that the wider the distribution, the larger the corresponding shift in the mean. This is a general property of asymmetric distributions like the log-normal and has important implications for interpreting means of scaling relations. This is clearly seen in Figures. 4.6,4.8 and 4.10 where removing one source of scatter reduces the mean proportionally to the width of the distribution.

If we instead add an additional term that is binomially distributed (for example, one choice for the escape fraction in Sec. 4.2.6) then the equation becomes:

$$\begin{aligned}\bar{\varepsilon} = \int \mathrm{d}M_h \frac{\mathrm{d}n(M_h)}{\mathrm{d}M_h} \int \mathrm{d}L \; L \; p(\log_{10} L | \log_{10} M_h) \times \\ \times \int \mathrm{d}f_{\mathrm{esc}} f_{\mathrm{esc}} \binom{n}{k} \mathcal{P}^n (1-\mathcal{P})^k,\end{aligned} \tag{4.15}$$

where we have explicitly written out the binomial distribution. In the above, $n = 1$, $k = 0$ and $\mathcal{P} = \bar{f}_{\mathrm{esc}}$ so the distribution trivially becomes:

$$\bar{\varepsilon} = \int \mathrm{d}M_h \frac{\mathrm{d}n(M_h)}{\mathrm{d}M_h} \int \mathrm{d}L \; L \; p(\log_{10} L | \log_{10} M_h) \bar{f}_{\mathrm{esc}}. \tag{4.16}$$

Therefore, the mean does not change if scatter is added following a binomial distribution. This is clearly seen in Sec. 4.5.

5

Conclusions and Future Directions

First galaxies are at the forefront of astrophysical research. These objects transformed the Universe from a neutral medium into the complex web of structures we see today. First galaxies also caused the last phase transition of the Universe – the EoR. Recent years have seen a dramatic increase in knowledge of this process, but there is still much to be learned. Key questions are: ***When did the EoR happen? How long did it last? How did it happen? What sources drove it?*** Helping to answer these questions has been the goal of this thesis.

Answering these questions has even bigger implications than just studying EoR. Due to its energetic nature, the EoR had a profound impact on structure formation up to the present day, with temperature, ionization, and low-mass structure formation being strongly affected. Understanding the EoR will help us understand how the first galaxies formed, and potentially help us answer some of the most important cosmological questions, such as what the elusive dark matter and dark energy are.

Recent decades have seen an increasing number of telescopes capable of probing this important epoch, including HST, JWST, ALMA, HERA, and others. The increase in quality and quantity of data will not stop soon, with future telescopes such as Euclid and SKA coming online. For that reason, it's important that our models of galaxy formation and the evolution of the IGM properly incorporate all of the physics that regulates the evolution of the Universe at high redshifts. This is connected to the inference techniques that are a necessary tool in astrophysics, where we don't have controlled experiments. In this thesis, I presented my contributions to these two problems.

In Chapter 2, I presented the work published in Nikolić, Andrei Mesinger, Qin, et al. (2023), where the patchy kSZ signal is added to the list of large-scale probes of the EoR. We demonstrated its utility in probing the evolution of the cosmic ionized fraction and found that current estimates point to a late and rapid end to the EoR, consistent with Lyman-α forest constraints from (Qin, Andrei Mesinger, Sarah E. I. Bosman, et al., 2021). We presented the astrophysical constraints in terms of scaling laws, showing how inferring the EoR is inherently linked to galaxy evolution, and how learning about the process can help us discover properties of galaxies too faint to be observed. The patchy kSZ signal will become even more useful with future telescopes targeting small-scale CMB anisotropies.

Despite the importance of understanding the evolution of the mean neutral hydrogen fraction, it's important to also probe the EoR morphology. This will help us understand the connection between galaxies and the IGM, helping us answer what galaxies sourced the reionization. An interesting method to do that is mapping the ionized bubbles around galaxies by observing their Lyman-α spectra. JWST is giving us a number of galaxies suited for the analysis, but a rigorous treatment of systematics is required, including stochasticity and the non-Gaussianity of the likelihood. In Chapter 3, we presented a framework that helps us infer the properties of ionized bubbles around a group of galaxies, taking into account stochasticity in both galaxy properties and the IGM. The pipeline consistently incorporates information from all of the galaxies in the group. We tested the framework on different Lyman-α distributions and on realistic EoR simulations. With the results, we developed a strategy for observations using JWST/NIRSpec, finding that an inference at the $\sim 10\%$ level is within reach.

An important aspect I assessed throughout the thesis is stochasticity. The scatter between any scaling relation in astrophysics, caused by the large dynamic range of astrophysical processes, is important to accurately model galaxy evolution. In Chapter 4, we developed a semi-analytic model that calculates various summaries—from emissivities at different wavelengths to luminosity functions and EoR histories—explicitly accounting for stochasticity. With this framework, we could isolate the most important sources of stochasticity at different scales, times, and wavelengths. Star-formation burstiness is the most important at all redshifts and wavelengths, while the distribution of escape fractions is important to account for when estimating ionizing emissivity and EoR history. We concluded that current levels of stochasticity are not enough to explain the JWST overabundance of photometric candidates at the highest redshifts.

What does the data tell us about stochasticity? The model developed in Chapter 4 is rooted in established relations and simulation predictions, which were tailored to match lower-redshift summaries. In the next section, we take the model presented there a step further by directly inferring the astrophysical relations and their associated scatters from data.

The work presented in this thesis is a step in the direction of inferring precise galaxy models with stochasticity included, using data from galaxy observations with JWST and other large-scale probes. There are many avenues for improvement, and in Section 5.2, we discuss some of them.

5.1 Using Current JWST Observations to Inform Our Galaxy Models

Here I outline the work currently in preparation, aiming to constrain stochasticity directly from the UV luminosity function (UV LF) and angular correlation functions using JWST data.

As mentioned in Chapter 4, stochasticity has a significant effect on galaxy populations. It has been invoked to explain the overabundance of JWST galaxies and other unexpected results (C. A. Mason, Trenti, and Tommaso Treu, 2023; Mirocha and Steven R. Furlanetto, 2023b; Guochao Sun et al., 2023; Nikolić, Andrei Mesinger, J. E. Davies, et al., 2024). However, the effect of stochasticity on the UV LF is degenerate with an increase in the amplitude of the luminosity–halo mass relation. In addition, simulations differ in the scaling of the stellar-to-halo mass relation (SHMR), the star-formation main sequence (SFMS), and luminosity scaling (Lovell et al., 2021; Ceverino, Klessen, and Glover, 2018; A. Pallottini, A. Ferrara, et al., 2022), as well as in the scatter around these relations.

Nevertheless, we can attempt to infer these scaling relations and their associated scatter directly from the data. The intrinsic degeneracy between the amplitude of a given scaling relation and its scatter can be broken using observations of the clustering of high-redshift galaxies (Muñoz, Mirocha, et al., 2023; Gelli, C. Mason, and Hayward, 2024). The clustering of galaxies is determined by the distribution of their dark matter halos. A high amplitude of the star-formation efficiency would imply that galaxies of a given mass reside in lower-mass halos, which would result in weaker clustering (since more massive halos are more strongly clustered). Higher stochasticity would also allow lower-mass halos and higher mass halos to participate in the correlation. These distinct responses to parameter changes enable us to constrain the level of stochasticity in galaxies observed by JWST.

In Section 5.1.1, I briefly describe the extensions to the galaxy model from Chapter 4, and in Section 5.1.2, I present preliminary results.

5.1.1 *Galaxy Model and Observables* We start by coupling our galaxy model from Nikolić, Andrei Mesinger, J. E. Davies, et al. (2024) to a halo occupation model. Halo occupation models describe the spatial distribution of galaxies within dark matter halos, categorizing galaxies as centrals (typically the most massive galaxy in the halo) and satellites (all others). We define the central occupation as (Z. Zheng et al., 2005):

$$N_c(M_h) = 0.5 \cdot \mathrm{erfc}\left(-\frac{\log(M_h) - \log(M_*^{-1}(M_{*,min}))}{\sigma_{\mathrm{SHMR}} \cdot \sqrt{2}}\right) \tag{5.1}$$

Similarly, we write the number of satellites as $N_s(M_h) = N_c \cdot (M_h - M_0)^{\alpha_{\mathrm{sat}}}$ where we set $M_0 = 10^{12} M_\odot$ and $\alpha_{\mathrm{sat}} = 1.0$, following inference results from Paquereau et al. (2025), Shuntov et al. (2025), and Bhowmick et al. (2018). We expect the satellite fraction among high-redshift galaxies to be low, and only constrain the 1-halo term of the correlation function. Therefore, our results are not sensitive to the precise choice of satellite parameters. From here we calculate the angular correlation function following S. G. Murray et al. (2021).

We apply the bias correction from Jose, C. G. Lacey, and Baugh (2016) on top of the large-scale halo bias from Tinker et al. (2010).

We compute the UV luminosity function from the model as:

$$\frac{\mathrm{d}n}{\mathrm{d}M_{\mathrm{UV}}}(M_{\mathrm{UV}}, z) = \int \mathrm{d}M_h \frac{\mathrm{d}n}{\mathrm{d}M_h} p(M_*|M_h) \int \mathrm{d}M_*\, p(\mathrm{SFR}|M_*) \int \mathrm{dSFR}\, p(M_{\mathrm{UV}}|\mathrm{SFR}), \tag{5.2}$$

which follows the same logic as in Sec.4.5. The galaxy model used is the same as in Chapter 4, but we now leave the scaling relations and their scatters as free parameters. The SHMR is written as:

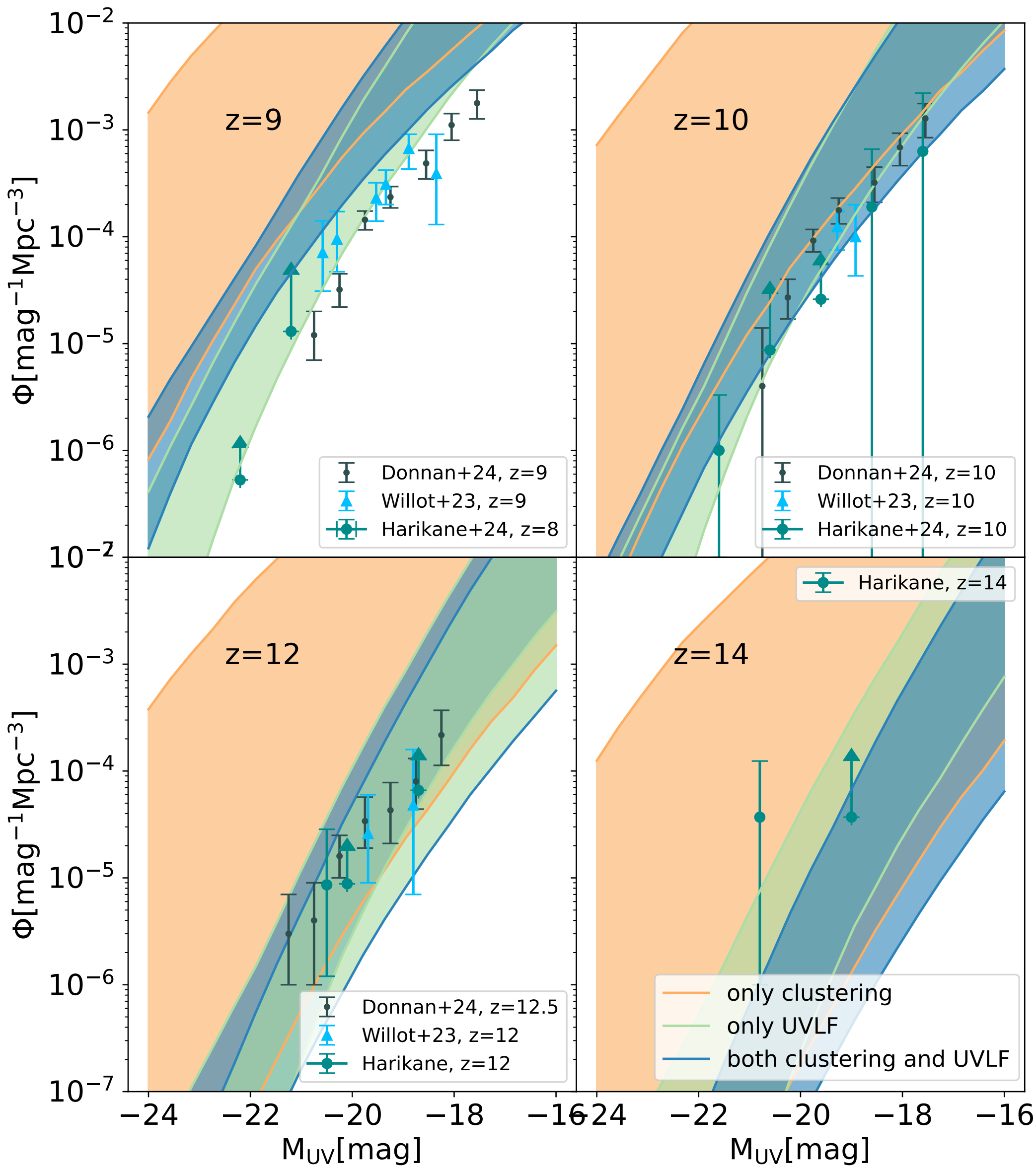

Figure 5.1: Posteriors in the space of UV luminosity functions. The orange contour represents the 95% C.I. when angular clustering is used in the likelihood; the green contour shows the same for UV LF alone; and the blue contour corresponds to the joint inference using both datasets. Each panel corresponds to a different redshift. Shown are also observational datasets from Donnan, McLure, et al. (2024) and Willott et al. (2024), and Harikane, Inoue, et al. (2025), the latter being used in the inference.

$$M_* = \frac{f_0}{\left(\frac{M_h}{M_{\mathrm{knee}}}\right)^{-\alpha_*} + \left(\frac{M_h}{M_{\mathrm{knee}}}\right)^{0.61}} M_h, \tag{5.3}$$

where the free parameters are the amplitude (f_0), low-mass slope (α_*), and the knee mass (M_{knee}). We do not vary the high-mass slope due to the scarcity of massive galaxies at high redshift. The scatter σ_{SHMR} is also allowed to vary.

The SFMS is parametrized as:

$$\mathrm{SFR}(z) = \frac{M_*}{t_* H^{-1}(z)} \tag{5.4}$$

with t_* as the free parameter. The scatter around the relation is mass-dependent:

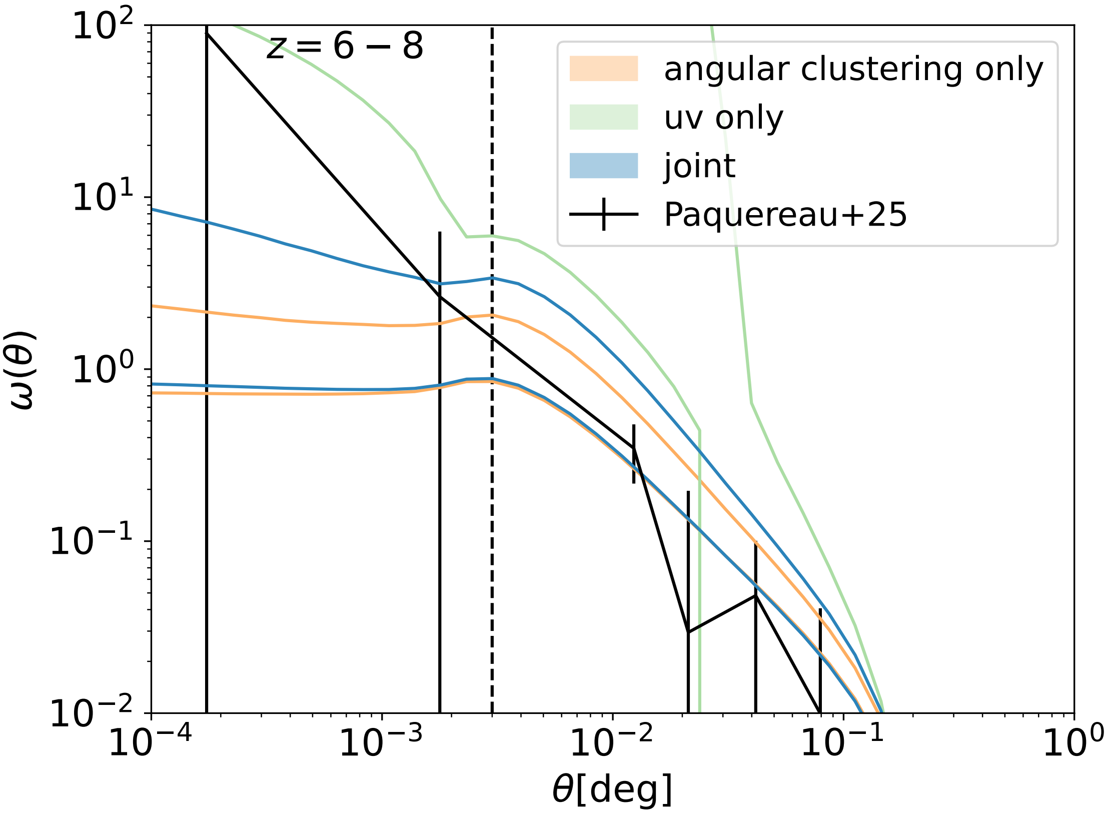

Figure 5.2: Same as Fig. 5.1, but for the angular correlation function at $z \sim 7$. Measurements from Paquereau et al. (2025) at $z \in [6, 8]$ are also shown, with the vertical dashed line marking the expected transition between 1-halo and 2-halo terms.

$$\sigma_{\text{SFMS}} = \begin{cases} a_{\sigma,\text{SFMS}} * \log\left(\frac{M_*}{10^{10} M_\odot}\right) + \sigma_{0,\text{SFMS}}, & \text{if } M_* \leq 10^{10} M_\odot \\ \sigma_{0,\text{SFMS}}, & \text{otherwise} \end{cases} \tag{5.5}$$

where both the amplitude ($\sigma_{0,\text{SFMS}}$) and the slope ($a_{\sigma,\text{SFMS}}$) are free parameters. We also use the fundamental mass–metallicity relation from Nikolić, Andrei Mesinger, J. E. Davies, et al. (2024), based on Curti, Filippo Mannucci, et al. (2020b) and Curti, Maiolino, et al. (2024). From metallicity and SFR, we calculate luminosities using BPASS (Eldridge et al., 2017), assuming an exponential SFH, following Nikolić, Andrei Mesinger, J. E. Davies, et al. (2024).

5.1.2 *Results* We constrain stochasticity using two observational datasets: the UV luminosity function (UV LF) and the angular correlation function. For the UV LF, we use spectroscopic estimates from Harikane, Inoue, et al. (2025), using 60 spectroscopically confirmed galaxies in the redshift range $7 \lesssim z \lesssim 14$ using JWST, ALMA, and Keck observations.

For the angular clustering measurements, we use estimates from the COSMOS-Web field presented in Paquereau et al. (2025), specifically their results for $z \in [5, 10]$ at $\log(M_*/M_\odot) = 9.0$. To avoid scales dominated by the 2-halo term, we restrict our inference to angular separations $\theta > 0.005$ deg.

Using our model, we perform inference on the seven parameters outlined above. We report results from three separate inferences: one using only UV LF data, one using only angular clustering data, and one combining both datasets.

The choice of priors is crucial in any Bayesian inference, particularly when the likelihood is only weakly constraining. This is generally the case at high redshift due to the faintness and scarcity of galaxies and the large uncertainties in observational summaries. In our case, the scaling relations are motivated by both lower-redshift observations (where data are more robust) and predictions from cosmological simulations. We construct priors from a broad set of simulations (Lovell et al., 2021; A. Pallottini, A. Ferrara, et al., 2022; Bird et al., 2022; Ceverino, Klessen, and Glover, 2018). Since

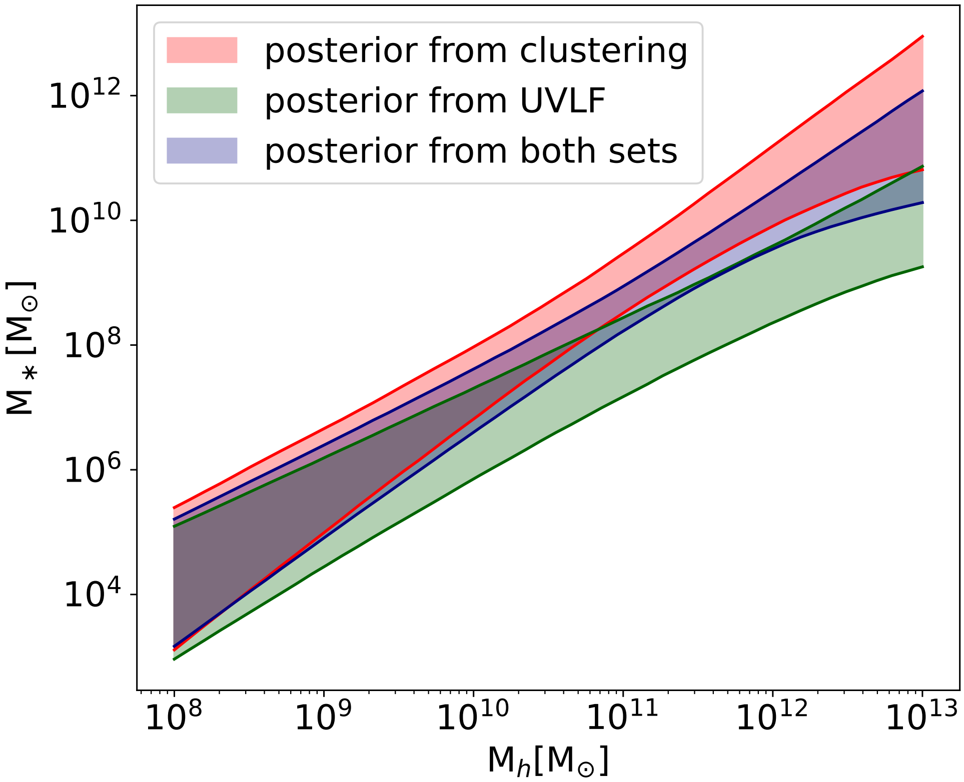

Figure 5.3: Posteriors in the space of SHMR. Red contour marks the 95% C.I. of the posterior distribution from the inference using only angular clustering measurement. Green contour represents the same, but for the inference using only UV LF, while the represents the posterior using both datasets.

there is no uniquely correct simulation, we adopt a multivariate Gaussian prior that encompasses the posteriors of each simulation. This approach balances conservativeness with the ability to disfavor unphysical scalings.

With these priors, we run three separate inferences using the MultiNest sampler (Feroz, Hobson, and Bridges, 2009):

- using only angular clustering data from Paquereau et al. (2025) in the likelihood
- using only UV LF data from Harikane, Inoue, et al. (2025) in the likelihood
- using both of the datasets

We show posterior distributions for all three inferences as 95% confidence contours in UV LF space at $z \in 9, 10, 12, 14$, angular clustering at $z = 7$, and in the SHMR and SFMS parameter spaces (Figs. 5.1, 5.2, 5.3, and 5.4). Our galaxy model successfully captures the UV LF evolution when Harikane, Inoue, et al. (2025) data are used in the likelihood. Other UV LF estimates shown in Fig. 5.1 are broadly consistent, though the posterior remains wide due to large measurement uncertainties – many data points are upper or lower limits. Despite this, even early spectroscopic confirmation of JWST galaxies helps constrain galaxy parameters significantly (as shown in Figs.5.3 and 5.4). However, the UV LF-derived posterior in the angular clustering space predicts stronger clustering than observed in Paquereau et al. (2025).

Conversely, the posterior obtained using only the clustering likelihood matches the observed angular correlation well. The posterior is again broad, reflecting measurement uncertainties. In contrast to the UV LF case, the clustering data favor larger SHMR amplitudes (to account for the lower observed clustering) and correspondingly higher UV LFs than the data suggest. Clustering also indicates a larger value of σ_{SHMR}.

The joint inference using both datasets yields a posterior that fits both UV LF and clustering data well. The inferred SHMR and SFMS lie between the two extremes, but clustering still points to higher levels of stochasticity than predicted by simulations. We find that inclusion of clustering is useful in distinguishing between various sources of scatter, unlike the UV LF which captures the total stochasticity present.

The work presented is still in progress and other observational datasets will be probed, as well as extensions to the model if the data requires it.

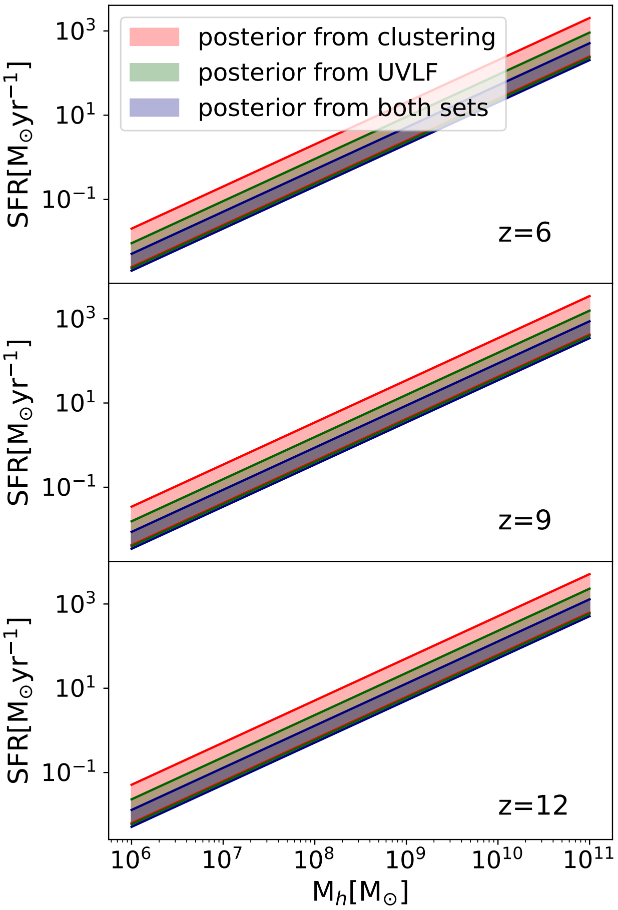

Figure 5.4: Same as Fig. 5.3, but for the SFMS, shown at $z \in 6, 9, 12$.

5.2 Future Directions

The work presented in this thesis naturally leads to several follow-up directions. Here I outline promising future research avenues I plan to pursue.

First, the JWST-based inference of galaxy properties described in the previous section can be significantly extended. Current JWST surveys cover relatively small sky areas and volumes, with each field sampling a distinct cosmic environment. This means inference from any single field is subject to cosmic variance. A potential way to mitigate this is to couple the galaxy model from Chapter 4 with a semi-numerical simulation (e.g., 21cmFAST; Andrei Mesinger, S. Furlanetto, and Cen 2011a) to simulate representative cosmic volumes. The galaxy model itself can also be extended to predict additional observables such as the UV continuum slope (β) and UV nebular lines. These quantities can be modeled either using CLOUDY (Ferland et al., 1998) or with simpler approximations. Including more observables will enable us to constrain additional galaxy properties—for example:

- Star-formation burstiness (via joint inference on UV and Hα),
- Escape fraction distributions (via nebular emission lines),
- Stellar ages, and others.

Performing this inference in a probabilistic framework would maximize the information we extract from JWST surveys.

Second, the work from Chapter 3 could be expanded in several ways. The simple analytic model of EoR morphology used there can be replaced with more physically one based on accurate semi-numerical simulations (again using 21cmFAST). Additionally, the current inference framework in Chapter 3 relies on maximum-likelihood estimates. This can be improved with a full posterior analysis that models the covariance between galaxies and imposes physically motivated priors on bubble sizes.

Furthermore, modern inference techniques from the simulation-based inference (SBI) domain could make posterior exploration more efficient. The long-term goal of this work remains to apply the pipeline to existing JWST datasets. Doing so would provide direct constraints on the EoR morphology and, in turn, on the galaxy–IGM connection.

Bibliography

Abazajian, Kevork N., Peter Adshead, Zeeshan Ahmed, et al. (Oct. 2016). “CMB-S4 Science Book, First Edition”. In: *arXiv e-prints*, arXiv:1610.02743, arXiv:1610.02743. arXiv: 1610.02743 [astro-ph.CO].

Abbott, T. M. C., M. Aguena, A. Alarcon, et al. (Jan. 2022). “Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing”. In: *Phys. Rev. D* 105.2, 023520, p. 023520. DOI: 10.1103/PhysRevD.105.023520. arXiv: 2105.13549 [astro-ph.CO].

Abdurashidova, Zara, James E. Aguirre, Paul Alexander, et al. (Jan. 2022). “HERA Phase I Limits on the Cosmic 21 cm Signal: Constraints on Astrophysics and Cosmology during the Epoch of Reionization”. In: *ApJ* 924.2, 51, p. 51. DOI: 10.3847/1538-4357/ac2ffc. arXiv: 2108.07282 [astro-ph.CO].

Abitbol, Maximilian H., Zeeshan Ahmed, Darcy Barron, et al. (June 2017). “CMB-S4 Technology Book, First Edition”. In: *arXiv e-prints*, arXiv:1706.02464, arXiv:1706.02464. arXiv: 1706.02464 [astro-ph.IM].

Adame, A. G., J. Aguilar, S. Ahlen, et al. (Feb. 2025). “DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations”. In: *JCAP* 2025.2, 021, p. 021. DOI: 10.1088/1475-7516/2025/02/021. arXiv: 2404.03002 [astro-ph.CO].

Adams, Nathan J., Christopher J. Conselice, Duncan Austin, et al. (Apr. 2024). “EPOCHS. II. The Ultraviolet Luminosity Function from 7.5 ¡ z ¡ 13.5 Using 180 arcmin^2 of Deep, Blank Fields from the PEARLS Survey and Public JWST Data”. In: *ApJ* 965.2, 169, p. 169. DOI: 10.3847/1538-4357/ad2a7b. arXiv: 2304.13721 [astro-ph.GA].

Ade, Peter, James Aguirre, Zeeshan Ahmed, et al. (Feb. 2019). “The Simons Observatory: science goals and forecasts”. In: *JCAP* 2019.2, 056, p. 056. DOI: 10.1088/1475-7516/2019/02/056. arXiv: 1808.07445 [astro-ph.CO].

Allende Prieto, C., L. Koesterke, I. Hubeny, et al. (Oct. 2018). “A collection of model stellar spectra for spectral types B to early-M”. In: *A&A* 618, A25, A25. DOI: 10.1051/0004-6361/201732484. arXiv: 1807.06049 [astro-ph.SR].

Almada Monter, Silvia and Max Gronke (Oct. 2024). “Crossing walls and windows: the curious escape of Lyman-α photons through ionized channels”. In: *MNRAS* 534.1, pp. L7–L13. DOI: 10.1093/mnrasl/slae074. arXiv: 2404.07169 [astro-ph.GA].

Alvarez, Marcelo A. (June 2016). “The Kinetic Sunyaev-Zel’dovich Effect from Reionization: Simulated Full-sky Maps at Arcminute Resolution”. In: *ApJ* 824.2, 118, p. 118. DOI: 10.3847/0004-637X/824/2/118. arXiv: 1511.02846 [astro-ph.CO].

Anau Montel, Noemi, James Alvey, and Christoph Weniger (June 2024). “Scalable inference with autoregressive neural ratio estimation”. In: *MNRAS* 530.4, pp. 4107–4124. DOI: 10.1093/mnras/stae1130. arXiv: 2308.08597 [astro-ph.IM].

Asplund, M., N. Grevesse, A. J. Sauval, et al. (Apr. 2004). “Line formation in solar granulation. IV. [O I], O I and OH lines and the photospheric O abundance”. In: *A&A* 417, pp. 751–768. DOI: 10.1051/0004-6361:20034328. arXiv: astro-ph/0312290 [astro-ph].

Atek, Hakim, Ivo Labbé, Lukas J. Furtak, et al. (Feb. 2024). “Most of the photons that reionized the Universe came from dwarf galaxies”. In: *Nature* 626.8001, pp. 975–978. DOI: 10.1038/s41586-024-07043-6. arXiv: 2308.08540 [astro-ph.GA].

Bañados, Eduardo, Bram P. Venemans, Chiara Mazzucchelli, et al. (Jan. 2018). “An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5”. In: *Nature* 553.7689, pp. 473–476. DOI: 10.1038/nature25180. arXiv: 1712.01860 [astro-ph.GA].

Barkana, R. and A. Loeb (July 2001). “In the beginning: the first sources of light and the reionization of the universe”. In: *Phys. Rep.* 349.2, pp. 125–238. DOI: 10.1016/S0370-1573(01)00019-9. arXiv: astro-ph/0010468 [astro-ph].

Barkana, Rennan and Abraham Loeb (July 2004). “Unusually Large Fluctuations in the Statistics of Galaxy Formation at High Redshift”. In: *ApJ* 609.2, pp. 474–481. DOI: 10.1086/421079. arXiv: astro-ph/0310338 [astro-ph].

Barlow, Roger (June 2004). "Asymmetric Statistical Errors". In: *arXiv e-prints*, physics/0406120, physics/0406120. arXiv: physics/0406120 [physics.data-an].

Barrow, Kirk S. S., Brant E. Robertson, Richard S. Ellis, et al. (Oct. 2020). "The Lyman Continuum Escape Survey: Connecting Time-dependent [O III] and [O II] Line Emission with Lyman Continuum Escape Fraction in Simulations of Galaxy Formation". In: *ApJ* 902.2, L39, p. L39. DOI: 10.3847/2041-8213/abbd8e. arXiv: 2010.00592 [astro-ph.GA].

Barrow, Kirk S. S., John H. Wise, Michael L. Norman, et al. (Aug. 2017). "First light: exploring the spectra of high-redshift galaxies in the Renaissance Simulations". In: *MNRAS* 469.4, pp. 4863–4878. DOI: 10.1093/mnras/stx1181. arXiv: 1701.02749 [astro-ph.IM].

Basu-Zych, Antara R., Bret D. Lehmer, Ann E. Hornschemeier, et al. (Sept. 2013). "Evidence for Elevated X-Ray Emission in Local Lyman Break Galaxy Analogs". In: *ApJ* 774.2, 152, p. 152. DOI: 10.1088/0004-637X/774/2/152. arXiv: 1306.0906 [astro-ph.CO].

Battaglia, N., A. Natarajan, H. Trac, et al. (Oct. 2013). "Reionization on Large Scales. III. Predictions for Low-l Cosmic Microwave Background Polarization and High-l Kinetic Sunyaev-Zel'dovich Observables". In: *ApJ* 776.2, 83, p. 83. DOI: 10.1088/0004-637X/776/2/83. arXiv: 1211.2832 [astro-ph.CO].

Becker, George D., James S. Bolton, Piero Madau, et al. (Mar. 2015). "Evidence of patchy hydrogen reionization from an extreme Lyα trough below redshift six". In: *MNRAS* 447.4, pp. 3402–3419. DOI: 10.1093/mnras/stu2646. arXiv: 1407.4850 [astro-ph.CO].

Becker, George D., Anson D'Aloisio, Holly M. Christenson, et al. (Dec. 2021). "The mean free path of ionizing photons at 5 ¡ z ¡ 6: evidence for rapid evolution near reionization". In: *MNRAS* 508.2, pp. 1853–1869. DOI: 10.1093/mnras/stab2696. arXiv: 2103.16610 [astro-ph.CO].

Becker, Robert H., Xiaohui Fan, Richard L. White, et al. (Dec. 2001). "Evidence for Reionization at z˜6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar". In: *AJ* 122.6, pp. 2850–2857. DOI: 10.1086/324231. arXiv: astro-ph/0108097 [astro-ph].

Bégin, Joëlle-Marie, Adrian Liu, and Adélie Gorce (Apr. 2022). "Joint constraints on reionization: A framework for combining the global 21 cm signal and the kinetic Sunyaev-Zel'dovich effect". In: *Phys. Rev. D* 105.8, 083503, p. 083503. DOI: 10.1103/PhysRevD.105.083503. arXiv: 2112.06933 [astro-ph.CO].

Behroozi, Peter, Risa H. Wechsler, Andrew P. Hearin, et al. (Sept. 2019). "UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10". In: *MNRAS* 488.3, pp. 3143–3194. DOI: 10.1093/mnras/stz1182. arXiv: 1806.07893 [astro-ph.GA].

Behroozi, Peter S., Risa H. Wechsler, and Charlie Conroy (June 2013). "The Average Star Formation Histories of Galaxies in Dark Matter Halos from z = 0-8". In: *ApJ* 770.1, 57, p. 57. DOI: 10.1088/0004-637X/770/1/57. arXiv: 1207.6105 [astro-ph.CO].

Bhatawdekar, Rachana, Christopher J. Conselice, Berta Margalef-Bentabol, et al. (July 2019). "Evolution of the galaxy stellar mass functions and UV luminosity functions at z = 6-9 in the Hubble Frontier Fields". In: *MNRAS* 486.3, pp. 3805–3830. DOI: 10.1093/mnras/stz866. arXiv: 1807.07580 [astro-ph.GA].

Bhowmick, Aklant K., Duncan Campbell, Tiziana Di Matteo, et al. (Nov. 2018). "Halo occupation distribution (HOD) modelling of high redshift galaxies using the BlueTides simulation". In: *MNRAS* 480.3, pp. 3177–3192. DOI: 10.1093/mnras/sty2128. arXiv: 1806.10612 [astro-ph.GA].

Bird, Simeon, Yueying Ni, Tiziana Di Matteo, et al. (May 2022). "The ASTRID simulation: galaxy formation and reionization". In: *MNRAS* 512.3, pp. 3703–3716. DOI: 10.1093/mnras/stac648. arXiv: 2111.01160 [astro-ph.GA].

Bolan, Patricia, Maruša Bradăc, Brian C. Lemaux, et al. (July 2024). "Ly α emission strength and stellar properties of faint galaxies from 5 ¡ z ¡ 8.2". In: *MNRAS* 531.3, pp. 2998–3010. DOI: 10.1093/mnras/stae1339. arXiv: 2403.00984 [astro-ph.GA].

Bolan, Patricia, Brian C. Lemaux, Charlotte Mason, et al. (Dec. 2022). "Inferring the intergalactic medium neutral fraction at z 6-8 with low-luminosity Lyman break galaxies". In: *MNRAS* 517.3, pp. 3263–3274. DOI: 10.1093/mnras/stac1963. arXiv: 2111.14912 [astro-ph.GA].

Bolton, James S., George D. Becker, J. Stuart B. Wyithe, et al. (July 2010). "A first direct measurement of the intergalactic medium temperature around a quasar at z = 6". In: *MNRAS* 406.1, pp. 612–625. DOI: 10.1111/j.1365-2966.2010.16701.x. arXiv: 1001.3415 [astro-ph.CO].

Bolton, James S. and Martin G. Haehnelt (Nov. 2007). "The observed ionization rate of the intergalactic medium and the ionizing emissivity at z ¿= 5: evidence for a photon-starved and extended epoch of reionization". In: *MNRAS* 382.1, pp. 325–341. DOI: 10.1111/j.1365-2966.2007.12372.x. arXiv: astro-ph/0703306 [astro-ph].

Bond, J. R., S. Cole, G. Efstathiou, et al. (Oct. 1991). “Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations”. In: *ApJ* 379, p. 440. DOI: 10.1086/170520.

Bond, J. R. and S. T. Myers (Mar. 1996). “The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms”. In: *ApJS* 103, p. 1. DOI: 10.1086/192267.

Bosman, Sarah E. I., Frederick B. Davies, George D. Becker, et al. (July 2022). “Hydrogen reionization ends by z = 5.3: Lyman-α optical depth measured by the XQR-30 sample”. In: *MNRAS* 514.1, pp. 55–76. DOI: 10.1093/mnras/stac1046. arXiv: 2108.03699 [astro-ph.CO].

Bosman, Sarah E. I., Dominika Ďurovčíková, Frederick B. Davies, et al. (May 2021). “A comparison of quasar emission reconstruction techniques for z $\geq$ 5.0 Lyman α and Lyman β transmission”. In: *MNRAS* 503.2, pp. 2077–2096. DOI: 10.1093/mnras/stab572. arXiv: 2006.10744 [astro-ph.GA].

Bosman, Sarah E. I., Xiaohui Fan, Linhua Jiang, et al. (Sept. 2018). “New constraints on Lyman-α opacity with a sample of 62 quasarsat z ¿ 5.7”. In: *MNRAS* 479.1, pp. 1055–1076. DOI: 10.1093/mnras/sty1344. arXiv: 1802.08177 [astro-ph.GA].

Bouwens, R. J., G. D. Illingworth, P. A. Oesch, et al. (Apr. 2015). “UV Luminosity Functions at Redshifts z $\sim$ 4 to z $\sim$ 10: 10,000 Galaxies from HST Legacy Fields”. In: *ApJ* 803.1, 34, p. 34. DOI: 10.1088/0004-637X/803/1/34. arXiv: 1403.4295 [astro-ph.CO].

Bouwens, R. J., G. D. Illingworth, P. G. van Dokkum, et al. (Mar. 2022). “Sizes of Lensed Lower-luminosity z = 4-8 Galaxies from the Hubble Frontier Field Program”. In: *ApJ* 927.1, 81, p. 81. DOI: 10.3847/1538-4357/ac4791. arXiv: 2112.02948 [astro-ph.GA].

Bouwens, R. J., P. A. Oesch, I. Labbé, et al. (Oct. 2016). “The Bright End of the z $\sim$ 9 and z $\sim$ 10 UV Luminosity Functions Using All Five CANDELS Fields*”. In: *ApJ* 830.2, 67, p. 67. DOI: 10.3847/0004-637X/830/2/67. arXiv: 1506.01035 [astro-ph.GA].

Bouwens, R. J., P. A. Oesch, M. Stefanon, et al. (Aug. 2021). “New Determinations of the UV Luminosity Functions from z 9 to 2 Show a Remarkable Consistency with Halo Growth and a Constant Star Formation Efficiency”. In: *AJ* 162.2, 47, p. 47. DOI: 10.3847/1538-3881/abf83e. arXiv: 2102.07775 [astro-ph.GA].

Bouwens, Rychard, Jorge González-López, Manuel Aravena, et al. (Oct. 2020). “The ALMA Spectroscopic Survey Large Program: The Infrared Excess of z = 1.5-10 UV-selected Galaxies and the Implied High-redshift Star Formation History”. In: *ApJ* 902.2, 112, p. 112. DOI: 10.3847/1538-4357/abb830. arXiv: 2009.10727 [astro-ph.GA].

Bouwens, Rychard J., Mauro Stefanon, Gabriel Brammer, et al. (July 2023a). “Evolution of the UV LF from z 15 to z 8 using new JWST NIRCam medium-band observations over the HUDF/XDF”. In: *MNRAS* 523.1, pp. 1036–1055. DOI: 10.1093/mnras/stad1145. arXiv: 2211.02607 [astro-ph.GA].

— (July 2023b). “Evolution of the UV LF from z 15 to z 8 using new JWST NIRCam medium-band observations over the HUDF/XDF”. In: *MNRAS* 523.1, pp. 1036–1055. DOI: 10.1093/mnras/stad1145. arXiv: 2211.02607 [astro-ph.GA].

Bowler, R. A. A., M. J. Jarvis, J. S. Dunlop, et al. (Apr. 2020). “A lack of evolution in the very bright end of the galaxy luminosity function from z $\simeq$ 8 to 10”. In: *MNRAS* 493.2, pp. 2059–2084. DOI: 10.1093/mnras/staa313. arXiv: 1911.12832 [astro-ph.GA].

Bowman, Judd D., Alan E. E. Rogers, Raul A. Monsalve, et al. (Mar. 2018). “An absorption profile centred at 78 megahertz in the sky-averaged spectrum”. In: *Nature* 555.7694, pp. 67–70. DOI: 10.1038/nature25792. arXiv: 1810.05912 [astro-ph.CO].

Brammer, Gabriel B., Pieter G. van Dokkum, and Paolo Coppi (Oct. 2008). “EAZY: A Fast, Public Photometric Redshift Code”. In: *ApJ* 686.2, pp. 1503–1513. DOI: 10.1086/591786. arXiv: 0807.1533 [astro-ph].

Braun, R., T. Bourke, J. A. Green, et al. (Apr. 2015). “Advancing Astrophysics with the Square Kilometre Array”. In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 174, p. 174. DOI: 10.22323/1.215.0174.

Breitman, Daniela, Andrei Mesinger, Steven G. Murray, et al. (Feb. 2024). “21CMEMU: an emulator of 21CMFAST summary observables”. In: *MNRAS* 527.4, pp. 9833–9852. DOI: 10.1093/mnras/stad3849. arXiv: 2309.05697 [astro-ph.CO].

Brinchmann, J., S. Charlot, S. D. M. White, et al. (July 2004). “The physical properties of star-forming galaxies in the low-redshift Universe”. In: *MNRAS* 351.4, pp. 1151–1179. DOI: 10.1111/j.1365-2966.2004.07881.x. arXiv: astro-ph/0311060 [astro-ph].

Bromm, Volker and Richard B. Larson (Sept. 2004). “The First Stars”. In: *ARA&A* 42.1, pp. 79–118. DOI: 10.1146/annurev.astro.42.053102.134034. arXiv: astro-ph/0311019 [astro-ph].

Brorby, M., P. Kaaret, A. Prestwich, et al. (Apr. 2016). “Enhanced X-ray emission from Lyman break analogues and a possible L_X-SFR-metallicity plane”. In: *MNRAS* 457.4, pp. 4081–4088. DOI: 10.1093/mnras/stw284. arXiv: 1602.01091 [astro-ph.HE].

Bruton, Sean, Claudia Scarlata, Francesco Haardt, et al. (Aug. 2023). “The Impact of Cosmic Variance on Inferences of Global Neutral Fraction Derived from Lyα Luminosity Functions during Reionization”. In: *ApJ* 953.1, 29, p. 29. DOI: 10.3847/1538-4357/acd179. arXiv: 2305.04949 [astro-ph.CO].

Bunker, Andrew J., Aayush Saxena, Alex J. Cameron, et al. (Sept. 2023). “JADES NIRSpec Spectroscopy of GN-z11: Lyman-α emission and possible enhanced nitrogen abundance in a z = 10.60 luminous galaxy”. In: *A&A* 677, A88, A88. DOI: 10.1051/0004-6361/202346159. arXiv: 2302.07256 [astro-ph.GA].

Byrne, C. M. and E. R. Stanway (June 2023). “On the impact of spectral template uncertainties in synthetic stellar populations”. In: *MNRAS* 521.4, pp. 4995–5012. DOI: 10.1093/mnras/stad832. arXiv: 2303.16920 [astro-ph.GA].

Byrne, C. M., E. R. Stanway, J. J. Eldridge, et al. (June 2022). “The dependence of theoretical synthetic spectra on α-enhancement in young, binary stellar populations”. In: *MNRAS* 512.4, pp. 5329–5338. DOI: 10.1093/mnras/stac807. arXiv: 2203.13275 [astro-ph.SR].

Byrohl, C. and M. Gronke (Oct. 2020). “Variations in shape among observed Lyman-α spectra due to intergalactic absorption”. In: *A&A* 642, L16, p. L16. DOI: 10.1051/0004-6361/202038685. arXiv: 2006.10041 [astro-ph.GA].

Cain, Christopher, Anson D’Aloisio, Nakul Gangolli, et al. (June 2023). “The morphology of reionization in a dynamically clumpy universe”. In: *MNRAS* 522.2, pp. 2047–2064. DOI: 10.1093/mnras/stad1057. arXiv: 2207.11266 [astro-ph.CO].

Calabrese, Erminia, Renée Hložek, Nick Battaglia, et al. (Aug. 2014). “Precision epoch of reionization studies with next-generation CMB experiments”. In: *JCAP* 2014.8, 010, p. 010. DOI: 10.1088/1475-7516/2014/08/010. arXiv: 1406.4794 [astro-ph.CO].

Cang, Junsong, Andrei Mesinger, Steven G. Murray, et al. (June 2025). “The EDGES measurement disfavors an excess radio background during the cosmic dawn”. In: *A&A* 698, A152, A152. DOI: 10.1051/0004-6361/202452982.

Carniani, Stefano, Francesco D’Eugenio, Xihan Ji, et al. (Apr. 2025). “The eventful life of a luminous galaxy at z = 14: metal enrichment, feedback, and low gas fraction?” In: *A&A* 696, A87, A87. DOI: 10.1051/0004-6361/202452451. arXiv: 2409.20533 [astro-ph.GA].

Carniani, Stefano, Kevin Hainline, Francesco D’Eugenio, et al. (Sept. 2024). “Spectroscopic confirmation of two luminous galaxies at a redshift of 14”. In: *Nature* 633.8029, pp. 318–322. DOI: 10.1038/s41586-024-07860-9. arXiv: 2405.18485 [astro-ph.GA].

Castellano, M., P. Dayal, L. Pentericci, et al. (Feb. 2016). “First Observational Support for Overlapping Reionized Bubbles Generated by a Galaxy Overdensity”. In: *ApJ* 818.1, L3, p. L3. DOI: 10.3847/2041-8205/818/1/L3. arXiv: 1601.03442 [astro-ph.GA].

Castellano, M., A. Fontana, E. Merlin, et al. (Apr. 2025). “Pushing JWST to the extremes: search and scrutiny of bright galaxy candidates at z≃15-30”. In: *arXiv e-prints*, arXiv:2504.05893, arXiv:2504.05893. DOI: 10.48550/arXiv.2504.05893. arXiv: 2504.05893 [astro-ph.GA].

Castellano, M., L. Pentericci, E. Vanzella, et al. (Aug. 2018). “Spectroscopic Investigation of a Reionized Galaxy Overdensity at z = 7”. In: *ApJ* 863.1, L3, p. L3. DOI: 10.3847/2041-8213/aad59b. arXiv: 1807.09277 [astro-ph.GA].

Castellano, Marco, Adriano Fontana, Tommaso Treu, et al. (May 2023). “Early Results from GLASS-JWST. XIX. A High Density of Bright Galaxies at z ≈ 10 in the A2744 Region”. In: *ApJ* 948.2, L14, p. L14. DOI: 10.3847/2041-8213/accea5. arXiv: 2212.06666 [astro-ph.GA].

Cen, Renyue and Taysun Kimm (Mar. 2015). “Quantifying Distributions of the Lyman Continuum Escape Fraction”. In: *ApJ* 801.2, L25, p. L25. DOI: 10.1088/2041-8205/801/2/L25. arXiv: 1502.04026 [astro-ph.GA].

Ceverino, Daniel, Ralf S. Klessen, and Simon C. O. Glover (Nov. 2018). “FirstLight - II. Star formation rates of primeval galaxies from z=5-15”. In: *MNRAS* 480.4, pp. 4842–4850. DOI: 10.1093/mnras/sty2124. arXiv: 1801.10382 [astro-ph.GA].

Chabrier, Gilles (July 2003). “Galactic Stellar and Substellar Initial Mass Function”. In: *PASP* 115.809, pp. 763–795. DOI: 10.1086/376392. arXiv: astro-ph/0304382 [astro-ph].

Chemerynska, Iryna, Hakim Atek, Pratika Dayal, et al. (Nov. 2024). “The Extreme Low-mass End of the Mass–Metallicity Relation at z ∼ 7”. In: *ApJ* 976.1, L15, p. L15. DOI: 10.3847/2041-8213/ad8dc9. arXiv: 2407.17110 [astro-ph.GA].

Chen, Nianyi, Hy Trac, Suvodip Mukherjee, et al. (Mar. 2022). "Patchy Kinetic Sunyaev-Zel'dovich Effect with Controlled Reionization History and Morphology". In: *arXiv e-prints*, arXiv:2203.04337, arXiv:2203.04337. arXiv: 2203.04337 [astro-ph.CO].

Chen, Zuyi, Daniel P. Stark, Charlotte Mason, et al. (Mar. 2024). "JWST spectroscopy of z 5-8 UV-selected galaxies: new constraints on the evolution of the Ly α escape fraction in the reionization era". In: *MNRAS* 528.4, pp. 7052–7075. DOI: 10.1093/mnras/stae455. arXiv: 2311.13683 [astro-ph.GA].

Chen, Zuyi, Daniel P. Stark, Charlotte A. Mason, et al. (May 2025). "The Impact of Galaxy Overdensities and Ionized Bubbles on Lyα Emission at $z \sim 7.0 - 8.5$". In: *arXiv e-prints*, arXiv:2505.24080, arXiv:2505.24080. DOI: 10.48550/arXiv.2505.24080. arXiv: 2505.24080 [astro-ph.GA].

Choudhury, T. Roy and A. Ferrara (Mar. 2006). "Physics of Cosmic Reionization". In: *arXiv e-prints*, astro-ph/0603149, astro–ph/0603149. arXiv: astro-ph/0603149 [astro-ph].

Choudhury, T. Roy, Aseem Paranjape, and Sarah E. I. Bosman (Mar. 2021). "Studying the Lyman α optical depth fluctuations at z $\sim$ 5.5 using fast semi-numerical methods". In: *MNRAS* 501.4, pp. 5782–5796. DOI: 10.1093/mnras/stab045. arXiv: 2003.08958 [astro-ph.CO].

Choudhury, Tirthankar Roy, Suvodip Mukherjee, and Sourabh Paul (Jan. 2021). "Cosmic microwave background constraints on a physical model of reionization". In: *MNRAS* 501.1, pp. L7–L11. DOI: 10.1093/mnrasl/slaa185. arXiv: 2007.03705 [astro-ph.CO].

Ciardi, B., A. Ferrara, and S. D. M. White (Sept. 2003). "Early reionization by the first galaxies". In: *MNRAS* 344.1, pp. L7–L11. DOI: 10.1046/j.1365-8711.2003.06976.x. arXiv: astro-ph/0302451 [astro-ph].

Ciesla, L., D. Elbaz, O. Ilbert, et al. (June 2024). "Identification of a transition from stochastic to secular star formation around z = 9 with JWST". In: *A&A* 686, A128, A128. DOI: 10.1051/0004-6361/202348091. arXiv: 2309.15720 [astro-ph.GA].

Cranmer, Kyle, Johann Brehmer, and Gilles Louppe (Dec. 2020). "The frontier of simulation-based inference". In: *Proceedings of the National Academy of Science* 117.48, pp. 30055–30062. DOI: 10.1073/pnas.1912789117. arXiv: 1911.01429 [stat.ML].

Crawford, T. M., K. K. Schaffer, S. Bhattacharya, et al. (Apr. 2014). "A Measurement of the Secondary-CMB and Millimeter-wave-foreground Bispectrum using 800 deg^2 of South Pole Telescope Data". In: *ApJ* 784.2, 143, p. 143. DOI: 10.1088/0004-637X/784/2/143. arXiv: 1303.3535 [astro-ph.CO].

Cueto, Elie R., Anne Hutter, Pratika Dayal, et al. (June 2024). "ASTRAEUS. IX. Impact of an evolving stellar initial mass function on early galaxies and reionisation". In: *A&A* 686, A138, A138. DOI: 10.1051/0004-6361/202349017. arXiv: 2312.12109 [astro-ph.GA].

Cullen, F., A. C. Carnall, D. Scholte, et al. (May 2025). "The JWST EXCELS survey: an extremely metal-poor galaxy at z = 8.271 hosting an unusual population of massive stars". In: *MNRAS*. DOI: 10.1093/mnras/staf838. arXiv: 2501.11099 [astro-ph.GA].

Curti, Mirko, Roberto Maiolino, Emma Curtis-Lake, et al. (Apr. 2024). "JADES: Insights into the low-mass end of the mass-metallicity-SFR relation at 3 ¡ z ¡ 10 from deep JWST/NIRSpec spectroscopy". In: *A&A* 684, A75, A75. DOI: 10.1051/0004-6361/202346698. arXiv: 2304.08516 [astro-ph.GA].

Curti, Mirko, Filippo Mannucci, Giovanni Cresci, et al. (Jan. 2020a). "The mass-metallicity and the fundamental metallicity relation revisited on a fully T$_e$-based abundance scale for galaxies". In: *MNRAS* 491.1, pp. 944–964. DOI: 10.1093/mnras/stz2910. arXiv: 1910.00597 [astro-ph.GA].

— (Jan. 2020b). "The mass-metallicity and the fundamental metallicity relation revisited on a fully T$_e$-based abundance scale for galaxies". In: *MNRAS* 491.1, pp. 944–964. DOI: 10.1093/mnras/stz2910. arXiv: 1910.00597 [astro-ph.GA].

Curtis-Lake, E., J. Chevallard, S. Charlot, et al. (May 2021). "Modelling the M*-SFR relation at high redshift: untangling factors driving biases in the intrinsic scatter measurement". In: *MNRAS* 503.4, pp. 4855–4877. DOI: 10.1093/mnras/stab698. arXiv: 2001.08560 [astro-ph.GA].

Curtis-Lake, Emma, Stefano Carniani, Alex Cameron, et al. (May 2023). "Spectroscopic confirmation of four metal-poor galaxies at z = 10.3-13.2". In: *Nature Astronomy* 7, pp. 622–632. DOI: 10.1038/s41550-023-01918-w. arXiv: 2212.04568 [astro-ph.GA].

D'Aloisio, Anson, Matthew McQuinn, Oliver Maupin, et al. (Apr. 2019a). "Heating of the Intergalactic Medium by Hydrogen Reionization". In: *ApJ* 874.2, 154, p. 154. DOI: 10.3847/1538-4357/ab0d83. arXiv: 1807.09282 [astro-ph.CO].

D'Aloisio, Anson, Matthew McQuinn, Oliver Maupin, et al. (Apr. 2019b). "Heating of the Intergalactic Medium by Hydrogen Reionization". In: *ApJ* 874.2, 154, p. 154. DOI: 10.3847/1538-4357/ab0d83.

D'Eugenio, Francesco, Roberto Maiolino, Stefano Carniani, et al. (Sept. 2024). "JADES: Carbon enrichment 350 Myr after the Big Bang". In: *A&A* 689, A152, A152. DOI: 10.1051/0004-6361/202348636. arXiv: 2311.09908 [astro-ph.GA].

D'Odorico, Valentina, E. Bañados, G. D. Becker, et al. (July 2023). "XQR-30: The ultimate XSHOOTER quasar sample at the reionization epoch". In: *MNRAS* 523.1, pp. 1399–1420. DOI: 10.1093/mnras/stad1468.

Das, Arpan, Andrei Mesinger, Andrea Pallottini, et al. (July 2017). "High-mass X-ray binaries and the cosmic 21-cm signal: impact of host galaxy absorption". In: *MNRAS* 469.1, pp. 1166–1174. DOI: 10.1093/mnras/stx943. arXiv: 1702.00409 [astro-ph.CO].

Das, Sudeep, Thibaut Louis, Michael R. Nolta, et al. (Apr. 2014). "The Atacama Cosmology Telescope: temperature and gravitational lensing power spectrum measurements from three seasons of data". In: *JCAP* 2014.4, 014, p. 014. DOI: 10.1088/1475-7516/2014/04/014. arXiv: 1301.1037 [astro-ph.CO].

Das, Sudeep, Tobias A. Marriage, Peter A. R. Ade, et al. (Mar. 2011). "The Atacama Cosmology Telescope: A Measurement of the Cosmic Microwave Background Power Spectrum at 148 and 218 GHz from the 2008 Southern Survey". In: *ApJ* 729.1, 62, p. 62. DOI: 10.1088/0004-637X/729/1/62. arXiv: 1009.0847 [astro-ph.CO].

Davé, Romeel, Daniel Anglés-Alcázar, Desika Narayanan, et al. (June 2019). "SIMBA: Cosmological simulations with black hole growth and feedback". In: *MNRAS* 486.2, pp. 2827–2849. DOI: 10.1093/mnras/stz937. arXiv: 1901.10203 [astro-ph.GA].

Davies, Frederick B., Sarah E. I. Bosman, and Steven R. Furlanetto (June 2024). "The Predicament of Absorption-dominated Reionization II: Observational Estimate of the Clumping Factor at the End of Reionization". In: *arXiv e-prints*, arXiv:2406.18186, arXiv:2406.18186. DOI: 10.48550/arXiv.2406.18186. arXiv: 2406.18186 [astro-ph.CO].

Davies, Frederick B., Sarah E. I. Bosman, Prakash Gaikwad, et al. (Apr. 2024). "Constraints on the Evolution of the Ionizing Background and Ionizing Photon Mean Free Path at the End of Reionization". In: *ApJ* 965.2, 134, p. 134. DOI: 10.3847/1538-4357/ad1d5d. arXiv: 2312.08464 [astro-ph.CO].

Davies, Frederick B. and Steven R. Furlanetto (Aug. 2016). "Large fluctuations in the hydrogen-ionizing background and mean free path following the epoch of reionization". In: *MNRAS* 460.2, pp. 1328–1339. DOI: 10.1093/mnras/stw931. arXiv: 1509.07131 [astro-ph.CO].

— (July 2022). "Improved treatments of the ionizing photon mean free path in seminumerical simulations of reionization". In: *MNRAS* 514.1, pp. 1302–1314. DOI: 10.1093/mnras/stac1005. arXiv: 2103.09821 [astro-ph.CO].

Davies, Frederick B., Joseph F. Hennawi, Eduardo Bañados, et al. (Sept. 2018). "Quantitative Constraints on the Reionization History from the IGM Damping Wing Signature in Two Quasars at z ¿ 7". In: *ApJ* 864.2, 142, p. 142. DOI: 10.3847/1538-4357/aad6dc. arXiv: 1802.06066 [astro-ph.CO].

Davies, James E., Simeon Bird, Simon Mutch, et al. (Oct. 2023). "Computationally efficient reionization in a large hydrodynamic galaxy formation simulation". In: *MNRAS* 525.2, pp. 2553–2564. DOI: 10.1093/mnras/stad2433. arXiv: 2306.07861 [astro-ph.CO].

Davies, James E., Andrei Mesinger, and Steven Murray (Apr. 2025). "Efficient simulation of discrete galaxy populations and associated radiation fields during the first billion years". In: *arXiv e-prints*, arXiv:2504.17254, arXiv:2504.17254. DOI: 10.48550/arXiv.2504.17254. arXiv: 2504.17254 [astro-ph.CO].

Dayal, Pratika and Andrea Ferrara (Dec. 2018). "Early galaxy formation and its large-scale effects". In: *Phys. Rep.* 780, pp. 1–64. DOI: 10.1016/j.physrep.2018.10.002. arXiv: 1809.09136 [astro-ph.GA].

Dayal, Pratika, Marta Volonteri, Jenny E. Greene, et al. (May 2025). "UNCOVERing the contribution of black holes to reionization". In: *A&A* 697, A211, A211. DOI: 10.1051/0004-6361/202449331. arXiv: 2401.11242 [astro-ph.GA].

De Barros, S., L. Pentericci, E. Vanzella, et al. (Dec. 2017). "VLT/FORS2 view at z 6: Lyman-α emitter fraction and galaxy physical properties at the edge of the epoch of cosmic reionization". In: *A&A* 608, A123, A123. DOI: 10.1051/0004-6361/201731476. arXiv: 1710.01784 [astro-ph.GA].

de Santi, Natalí S. M., Francisco Villaescusa-Navarro, L. Raul Abramo, et al. (Oct. 2023). "Field-level simulation-based inference with galaxy catalogs: the impact of systematic effects". In: *arXiv e-prints*, arXiv:2310.15234, arXiv:2310.15234. DOI: 10.48550/arXiv.2310.15234. arXiv: 2310.15234 [astro-ph.CO].

Di Cesare, C., L. Graziani, R. Schneider, et al. (Mar. 2023). "The assembly of dusty galaxies at z $\geq$ 4: the build-up of stellar mass and its scaling relations with hints from early JWST data". In: *MNRAS* 519.3, pp. 4632–4650. DOI: 10.1093/mnras/stac3702. arXiv: 2209.05496 [astro-ph.GA].

Dijkstra, Mark (Oct. 2014). "Lyα Emitting Galaxies as a Probe of Reionisation". In: *PASA* 31, e040, e040. DOI: 10.1017/pasa.2014.33. arXiv: 1406.7292 [astro-ph.CO].

Dijkstra, Mark, Zoltán Haiman, Andrei Mesinger, et al. (Dec. 2008). "Fluctuations in the high-redshift Lyman-Werner background: close halo pairs as the origin of supermassive black holes". In: *MNRAS* 391.4, pp. 1961–1972. DOI: 10.1111/j.1365-2966.2008.14031.x. arXiv: 0810.0014 [astro-ph].

Dijkstra, Mark, Andrei Mesinger, and J. Stuart B. Wyithe (July 2011). "The detectability of Lyα emission from galaxies during the epoch of reionization". In: *MNRAS* 414.3, pp. 2139–2147. DOI: 10.1111/j.1365-2966.2011.18530.x. arXiv: 1101.5160 [astro-ph.CO].

Dijkstra, Mark and J. Stuart B. Wyithe (Feb. 2012). "An empirical study of the relationship between Lyα and UV-selected galaxies: do theorists and observers 'select' the same objects?" In: *MNRAS* 419.4, pp. 3181–3193. DOI: 10.1111/j.1365-2966.2011.19958.x. arXiv: 1108.3840 [astro-ph.CO].

Dixon, Keri L., Ilian T. Iliev, Garrelt Mellema, et al. (Mar. 2016). "The large-scale observational signatures of low-mass galaxies during reionization". In: *MNRAS* 456.3, pp. 3011–3029. DOI: 10.1093/mnras/stv2887. arXiv: 1512.03836 [astro-ph.CO].

Dodelson, Scott (2003). *Modern Cosmology*.

Donnan, C. T., D. J. McLeod, J. S. Dunlop, et al. (Feb. 2023). "The evolution of the galaxy UV luminosity function at redshifts z $\simeq$ 8 - 15 from deep JWST and ground-based near-infrared imaging". In: *MNRAS* 518.4, pp. 6011–6040. DOI: 10.1093/mnras/stac3472. arXiv: 2207.12356 [astro-ph.GA].

Donnan, C. T., R. J. McLure, J. S. Dunlop, et al. (Sept. 2024). "JWST PRIMER: a new multifield determination of the evolving galaxy UV luminosity function at redshifts z $\simeq$ 9 - 15". In: *MNRAS* 533.3, pp. 3222–3237. DOI: 10.1093/mnras/stae2037. arXiv: 2403.03171 [astro-ph.GA].

Doussot, Aristide and Benoı ̂ t Semelin (Nov. 2022). "A bubble size distribution model for the Epoch of Reionization". In: *A&A* 667, A118, A118. DOI: 10.1051/0004-6361/202244108. arXiv: 2208.14044 [astro-ph.CO].

Duan, Qiao, Christopher J. Conselice, Qiong Li, et al. (June 2025). "Galaxy mergers in the epoch of reionization – I. A JWST study of pair fractions, merger rates, and stellar mass accretion rates at z = 4.5–11.5". In: *MNRAS* 540.1, pp. 774–805. DOI: 10.1093/mnras/staf638. arXiv: 2407.09472 [astro-ph.GA].

Dunkley, J., E. Calabrese, J. Sievers, et al. (July 2013). "The Atacama Cosmology Telescope: likelihood for small-scale CMB data". In: *JCAP* 2013.7, 025, p. 025. DOI: 10.1088/1475-7516/2013/07/025. arXiv: 1301.0776 [astro-ph.CO].

Ďurovčíková, Dominika, Anna-Christina Eilers, Huanqing Chen, et al. (July 2024). "Chronicling the Reionization History at 6 $\lesssim$ z $\lesssim$ 7 with Emergent Quasar Damping Wings". In: *ApJ* 969.2, 162, p. 162. DOI: 10.3847/1538-4357/ad4888. arXiv: 2401.10328 [astro-ph.CO].

Efstathiou, George (Aug. 2021). "To H_0 or not to H_0?" In: *MNRAS* 505.3, pp. 3866–3872. DOI: 10.1093/mnras/stab1588. arXiv: 2103.08723 [astro-ph.CO].

Eide, Marius B., Luca Graziani, Benedetta Ciardi, et al. (May 2018). "The epoch of cosmic heating by early sources of X-rays". In: *MNRAS* 476.1, pp. 1174–1190. DOI: 10.1093/mnras/sty272. arXiv: 1801.09719 [astro-ph.CO].

Einstein, Albert (Jan. 1915). "Zur allgemeinen Relativitätstheorie". In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pp. 778–786.

Eisenstein, Daniel J., Chris Willott, Stacey Alberts, et al. (June 2023). "Overview of the JWST Advanced Deep Extragalactic Survey (JADES)". In: *arXiv e-prints*, arXiv:2306.02465, arXiv:2306.02465. DOI: 10.48550/arXiv.2306.02465. arXiv: 2306.02465 [astro-ph.GA].

Eke, Vincent R., Shaun Cole, and Carlos S. Frenk (Sept. 1996). "Cluster evolution as a diagnostic for Omega". In: *MNRAS* 282, pp. 263–280. DOI: 10.1093/mnras/282.1.263. arXiv: astro-ph/9601088 [astro-ph].

Eldridge, J. J., E. R. Stanway, L. Xiao, et al. (Nov. 2017). "Binary Population and Spectral Synthesis Version 2.1: Construction, Observational Verification, and New Results". In: *PASA* 34, e058, e058. DOI: 10.1017/pasa.2017.51. arXiv: 1710.02154 [astro-ph.SR].

Endsley, Ryan and Daniel P. Stark (Apr. 2022). "Strong Lyman-α emission in an overdense region at z = 6.8: a very large (R 3 physical Mpc) ionized bubble in COSMOS?" In: *MNRAS* 511.4, pp. 6042–6054. DOI: 10.1093/mnras/stac524. arXiv: 2112.14779 [astro-ph.GA].

Endsley, Ryan, Daniel P. Stark, Rychard J. Bouwens, et al. (Dec. 2022). “The REBELS ALMA Survey: efficient Ly α transmission of UV-bright z $\simeq$ 7 galaxies from large velocity offsets and broad line widths”. In: *MNRAS* 517.4, pp. 5642–5659. DOI: 10.1093/mnras/stac3064. arXiv: 2202.01219 [astro-ph.GA].

Endsley, Ryan, Daniel P. Stark, Lily Whitler, Michael W. Topping, Zuyi Chen, et al. (Sept. 2023). “A JWST/NIRCam study of key contributors to reionization: the star-forming and ionizing properties of UV-faint z 7-8 galaxies”. In: *MNRAS* 524.2, pp. 2312–2330. DOI: 10.1093/mnras/stad1919. arXiv: 2208.14999 [astro-ph.GA].

Endsley, Ryan, Daniel P. Stark, Lily Whitler, Michael W. Topping, Benjamin D. Johnson, et al. (Sept. 2024). “The star-forming and ionizing properties of dwarf z 6-9 galaxies in JADES: insights on bursty star formation and ionized bubble growth”. In: *MNRAS* 533.1, pp. 1111–1142. DOI: 10.1093/mnras/stae1857. arXiv: 2306.05295 [astro-ph.GA].

Fathivavsari, Hassan (Aug. 2020). “Deep Learning Prediction of the Broad Lyα Emission Line of Quasars”. In: *ApJ* 898.2, 114, p. 114. DOI: 10.3847/1538-4357/ab9b7d. arXiv: 2006.05124 [astro-ph.GA].

Faucher-Giguère, Claude-André (Jan. 2018). “A model for the origin of bursty star formation in galaxies”. In: *MNRAS* 473.3, pp. 3717–3731. DOI: 10.1093/mnras/stx2595. arXiv: 1701.04824 [astro-ph.GA].

Faucher-Giguère, Claude-André, Dušan Kereš, and Chung-Pei Ma (Nov. 2011). “The baryonic assembly of dark matter haloes”. In: *MNRAS* 417.4, pp. 2982–2999. DOI: 10.1111/j.1365-2966.2011.19457.x. arXiv: 1103.0001 [astro-ph.CO].

Ferland, G. J., K. T. Korista, D. A. Verner, et al. (July 1998). “CLOUDY 90: Numerical Simulation of Plasmas and Their Spectra”. In: *PASP* 110.749, pp. 761–778. DOI: 10.1086/316190.

Feroz, F., M. P. Hobson, and M. Bridges (Oct. 2009). “MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics”. In: *MNRAS* 398.4, pp. 1601–1614. DOI: 10.1111/j.1365-2966.2009.14548.x. arXiv: 0809.3437 [astro-ph].

Ferrara, Andrea, Andrea Pallottini, and Pratika Dayal (July 2023). “On the stunning abundance of super-early, luminous galaxies revealed by JWST”. In: *MNRAS* 522.3, pp. 3986–3991. DOI: 10.1093/mnras/stad1095. arXiv: 2208.00720 [astro-ph.GA].

Finkelstein, Steven L., Micaela B. Bagley, Henry C. Ferguson, et al. (Mar. 2023a). “CEERS Key Paper. I. An Early Look into the First 500 Myr of Galaxy Formation with JWST”. In: *ApJ* 946.1, L13, p. L13. DOI: 10.3847/2041-8213/acade4. arXiv: 2211.05792 [astro-ph.GA].

— (Mar. 2023b). “CEERS Key Paper. I. An Early Look into the First 500 Myr of Galaxy Formation with JWST”. In: *ApJ* 946.1, L13, p. L13. DOI: 10.3847/2041-8213/acade4. arXiv: 2211.05792 [astro-ph.GA].

Finkelstein, Steven L., Jr. Ryan Russell E., Casey Papovich, et al. (Sept. 2015). “The Evolution of the Galaxy Rest-frame Ultraviolet Luminosity Function over the First Two Billion Years”. In: *ApJ* 810.1, 71, p. 71. DOI: 10.1088/0004-637X/810/1/71. arXiv: 1410.5439 [astro-ph.GA].

Flury, Sophia R., Anne E. Jaskot, Harry C. Ferguson, et al. (May 2022). “The Low-redshift Lyman Continuum Survey. I. New, Diverse Local Lyman Continuum Emitters”. In: *ApJS* 260.1, 1, p. 1. DOI: 10.3847/1538-4365/ac5331. arXiv: 2201.11716 [astro-ph.GA].

Fornasini, Francesca M., Mariska Kriek, Ryan L. Sanders, et al. (Nov. 2019). “The MOSDEF Survey: The Metallicity Dependence of X-Ray Binary Populations at z $\sim$ 2”. In: *ApJ* 885.1, 65, p. 65. DOI: 10.3847/1538-4357/ab4653. arXiv: 1909.08635 [astro-ph.HE].

Fragos, T., B. D. Lehmer, S. Naoz, et al. (Oct. 2013). “Energy Feedback from X-Ray Binaries in the Early Universe”. In: *ApJ* 776.2, L31, p. L31. DOI: 10.1088/2041-8205/776/2/L31. arXiv: 1306.1405 [astro-ph.CO].

Fragos, Tassos, Jeff J. Andrews, Simone S. Bavera, et al. (Feb. 2023). “POSYDON: A General-purpose Population Synthesis Code with Detailed Binary-evolution Simulations”. In: *ApJS* 264.2, 45, p. 45. DOI: 10.3847/1538-4365/ac90c1. arXiv: 2202.05892 [astro-ph.SR].

Friedmann, A. (Jan. 1922). “Über die Krümmung des Raumes”. In: *Zeitschrift fur Physik* 10, pp. 377–386. DOI: 10.1007/BF01332580.

Fujimoto, Seiji, Rohan P. Naidu, John Chisholm, et al. (Jan. 2025). “GLIMPSE: An ultra-faint $\simeq 10^5$ $M_\odot$ Pop III Galaxy Candidate and First Constraints on the Pop III UV Luminosity Function at $z \simeq 6-7$”. In: *arXiv e-prints*, arXiv:2501.11678, arXiv:2501.11678. DOI: 10.48550/arXiv.2501.11678. arXiv: 2501.11678 [astro-ph.GA].

Fujimoto, Seiji, Bingjie Wang, John R. Weaver, et al. (Dec. 2024). “UNCOVER: A NIRSpec Census of Lensed Galaxies at z = 8.50–13.08 Probing a High-AGN Fraction and Ionized Bubbles in the Shadow”. In: *ApJ* 977.2, 250, p. 250. DOI: 10.3847/1538-4357/ad9027. arXiv: 2308.11609 [astro-ph.GA].

Furlanetto, Steven R., Matthew McQuinn, and Lars Hernquist (Jan. 2006). "Characteristic scales during reionization". In: *MNRAS* 365.1, pp. 115–126. DOI: 10.1111/j.1365-2966.2005.09687.x. arXiv: astro-ph/0507524 [astro-ph].

Furlanetto, Steven R. and Jordan Mirocha (Apr. 2022). "Bursty star formation during the Cosmic Dawn driven by delayed stellar feedback". In: *MNRAS* 511.3, pp. 3895–3909. DOI: 10.1093/mnras/stac310. arXiv: 2109.04488 [astro-ph.GA].

Furlanetto, Steven R., Matias Zaldarriaga, and Lars Hernquist (Sept. 2004). "The Growth of H II Regions During Reionization". In: *ApJ* 613.1, pp. 1–15. DOI: 10.1086/423025. arXiv: astro-ph/0403697 [astro-ph].

Gaikwad, Prakash, Martin G. Haehnelt, Fredrick B. Davies, et al. (Nov. 2023). "Measuring the photoionization rate, neutral fraction, and mean free path of H I ionizing photons at $4.9 \leq z \leq 6.0$ from a large sample of XShooter and ESI spectra". In: *MNRAS* 525.3, pp. 4093–4120. DOI: 10.1093/mnras/stad2566. arXiv: 2304.02038 [astro-ph.CO].

Gaikwad, Prakash, Michael Rauch, Martin G. Haehnelt, et al. (June 2020). "Probing the thermal state of the intergalactic medium at z ¿ 5 with the transmission spikes in high-resolution Ly α forest spectra". In: *MNRAS* 494.4, pp. 5091–5109. DOI: 10.1093/mnras/staa907. arXiv: 2001.10018 [astro-ph.CO].

Garcia, Alex M., Paul Torrey, Sara Ellison, et al. (June 2024). "Does the fundamental metallicity relation evolve with redshift? I: the correlation between offsets from the mass-metallicity relation and star formation rate". In: *MNRAS* 531.1, pp. 1398–1408. DOI: 10.1093/mnras/stae1252. arXiv: 2403.08856 [astro-ph.GA].

Geda, Robel, Andy D. Goulding, Bret D. Lehmer, et al. (Apr. 2024). "The High-mass X-Ray Binary Luminosity Functions of Dwarf Galaxies". In: *ApJ* 965.1, 67, p. 67. DOI: 10.3847/1538-4357/ad2fc0. arXiv: 2401.14477 [astro-ph.GA].

Gelli, Viola, Charlotte Mason, and Christopher C. Hayward (May 2024). "The impact of mass-dependent stochasticity at cosmic dawn". In: *arXiv e-prints*, arXiv:2405.13108, arXiv:2405.13108. DOI: 10.48550/arXiv.2405.13108. arXiv: 2405.13108 [astro-ph.GA].

George, E. M., C. L. Reichardt, K. A. Aird, et al. (Feb. 2015). "A Measurement of Secondary Cosmic Microwave Background Anisotropies from the 2500 Square-degree SPT-SZ Survey". In: *ApJ* 799.2, 177, p. 177. DOI: 10.1088/0004-637X/799/2/177. arXiv: 1408.3161 [astro-ph.CO].

Ghara, Raghunath, Garrelt Mellema, Sambit K. Giri, et al. (May 2018). "Prediction of the 21-cm signal from reionization: comparison between 3D and 1D radiative transfer schemes". In: *MNRAS* 476.2, pp. 1741–1755. DOI: 10.1093/mnras/sty314. arXiv: 1710.09397 [astro-ph.CO].

Gillet, Nicolas J. F., Andrei Mesinger, and Jaehong Park (Jan. 2020). "Combining high-z galaxy luminosity functions with Bayesian evidence". In: *MNRAS* 491.2, pp. 1980–1997. DOI: 10.1093/mnras/stz2988. arXiv: 1906.06296 [astro-ph.GA].

Giri, Sambit K., Garrelt Mellema, Keri L. Dixon, et al. (Jan. 2018). "Bubble size statistics during reionization from 21-cm tomography". In: *MNRAS* 473.3, pp. 2949–2964. DOI: 10.1093/mnras/stx2539. arXiv: 1706.00665 [astro-ph.CO].

Gorce, A., M. Douspis, N. Aghanim, et al. (Aug. 2018). "Observational constraints on key-parameters of cosmic reionisation history". In: *A&A* 616, A113, A113. DOI: 10.1051/0004-6361/201629661. arXiv: 1710.04152 [astro-ph.CO].

Gorce, A., S. Ilić, M. Douspis, et al. (Aug. 2020). "Improved constraints on reionisation from CMB observations: A parameterisation of the kSZ effect". In: *A&A* 640, A90, A90. DOI: 10.1051/0004-6361/202038170. arXiv: 2004.06616 [astro-ph.CO].

Gorce, Adélie, Marian Douspis, and Laura Salvati (Feb. 2022). "Retrieving cosmological information from small-scale CMB foregrounds II. The kinetic Sunyaev Zel'dovich effect". In: *arXiv e-prints*, arXiv:2202.08698, arXiv:2202.08698. arXiv: 2202.08698 [astro-ph.CO].

Grazian, A., E. Giallongo, D. Paris, et al. (June 2017). "Lyman continuum escape fraction of faint galaxies at z 3.3 in the CANDELS/GOODS-North, EGS, and COSMOS fields with LBC". In: *A&A* 602, A18, A18. DOI: 10.1051/0004-6361/201730447. arXiv: 1703.00354 [astro-ph.GA].

Greig, Bradley, S. E. I. Bosman, F. B. Davies, et al. (Sept. 2024). "Blind QSO reconstruction challenge: exploring methods to reconstruct the Ly α emission line of QSOs". In: *MNRAS* 533.3, pp. 3312–3343. DOI: 10.1093/mnras/stae1985. arXiv: 2404.01556 [astro-ph.CO].

Greig, Bradley and Andrei Mesinger (June 2015). "21CMMC: an MCMC analysis tool enabling astrophysical parameter studies of the cosmic 21 cm signal". In: *MNRAS* 449.4, pp. 4246–4263. DOI: 10.1093/mnras/stv571. arXiv: 1501.06576 [astro-ph.CO].

— (Mar. 2017). "The global history of reionization". In: *MNRAS* 465.4, pp. 4838–4852. DOI: 10.1093/mnras/stw3026. arXiv: 1605.05374 [astro-ph.CO].

— (July 2018). "21CMMC with a 3D light-cone: the impact of the co-evolution approximation on the astrophysics of reionization and cosmic dawn". In: *MNRAS* 477.3, pp. 3217–3229. DOI: 10.1093/mnras/sty796. arXiv: 1801.01592 [astro-ph.CO].

Greig, Bradley, Cathryn M. Trott, Nichole Barry, et al. (Jan. 2021). "Exploring reionization and high-z galaxy observables with recent multiredshift MWA upper limits on the 21-cm signal". In: *MNRAS* 500.4, pp. 5322–5335. DOI: 10.1093/mnras/staa3494. arXiv: 2008.02639 [astro-ph.CO].

Gunn, James E. and Bruce A. Peterson (Nov. 1965). "On the Density of Neutral Hydrogen in Intergalactic Space." In: *ApJ* 142, pp. 1633–1636. DOI: 10.1086/148444.

Guth, Alan H. (Jan. 1981). "Inflationary universe: A possible solution to the horizon and flatness problems". In: *Phys. Rev. D* 23.2, pp. 347–356. DOI: 10.1103/PhysRevD.23.347.

Haiman, Zoltan, Martin J. Rees, and Abraham Loeb (Aug. 1996). "H 2 Cooling of Primordial Gas Triggered by UV Irradiation". In: *ApJ* 467, p. 522. DOI: 10.1086/177628. arXiv: astro-ph/9511126 [astro-ph].

Haiman, Zoltán, Tom Abel, and Martin J. Rees (May 2000). "The Radiative Feedback of the First Cosmological Objects". In: *ApJ* 534.1, pp. 11–24. DOI: 10.1086/308723. arXiv: astro-ph/9903336 [astro-ph].

Harikane, Yuichi, Akio K. Inoue, Richard S. Ellis, et al. (Feb. 2025). "JWST, ALMA, and Keck Spectroscopic Constraints on the UV Luminosity Functions at z $\sim$ 7–14: Clumpiness and Compactness of the Brightest Galaxies in the Early Universe". In: *ApJ* 980.1, 138, p. 138. DOI: 10.3847/1538-4357/ad9b2c. arXiv: 2406.18352 [astro-ph.GA].

Harikane, Yuichi, Kimihiko Nakajima, Masami Ouchi, et al. (Jan. 2024). "Pure Spectroscopic Constraints on UV Luminosity Functions and Cosmic Star Formation History from 25 Galaxies at z $_{spec}$ = 8.61-13.20 Confirmed with JWST/NIRSpec". In: *ApJ* 960.1, 56, p. 56. DOI: 10.3847/1538-4357/ad0b7e. arXiv: 2304.06658 [astro-ph.GA].

Harikane, Yuichi, Masami Ouchi, Masamune Oguri, et al. (Mar. 2023). "A Comprehensive Study of Galaxies at z 9-16 Found in the Early JWST Data: Ultraviolet Luminosity Functions and Cosmic Star Formation History at the Pre-reionization Epoch". In: *ApJS* 265.1, 5, p. 5. DOI: 10.3847/1538-4365/acaaa9. arXiv: 2208.01612 [astro-ph.GA].

Harikane, Yuichi, Masami Ouchi, Yoshiaki Ono, et al. (Apr. 2016). "Evolution of Stellar-to-Halo Mass Ratio at z = 0 - 7 Identified by Clustering Analysis with the Hubble Legacy Imaging and Early Subaru/Hyper Suprime-Cam Survey Data". In: *ApJ* 821.2, 123, p. 123. DOI: 10.3847/0004-637X/821/2/123. arXiv: 1511.07873 [astro-ph.GA].

Harikane, Yuichi, Yechi Zhang, Kimihiko Nakajima, et al. (Mar. 2023). "JWST/NIRSpec First Census of Broad-Line AGNs at z=4-7: Detection of 10 Faint AGNs with M_BH~106̂-107̂ M_sun and Their Host Galaxy Properties". In: *arXiv e-prints*, arXiv:2303.11946, arXiv:2303.11946. DOI: 10.48550/arXiv.2303.11946. arXiv: 2303.11946 [astro-ph.GA].

Hassan, Sultan, Romeel Davé, Matthew McQuinn, et al. (May 2022). "Reionization with SIMBA: How Much Does Astrophysics Matter in Modeling Cosmic Reionization?" In: *ApJ* 931.1, 62, p. 62. DOI: 10.3847/1538-4357/ac69e2. arXiv: 2109.03840 [astro-ph.CO].

Hayes, Matthew J., Axel Runnholm, Claudia Scarlata, et al. (Apr. 2023). "Spectral shapes of the Ly α emission from galaxies - II. The influence of stellar properties and nebular conditions on the emergent Ly α profiles". In: *MNRAS* 520.4, pp. 5903–5927. DOI: 10.1093/mnras/stad477. arXiv: 2302.04875 [astro-ph.GA].

Hayes, Matthew J. and Claudia Scarlata (Sept. 2023). "On the Sizes of Ionized Bubbles Around Galaxies During the Reionization Epoch. The Spectral Shapes of the Lyα Emission from Galaxies". In: *ApJ* 954.1, L14, p. L14. DOI: 10.3847/2041-8213/acee6a. arXiv: 2303.03160 [astro-ph.GA].

Heinrich, Chen and Wayne Hu (Sept. 2021). "Reionization effective likelihood from Planck 2018 data". In: *Phys. Rev. D* 104.6, 063505, p. 063505. DOI: 10.1103/PhysRevD.104.063505. arXiv: 2104.13998 [astro-ph.CO].

Hirano, Shingo, Takashi Hosokawa, Naoki Yoshida, et al. (Feb. 2014). "One Hundred First Stars: Protostellar Evolution and the Final Masses". In: *ApJ* 781.2, 60, p. 60. DOI: 10.1088/0004-637X/781/2/60. arXiv: 1308.4456 [astro-ph.CO].

Hoag, A., M. Bradač, K. Huang, et al. (June 2019). "Constraining the Neutral Fraction of Hydrogen in the IGM at Redshift 7.5". In: *ApJ* 878.1, 12, p. 12. DOI: `10.3847/1538-4357/ab1de7`. arXiv: `1901.09001 [astro-ph.GA]`.

Holzbauer, Lauren N. and Steven R. Furlanetto (Jan. 2012). "Fluctuations in the high-redshift Lyman-Werner and Lyα radiation backgrounds". In: *MNRAS* 419.1, pp. 718–731. DOI: `10.1111/j.1365-2966.2011.19752.x`. arXiv: `1105.5648 [astro-ph.CO]`.

Hsiao, Tiger Yu-Yang, Abdurro'uf, Dan Coe, et al. (Sept. 2024). "JWST NIRSpec Spectroscopy of the Triply Lensed z = 10.17 Galaxy MACS0647–JD". In: *ApJ* 973.1, 8, p. 8. DOI: `10.3847/1538-4357/ad5da8`. arXiv: `2305.03042 [astro-ph.GA]`.

Hu, Weida, Crystal L. Martin, Max Gronke, et al. (Oct. 2023). "CLASSY VII Lyα Profiles: The Structure and Kinematics of Neutral Gas and Implications for LyC Escape in Reionization-era Analogs". In: *ApJ* 956.1, 39, p. 39. DOI: `10.3847/1538-4357/aceefd`. arXiv: `2307.04911 [astro-ph.GA]`.

Hubble, Edwin (Mar. 1929). "A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae". In: *Proceedings of the National Academy of Science* 15.3, pp. 168–173. DOI: `10.1073/pnas.15.3.168`.

Hutter, Anne, Pratika Dayal, Gustavo Yepes, et al. (May 2021). "Astraeus I: the interplay between galaxy formation and reionization". In: *MNRAS* 503.3, pp. 3698–3723. DOI: `10.1093/mnras/stab602`. arXiv: `2004.08401 [astro-ph.GA]`.

Hutter, Anne, Maxime Trebitsch, Pratika Dayal, et al. (Oct. 2023). "ASTRAEUS - VIII. A new framework for Lyman-α emitters applied to different reionization scenarios". In: *MNRAS* 524.4, pp. 6124–6148. DOI: `10.1093/mnras/stad2230`. arXiv: `2209.14592 [astro-ph.GA]`.

Iliev, Ilian T., Garrelt Mellema, Kyungjin Ahn, et al. (Mar. 2014). "Simulating cosmic reionization: how large a volume is large enough?" In: *MNRAS* 439.1, pp. 725–743. DOI: `10.1093/mnras/stt2497`. arXiv: `1310.7463 [astro-ph.CO]`.

Iliev, Ilian T., Ue-Li Pen, J. Richard Bond, et al. (May 2007). "The Kinetic Sunyaev-Zel'dovich Effect from Radiative Transfer Simulations of Patchy Reionization". In: *ApJ* 660.2, pp. 933–944. DOI: `10.1086/513687`. arXiv: `astro-ph/0609592 [astro-ph]`.

Inoue, Akio K., Kenji Hasegawa, Tomoaki Ishiyama, et al. (June 2018). "SILVERRUSH. VI. A simulation of Lyα emitters in the reionization epoch and a comparison with Subaru Hyper Suprime-Cam survey early data". In: *PASJ* 70.3, 55, p. 55. DOI: `10.1093/pasj/psy048`. arXiv: `1801.00067 [astro-ph.GA]`.

Iršič, Vid, Matteo Viel, Trystyn A. M. Berg, et al. (Apr. 2017). "The Lyman α forest power spectrum from the XQ-100 Legacy Survey". In: *MNRAS* 466.4, pp. 4332–4345. DOI: `10.1093/mnras/stw3372`. arXiv: `1702.01761 [astro-ph.CO]`.

Ishigaki, Masafumi, Ryota Kawamata, Masami Ouchi, et al. (Feb. 2018). "Full-data Results of Hubble Frontier Fields: UV Luminosity Functions at z $\sim$ 6-10 and a Consistent Picture of Cosmic Reionization". In: *ApJ* 854.1, 73, p. 73. DOI: `10.3847/1538-4357/aaa544`. arXiv: `1702.04867 [astro-ph.GA]`.

Izotov, Y. I., D. Schaerer, T. X. Thuan, et al. (Oct. 2016). "Detection of high Lyman continuum leakage from four low-redshift compact star-forming galaxies". In: *MNRAS* 461.4, pp. 3683–3701. DOI: `10.1093/mnras/stw1205`. arXiv: `1605.05160 [astro-ph.GA]`.

Jakobsen, P., P. Ferruit, C. Alves de Oliveira, et al. (May 2022). "The Near-Infrared Spectrograph (NIRSpec) on the James Webb Space Telescope. I. Overview of the instrument and its capabilities". In: *A&A* 661, A80, A80. DOI: `10.1051/0004-6361/202142663`. arXiv: `2202.03305 [astro-ph.IM]`.

Jaskot, Anne E., Anneliese C. Silveyra, Anna Plantinga, et al. (Sept. 2024). "Multivariate Predictors of Lyman Continuum Escape. I. A Survival Analysis of the Low-redshift Lyman Continuum Survey". In: *ApJ* 972.1, 92, p. 92. DOI: `10.3847/1538-4357/ad58b9`. arXiv: `2406.10171 [astro-ph.GA]`.

Jenkins, A., C. S. Frenk, S. D. M. White, et al. (Feb. 2001). "The mass function of dark matter haloes". In: *MNRAS* 321.2, pp. 372–384. DOI: `10.1046/j.1365-8711.2001.04029.x`. arXiv: `astro-ph/0005260 [astro-ph]`.

Jin, Xiangyu, Jinyi Yang, Xiaohui Fan, et al. (Jan. 2023). "(Nearly) Model-independent Constraints on the Neutral Hydrogen Fraction in the Intergalactic Medium at z 5-7 Using Dark Pixel Fractions in Lyα and Lyβ Forests". In: *ApJ* 942.2, 59, p. 59. DOI: `10.3847/1538-4357/aca678`. arXiv: `2211.12613 [astro-ph.CO]`.

Jones, Gareth C., Andrew J. Bunker, Aayush Saxena, Santiago Arribas, et al. (Jan. 2025). "JADES: measuring reionization properties using Lyman-alpha emission". In: *MNRAS* 536.3, pp. 2355–2380. DOI: `10.1093/mnras/stae2670`. arXiv: `2409.06405 [astro-ph.GA]`.

Jones, Gareth C., Andrew J. Bunker, Aayush Saxena, Joris Witstok, et al. (Mar. 2024). “JADES: The emergence and evolution of Lyα emission and constraints on the intergalactic medium neutral fraction”. In: *A&A* 683, A238, A238. DOI: 10.1051/0004-6361/202347099. arXiv: 2306.02471 [astro-ph.GA].

Jose, Charles, Cedric G. Lacey, and Carlton M. Baugh (Nov. 2016). “The clustering of dark matter haloes: scale-dependent bias on quasi-linear scales”. In: *MNRAS* 463.1, pp. 270–281. DOI: 10.1093/mnras/stw1702. arXiv: 1509.06715 [astro-ph.CO].

Jung, Intae, Steven L. Finkelstein, Mark Dickinson, et al. (Dec. 2020). “Texas Spectroscopic Search for Lyα Emission at the End of Reionization. III. The Lyα Equivalent-width Distribution and Ionized Structures at z ¿ 7”. In: *ApJ* 904.2, 144, p. 144. DOI: 10.3847/1538-4357/abbd44. arXiv: 2009.10092 [astro-ph.GA].

Jung, Intae, Steven L. Finkelstein, Rebecca L. Larson, et al. (Dec. 2022). “New $z > 7$ Lyman-alpha Emitters in EGS: Evidence of an Extended Ionized Structure at $z \sim 7.7$”. In: *arXiv e-prints*, arXiv:2212.09850, arXiv:2212.09850. DOI: 10.48550/arXiv.2212.09850. arXiv: 2212.09850 [astro-ph.GA].

Kageura, Yuta, Masami Ouchi, Minami Nakane, et al. (June 2025). “Census of Lyα Emission from ~600 Galaxies at z = 5–14: Evolution of the Lyα Luminosity Function and a Late Sharp Cosmic Reionization”. In: *ApJS* 278.2, 33, p. 33. DOI: 10.3847/1538-4365/adc690. arXiv: 2501.05834 [astro-ph.GA].

Kannan, R., E. Garaldi, A. Smith, et al. (Apr. 2022). “Introducing the THESAN project: radiation-magnetohydrodynamic simulations of the epoch of reionization”. In: *MNRAS* 511.3, pp. 4005–4030. DOI: 10.1093/mnras/stab3710. arXiv: 2110.00584 [astro-ph.GA].

Kauffmann, O. B., O. Ilbert, J. R. Weaver, et al. (Nov. 2022). “COSMOS2020: UV-selected galaxies at z ≥ 7.5”. In: *A&A* 667, A65, A65. DOI: 10.1051/0004-6361/202243088. arXiv: 2207.11740 [astro-ph.GA].

Kaur, Harman Deep, Nicolas Gillet, and Andrei Mesinger (June 2020). “Minimum size of 21-cm simulations”. In: *MNRAS* 495.2, pp. 2354–2362. DOI: 10.1093/mnras/staa1323. arXiv: 2004.06709 [astro-ph.CO].

Kaur, Harman Deep, Yuxiang Qin, Andrei Mesinger, et al. (July 2022). “The 21-cm signal from the cosmic dawn: metallicity dependence of high-mass X-ray binaries”. In: *MNRAS* 513.4, pp. 5097–5108. DOI: 10.1093/mnras/stac1226. arXiv: 2203.10851 [astro-ph.GA].

Kaurov, Alexander A., Tejaswi Venumadhav, Liang Dai, et al. (Sept. 2018). “Implication of the Shape of the EDGES Signal for the 21 cm Power Spectrum”. In: *ApJ* 864.1, L15, p. L15. DOI: 10.3847/2041-8213/aada4c. arXiv: 1805.03254 [astro-ph.CO].

Keating, Laura C., Ewald Puchwein, James S. Bolton, et al. (June 2024). “The origin of the characteristic shape and scatter of intergalactic damping wings during reionization”. In: *MNRAS* 531.1, pp. L34–L39. DOI: 10.1093/mnrasl/slae022. arXiv: 2308.11709 [astro-ph.CO].

Kennicutt Jr., Robert C. (Jan. 1998). “Star Formation in Galaxies Along the Hubble Sequence”. In: *ARA&A* 36, pp. 189–232. DOI: 10.1146/annurev.astro.36.1.189. arXiv: astro-ph/9807187 [astro-ph].

Kimm, Taysun, Renyue Cen, Julien Devriendt, et al. (Aug. 2015). “Towards simulating star formation in turbulent high-z galaxies with mechanical supernova feedback”. In: *MNRAS* 451.3, pp. 2900–2921. DOI: 10.1093/mnras/stv1211. arXiv: 1501.05655 [astro-ph.GA].

Kimm, Taysun, Harley Katz, Martin Haehnelt, et al. (Apr. 2017). “Feedback-regulated star formation and escape of LyC photons from mini-haloes during reionization”. In: *MNRAS* 466.4, pp. 4826–4846. DOI: 10.1093/mnras/stx052. arXiv: 1608.04762 [astro-ph.GA].

Klessen, Ralf S. and Simon C. O. Glover (Aug. 2023). “The First Stars: Formation, Properties, and Impact”. In: *ARA&A* 61, pp. 65–130. DOI: 10.1146/annurev-astro-071221-053453. arXiv: 2303.12500 [astro-ph.CO].

Komatsu, E., K. M. Smith, J. Dunkley, et al. (Feb. 2011). “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation”. In: *ApJS* 192.2, 18, p. 18. DOI: 10.1088/0067-0049/192/2/18. arXiv: 1001.4538 [astro-ph.CO].

Konno, Akira, Masami Ouchi, Takatoshi Shibuya, et al. (Jan. 2018). “SILVERRUSH. IV. Lyα luminosity functions at z = 5.7 and 6.6 studied with ~1300 Lyα emitters on the 14-21 deg^2 sky”. In: *PASJ* 70, S16, S16. DOI: 10.1093/pasj/psx131. arXiv: 1705.01222 [astro-ph.GA].

Koopmans, L., J. Pritchard, G. Mellema, et al. (Apr. 2015). “The Cosmic Dawn and Epoch of Reionisation with SKA”. In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 1, p. 1. DOI: 10.22323/1.215.0001. arXiv: 1505.07568 [astro-ph.CO].

Kostyuk, Ivan and Benedetta Ciardi (Dec. 2024). “Influence of mergers on LyC escape of high redshift galaxies”. In: *arXiv e-prints*, arXiv:2412.04348, arXiv:2412.04348. DOI: 10.48550/arXiv.2412.04348. arXiv: 2412.04348 [astro-ph.GA].

Kostyuk, Ivan, Dylan Nelson, Benedetta Ciardi, et al. (May 2023). “Ionizing photon production and escape fractions during cosmic reionization in the TNG50 simulation”. In: *MNRAS* 521.2, pp. 3077–3097. DOI: `10.1093/mnras/stad677`. arXiv: `2207.11278 [astro-ph.GA]`.

Kragh Jespersen, Christian, Charles L. Steinhardt, Rachel S. Somerville, et al. (Feb. 2024). “On the Significance of Rare Objects at High Redshift: The Impact of Cosmic Variance”. In: *arXiv e-prints*, arXiv:2403.00050, arXiv:2403.0005 DOI: `10.48550/arXiv.2403.00050`. arXiv: `2403.00050 [astro-ph.GA]`.

Kravtsov, Andrey and Vasily Belokurov (May 2024). “Stochastic star formation and the abundance of $z > 10$ UV-bright galaxies”. In: *arXiv e-prints*, arXiv:2405.04578, arXiv:2405.04578. DOI: `10.48550/arXiv.2405.04578`. arXiv: `2405.04578 [astro-ph.GA]`.

Kreilgaard, Kimi C., Charlotte A. Mason, Fergus Cullen, et al. (May 2024). “Inferring the Distribution of the Ionising Photon Escape Fraction”. In: *A&A, in press* 10.1051/0004-6361/202450747. DOI: `10.48550/arXiv.2405.10364`.

Lacey, Cedric and Shaun Cole (June 1993). “Merger rates in hierarchical models of galaxy formation”. In: *MNRAS* 262.3, pp. 627–649. DOI: `10.1093/mnras/262.3.627`.

Larson, Rebecca L., Steven L. Finkelstein, Taylor A. Hutchison, et al. (May 2022). “Searching for Islands of Reionization: A Potential Ionized Bubble Powered by a Spectroscopic Overdensity at $z = 8.7$”. In: *ApJ* 930.2, 104, p. 104. DOI: `10.3847/1538-4357/ac5dbd`. arXiv: `2203.08461 [astro-ph.GA]`.

Larson, Richard B. (Oct. 2003). “The physics of star formation”. In: *Reports on Progress in Physics* 66.10, pp. 1651–1697. DOI: `10.1088/0034-4885/66/10/R03`. arXiv: `astro-ph/0306595 [astro-ph]`.

Laursen, Peter, Jesper Sommer-Larsen, and Alexei O. Razoumov (Feb. 2011). “Intergalactic Transmission and Its Impact on the Lyα Line”. In: *ApJ* 728.1, 52, p. 52. DOI: `10.1088/0004-637X/728/1/52`. arXiv: `1009.1384 [astro-ph.CO]`.

Leethochawalit, Nicha, Guido Roberts-Borsani, Takahiro Morishita, et al. (Oct. 2023). “The UV luminosity functions of bright z ¿ 8 galaxies: determination from 0.41 deg^2 of HST observations along 300 independent sightlines”. In: *MNRAS* 524.4, pp. 5454–5467. DOI: `10.1093/mnras/stad2202`. arXiv: `2205.15388 [astro-ph.GA]`.

Lehmer, B. D., A. R. Basu-Zych, S. Mineo, et al. (July 2016). “The Evolution of Normal Galaxy X-Ray Emission through Cosmic History: Constraints from the 6 MS Chandra Deep Field-South”. In: *ApJ* 825.1, 7, p. 7. DOI: `10.3847/0004-637X/825/1/7`. arXiv: `1604.06461 [astro-ph.GA]`.

Lehmer, Bret D., Rafael T. Eufrasio, Antara Basu-Zych, et al. (Jan. 2021). “The Metallicity Dependence of the High-mass X-Ray Binary Luminosity Function”. In: *ApJ* 907.1, 17, p. 17. DOI: `10.3847/1538-4357/abcec1`. arXiv: `2011.09476 [astro-ph.GA]`.

Leitherer, Claus and Timothy M. Heckman (Jan. 1995). “Synthetic Properties of Starburst Galaxies”. In: *ApJS* 96, p. 9. DOI: `10.1086/192112`.

Lemaître, G. (Jan. 1927). “Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques”. In: *Annales de la Société Scientifique de Bruxelles* 47, pp. 49–59.

Lemos, Pablo, Liam H. Parker, ChangHoon Hahn, et al. (July 2023). “SimBIG: Field-level Simulation-based Inference of Large-scale Structure”. In: *Machine Learning for Astrophysics*, 18, p. 18. DOI: `10.48550/arXiv.2310.15256`. arXiv: `2310.15256 [astro-ph.CO]`.

Leonova, E., P. A. Oesch, Y. Qin, et al. (Oct. 2022a). “The prevalence of galaxy overdensities around UV-luminous Lyman α emitters in the Epoch of Reionization”. In: *MNRAS* 515.4, pp. 5790–5801. DOI: `10.1093/mnras/stac1908`. arXiv: `2112.07675 [astro-ph.GA]`.

— (Oct. 2022b). “The prevalence of galaxy overdensities around UV-luminous Lyman α emitters in the Epoch of Reionization”. In: *MNRAS* 515.4, pp. 5790–5801. DOI: `10.1093/mnras/stac1908`. arXiv: `2112.07675 [astro-ph.GA]`.

Lewis, Joseph S. W., Pierre Ocvirk, Dominique Aubert, et al. (Aug. 2020). “Galactic ionizing photon budget during the epoch of reionization in the Cosmic Dawn II simulation”. In: *MNRAS* 496.4, pp. 4342–4357. DOI: `10.1093/mnras/staa1748`. arXiv: `2001.07785 [astro-ph.GA]`.

Lewis, Joseph S. W., Pierre Ocvirk, Jenny G. Sorce, et al. (Nov. 2022). “The short ionizing photon mean free path at $z = 6$ in Cosmic Dawn III, a new fully coupled radiation-hydrodynamical simulation of the Epoch of Reionization”. In: *MNRAS* 516.3, pp. 3389–3397. DOI: `10.1093/mnras/stac2383`. arXiv: `2202.05869 [astro-ph.CO]`.

Lin, Xiaojing, Zheng Cai, Yunjing Wu, et al. (June 2024). “Quantifying the Escape of Lyα at $z \approx 5$–6: A Census of Lyα Escape Fraction with Hα-emitting Galaxies Spectroscopically Confirmed by JWST and VLT/MUSE”. In: *ApJS* 272.2, 33, p. 33. DOI: `10.3847/1538-4365/ad3e7d`. arXiv: `2401.09532 [astro-ph.GA]`.

Lin, Yin, S. Peng Oh, Steven R. Furlanetto, et al. (Sept. 2016). “The distribution of bubble sizes during reionization”. In: *MNRAS* 461.3, pp. 3361–3374. DOI: `10.1093/mnras/stw1542`. arXiv: `1511.01506 [astro-ph.CO]`.

Liu, Bin and Rongmon Bordoloi (Apr. 2021). “A deep learning approach to quasar continuum prediction”. In: *MNRAS* 502.3, pp. 3510–3532. DOI: 10.1093/mnras/stab177. arXiv: 2006.04814 [astro-ph.GA].

Livermore, R. C., S. L. Finkelstein, and J. M. Lotz (Feb. 2017). “Directly Observing the Galaxies Likely Responsible for Reionization”. In: *ApJ* 835.2, 113, p. 113. DOI: 10.3847/1538-4357/835/2/113. arXiv: 1604.06799 [astro-ph.GA].

Llerena, M., L. Pentericci, L. Napolitano, et al. (Dec. 2024). “The ionizing photon production efficiency of star-forming galaxies at $z \sim 4-10$”. In: *arXiv e-prints*, arXiv:2412.01358, arXiv:2412.01358. DOI: 10.48550/arXiv.2412.01358. arXiv: 2412.01358 [astro-ph.GA].

Lovell, Christopher C., Aswin P. Vijayan, Peter A. Thomas, et al. (Jan. 2021). “First Light And Reionization Epoch Simulations (FLARES) - I. Environmental dependence of high-redshift galaxy evolution”. In: *MNRAS* 500.2, pp. 2127–2145. DOI: 10.1093/mnras/staa3360. arXiv: 2004.07283 [astro-ph.GA].

Lu, Ting-Yi, Charlotte A. Mason, Anne Hutter, et al. (Mar. 2024). “The reionizing bubble size distribution around galaxies”. In: *MNRAS* 528.3, pp. 4872–4890. DOI: 10.1093/mnras/stae266. arXiv: 2304.11192 [astro-ph.GA].

Lu, Ting-Yi, Charlotte A. Mason, Andrei Mesinger, et al. (May 2025). “Mapping reionization bubbles in JWST era: I. Empirical edge detection with Lyman alpha emission from galaxies”. In: *A&A* 697, A69, A69. DOI: 10.1051/0004-6361/202452912. arXiv: 2411.04176 [astro-ph.GA].

Lueker, M., C. L. Reichardt, K. K. Schaffer, et al. (Aug. 2010). “Measurements of Secondary Cosmic Microwave Background Anisotropies with the South Pole Telescope”. In: *ApJ* 719.2, pp. 1045–1066. DOI: 10.1088/0004-637X/719/2/1045. arXiv: 0912.4317 [astro-ph.CO].

Ma, Chung-Pei and J. N. Fry (May 2002). “Nonlinear Kinetic Sunyaev-Zeldovich Effect”. In: *Phys. Rev. Lett.* 88.21, 211301, p. 211301. DOI: 10.1103/PhysRevLett.88.211301. arXiv: astro-ph/0106342 [astro-ph].

Ma, Xiangcheng, Philip F. Hopkins, Shea Garrison-Kimmel, et al. (Aug. 2018). “Simulating galaxies in the reionization era with FIRE-2: galaxy scaling relations, stellar mass functions, and luminosity functions”. In: *MNRAS* 478.2, pp. 1694–1715. DOI: 10.1093/mnras/sty1024. arXiv: 1706.06605 [astro-ph.GA].

Ma, Xiangcheng, Daniel Kasen, Philip F. Hopkins, et al. (Oct. 2015). “The difficulty of getting high escape fractions of ionizing photons from high-redshift galaxies: a view from the FIRE cosmological simulations”. In: *MNRAS* 453.1, pp. 960–975. DOI: 10.1093/mnras/stv1679. arXiv: 1503.07880 [astro-ph.GA].

Ma, Xiangcheng, Eliot Quataert, Andrew Wetzel, et al. (Oct. 2020). “No missing photons for reionization: moderate ionizing photon escape fractions from the FIRE-2 simulations”. In: *MNRAS* 498.2, pp. 2001–2017. DOI: 10.1093/mnras/staa2404. arXiv: 2003.05945 [astro-ph.GA].

Madau, Piero and Mark Dickinson (Aug. 2014). “Cosmic Star-Formation History”. In: *ARA&A* 52, pp. 415–486. DOI: 10.1146/annurev-astro-081811-125615.

Madau, Piero and Tassos Fragos (May 2017). “Radiation Backgrounds at Cosmic Dawn: X-Rays from Compact Binaries”. In: *ApJ* 840.1, 39, p. 39. DOI: 10.3847/1538-4357/aa6af9. arXiv: 1606.07887 [astro-ph.GA].

Madau, Piero, Francesco Haardt, and Martin J. Rees (Apr. 1999). “Radiative Transfer in a Clumpy Universe. III. The Nature of Cosmological Ionizing Sources”. In: *ApJ* 514.2, pp. 648–659. DOI: 10.1086/306975. arXiv: astro-ph/9809058 [astro-ph].

Magg, Mattis, Itamar Reis, Anastasia Fialkov, et al. (Aug. 2022). “Effect of the cosmological transition to metal-enriched star formation on the hydrogen 21-cm signal”. In: *MNRAS* 514.3, pp. 4433–4449. DOI: 10.1093/mnras/stac1664. arXiv: 2110.15948 [astro-ph.CO].

Mannucci, F., G. Cresci, R. Maiolino, et al. (Nov. 2010). “A fundamental relation between mass, star formation rate and metallicity in local and high-redshift galaxies”. In: *MNRAS* 408.4, pp. 2115–2127. DOI: 10.1111/j.1365-2966.2010.17291.x. arXiv: 1005.0006 [astro-ph.CO].

Mascia, S., L. Pentericci, A. Calabrò, et al. (Apr. 2023). “Closing in on the sources of cosmic reionization: First results from the GLASS-JWST program”. In: *A&A* 672, A155, A155. DOI: 10.1051/0004-6361/202345866. arXiv: 2301.02816 [astro-ph.GA].

Mascia, S., L. Pentericci, M. Llerena, et al. (Jan. 2025). “Little impact of mergers and galaxy morphology on the production and escape of ionizing photons in the early Universe”. In: *arXiv e-prints*, arXiv:2501.08268, arXiv:2501.08268. DOI: 10.48550/arXiv.2501.08268. arXiv: 2501.08268 [astro-ph.GA].

Mason, Charlotte A., Zuyi Chen, Daniel P. Stark, et al. (Jan. 2025). “Constraints on the $z \sim 6-13$ intergalactic medium from JWST spectroscopy of Lyman-alpha damping wings in galaxies”. In: *arXiv e-prints*, arXiv:2501.11702, arXiv:2501.11702. DOI: 10.48550/arXiv.2501.11702. arXiv: 2501.11702 [astro-ph.GA].

Mason, Charlotte A., Adriano Fontana, Tommaso Treu, et al. (May 2019). “Inferences on the timeline of reionization at z $\sim$ 8 from the KMOS Lens-Amplified Spectroscopic Survey”. In: *MNRAS* 485.3, pp. 3947–3969. DOI: `10.1093/mnras/stz632`. arXiv: `1901.11045 [astro-ph.CO]`.

Mason, Charlotte A., Michele Trenti, and Tommaso Treu (May 2023). “The brightest galaxies at cosmic dawn”. In: *MNRAS* 521.1, pp. 497–503. DOI: `10.1093/mnras/stad035`. arXiv: `2207.14808 [astro-ph.GA]`.

Mason, Charlotte A., Tommaso Treu, Mark Dijkstra, et al. (Mar. 2018). “The Universe Is Reionizing at z $\sim$ 7: Bayesian Inference of the IGM Neutral Fraction Using Lyα Emission from Galaxies”. In: *ApJ* 856.1, 2, p. 2. DOI: `10.3847/1538-4357/aab0a7`. arXiv: `1709.05356 [astro-ph.CO]`.

McGreer, Ian D., Andrei Mesinger, and Valentina D’Odorico (Feb. 2015). “Model-independent evidence in favour of an end to reionization by z $\approx$ 6”. In: *MNRAS* 447.1, pp. 499–505. DOI: `10.1093/mnras/stu2449`. arXiv: `1411.5375 [astro-ph.CO]`.

McGreer, Ian D., Andrei Mesinger, and Xiaohui Fan (Aug. 2011). “The first (nearly) model-independent constraint on the neutral hydrogen fraction at z $\sim$ 6”. In: *MNRAS* 415.4, pp. 3237–3246. DOI: `10.1111/j.1365-2966.2011.18935.x`. arXiv: `1101.3314 [astro-ph.CO]`.

McKee, Christopher F. and Eve C. Ostriker (Sept. 2007). “Theory of Star Formation”. In: *ARA&A* 45.1, pp. 565–687. DOI: `10.1146/annurev.astro.45.051806.110602`. arXiv: `0707.3514 [astro-ph]`.

McLeod, D. J., C. T. Donnan, R. J. McLure, et al. (Jan. 2024). “The galaxy UV luminosity function at z $\simeq$ 11 from a suite of public JWST ERS, ERO, and Cycle-1 programs”. In: *MNRAS* 527.3, pp. 5004–5022. DOI: `10.1093/mnras/stad3471`. arXiv: `2304.14469 [astro-ph.GA]`.

McQuinn, Matthew (Sept. 2016). “The Evolution of the Intergalactic Medium”. In: *ARA&A* 54, pp. 313–362. DOI: `10.1146/annurev-astro-082214-122355`. arXiv: `1512.00086 [astro-ph.CO]`.

McQuinn, Matthew, Steven R. Furlanetto, Lars Hernquist, et al. (Sept. 2005). “The Kinetic Sunyaev-Zel’dovich Effect from Reionization”. In: *ApJ* 630.2, pp. 643–656. DOI: `10.1086/432049`. arXiv: `astro-ph/0504189 [astro-ph]`.

McQuinn, Matthew, Adam Lidz, Oliver Zahn, et al. (May 2007). “The morphology of HII regions during reionization”. In: *MNRAS* 377.3, pp. 1043–1063. DOI: `10.1111/j.1365-2966.2007.11489.x`. arXiv: `astro-ph/0610094 [astro-ph]`.

McQuinn, Matthew, S. Peng Oh, and Claude-André Faucher-Giguère (Dec. 2011). “On Lyman-limit Systems and the Evolution of the Intergalactic Ionizing Background”. In: *ApJ* 743.1, 82, p. 82. DOI: `10.1088/0004-637X/743/1/82`. arXiv: `1101.1964 [astro-ph.CO]`.

McQuinn, Matthew, Oliver Zahn, Matias Zaldarriaga, et al. (Dec. 2006). “Cosmological Parameter Estimation Using 21 cm Radiation from the Epoch of Reionization”. In: *ApJ* 653.2, pp. 815–834. DOI: `10.1086/505167`. arXiv: `astro-ph/0512263 [astro-ph]`.

Meriot, R. and B. Semelin (Mar. 2024). “The LORELI database: 21 cm signal inference with 3D radiative hydrodynamics simulations”. In: *A&A* 683, A24, A24. DOI: `10.1051/0004-6361/202347591`.

Mertens, F. G., M. Mevius, L. V. E. Koopmans, et al. (Apr. 2020). “Improved upper limits on the 21 cm signal power spectrum of neutral hydrogen at z $\approx$ 9.1 from LOFAR”. In: *MNRAS* 493.2, pp. 1662–1685. DOI: `10.1093/mnras/staa327`. arXiv: `2002.07196 [astro-ph.CO]`.

Mesinger, Andrei (Sept. 2010). “Was reionization complete by z ~5-6?” In: *MNRAS* 407.2, pp. 1328–1337. DOI: `10.1111/j.1365-2966.2010.16995.x`. arXiv: `0910.4161 [astro-ph.CO]`.

— ed. (Jan. 2016). *Understanding the Epoch of Cosmic Reionization*. Vol. 423. Astrophysics and Space Science Library. DOI: `10.1007/978-3-319-21957-8`.

— (2019). *The Cosmic 21-cm Revolution; Charting the first billion years of our universe*. DOI: `10.1088/2514-3433/ab4a73`.

Mesinger, Andrei, Aycin Aykutalp, Eros Vanzella, et al. (Jan. 2015). “Can the intergalactic medium cause a rapid drop in Lyα emission at z ¿ 6?” In: *MNRAS* 446.1, pp. 566–577. DOI: `10.1093/mnras/stu2089`. arXiv: `1406.6373 [astro-ph.CO]`.

Mesinger, Andrei, Andrea Ferrara, and David S. Spiegel (May 2013). “Signatures of X-rays in the early Universe”. In: *MNRAS* 431.1, pp. 621–637. DOI: `10.1093/mnras/stt198`. arXiv: `1210.7319 [astro-ph.CO]`.

Mesinger, Andrei and Steven Furlanetto (Nov. 2007a). “Efficient Simulations of Early Structure Formation and Reionization”. In: *ApJ* 669.2, pp. 663–675. DOI: `10.1086/521806`. arXiv: `0704.0946 [astro-ph]`.

— (Nov. 2007b). “Efficient Simulations of Early Structure Formation and Reionization”. In: *ApJ* 669.2, pp. 663–675. DOI: `10.1086/521806`. arXiv: `0704.0946 [astro-ph]`.

Mesinger, Andrei, Steven Furlanetto, and Renyue Cen (Feb. 2011a). "21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal". In: *MNRAS* 411.2, pp. 955–972. DOI: 10.1111/j.1365-2966.2010.17731.x. arXiv: 1003.3878 [astro-ph.CO].

— (Feb. 2011b). "21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal". In: *MNRAS* 411.2, pp. 955–972. DOI: 10.1111/j.1365-2966.2010.17731.x. arXiv: 1003.3878 [astro-ph.CO].

Mesinger, Andrei and Steven R. Furlanetto (Apr. 2008a). "Lyα damping wing constraints on inhomogeneous reionization". In: *MNRAS* 385.3, pp. 1348–1358. DOI: 10.1111/j.1365-2966.2007.12836.x. arXiv: 0710.0371 [astro-ph].

— (June 2008b). "Lyα emitters during the early stages of reionization". In: *MNRAS* 386.4, pp. 1990–2002. DOI: 10.1111/j.1365-2966.2008.13039.x. arXiv: 0708.0006 [astro-ph].

Mesinger, Andrei, Bradley Greig, and Emanuele Sobacchi (July 2016). "The Evolution Of 21 cm Structure (EOS): public, large-scale simulations of Cosmic Dawn and reionization". In: *MNRAS* 459.3, pp. 2342–2353. DOI: 10.1093/mnras/stw831. arXiv: 1602.07711 [astro-ph.CO].

Mesinger, Andrei and Zoltán Haiman (Aug. 2004). "Evidence of a Cosmological Strömgren Surface and of Significant Neutral Hydrogen Surrounding the Quasar SDSS J1030+0524". In: *ApJ* 611.2, pp. L69–L72. DOI: 10.1086/423935. arXiv: astro-ph/0406188 [astro-ph].

Mesinger, Andrei, Matthew McQuinn, and David N. Spergel (May 2012). "The kinetic Sunyaev-Zel'dovich signal from inhomogeneous reionization: a parameter space study". In: *MNRAS* 422.2, pp. 1403–1417. DOI: 10.1111/j.1365-2966.2012.20713.x. arXiv: 1112.1820 [astro-ph.CO].

Meyer, R. A., P. A. Oesch, E. Giovinazzo, et al. (Nov. 2024). "JWST FRESCO: a comprehensive census of H β + [O III] emitters at 6.8 ¡ z ¡ 9.0 in the GOODS fields". In: *MNRAS* 535.1, pp. 1067–1094. DOI: 10.1093/mnras/stae2353. arXiv: 2405.05111 [astro-ph.GA].

Meyer, Romain A., Nicolas Laporte, Richard S. Ellis, et al. (Jan. 2021). "Double-peaked Lyman α emission at z = 6.803: a reionization-era galaxy self-ionizing its local H II bubble". In: *MNRAS* 500.1, pp. 558–564. DOI: 10.1093/mnras/staa3216. arXiv: 2010.06241 [astro-ph.GA].

Milosavljević, Miloš and Chalence Safranek-Shrader (Jan. 2016). "Star Formation for Predictive Primordial Galaxy Formation". In: *Understanding the Epoch of Cosmic Reionization: Challenges and Progress*. Ed. by Andrei Mesinger. Vol. 423. Astrophysics and Space Science Library, p. 65. DOI: 10.1007/978-3-319-21957-8_3. arXiv: 1511.01110 [astro-ph.GA].

Mineo, S., M. Gilfanov, and R. Sunyaev (Jan. 2012). "X-ray emission from star-forming galaxies - I. High-mass X-ray binaries". In: *MNRAS* 419.3, pp. 2095–2115. DOI: 10.1111/j.1365-2966.2011.19862.x. arXiv: 1105.4610 [astro-ph.HE].

Miralda-Escudé, J. and M. J. Rees (Jan. 1994). "Reionization and thermal evolution of a photoionized intergalactic medium." In: *MNRAS* 266, pp. 343–352. DOI: 10.1093/mnras/266.2.343.

Miralda-Escudé, Jordi (July 1998). "Reionization of the Intergalactic Medium and the Damping Wing of the Gunn-Peterson Trough". In: *ApJ* 501.1, pp. 15–22. DOI: 10.1086/305799. arXiv: astro-ph/9708253 [astro-ph].

Mirocha, Jordan and Steven R. Furlanetto (Feb. 2023a). "Balancing the efficiency and stochasticity of star formation with dust extinction in z $\gtrsim$ 10 galaxies observed by JWST". In: *MNRAS* 519.1, pp. 843–853. DOI: 10.1093/mnras/stac3578. arXiv: 2208.12826 [astro-ph.GA].

— (Feb. 2023b). "Balancing the efficiency and stochasticity of star formation with dust extinction in z $\gtrsim$ 10 galaxies observed by JWST". In: *MNRAS* 519.1, pp. 843–853. DOI: 10.1093/mnras/stac3578. arXiv: 2208.12826 [astro-ph.GA].

Mirocha, Jordan, Steven R. Furlanetto, and Guochao Sun (Jan. 2017). "The global 21-cm signal in the context of the high- z galaxy luminosity function". In: *MNRAS* 464.2, pp. 1365–1379. DOI: 10.1093/mnras/stw2412. arXiv: 1607.00386 [astro-ph.GA].

Mirocha, Jordan, Paul La Plante, and Adrian Liu (Nov. 2021). "The importance of galaxy formation histories in models of reionization". In: *MNRAS* 507.3, pp. 3872–3887. DOI: 10.1093/mnras/stab1871. arXiv: 2012.09189 [astro-ph.CO].

Mo, H. J. and S. D. M. White (Sept. 1996). "An analytic model for the spatial clustering of dark matter haloes". In: *MNRAS* 282.2, pp. 347–361. DOI: 10.1093/mnras/282.2.347. arXiv: astro-ph/9512127 [astro-ph].

Morishita, Takahiro, Charlotte A. Mason, Kimi C. Kreilgaard, et al. (Dec. 2024). "BEACON: JWST NIRCam Pure-parallel Imaging Survey. I. Survey Design and Initial Results". In: *arXiv e-prints*, arXiv:2412.04211, arXiv:2412.04211. DOI: 10.48550/arXiv.2412.04211. arXiv: 2412.04211 [astro-ph.GA].

Muñoz, Julian B., Jordan Mirocha, Steven Furlanetto, et al. (Nov. 2023). “Breaking degeneracies in the first galaxies with clustering”. In: *MNRAS* 526.1, pp. L47–L55. DOI: `10.1093/mnrasl/slad115`. arXiv: `2306.09403 [astro-ph.CO]`.

Muñoz, Julian B., Yuxiang Qin, Andrei Mesinger, et al. (Apr. 2022). “The impact of the first galaxies on cosmic dawn and reionization”. In: *MNRAS* 511.3, pp. 3657–3681. DOI: `10.1093/mnras/stac185`. arXiv: `2110.13919 [astro-ph.CO]`.

Murmu, Chandra Shekhar, Kanan K. Datta, Suman Majumdar, et al. (Aug. 2024). “Impact of astrophysical scatter on the epoch of reionization [H I]$_{21}$ bispectrum”. In: *JCAP* 2024.8, 032, p. 032. DOI: `10.1088/1475-7516/2024/08/032`. arXiv: `2311.17062 [astro-ph.CO]`.

Murray, S. G., B. Diemer, Z. Chen, et al. (July 2021). “THEHALOMOD: An online calculator for the halo model”. In: *Astronomy and Computing* 36, 100487, p. 100487. DOI: `10.1016/j.ascom.2021.100487`. arXiv: `2009.14066 [astro-ph.CO]`.

Murray, Steven, Bradley Greig, Andrei Mesinger, et al. (Oct. 2020a). “21cmFAST v3: A Python-integrated C code for generating 3D realizations of the cosmic 21cm signal.” In: *The Journal of Open Source Software* 5.54, 2582, p. 2582. DOI: `10.21105/joss.02582`. arXiv: `2010.15121 [astro-ph.IM]`.

— (Oct. 2020b). “21cmFAST v3: A Python-integrated C code for generating 3D realizations of the cosmic 21cm signal.” In: *The Journal of Open Source Software* 5.54, 2582, p. 2582. DOI: `10.21105/joss.02582`. arXiv: `2010.15121 [astro-ph.IM]`.

Mutch, Simon J., Paul M. Geil, Gregory B. Poole, et al. (Oct. 2016). “Dark-ages reionization and galaxy formation simulation - III. Modelling galaxy formation and the epoch of reionization”. In: *MNRAS* 462.1, pp. 250–276. DOI: `10.1093/mnras/stw1506`. arXiv: `1512.00562 [astro-ph.GA]`.

Naidu, Rohan P., Pascal A. Oesch, Pieter van Dokkum, et al. (Nov. 2022). “Two Remarkably Luminous Galaxy Candidates at z $\approx$ 10-12 Revealed by JWST”. In: *ApJ* 940.1, L14, p. L14. DOI: `10.3847/2041-8213/ac9b22`. arXiv: `2207.09434 [astro-ph.GA]`.

Nakajima, Kimihiko, Masami Ouchi, Yuki Isobe, et al. (Jan. 2023). “JWST Census for the Mass-Metallicity Star-Formation Relations at z=4-10 with the Self-Consistent Flux Calibration and the Proper Metallicity Calibrators”. In: *arXiv e-prints*, arXiv:2301.12825, arXiv:2301.12825. DOI: `10.48550/arXiv.2301.12825`. arXiv: `2301.12825 [astro-ph.GA]`.

Nakamura, Fumitaka and Masayuki Umemura (Feb. 2001). “On the Initial Mass Function of Population III Stars”. In: *ApJ* 548.1, pp. 19–32. DOI: `10.1086/318663`. arXiv: `astro-ph/0010464 [astro-ph]`.

Nakane, Minami, Masami Ouchi, Kimihiko Nakajima, et al. (May 2024). “Lyα Emission at z = 7–13: Clear Evolution of Lyα Equivalent Width Indicating a Late Cosmic Reionization History”. In: *ApJ* 967.1, 28, p. 28. DOI: `10.3847/1538-4357/ad38c2`. arXiv: `2312.06804 [astro-ph.GA]`.

Napolitano, L., M. Castellano, L. Pentericci, et al. (Jan. 2025). “Seven wonders of Cosmic Dawn: JWST confirms a high abundance of galaxies and AGN at z $\simeq$ 9–11 in the GLASS field”. In: *A&A* 693, A50, A50. DOI: `10.1051/0004-6361/202452090`. arXiv: `2410.10967 [astro-ph.GA]`.

Napolitano, L., L. Pentericci, P. Santini, et al. (Aug. 2024). “Peering into cosmic reionization: Lyα visibility evolution from galaxies at z = 4.5-8.5 with JWST”. In: *A&A* 688, A106, A106. DOI: `10.1051/0004-6361/202449644`. arXiv: `2402.11220 [astro-ph.GA]`.

Neistein, Eyal, Frank C. van den Bosch, and Avishai Dekel (Oct. 2006). “Natural downsizing in hierarchical galaxy formation”. In: *MNRAS* 372.2, pp. 933–948. DOI: `10.1111/j.1365-2966.2006.10918.x`. arXiv: `astro-ph/0605045 [astro-ph]`.

Neufeld, David A. (Feb. 1990). “The Transfer of Resonance-Line Radiation in Static Astrophysical Media”. In: *ApJ* 350, p. 216. DOI: `10.1086/168375`.

Neyer, Meredith, Aaron Smith, Rahul Kannan, et al. (July 2024). “The THESAN project: connecting ionized bubble sizes to their local environments during the Epoch of Reionization”. In: *MNRAS* 531.3, pp. 2943–2957. DOI: `10.1093/mnras/stae1325`. arXiv: `2310.03783 [astro-ph.GA]`.

Nikolić, Ivan, Andrei Mesinger, James E. Davies, et al. (Dec. 2024). “The importance of stochasticity in determining galaxy emissivities and UV LFs during cosmic dawn and reionization”. In: *A&A* 692, A142, A142. DOI: `10.1051/0004-6361/202451213`. arXiv: `2406.15237 [astro-ph.CO]`.

Nikolić, Ivan, Andrei Mesinger, Charlotte A. Mason, et al. (Jan. 2025). “Mapping reionization bubbles in the JWST era II: inferring the position and characteristic size of individual bubbles”. In: *arXiv e-prints*, arXiv:2501.07980, arXiv:2501.07980. DOI: `10.48550/arXiv.2501.07980`. arXiv: `2501.07980 [astro-ph.GA]`.

Nikolić, Ivan, Andrei Mesinger, Yuxiang Qin, et al. (Dec. 2023). "Inferring reionization and galaxy properties from the patchy kinetic Sunyaev-Zel'dovich signal". In: *MNRAS* 526.2, pp. 3170–3183. DOI: `10.1093/mnras/stad2961`. arXiv: `2307.01265 [astro-ph.CO]`.

Nunhokee, C. D., D. Null, C. M Trott, et al. (May 2025). "Limits on the 21 cm power spectrum at z=6.5-7.0 from MWA observations". In: *arXiv e-prints*, arXiv:2505.09097, arXiv:2505.09097. DOI: `10.48550/arXiv.2505.09097`. arXiv: `2505.09097 [astro-ph.CO]`.

Oesch, P. A., R. J. Bouwens, G. D. Illingworth, et al. (Mar. 2018a). "The Dearth of z $\sim$ 10 Galaxies in All HST Legacy Fields—The Rapid Evolution of the Galaxy Population in the First 500 Myr". In: *ApJ* 855.2, 105, p. 105. DOI: `10.3847/1538-4357/aab03f`. arXiv: `1710.11131 [astro-ph.GA]`.

— (Mar. 2018b). "The Dearth of z $\sim$ 10 Galaxies in All HST Legacy Fields—The Rapid Evolution of the Galaxy Population in the First 500 Myr". In: *ApJ* 855.2, 105, p. 105. DOI: `10.3847/1538-4357/aab03f`. arXiv: `1710.11131 [astro-ph.GA]`.

Oesch, P. A., G. Brammer, P. G. van Dokkum, et al. (Mar. 2016). "A Remarkably Luminous Galaxy at z=11.1 Measured with Hubble Space Telescope Grism Spectroscopy". In: *ApJ* 819.2, 129, p. 129. DOI: `10.3847/0004-637X/819/2/129`. arXiv: `1603.00461 [astro-ph.GA]`.

Oesch, P. A., P. G. van Dokkum, G. D. Illingworth, et al. (May 2015). "A Spectroscopic Redshift Measurement for a Luminous Lyman Break Galaxy at z = 7.730 Using Keck/MOSFIRE". In: *ApJ* 804.2, L30, p. L30. DOI: `10.1088/2041-8205/804/2/L30`. arXiv: `1502.05399 [astro-ph.GA]`.

Oh, S. Peng (June 2001). "Reionization by Hard Photons. I. X-Rays from the First Star Clusters". In: *ApJ* 553.2, pp. 499–512. DOI: `10.1086/320957`. arXiv: `astro-ph/0005262 [astro-ph]`.

Oke, J. B. and J. E. Gunn (Mar. 1983). "Secondary standard stars for absolute spectrophotometry." In: *ApJ* 266, pp. 713–717. DOI: `10.1086/160817`.

Orlitová, I., A. Verhamme, A. Henry, et al. (Aug. 2018). "Puzzling Lyman-alpha line profiles in green pea galaxies". In: *A&A* 616, A60, A60. DOI: `10.1051/0004-6361/201732478`. arXiv: `1806.01027 [astro-ph.GA]`.

Ota, Kazuaki, Masanori Iye, Nobunari Kashikawa, et al. (July 2017). "A New Constraint on Reionization from the Evolution of the Lyα Luminosity Function at z $\sim$ 6-7 Probed by a Deep Census of z = 7.0 Lyα Emitter Candidates to 0.3L *". In: *ApJ* 844.1, 85, p. 85. DOI: `10.3847/1538-4357/aa7a0a`. arXiv: `1703.02501 [astro-ph.GA]`.

Paardekooper, Jan-Pieter, Sadegh Khochfar, and Claudio Dalla Vecchia (Aug. 2015). "The First Billion Years project: the escape fraction of ionizing photons in the epoch of reionization". In: *MNRAS* 451.3, pp. 2544–2563. DOI: `10.1093/mnras/stv1114`. arXiv: `1501.01967 [astro-ph.CO]`.

Pacucci, Fabio, Pratika Dayal, Yuichi Harikane, et al. (July 2022). "Are the newly-discovered z 13 drop-out sources starburst galaxies or quasars?" In: *MNRAS* 514.1, pp. L6–L10. DOI: `10.1093/mnrasl/slac035`. arXiv: `2201.00823 [astro-ph.GA]`.

Pacucci, Fabio, Andrei Mesinger, Stefano Mineo, et al. (Sept. 2014). "The X-ray spectra of the first galaxies: 21 cm signatures". In: *MNRAS* 443.1, pp. 678–686. DOI: `10.1093/mnras/stu1240`. arXiv: `1403.6125 [astro-ph.CO]`.

Padmanabhan, T. (1993). *Structure Formation in the Universe.*

Pahl, Anthony J., Alice Shapley, Charles C. Steidel, et al. (May 2023). "The connection between the escape of ionizing radiation and galaxy properties at z $\sim$ 3 in the Keck Lyman continuum spectroscopic survey". In: *MNRAS* 521.3, pp. 3247–3259. DOI: `10.1093/mnras/stad774`. arXiv: `2210.16697 [astro-ph.GA]`.

Pallottini, A. and A. Ferrara (Sept. 2023). "Stochastic star formation in early galaxies: Implications for the James Webb Space Telescope". In: *A&A* 677, L4, p. L4. DOI: `10.1051/0004-6361/202347384`. arXiv: `2307.03219 [astro-ph.GA]`.

Pallottini, A., A. Ferrara, S. Gallerani, et al. (July 2022). "A survey of high-z galaxies: SERRA simulations". In: *MNRAS* 513.4, pp. 5621–5641. DOI: `10.1093/mnras/stac1281`. arXiv: `2201.02636 [astro-ph.GA]`.

Paquereau, Louise, Clotilde Laigle, Henry Joy McCracken, et al. (Jan. 2025). "Tracing the galaxy-halo connection with galaxy clustering in COSMOS-Web from z = 0.1 to z ~12". In: *arXiv e-prints*, arXiv:2501.11674, arXiv:2501.11674. DOI: `10.48550/arXiv.2501.11674`. arXiv: `2501.11674 [astro-ph.GA]`.

Park, Hyunbae, Paul R. Shapiro, Eiichiro Komatsu, et al. (June 2013). "The Kinetic Sunyaev-Zel'dovich Effect as a Probe of the Physics of Cosmic Reionization: The Effect of Self-regulated Reionization". In: *ApJ* 769.2, 93, p. 93. DOI: `10.1088/0004-637X/769/2/93`. arXiv: `1301.3607 [astro-ph.CO]`.

Park, Jaehong, Nicolas Gillet, Andrei Mesinger, et al. (Jan. 2020). "Properties of reionization-era galaxies from JWST luminosity functions and 21-cm interferometry". In: *MNRAS* 491.3, pp. 3891–3899. DOI: `10.1093/mnras/stz3278`. arXiv: `1909.01348 [astro-ph.CO]`.

Park, Jaehong, Bradley Greig, and Andrei Mesinger (Nov. 2022). "Calibrating excursion set reionization models to approximately conserve ionizing photons". In: *MNRAS* 517.1, pp. 192–200. DOI: 10.1093/mnras/stac2756. arXiv: 2112.05184 [astro-ph.CO].

Park, Jaehong, Andrei Mesinger, Bradley Greig, et al. (Mar. 2019). "Inferring the astrophysics of reionization and cosmic dawn from galaxy luminosity functions and the 21-cm signal". In: *MNRAS* 484.1, pp. 933–949. DOI: 10.1093/mnras/stz032. arXiv: 1809.08995 [astro-ph.GA].

Paul, Sourabh, Suvodip Mukherjee, and Tirthankar Roy Choudhury (Jan. 2021). "Inevitable imprints of patchy reionization on the cosmic microwave background anisotropy". In: *MNRAS* 500.1, pp. 232–246. DOI: 10.1093/mnras/staa3221. arXiv: 2005.05327 [astro-ph.CO].

Pentericci, L., E. Vanzella, A. Fontana, et al. (Oct. 2014). "New Observations of z ~7 Galaxies: Evidence for a Patchy Reionization". In: *ApJ* 793.2, 113, p. 113. DOI: 10.1088/0004-637X/793/2/113. arXiv: 1403.5466 [astro-ph.CO].

Pérez-González, Pablo G., Luca Costantin, Danial Langeroodi, et al. (July 2023). "Life beyond 30: Probing the -20 ¡ M$_{UV}$ ¡ -17 Luminosity Function at 8 ¡ z ¡ 13 with the NIRCam Parallel Field of the MIRI Deep Survey". In: *ApJ* 951.1, L1, p. L1. DOI: 10.3847/2041-8213/acd9d0. arXiv: 2302.02429 [astro-ph.GA].

Pérez-González, Pablo G., Göran Östlin, Luca Costantin, et al. (Mar. 2025). "The rise of the galactic empire: luminosity functions at $z \sim 17$ and $z \sim 25$ estimated with the MIDIS+NGDEEP ultra-deep JWST/NIRCam dataset". In: *arXiv e-prints*, arXiv:2503.15594, arXiv:2503.15594. DOI: 10.48550/arXiv.2503.15594. arXiv: 2503.15594 [astro-ph.GA].

Perlmutter, S., G. Aldering, G. Goldhaber, et al. (June 1999). "Measurements of Ω and Λ from 42 High-Redshift Supernovae". In: *ApJ* 517.2, pp. 565–586. DOI: 10.1086/307221. arXiv: astro-ph/9812133 [astro-ph].

Philcox, Oliver H. E. and Mikhail M. Ivanov (Feb. 2022). "BOSS DR12 full-shape cosmology: Λ CDM constraints from the large-scale galaxy power spectrum and bispectrum monopole". In: *Phys. Rev. D* 105.4, 043517, p. 043517. DOI: 10.1103/PhysRevD.105.043517. arXiv: 2112.04515 [astro-ph.CO].

Planck Collaboration, R. Adam, N. Aghanim, et al. (Dec. 2016). "Planck intermediate results. XLVII. Planck constraints on reionization history". In: *A&A* 596, A108, A108. DOI: 10.1051/0004-6361/201628897. arXiv: 1605.03507 [astro-ph.CO].

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. (Sept. 2016). "Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts". In: *A&A* 594, A24, A24. DOI: 10.1051/0004-6361/201525833. arXiv: 1502.01597 [astro-ph.CO].

Planck Collaboration, N. Aghanim, Y. Akrami, et al. (Sept. 2020). "Planck 2018 results. VI. Cosmological parameters". In: *A&A* 641, A6, A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

Planck Collaboration, N. Aghanim, M. Ashdown, et al. (Dec. 2016). "Planck intermediate results. XLVI. Reduction of large-scale systematic effects in HFI polarization maps and estimation of the reionization optical depth". In: *A&A* 596, A107, A107. DOI: 10.1051/0004-6361/201628890. arXiv: 1605.02985 [astro-ph.CO].

Popesso, P., A. Concas, G. Cresci, et al. (Feb. 2023). "The main sequence of star-forming galaxies across cosmic times". In: *MNRAS* 519.1, pp. 1526–1544. DOI: 10.1093/mnras/stac3214. arXiv: 2203.10487 [astro-ph.GA].

Prelogović, David and Andrei Mesinger (Sept. 2023). "Exploring the likelihood of the 21-cm power spectrum with simulation-based inference". In: *MNRAS* 524.3, pp. 4239–4255. DOI: 10.1093/mnras/stad2027. arXiv: 2305.03074 [astro-ph.CO].

Prelogović, David, Andrei Mesinger, Steven Murray, et al. (Jan. 2022). "Machine learning astrophysics from 21 cm lightcones: impact of network architectures and signal contamination". In: *MNRAS* 509.3, pp. 3852–3867. DOI: 10.1093/mnras/stab3215. arXiv: 2107.00018 [astro-ph.CO].

Press, William H. and Paul Schechter (Feb. 1974). "Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation". In: *ApJ* 187, pp. 425–438. DOI: 10.1086/152650.

Qin, Yuxiang, Sreedhar Balu, and J. Stuart B. Wyithe (Nov. 2023). "Implications of z $\gtrsim$ 12 JWST galaxies for galaxy formation at high redshift". In: *MNRAS* 526.1, pp. 1324–1342. DOI: 10.1093/mnras/stad2448. arXiv: 2305.17959 [astro-ph.GA].

Qin, Yuxiang, Andrei Mesinger, Sarah E. I. Bosman, et al. (Sept. 2021). "Reionization and galaxy inference from the high-redshift Ly α forest". In: *MNRAS* 506.2, pp. 2390–2407. DOI: 10.1093/mnras/stab1833. arXiv: 2101.09033 [astro-ph.CO].

Qin, Yuxiang, Andrei Mesinger, Bradley Greig, et al. (Mar. 2021). "A tale of two sites - II. Inferring the properties of minihalo-hosted galaxies with upcoming 21-cm interferometers". In: *MNRAS* 501.4, pp. 4748–4758. DOI: 10.1093/mnras/staa3408. arXiv: 2009.11493 [astro-ph.CO].

Qin, Yuxiang, Andrei Mesinger, David Prelogović, et al. (Dec. 2024). "Percent-level timing of reionization: self-consistent, implicit-likelihood inference from XQR-30+ Lyα forest data". In: *arXiv e-prints*, arXiv:2412.00799, arXiv:2412.00799. arXiv: 2412.00799 [astro-ph.CO].

Qin, Yuxiang, Simon J. Mutch, Gregory B. Poole, et al. (Dec. 2017). "Dark-ages reionization and galaxy formation simulation - X. The small contribution of quasars to reionization". In: *MNRAS* 472.2, pp. 2009–2027. DOI: 10.1093/mnras/stx1909. arXiv: 1703.04895 [astro-ph.CO].

Qin, Yuxiang, Vivian Poulin, Andrei Mesinger, et al. (Nov. 2020). "Reionization inference from the CMB optical depth and E-mode polarization power spectra". In: *MNRAS* 499.1, pp. 550–558. DOI: 10.1093/mnras/staa2797. arXiv: 2006.16828 [astro-ph.CO].

Reed, Darren S., Richard Bower, Carlos S. Frenk, et al. (Jan. 2007). "The halo mass function from the dark ages through the present day". In: *MNRAS* 374.1, pp. 2–15. DOI: 10.1111/j.1365-2966.2006.11204.x. arXiv: astro-ph/0607150 [astro-ph].

Reichardt, C. L., S. Patil, P. A. R. Ade, et al. (Feb. 2021). "An Improved Measurement of the Secondary Cosmic Microwave Background Anisotropies from the SPT-SZ + SPTpol Surveys". In: *ApJ* 908.2, 199, p. 199. DOI: 10.3847/1538-4357/abd407. arXiv: 2002.06197 [astro-ph.CO].

Reichardt, C. L., L. Shaw, O. Zahn, et al. (Aug. 2012). "A Measurement of Secondary Cosmic Microwave Background Anisotropies with Two Years of South Pole Telescope Observations". In: *ApJ* 755.1, 70, p. 70. DOI: 10.1088/0004-637X/755/1/70. arXiv: 1111.0932 [astro-ph.CO].

Reichardt, Christian L. (Jan. 2016). "Observing the Epoch of Reionization with the Cosmic Microwave Background". In: *Understanding the Epoch of Cosmic Reionization: Challenges and Progress*. Ed. by Andrei Mesinger. Vol. 423. Astrophysics and Space Science Library, p. 227. DOI: 10.1007/978-3-319-21957-8_8. arXiv: 1511.01117 [astro-ph.CO].

Reis, Itamar, Rennan Barkana, and Anastasia Fialkov (Apr. 2022). "Shot noise and scatter in the star formation efficiency as a source of 21-cm fluctuations". In: *MNRAS* 511.4, pp. 5265–5273. DOI: 10.1093/mnras/stac411. arXiv: 2106.13111 [astro-ph.CO].

Riess, Adam G., Alexei V. Filippenko, Peter Challis, et al. (Sept. 1998). "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: *AJ* 116.3, pp. 1009–1038. DOI: 10.1086/300499. arXiv: astro-ph/9805201 [astro-ph].

Roberts-Borsani, G. W., R. J. Bouwens, P. A. Oesch, et al. (June 2016). "$z \gtrsim 7$ Galaxies with Red Spitzer/IRAC [3.6]-[4.5] Colors in the Full CANDELS Data Set: The Brightest-Known Galaxies at z ~7-9 and a Probable Spectroscopic Confirmation at $z = 7.48$". In: *ApJ* 823.2, 143, p. 143. DOI: 10.3847/0004-637X/823/2/143. arXiv: 1506.00854 [astro-ph.GA].

Roberts-Borsani, Guido, Takahiro Morishita, Tommaso Treu, et al. (Oct. 2022). "Early Results from GLASS-JWST. I: Confirmation of Lensed $z \geq 7$ Lyman-break Galaxies behind the Abell 2744 Cluster with NIRISS". In: *ApJ* 938.2, L13, p. L13. DOI: 10.3847/2041-8213/ac8e6e. arXiv: 2207.11387 [astro-ph.GA].

Roberts-Borsani, Guido, Tommaso Treu, Alice Shapley, et al. (Dec. 2024). "Between the Extremes: A JWST Spectroscopic Benchmark for High-redshift Galaxies Using ~500 Confirmed Sources at $z \geq 5$". In: *ApJ* 976.2, 193, p. 193. DOI: 10.3847/1538-4357/ad85d3. arXiv: 2403.07103 [astro-ph.GA].

Robertson, Brant, Benjamin D. Johnson, Sandro Tacchella, et al. (July 2024). "Earliest Galaxies in the JADES Origins Field: Luminosity Function and Cosmic Star Formation Rate Density 300 Myr after the Big Bang". In: *ApJ* 970.1, 31, p. 31. DOI: 10.3847/1538-4357/ad463d. arXiv: 2312.10033 [astro-ph.GA].

Ross, Hannah E., Keri L. Dixon, Ilian T. Iliev, et al. (July 2017). "Simulating the impact of X-ray heating during the cosmic dawn". In: *MNRAS* 468.4, pp. 3785–3797. DOI: 10.1093/mnras/stx649. arXiv: 1607.06282 [astro-ph.CO].

Roy, Namrata, Timothy Heckman, Alaina Henry, et al. (Oct. 2024). "Lyman Continuum leakage from massive leaky starbursts: A different class of emitters?" In: *arXiv e-prints*, arXiv:2410.13254, arXiv:2410.13254. DOI: 10.48550/arXiv.2410.13254. arXiv: 2410.13254 [astro-ph.GA].

Rubin, V. C., W. K. Ford Jr., and N. Thonnard (June 1980). "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc)." In: *ApJ* 238, pp. 471–487. DOI: 10.1086/158003.

Salpeter, Edwin E. (Jan. 1955). "The Luminosity Function and Stellar Evolution." In: *ApJ* 121, p. 161. DOI: 10.1086/145971.

Santini, Paola, Adriano Fontana, Marco Castellano, et al. (Sept. 2017). “The Star Formation Main Sequence in the Hubble Space Telescope Frontier Fields”. In: *ApJ* 847.1, 76, p. 76. DOI: `10.3847/1538-4357/aa8874`. arXiv: `1706.07059 [astro-ph.GA]`.

Saxena, Aayush, Brant E. Robertson, Andrew J. Bunker, et al. (Oct. 2023). “JADES: Discovery of extremely high equivalent width Lyman-α emission from a faint galaxy within an ionized bubble at z = 7.3”. In: *A&A* 678, A68, A68. DOI: `10.1051/0004-6361/202346245`. arXiv: `2302.12805 [astro-ph.GA]`.

Schaeffer, Timothée, Sambit K. Giri, and Aurel Schneider (Dec. 2023). “BEORN: a fast and flexible framework to simulate the epoch of reionization and cosmic dawn”. In: *MNRAS* 526.2, pp. 2942–2959. DOI: `10.1093/mnras/stad2937`. arXiv: `2305.15466 [astro-ph.CO]`.

Schenker, Matthew A., Richard S. Ellis, Nick P. Konidaris, et al. (Nov. 2014). “Line-emitting Galaxies beyond a Redshift of 7: An Improved Method for Estimating the Evolving Neutrality of the Intergalactic Medium”. In: *ApJ* 795.1, 20, p. 20. DOI: `10.1088/0004-637X/795/1/20`. arXiv: `1404.4632 [astro-ph.CO]`.

Scoccimarro, Roman (Oct. 1998). “Transients from initial conditions: a perturbative analysis”. In: *MNRAS* 299.4, pp. 1097–1118. DOI: `10.1046/j.1365-8711.1998.01845.x`. arXiv: `astro-ph/9711187 [astro-ph]`.

Sehgal, Neelima, Simone Aiola, Yashar Akrami, et al. (Sept. 2019). “CMB-HD: An Ultra-Deep, High-Resolution Millimeter-Wave Survey Over Half the Sky”. In: *Bulletin of the American Astronomical Society*. Vol. 51, 6, p. 6. arXiv: `1906.10134 [astro-ph.CO]`.

Seiler, Jacob, Anne Hutter, Manodeep Sinha, et al. (Aug. 2019). “The escape fraction of ionizing photons during the Epoch of Reionization: observability with the Square Kilometre Array”. In: *MNRAS* 487.4, pp. 5739–5752. DOI: `10.1093/mnras/stz1663`. arXiv: `1902.01611 [astro-ph.GA]`.

Seljak, Uros and Matias Zaldarriaga (Oct. 1996). “A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies”. In: *ApJ* 469, p. 437. DOI: `10.1086/177793`. arXiv: `astro-ph/9603033 [astro-ph]`.

Semelin, B., R. Mériot, A. Mishra, et al. (June 2025). “Combining summary statistics with simulation-based inference for the 21 cm signal from the Epoch of Reionisation”. In: *A&A* 698, A35, A35. DOI: `10.1051/0004-6361/202453115`. arXiv: `2411.14419 [astro-ph.CO]`.

Shaw, Laurie D., Douglas H. Rudd, and Daisuke Nagai (Sept. 2012). “Deconstructing the Kinetic SZ Power Spectrum”. In: *ApJ* 756.1, 15, p. 15. DOI: `10.1088/0004-637X/756/1/15`. arXiv: `1109.0553 [astro-ph.CO]`.

Shen, Xuejian, Mark Vogelsberger, Michael Boylan-Kolchin, et al. (Nov. 2023). “The impact of UV variability on the abundance of bright galaxies at $z \geq 9$”. In: *MNRAS* 525.3, pp. 3254–3261. DOI: `10.1093/mnras/stad2508`. arXiv: `2305.05679 [astro-ph.GA]`.

Sheth, Ravi K. (Nov. 1998). “An excursion set model for the distribution of dark matter and dark matter haloes”. In: *MNRAS* 300.4, pp. 1057–1070. DOI: `10.1046/j.1365-8711.1998.01976.x`. arXiv: `astro-ph/9805319 [astro-ph]`.

Sheth, Ravi K., H. J. Mo, and Giuseppe Tormen (May 2001). “Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes”. In: *MNRAS* 323.1, pp. 1–12. DOI: `10.1046/j.1365-8711.2001.04006.x`. arXiv: `astro-ph/9907024 [astro-ph]`.

Sheth, Ravi K. and Giuseppe Tormen (Sept. 1999). “Large-scale bias and the peak background split”. In: *MNRAS* 308.1, pp. 119–126. DOI: `10.1046/j.1365-8711.1999.02692.x`. arXiv: `astro-ph/9901122 [astro-ph]`.

Shirokoff, E., C. L. Reichardt, L. Shaw, et al. (July 2011). “Improved Constraints on Cosmic Microwave Background Secondary Anisotropies from the Complete 2008 South Pole Telescope Data”. In: *ApJ* 736.1, 61, p. 61. DOI: `10.1088/0004-637X/736/1/61`. arXiv: `1012.4788 [astro-ph.CO]`.

Shuntov, Marko, Pascal A. Oesch, Sune Toft, et al. (Mar. 2025). “Constraints on the early Universe star formation efficiency from galaxy clustering and halo modeling of Hα and [O III] emitters”. In: *arXiv e-prints*, arXiv:2503.14280, arXiv:2503.14280. DOI: `10.48550/arXiv.2503.14280`. arXiv: `2503.14280 [astro-ph.GA]`.

Singh, Saurabh, Nambissan T. Jishnu, Ravi Subrahmanyan, et al. (Feb. 2022). “On the detection of a cosmic dawn signal in the radio background”. In: *Nature Astronomy* 6, pp. 607–617. DOI: `10.1038/s41550-022-01610-5`. arXiv: `2112.06778 [astro-ph.CO]`.

Sobacchi, E. and A. Mesinger (May 2013). “The depletion of gas in high-redshift dwarf galaxies from an inhomogeneous reionization.” In: *MNRAS* 432, pp. L51–L55. DOI: `10.1093/mnrasl/slt035`. arXiv: `1301.6776 [astro-ph.CO]`.

Sobacchi, Emanuele and Andrei Mesinger (May 2014). “Inhomogeneous recombinations during cosmic reionization”. In: *MNRAS* 440.2, pp. 1662–1673. DOI: `10.1093/mnras/stu377`. arXiv: `1402.2298 [astro-ph.CO]`.

— (Oct. 2015). “The clustering of Lyman α emitters at $z \approx 7$: implications for reionization and host halo masses”. In: *MNRAS* 453.2, pp. 1843–1854. DOI: `10.1093/mnras/stv1751`. arXiv: `1505.02787 [astro-ph.CO]`.

Somerville, Rachel S., Philip F. Hopkins, Thomas J. Cox, et al. (Dec. 2008). “A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei”. In: *MNRAS* 391.2, pp. 481–506. DOI: `10.1111/j.1365-2966.2008.13805.x`. arXiv: `0808.1227 [astro-ph]`.

Somerville, Rachel S. and Tsafrir S. Kolatt (May 1999). “How to plant a merger tree”. In: *MNRAS* 305.1, pp. 1–14. DOI: `10.1046/j.1365-8711.1999.02154.x`. arXiv: `astro-ph/9711080 [astro-ph]`.

Spergel, D. N., R. Bean, O. Doré, et al. (June 2007). “Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology”. In: *ApJS* 170.2, pp. 377–408. DOI: `10.1086/513700`. arXiv: `astro-ph/0603449 [astro-ph]`.

Spina, Benedetta, Sarah E. I. Bosman, Frederick B. Davies, et al. (Aug. 2024). “Damping wings in the Lyman-α forest: A model-independent measurement of the neutral fraction at 5.4 ¡ z ¡ 6.1”. In: *A&A* 688, L26, p. L26. DOI: `10.1051/0004-6361/202450798`. arXiv: `2405.12273 [astro-ph.CO]`.

Stanway, E. R., A. A. Chrimes, J. J. Eldridge, et al. (July 2020). “Evaluating the impact of binary parameter uncertainty on stellar population properties”. In: *MNRAS* 495.4, pp. 4605–4621. DOI: `10.1093/mnras/staa1166`. arXiv: `2004.11913 [astro-ph.GA]`.

Stanway, E. R. and J. J. Eldridge (Sept. 2018). “Re-evaluating old stellar populations”. In: *MNRAS* 479.1, pp. 75–93. DOI: `10.1093/mnras/sty1353`. arXiv: `1805.08784 [astro-ph.GA]`.

Stark, Daniel P., Richard S. Ellis, Stéphane Charlot, et al. (Jan. 2017a). “Lyα and C III] emission in z = 7-9 Galaxies: accelerated reionization around luminous star-forming systems?” In: *MNRAS* 464.1, pp. 469–479. DOI: `10.1093/mnras/stw2233`. arXiv: `1606.01304 [astro-ph.GA]`.

— (Jan. 2017b). “Lyα and C III] emission in z = 7-9 Galaxies: accelerated reionization around luminous star-forming systems?” In: *MNRAS* 464.1, pp. 469–479. DOI: `10.1093/mnras/stw2233`. arXiv: `1606.01304 [astro-ph.GA]`.

Stark, Daniel P., Richard S. Ellis, Kuenley Chiu, et al. (Nov. 2010). “Keck spectroscopy of faint 3 ¡ z ¡ 7 Lyman break galaxies - I. New constraints on cosmic reionization from the luminosity and redshift-dependent fraction of Lyman α emission”. In: *MNRAS* 408.3, pp. 1628–1648. DOI: `10.1111/j.1365-2966.2010.17227.x`. arXiv: `1003.5244 [astro-ph.CO]`.

Stark, Daniel P., Richard S. Ellis, and Masami Ouchi (Feb. 2011). “Keck Spectroscopy of Faint 3¿z¿7 Lyman Break Galaxies: A High Fraction of Line Emitters at Redshift Six”. In: *ApJ* 728.1, L2, p. L2. DOI: `10.1088/2041-8205/728/1/L2`. arXiv: `1009.5471 [astro-ph.CO]`.

Stefanon, Mauro, Rychard J. Bouwens, Ivo Labbé, et al. (Nov. 2021). “Galaxy Stellar Mass Functions from z 10 to z 6 using the Deepest Spitzer/Infrared Array Camera Data: No Significant Evolution in the Stellar-to-halo Mass Ratio of Galaxies in the First Gigayear of Cosmic Time”. In: *ApJ* 922.1, 29, p. 29. DOI: `10.3847/1538-4357/ac1bb6`. arXiv: `2103.16571 [astro-ph.GA]`.

Stefanon, Mauro, Ivo Labbé, Rychard J. Bouwens, et al. (Sept. 2019). “The Brightest z $\gtrsim$ 8 Galaxies over the COSMOS UltraVISTA Field”. In: *ApJ* 883.1, 99, p. 99. DOI: `10.3847/1538-4357/ab3792`. arXiv: `1902.10713 [astro-ph.GA]`.

Steidel, Charles C., Milan Bogosavljević, Alice E. Shapley, et al. (Dec. 2018). “The Keck Lyman Continuum Spectroscopic Survey (KLCS): The Emergent Ionizing Spectrum of Galaxies at z $\sim$ 3”. In: *ApJ* 869.2, 123, p. 123. DOI: `10.3847/1538-4357/aaed28`. arXiv: `1805.06071 [astro-ph.GA]`.

Steidel, Charles C., Gwen C. Rudie, Allison L. Strom, et al. (Nov. 2014). “Strong Nebular Line Ratios in the Spectra of z ~2-3 Star Forming Galaxies: First Results from KBSS-MOSFIRE”. In: *ApJ* 795.2, 165, p. 165. DOI: `10.1088/0004-637X/795/2/165`. arXiv: `1405.5473 [astro-ph.GA]`.

Strom, Allison L., Charles C. Steidel, Gwen C. Rudie, et al. (Dec. 2018). “Measuring the Physical Conditions in High-redshift Star-forming Galaxies: Insights from KBSS-MOSFIRE”. In: *ApJ* 868.2, 117, p. 117. DOI: `10.3847/1538-4357/aae1a5`. arXiv: `1711.08820 [astro-ph.GA]`.

Sun, G. and S. R. Furlanetto (July 2016). “Constraints on the star formation efficiency of galaxies during the epoch of reionization”. In: *MNRAS* 460.1, pp. 417–433. DOI: `10.1093/mnras/stw980`. arXiv: `1512.06219 [astro-ph.GA]`.

Sun, Guochao, Claude-André Faucher-Giguère, Christopher C. Hayward, et al. (Oct. 2023). “Bursty Star Formation Naturally Explains the Abundance of Bright Galaxies at Cosmic Dawn”. In: *ApJ* 955.2, L35, p. L35. DOI: `10.3847/2041-8213/acf85a`. arXiv: `2307.15305 [astro-ph.GA]`.

Sun, Zechang, Yuan-Sen Ting, and Zheng Cai (Nov. 2023). “Quasar Factor Analysis-An Unsupervised and Probabilistic Quasar Continuum Prediction Algorithm with Latent Factor Analysis”. In: *ApJS* 269.1, 4, p. 4. DOI: `10.3847/1538-4365/acf2f1`. arXiv: `2211.11784 [astro-ph.CO]`.

Sunyaev, R. A. and Ya. B. Zeldovich (Apr. 1970). “Small-Scale Fluctuations of Relic Radiation”. In: *Ap&SS* 7.1, pp. 3–19. DOI: 10.1007/BF00653471.

Tacchella, Sandro, Sownak Bose, Charlie Conroy, et al. (Dec. 2018). “A Redshift-independent Efficiency Model: Star Formation and Stellar Masses in Dark Matter Halos at z $\gtrsim$ 4”. In: *ApJ* 868.2, 92, p. 92. DOI: 10.3847/1538-4357/aae8e0. arXiv: 1806.03299 [astro-ph.GA].

Tacchella, Sandro, Avishai Dekel, C. Marcella Carollo, et al. (Apr. 2016). “The confinement of star-forming galaxies into a main sequence through episodes of gas compaction, depletion and replenishment”. In: *MNRAS* 457.3, pp. 2790–2813. DOI: 10.1093/mnras/stw131. arXiv: 1509.02529 [astro-ph.GA].

Tacchella, Sandro, Daniel J. Eisenstein, Kevin Hainline, et al. (July 2023). “JADES Imaging of GN-z11: Revealing the Morphology and Environment of a Luminous Galaxy 430 Myr after the Big Bang”. In: *ApJ* 952.1, 74, p. 74. DOI: 10.3847/1538-4357/acdbc6. arXiv: 2302.07234 [astro-ph.GA].

Tang, Mengtao, Daniel P. Stark, Zuyi Chen, et al. (Dec. 2023). “JWST/NIRSpec spectroscopy of z = 7-9 star-forming galaxies with CEERS: new insight into bright Lyα emitters in ionized bubbles”. In: *MNRAS* 526.2, pp. 1657–1686. DOI: 10.1093/mnras/stad2763. arXiv: 2301.07072 [astro-ph.GA].

Tang, Mengtao, Daniel P. Stark, Richard S. Ellis, et al. (June 2024). “Lyα emission in galaxies at z $\simeq$ 5-6: new insight from JWST into the statistical distributions of Lyα properties at the end of reionization”. In: *MNRAS* 531.2, pp. 2701–2730. DOI: 10.1093/mnras/stae1338. arXiv: 2402.06070 [astro-ph.GA].

Tang, Mengtao, Daniel P. Stark, Michael W. Topping, et al. (Aug. 2024). “JWST/NIRSpec Observations of Lyα Emission in Star Forming Galaxies at $6.5 \lesssim z \lesssim 13$”. In: *arXiv e-prints*, arXiv:2408.01507, arXiv:2408.01507. DOI: 10.48550/arXiv.2408.01507. arXiv: 2408.01507 [astro-ph.GA].

Tegmark, Max, Joseph Silk, Martin J. Rees, et al. (Jan. 1997). “How Small Were the First Cosmological Objects?” In: *ApJ* 474, p. 1. DOI: 10.1086/303434. arXiv: astro-ph/9603007 [astro-ph].

The HERA collaboration, Zara Abdurashidova, Tyrone Adams, et al. (Mar. 2023). “Improved Constraints on the 21 cm EoR Power Spectrum and the X-Ray Heating of the IGM with HERA Phase I Observations”. In: *ApJ* 945.2, 124, p. 124. DOI: 10.3847/1538-4357/acaf50.

The HERA collaboration Abdurashidova, Zara, James E. Aguirre, Paul Alexander, et al. (Jan. 2022). “HERA Phase I Limits on the Cosmic 21 cm Signal: Constraints on Astrophysics and Cosmology during the Epoch of Reionization”. In: *ApJ* 924.2, 51, p. 51. DOI: 10.3847/1538-4357/ac2ffc. arXiv: 2108.07282 [astro-ph.CO].

Tilvi, V., S. Malhotra, J. E. Rhoads, et al. (Mar. 2020). “Onset of Cosmic Reionization: Evidence of an Ionized Bubble Merely 680 Myr after the Big Bang”. In: *ApJ* 891.1, L10, p. L10. DOI: 10.3847/2041-8213/ab75ec. arXiv: 2001.00873 [astro-ph.GA].

Tilvi, V., C. Papovich, S. L. Finkelstein, et al. (Oct. 2014). “Rapid Decline of Lyα Emission toward the Reionization Era”. In: *ApJ* 794.1, 5, p. 5. DOI: 10.1088/0004-637X/794/1/5. arXiv: 1405.4869 [astro-ph.CO].

Tinker, Jeremy L., Brant E. Robertson, Andrey V. Kravtsov, et al. (Dec. 2010). “The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests”. In: *ApJ* 724.2, pp. 878–886. DOI: 10.1088/0004-637X/724/2/878. arXiv: 1001.3162 [astro-ph.CO].

Trapp, A. C. and Steven R. Furlanetto (Dec. 2020). “A flexible analytic model of cosmic variance in the first billion years”. In: *MNRAS* 499.2, pp. 2401–2415. DOI: 10.1093/mnras/staa2828. arXiv: 2009.05059 [astro-ph.GA].

Trebitsch, Maxime, Yohan Dubois, Marta Volonteri, et al. (Sept. 2021). “The OBELISK simulation: Galaxies contribute more than AGN to H I reionization of protoclusters”. In: *A&A* 653, A154, A154. DOI: 10.1051/0004-6361/202037698. arXiv: 2002.04045 [astro-ph.GA].

Treu, T., G. Roberts-Borsani, M. Bradac, et al. (Aug. 2022). “The GLASS-JWST Early Release Science Program. I. Survey Design and Release Plans”. In: *ApJ* 935.2, 110, p. 110. DOI: 10.3847/1538-4357/ac8158. arXiv: 2206.07978 [astro-ph.GA].

Treu, Tommaso, Michele Trenti, Massimo Stiavelli, et al. (Mar. 2012). “Inferences on the Distribution of Lyα Emission of z ~7 and z ~8 Galaxies”. In: *ApJ* 747.1, 27, p. 27. DOI: 10.1088/0004-637X/747/1/27. arXiv: 1201.0016 [astro-ph.CO].

Trott, Cathryn M., C. H. Jordan, S. Midgley, et al. (Apr. 2020). “Deep multiredshift limits on Epoch of Reionization 21 cm power spectra from four seasons of Murchison Widefield Array observations”. In: *MNRAS* 493.4, pp. 4711–4727. DOI: 10.1093/mnras/staa414. arXiv: 2002.02575 [astro-ph.CO].

Trotta, Roberto (Jan. 2017). “Bayesian Methods in Cosmology”. In: *arXiv e-prints*, arXiv:1701.01467, arXiv:1701.01467. arXiv: 1701.01467 [astro-ph.CO].

Umeda, Hiroya, Masami Ouchi, Kimihiko Nakajima, et al. (Aug. 2024). “JWST Measurements of Neutral Hydrogen Fractions and Ionized Bubble Sizes at z = 7–12 Obtained with Lyα Damping Wing Absorptions in 27 Bright Continuum Galaxies”. In: *ApJ* 971.2, 124, p. 124. DOI: `10.3847/1538-4357/ad554e`. arXiv: `2306.00487 [astro-ph.GA]`.

Verhamme, A., T. Garel, E. Ventou, et al. (July 2018). “Recovering the systemic redshift of galaxies from their Lyman alpha line profile”. In: *MNRAS* 478.1, pp. L60–L65. DOI: `10.1093/mnrasl/sly058`. arXiv: `1804.01883 [astro-ph.GA]`.

Viel, Matteo, George D. Becker, James S. Bolton, et al. (Aug. 2013). “Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data”. In: *Phys. Rev. D* 88.4, 043502, p. 043502. DOI: `10.1103/PhysRevD.88.043502`. arXiv: `1306.2314 [astro-ph.CO]`.

Volonteri, Marta, Piero Madau, Eliot Quataert, et al. (Feb. 2005). “The Distribution and Cosmic Evolution of Massive Black Hole Spins”. In: *ApJ* 620.1, pp. 69–77. DOI: `10.1086/426858`. arXiv: `astro-ph/0410342 [astro-ph]`.

Wald, Robert M. (1984). *General Relativity.*

Wang, Feige, Frederick B. Davies, Jinyi Yang, et al. (June 2020). “A Significantly Neutral Intergalactic Medium Around the Luminous z = 7 Quasar J0252-0503”. In: *ApJ* 896.1, 23, p. 23. DOI: `10.3847/1538-4357/ab8c45`. arXiv: `2004.10877 [astro-ph.GA]`.

Watson, William A., Ilian T. Iliev, Anson D’Aloisio, et al. (Aug. 2013). “The halo mass function through the cosmic ages”. In: *MNRAS* 433.2, pp. 1230–1245. DOI: `10.1093/mnras/stt791`. arXiv: `1212.0095 [astro-ph.CO]`.

Wechsler, Risa H. and Jeremy L. Tinker (Sept. 2018). “The Connection Between Galaxies and Their Dark Matter Halos”. In: *ARA&A* 56, pp. 435–487. DOI: `10.1146/annurev-astro-081817-051756`. arXiv: `1804.03097 [astro-ph.GA]`.

Whitler, Lily R., Charlotte A. Mason, Keven Ren, et al. (July 2020). “The impact of scatter in the galaxy UV luminosity to halo mass relation on Ly α visibility during the epoch of reionization”. In: *MNRAS* 495.4, pp. 3602–3613. DOI: `10.1093/mnras/staa1178`. arXiv: `1911.03499 [astro-ph.CO]`.

Wilkins, Stephen M., Christopher C. Lovell, and Elizabeth R. Stanway (Dec. 2019). “Recalibrating the cosmic star formation history”. In: *MNRAS* 490.4, pp. 5359–5365. DOI: `10.1093/mnras/stz2894`. arXiv: `1910.05220 [astro-ph.GA]`.

Willott, Chris J., Guillaume Desprez, Yoshihisa Asada, et al. (May 2024). “A Steep Decline in the Galaxy Space Density beyond Redshift 9 in the CANUCS UV Luminosity Function”. In: *ApJ* 966.1, 74, p. 74. DOI: `10.3847/1538-4357/ad35bc`. arXiv: `2311.12234 [astro-ph.GA]`.

Witstok, Joris, Peter Jakobsen, Roberto Maiolino, et al. (Mar. 2025). “Witnessing the onset of reionization through Lyman-α emission at redshift 13”. In: *Nature* 639.8056, pp. 897–901. DOI: `10.1038/s41586-025-08779-5`. arXiv: `2408.16608 [astro-ph.GA]`.

Witstok, Joris, Roberto Maiolino, Renske Smit, et al. (Apr. 2024). “JADES: Primeval Lyman-α emitting galaxies reveal early sites of reionisation out to redshift $z \sim 9$”. In: *arXiv e-prints*, arXiv:2404.05724, arXiv:2404.05724. DOI: `10.48550/arXiv.2404.05724`. arXiv: `2404.05724 [astro-ph.GA]`.

Witstok, Joris, Renske Smit, Aayush Saxena, et al. (Feb. 2024). “Inside the bubble: exploring the environments of reionisation-era Lyman-α emitting galaxies with JADES and FRESCO”. In: *A&A* 682, A40, A40. DOI: `10.1051/0004-6361/202347176`. arXiv: `2306.04627 [astro-ph.GA]`.

Witten, Callum, Nicolas Laporte, Sergio Martin-Alvarez, et al. (Mar. 2024). “Deciphering Lyman-α emission deep into the epoch of reionization”. In: *Nature Astronomy* 8, pp. 384–396. DOI: `10.1038/s41550-023-02179-3`. arXiv: `2303.16225 [astro-ph.GA]`.

Wolcott-Green, J., Z. Haiman, and G. L. Bryan (Dec. 2011). “Photodissociation of H_2 in protogalaxies: modelling self-shielding in three-dimensional simulations”. In: *MNRAS* 418.2, pp. 838–852. DOI: `10.1111/j.1365-2966.2011.19538.x`. arXiv: `1106.3523 [astro-ph.CO]`.

Wolcott-Green, Jemma, Zoltán Haiman, and Greg L. Bryan (Jan. 2021). “Suppression of H_2-cooling in protogalaxies aided by trapped Lyα cooling radiation”. In: *MNRAS* 500.1, pp. 138–144. DOI: `10.1093/mnras/staa3057`. arXiv: `2001.05498 [astro-ph.GA]`.

Wyithe, J. Stuart B. and Abraham Loeb (Jan. 2013). “A suppressed contribution of low-mass galaxies to reionization due to supernova feedback”. In: *MNRAS* 428.3, pp. 2741–2754. DOI: `10.1093/mnras/sts242`. arXiv: `1209.2215 [astro-ph.CO]`.

Xu, Hao, Kyungjin Ahn, John H. Wise, et al. (Aug. 2014). “Heating the Intergalactic Medium by X-Rays from Population III Binaries in High-redshift Galaxies”. In: *ApJ* 791.2, 110, p. 110. DOI: 10.1088/0004-637X/791/2/110. arXiv: 1404.6555 [astro-ph.CO].

Xu, Hao, John H. Wise, Michael L. Norman, et al. (Dec. 2016). “Galaxy Properties and UV Escape Fractions during the Epoch of Reionization: Results from the Renaissance Simulations”. In: *ApJ* 833.1, 84, p. 84. DOI: 10.3847/1538-4357/833/1/84. arXiv: 1604.07842 [astro-ph.GA].

Xu, Xinfeng, Alaina Henry, Timothy Heckman, et al. (Feb. 2023). “The Low-redshift Lyman Continuum Survey: Optically Thin and Thick Mg II Lines as Probes of Lyman Continuum Escape”. In: *ApJ* 943.2, 94, p. 94. DOI: 10.3847/1538-4357/aca89a. arXiv: 2301.04087 [astro-ph.GA].

Yamada, T., Y. Matsuda, K. Kousai, et al. (May 2012). “Profiles of Lyα Emission Lines of the Emitters at z = 3.1”. In: *ApJ* 751.1, 29, p. 29. DOI: 10.1088/0004-637X/751/1/29. arXiv: 1203.3633 [astro-ph.CO].

Yeh, Jessica Y. -C., Aaron Smith, Rahul Kannan, et al. (May 2022). “The THESAN project: ionizing escape fractions of reionization-era galaxies”. In: *arXiv e-prints*, arXiv:2205.02238, arXiv:2205.02238. arXiv: 2205.02238 [astro-ph.GA].

Yuan, Fang-Ting, Zhen-Ya Zheng, Chunyan Jiang, et al. (Nov. 2024). “Merging Signatures in an Offset Lyman Continuum Emitter at Redshift 3.8”. In: *ApJ* 975.1, 53, p. 53. DOI: 10.3847/1538-4357/ad75ff. arXiv: 2409.20352 [astro-ph.GA].

Yung, L. Y. Aaron, Rachel S. Somerville, Steven L. Finkelstein, et al. (Jan. 2024). “Are the ultra-high-redshift galaxies at z ¿ 10 surprising in the context of standard galaxy formation models?” In: *MNRAS* 527.3, pp. 5929–5948. DOI: 10.1093/mnras/stad3484. arXiv: 2304.04348 [astro-ph.GA].

Yung, L. Y. Aaron, Rachel S. Somerville, Gergö Popping, et al. (Dec. 2019). “Semi-analytic forecasts for JWST - II. Physical properties and scaling relations for galaxies at z = 4-10”. In: *MNRAS* 490.2, pp. 2855–2879. DOI: 10.1093/mnras/stz2755. arXiv: 1901.05964 [astro-ph.GA].

Zahn, O., C. L. Reichardt, L. Shaw, et al. (Sept. 2012). “Cosmic Microwave Background Constraints on the Duration and Timing of Reionization from the South Pole Telescope”. In: *ApJ* 756.1, 65, p. 65. DOI: 10.1088/0004-637X/756/1/65. arXiv: 1111.6386 [astro-ph.CO].

Zahn, Oliver, Andrei Mesinger, Matthew McQuinn, et al. (June 2011). “Comparison of reionization models: radiative transfer simulations and approximate, seminumeric models”. In: *MNRAS* 414.1, pp. 727–738. DOI: 10.1111/j.1365-2966.2011.18439.x. arXiv: 1003.3455 [astro-ph.CO].

Zahn, Oliver, Matias Zaldarriaga, Lars Hernquist, et al. (Sept. 2005). “The Influence of Nonuniform Reionization on the CMB”. In: *ApJ* 630.2, pp. 657–666. DOI: 10.1086/431947. arXiv: astro-ph/0503166 [astro-ph].

Zaroubi, Saleem (Jan. 2013). “The Epoch of Reionization”. In: *The First Galaxies*. Ed. by Tommy Wiklind, Bahram Mobasher, and Volker Bromm. Vol. 396. Astrophysics and Space Science Library, p. 45. DOI: 10.1007/978-3-642-32362-1_2. arXiv: 1206.0267 [astro-ph.CO].

Zheng, Zhen-Ya, Junxian Wang, James Rhoads, et al. (June 2017). “First Results from the Lyman Alpha Galaxies in the Epoch of Reionization (LAGER) Survey: Cosmological Reionization at z $\sim$ 7”. In: *ApJ* 842.2, L22, p. L22. DOI: 10.3847/2041-8213/aa794f. arXiv: 1703.02985 [astro-ph.GA].

Zheng, Zheng, Andreas A. Berlind, David H. Weinberg, et al. (Nov. 2005). “Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies”. In: *ApJ* 633.2, pp. 791–809. DOI: 10.1086/466510. arXiv: astro-ph/0408564 [astro-ph].

Zhu, Yongda, George D. Becker, Holly M. Christenson, et al. (Oct. 2023). “Probing Ultralate Reionization: Direct Measurements of the Mean Free Path over 5 ¡ z ¡ 6”. In: *ApJ* 955.2, 115, p. 115. DOI: 10.3847/1538-4357/aceef4. arXiv: 2308.04614 [astro-ph.CO].

Zitrin, Adi, Ivo Labbé, Sirio Belli, et al. (Sept. 2015). “Lymanα Emission from a Luminous z = 8.68 Galaxy: Implications for Galaxies as Tracers of Cosmic Reionization”. In: *ApJ* 810.1, L12, p. L12. DOI: 10.1088/2041-8205/810/1/L12. arXiv: 1507.02679 [astro-ph.GA].

List of Figures