Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **526,** 3170–3183 (2023) https://doi.org/10.1093/mnras/stad2961
Advance Access publication 2023 September 28

Inferring reionization and galaxy properties from the patchy kinetic Sunyaev–Zel'dovich signal

Ivan Nikolić,[1]★ Andrei Mesinger,[1] Yuxiang Qin[2,3] and Adélie Gorce[4]

[1]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, PI, Italy*
[2]*School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[3]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D)*
[4]*Department of Physics and Trottier Space Institute, McGill University, Montreal, QC H3A 2T8, Canada*

Accepted 2023 September 25. Received 2023 September 25; in original form 2023 July 4

ABSTRACT

The patchy kinetic Sunyaev–Zel'dovich (kSZ) signal is an integral probe of the timing and morphology of the epoch of reionization (EoR). Recent observations have claimed a low signal-to-noise (S/N) measurement, with a dramatic increase in S/N expected in the near future. In this work, we quantify what we can learn about the EoR from the kSZ signal. We perform Bayesian inference by sampling galaxy properties and using forward-models of the kSZ as well as other EoR and galaxy observations in the likelihood. Including the recent kSZ measurement obtained by the *South Pole Telescope* ($\mathcal{D}_{3000}^{\rm pkSZ} = 1.1^{+1.1}_{-0.7}\mu$ K^2) shifts the posterior distribution in favour of faster and later reionization models, resulting in lower values of the optical depth to the cosmic microwave background: $\tau_e = 0.052^{+0.009}_{-0.008}$ with a 68 per cent confidence interval (CI). The combined EoR and ultraviolet luminosity function observations also imply a typical ionizing escape fraction of $0.04^{+0.05}_{-0.03}$ (95 per cent CI), without a strong dependence on halo mass. We show how the patchy kSZ power from our posterior depends on the midpoint and duration of reionization: a popular parametrization of EoR timing. For a given midpoint and duration, the EoR morphology only has a few per cent impact on the patchy kSZ power in our posterior. However, a physical model is needed to obtain tight constraints from the current low S/N patchy kSZ measurement, as it allows us to take advantage of complimentary high-z observations. Future high S/N detections of the patchy kSZ should decrease the current uncertainties on the timing of the EoR by factors of ∼2–3.

Key words: cosmic background radiation – dark ages, reionization, first stars – early Universe – large-scale structure of Universe – galaxies: high-redshift.

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 et al. 2010; Becker et al. 2015; Iršič et al. 2017; Bosman et al. 2018; Gaikwad et al. 2020); (ii) Ly α emitting galaxies (e.g. Stark et al. 2010, 2017; Konno et al. 2018; Hoag et al. 2019; Mason et al. 2019; Endsley et al. 2022; Leonova et al. 2022); (iii) the optical depth to the cosmic microwave background (e.g. CMB; Planck Collaboration 2016, 2020; Heinrich & Hu 2021); (iv) ultraviolet (UV) luminosity functions (LF; e.g. Bouwens et al. 2015, 2016, 2023; Finkelstein et al. 2015; Oesch et al. 2018; Bhatawdekar et al. 2019; Harikane et al. 2023b); (v) preliminary upper limits on the 21-cm power spectrum (Mertens et al. 2020; Trott et al. 2020; The HERA Collaboration 2022, 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.[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 et al. 2007; Mesinger, McQuinn & Spergel 2012; Park et al. 2013; Choudhury, Mukherjee & Paul 2021a; Bégin, Liu & 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; Das et al. 2011, 2014; Shirokoff et al. 2011; Reichardt et al. 2012; Dunkley et al. 2013; George et al. 2015).

★ E-mail: ivan.nikolic4848@gmail.com
[1]https://www.skatelescope.org/

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 per cent CI; Reichardt et al. 2021). These relatively low values qualitatively point to a much later and more rapid EoR compared to original estimates (e.g. McQuinn et al. 2005; Zahn et al. 2005, 2012; Iliev et al. 2007; Mesinger et al. 2012; Park et al. 2013; Calabrese et al. 2014; Alvarez 2016; Paul, Mukherjee & Choudhury 2021). Future telescopes, such as the *Simons Observatory*[2] (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. First, 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 Section 2 we discuss how we compute the patchy kSZ signal. Our Bayesian framework, combining the kSZ with complementary observations, is summarized in Section 3. We present and discuss our results in Section 4. In Section 5 we quantify how accurately the midpoint and duration of reionization can predict the patchy kSZ at $l = 3000$. Finally, we conclude in Section 6. Throughout this work, we assume standard ΛCDM cosmological parameters ($\Omega_{\rm m}, \Omega_{\rm b}, \Omega_\Lambda, h, \sigma_8, n_{\rm s} = 0.321, 0.049, 0.679, 0.67, 0.81, 0.963$), consistent with the latest estimates from Planck Collaboration (2020). Unless stated otherwise, we quote all quantities in comoving units.

2 THE PATCHY KSZ SIGNAL

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

$$\begin{aligned}\delta T_{\rm kSZ} &= \frac{\Delta T}{T}(\hat{\mathbf{u}}) \\ &= \sigma_{\rm T}\int {\rm d}z \left(\frac{{\rm d}t}{{\rm d}z}\right) e^{-\tau_e(z)} n_e\, \hat{\mathbf{u}}\cdot\mathbf{v} \\ &= \sigma_{\rm T}\int {\rm d}z \left(\frac{{\rm d}t}{{\rm d}z}\right) e^{-\tau_e(z)}\, x_e\, n_b\, \hat{\mathbf{u}}\cdot\mathbf{v}. \end{aligned} \quad (1)$$

Here, $\sigma_{\rm T}$ 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,

$$\tau_e(z) = \sigma_{\rm T}\int_0^z {\rm d}z'\, c\frac{{\rm d}t}{{\rm d}z'} n_e. \quad (2)$$

The redshift integral in equation (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: $\mathcal{D}_l^{\rm kSZ} = \mathcal{D}_l^{\rm hkSZ} + \mathcal{D}_l^{\rm pkSZ}$. The post-reionization kSZ power spectrum, $\mathcal{D}_l^{\rm hkSZ}$, is dominated by fluctuations in n_b during the era of cluster formation at $z \lesssim 1$ (e.g. Shaw, Rudd & Nagai 2012), while the patchy kSZ power spectrum, $\mathcal{D}_l^{\rm pkSZ}$, 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 Section 2.2. Because we do not know *a priori* the reionization redshift, here we *define* the patchy kSZ component as the contribution to equation (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 motivated by recent Lyα forest data whose interpretation requires a late reionization, ending at $5.3 \lesssim z \lesssim 5.6$ (Choudhury, Paranjape & Bosman 2021b; Qin et al. 2021b; Bosman et al. 2022, Qin et al., in preparation).

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. Reichardt et al. 2012; Zahn et al. 2012; Alvarez 2016). During the EoR, $l = 3000$ roughly corresponds to physical scales of $\sim$20 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 cMpc are sufficient to capture the ionization power spectra on those scales (e.g. Iliev et al. 2014; Kaur, Gillet & 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. Shaw et al. 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. Park et al. 2013; Gorce et al. 2020) but at the price of neglecting the contribution of higher order correlations (above two points) which can represent up to 10 per cent 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 seminumerical simulations; e.g. Mesinger et al. 2012; Zahn et al.

[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 $\mathcal{D}_l^{\rm kSZ} = \frac{l(l+1)}{2\pi} C_l^{\rm kSZ}$, where $C_l^{\rm kSZ} = T_{\rm CMB}^2 |\overline{\delta T}_{\rm kSZ}(k)|^2$, $T_{\rm CMB}$ is the mean CMB temperature, and $\overline{\delta T}_{\rm kSZ}(k)$ is the Fourier transform of $\delta T_{\rm kSZ}$.

2012; Battaglia et al. 2013; Seiler et al. 2019; Choudhury et al. 2021a; Chen et al. 2023). Our self-consistent approach allows us to incorporate multifrequency observations of the EoR and high-z galaxies in the likelihood. We discuss how this is done in the following section.

2.1 Computing the patchy kSZ from galaxy-driven EoR simulations

In this work we extend the public simulation package 21CMFAST[7] (e.g. Mesinger & Furlanetto 2007; Mesinger, Furlanetto & Cen 2011; Murray et al. 2020) to forward-model the patchy kSZ signal together with other observables. 21CMFAST is a seminumerical 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. Furlanetto, Zaldarriaga & Hernquist 2004). Specifically, a cell is marked as ionized if at any radius,

$$n_{\rm ion} \geq (1 + n_{\rm rec})(1 - x_e), \tag{3}$$

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

$$n_{\rm ion} = \frac{1}{\rho_b} \int_0^\infty {\rm d}M_{\rm h}\, \frac{{\rm d}n(M_{\rm h}, z|R, \delta_R)}{{\rm d}M_{\rm h}}\, f_{\rm duty}\, M_*\, f_{\rm esc}\, N_{\gamma/b}. \tag{4}$$

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

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

$$f_{\rm duty}(M_{\rm h}) = \exp\left(-\frac{M_{\rm h}}{M_{\rm turn}}\right). \tag{5}$$

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

$$f_{\rm esc}(M_{\rm h}) = f_{\rm esc,10} \left(\frac{M_{\rm h}}{10^{10}{\rm M}_\odot}\right)^{\alpha_{\rm esc}}, \tag{6}$$

$$M_*(M_{\rm h}) = f_{*,10} \left(\frac{M_{\rm h}}{10^{10}{\rm M}_\odot}\right)^{\alpha_*} \left(\frac{\Omega_b}{\Omega_m}\right) M_{\rm h}. \tag{7}$$

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 $f_* \equiv f_{*,10}\left(\frac{M_{\rm h}}{10^{10}{\rm M}_\odot}\right)^{\alpha_*}$ have a physical upper limit of 1. This model also assumes that the star formation rate (SFR) can be expressed on average as the stellar mass divided by some characteristic time-scale,

$$\dot{M}_*(M_{\rm h}, z) = \frac{M_*}{H^{-1}(z) t_*}, \tag{8}$$

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

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. Ishigaki et al. 2018; Behroozi et al. 2019; Ma et al. 2020; Park et al. 2020; Bouwens et al. 2022). Further details about the code and the parametrization can be found in Mesinger et al. (2011), Park et al. (2019), and Murray et al. (2020).

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 (1), accounting also for the angular evolution of $\hat{\mathbf{u}}$ (Mesinger et al. 2012).

In Fig. 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 et al. (2021b; 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 $\Delta_z \equiv z(\bar{x}_{\rm HI} = 0.75) - z(\bar{x}_{\rm HI} = 0.25)$, 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 Reichardt 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 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 (Reichardt et al. 2012; Dunkley et al. 2013; Das et al. 2014; 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), Reichardt et al. (2021) recently claimed a 3σ measurement of the total kSZ power: $\mathcal{D}^{\rm kSZ}_{3000} = 3.0 \pm 1.0\ \mu{\rm K}^2$ (68 per cent CI).

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 Shaw et al. (2012): $\mathcal{D}^{\rm hkSZ}_{3000} = 1.65\mu{\rm K}^2$. 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, Reichardt et al. (2021) estimate the patchy kSZ power at $l = 3000$ to be $\mathcal{D}^{\rm pkSZ}_{3000} = 1.1^{+1.0}_{-0.7}\ \mu{\rm K}^2$ (68 per cent CI). 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 Reichardt 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 Shaw et al. (2012) and fig. 5 in Mesinger et al. (2012)]. Therefore we use the following observational constraint when

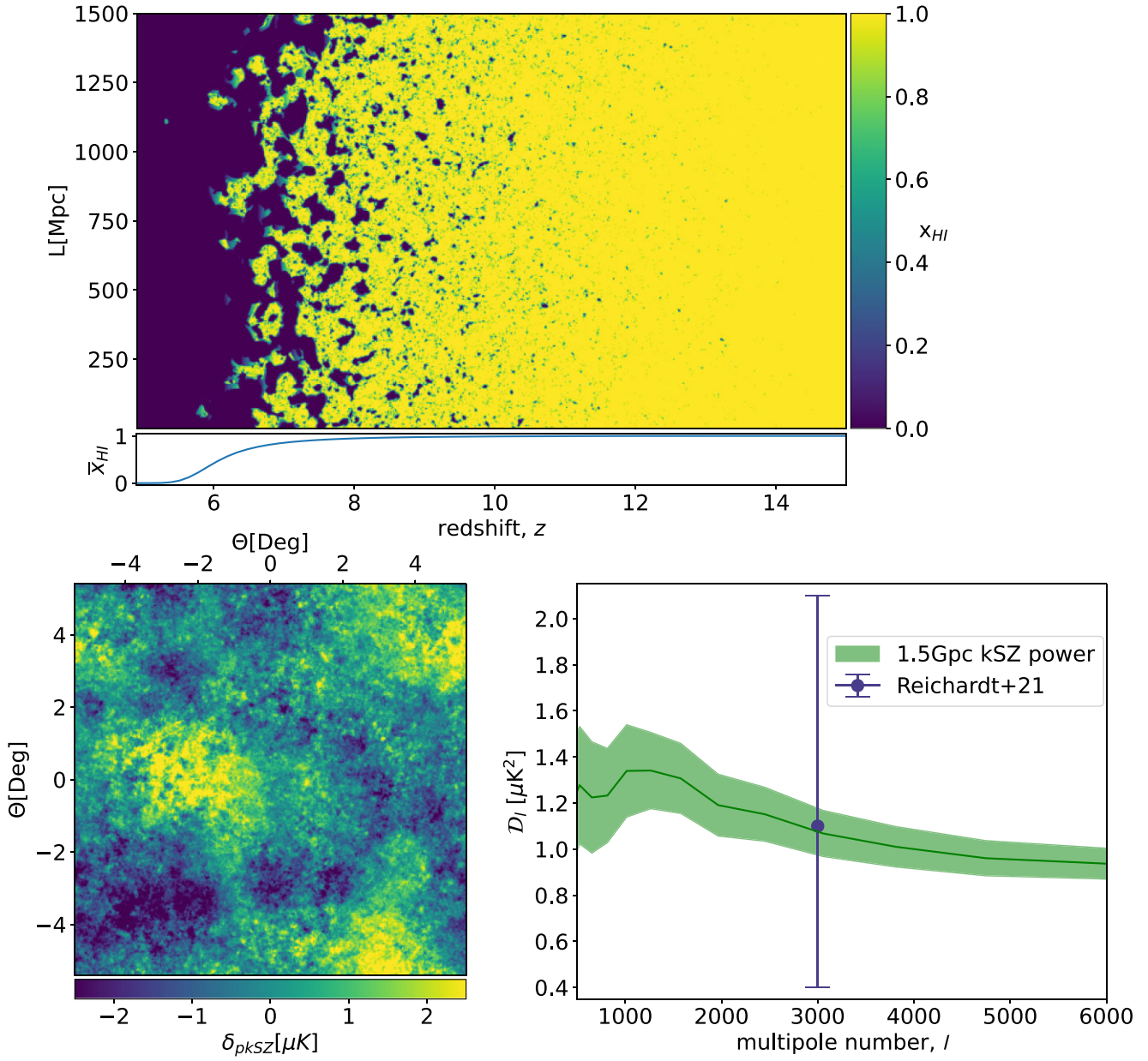

Figure 1. Upper panel: 2D slice through the neutral hydrogen fraction lightcone together with its mean evolution on the bottom. The lightcone slice is 1.5 Gpc in height and 1.4 Mpc thick. Lower left panel: Map of the patchy kinetic Sunyaev–Zel'dovich signal, defined as being sourced by redshifts greater 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 Reichardt 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 $t_* = 0.392$.

performing inference in Section 4.1: $\mathcal{D}_{3000}^{\rm pkSZ} = 1.2^{+1.0}_{-0.7}$ μK^2 (68 per cent CI).

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

[8]As this project was started before the publication of Gorce et al. (2022), here we use the original patchy kSZ estimate by Reichardt et al. (2021). The estimate in Gorce et al. (2022) would imply an even later reionization than shown here, also consistent with the newest analysis of the Lyman alpha forest spectra (Qin et al., in preparation) as well as the forest dark fraction (Jin et al. 2023; Campo et al., in preparation). We aim to revisit this in future work when more of these new constraints become public.

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 Section 4.3.

3 COMPLEMENTARY EOR AND GALAXY OBSERVATIONS

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

(i) *Ly 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 Bosman et al. (2018). Qin et al. (2021b) showed that this data require reionization to end late, $z \leq 5.6$ (see also Choudhury et al. 2021b).

(ii) *Dark fraction in the Ly α and Ly β forests*: The fraction of QSO spectral pixels that are dark (zero transmission) in both Ly α and Ly β from the sample in McGreer et al. (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 data set favours earlier reionization models.

(iii) *High-redshift galaxy UV LFs*: The 1500 Å restframe UV LFs at $z = 6$–10, estimated by Bouwens et al. (2015, 2016) and Oesch et al. (2018). To constrain our models, we assume a conversion factor between the SFR and UV luminosity, $\dot{M}_* = \mathcal{K}_{\rm UV} L_{\rm UV}$, and take $\mathcal{K}_{\rm UV} = 1.15 \cdot 10^{-28}\,{\rm M_\odot yr^{-1} erg^{-1}\ s\ Hz}$, following Sun & Furlanetto (2016).[9] UV luminosities are then related to magnitudes using the AB magnitude relation (Oke & Gunn 1983): $\log_{10}\left(\frac{L_{\rm UV}}{\rm ergs^{-1}Hz^{-1}}\right) = 0.4 \times (51.63 - M_{1500})$. 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.

(iv) *The CMB optical depth*: The Thomson scattering optical depth of CMB photons as computed by Planck Collaboration (2020), $\tau_e = 0.0561 \pm 0.071$ (1σ). Although it is more accurate to directly forward model the CMB EE power spectra, Qin 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 data sets are used in all of our inferences, each contributing a factor in the final likelihood. We write out explicitly all likelihood terms in Appendix B. For further details, we refer the interested reader to Qin et al. (2021b).

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 (Reichardt 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.

4.1 Inference set-up

To perform Bayesian inference, we use 21CMMC[10] (Greig & Mesinger 2015, 2018), a public Monte Carlo sampler of 21CMFAST. For each set of model parameters (see Section 2.1), 21CMMC computes a 3D lightcone realization of cosmological fields, comparing the model to the observations (see Section 2.2 and 3). Here we use the MultiNest (Feroz, Hobson & Bridges 2009; Qin et al. 2021a) sampler, which is fully implemented in 21CMMC and scales well to high-dimensional inference (e.g. The HERA Collaboration 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 ∼45k samples, taking ∼260k core hours.

Unfortunately, due to computational limitations, we cannot use ultra-large simulations (e.g. Fig. 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 et al. 2007; Shaw et al. 2012; Park et al. 2013; Alvarez 2016). Specifically, we use simulations of $(500\ {\rm Mpc})^3$ 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. Mesinger et al. 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, $f^{\rm 1.5Gpc}_{\rm 0.5Gpc} \equiv \mathcal{D}^{\rm pkSZ,1.5Gpc}_{3000}/\mathcal{D}^{\rm pkSZ,500Mpc}_{3000}$, adjusting our forward models by this factor and including the corresponding variance in the denominator of the likelihood. We obtain $f^{\rm 1.5Gpc}_{\rm 0.5Gpc} = 1.27 \pm 0.19$. Further details on this calibration procedure can be found in Appendix A.

4.2 Inference results using the recent *SPT* measurement

We compute two posteriors:

(i) *Without kSZ*: This corresponds to the posterior based on the observational data (i)–(iv) from the previous section, i.e. large-scale Ly α forest opacity PDFs, the forest dark fraction, UV LFs, and the CMB optical depth.[11]

(ii) *With kSZ*: This is the same as *without kSZ*, but including an additional factor in the likelihood, $\mathcal{L}_{\rm kSZ}$ (see Appendix A for details) corresponding to the patchy kSZ measurement by Reichardt et al.

[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 & Dickinson 2014).

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

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

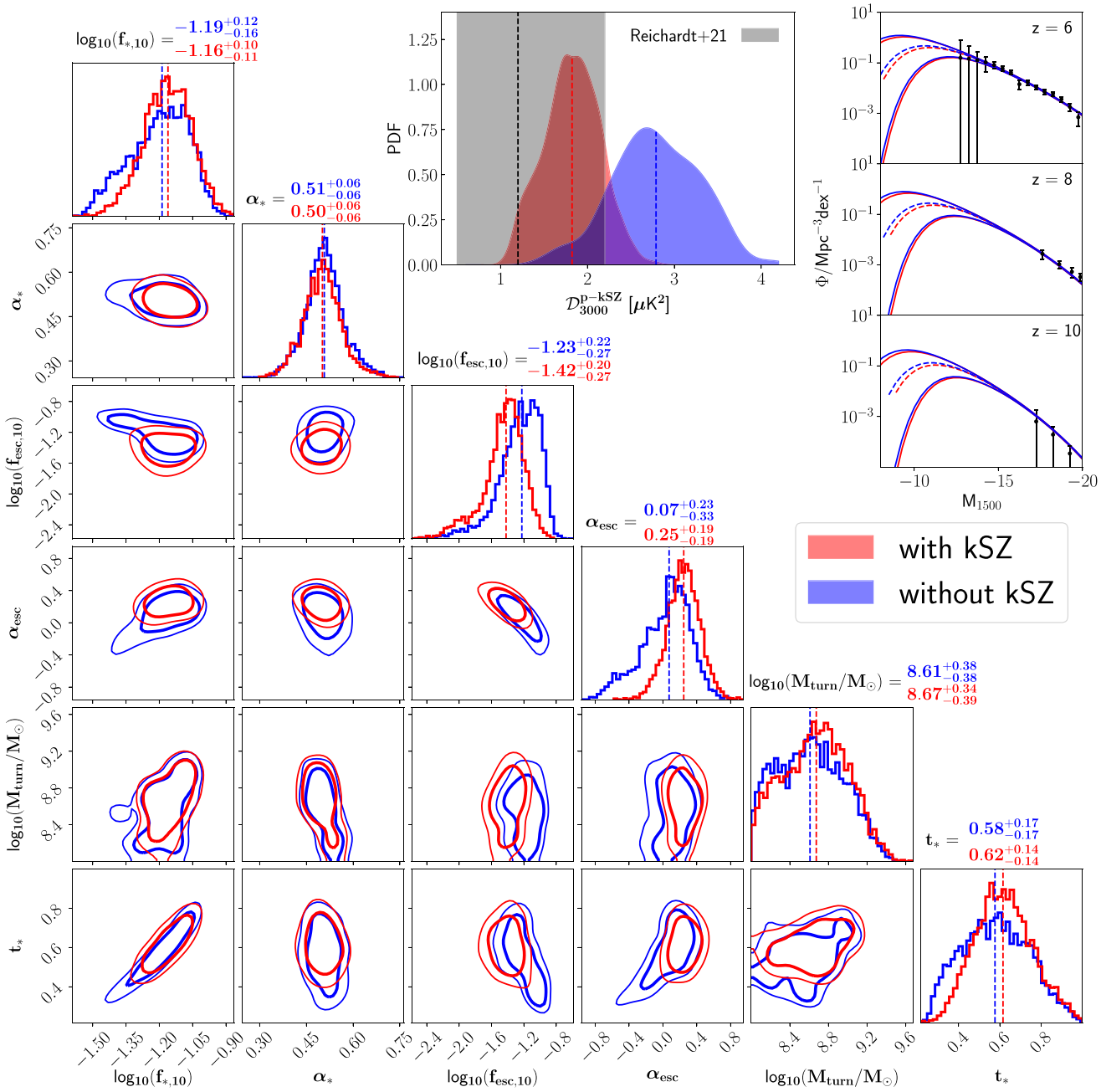

Figure 2. Marginalized posteriors of *without kSZ* (*blue*) and *with kSZ* (*red*). As discussed in the text, *without kSZ* is constrained using large-scale Ly α forest opacity probability density functions (PDFs), the forest dark fraction, UV LFs, and the CMB optical depth, while *with kSZ* additionally 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 per cent and 68 per cent credible intervals (CI), respectively. The marginalized median values (shown also as dashed lines) with the 68 per cent central CI 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 Reichardt et al. (2021) observational estimate in grey. Also shown are the median and [14, 86] per cent CI (dashed and solid lines, respectively) of the inferred UV LFs at $z = 6$, 8, and 10. Black points with error bars are UV LF observations used for the inference from Bouwens et al. (2015, 2016) and Oesch et al. (2018).

(2021), adjusted for the slightly different lower redshift bound as discussed above: $\mathcal{D}_{3000}^{\rm pkSZ} = 1.2^{+1.0}_{-0.7}$ μK^2 (68 per cent CI).

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.

4.2.1 *Galaxy parameters and EoR history*

In the bottom left of Fig. 2, we show the resulting 2D and 1D 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 centre*), and the UV LFs at $z = 6$, 8, 10 (*top right*).

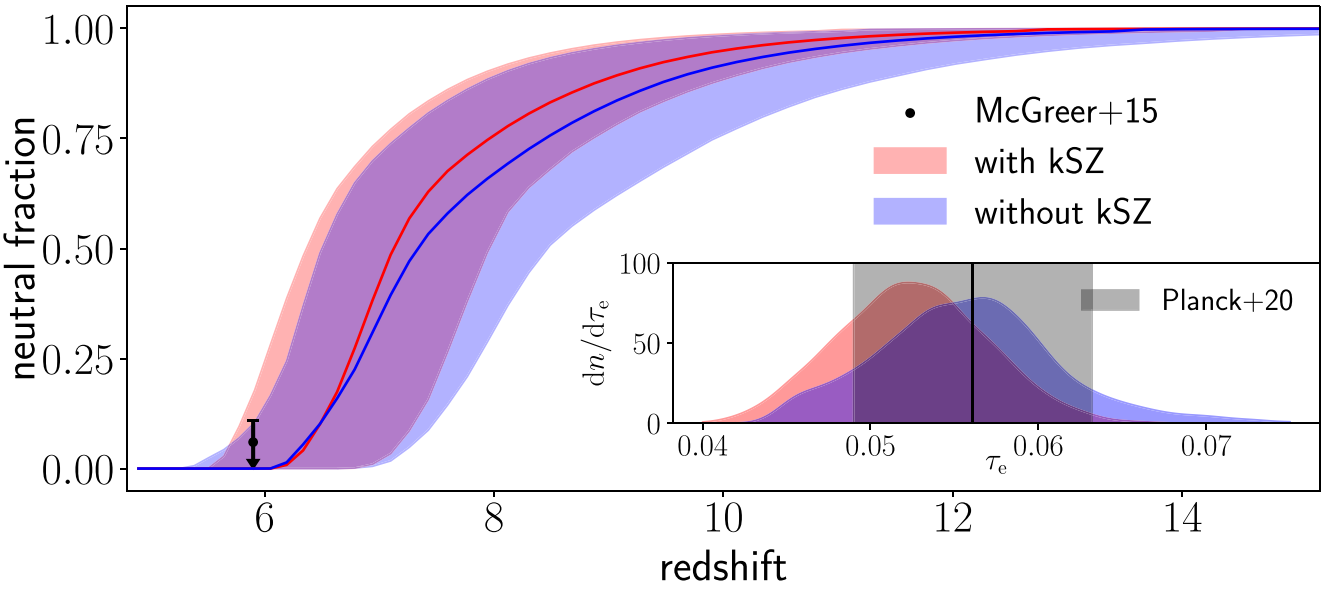

Figure 3. Median EoR histories with 95 per cent CI for the *without kSZ* (*blue*) and *with kSZ* (*red*) posteriors. Also shown is the upper limit from McGreer, Mesinger & D'Odorico (2015) at $z \sim 5.9$ that is used in the likelihood (Section 3) for both posteriors. The insert shows probability density distributions of τ_e for the *with kSZ* and *without kSZ* posteriors. The vertical black line and grey shaded region correspond to the Planck Collaboration (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.

From the $\mathcal{D}_{3000}^{\rm pkSZ}$ PDFs shown in the top centre panel, we see that the recent measurement by Reichardt et al. (2021) is in mild tension with the *without kSZ* posterior: the kSZ data favour 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 favour of smaller kSZ power. Most of the *with kSZ* posterior is still above the mean estimate of the kSZ power by Reichardt 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 $f_{\rm esc,\,10}$ and $\alpha_{\rm esc}$ (see equation 6). The *SPT* measurement favours slightly lower values of $f_{\rm esc,\,10}$ and higher values of $\alpha_{\rm esc}$. 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. 3. Including the relatively low patchy kSZ amplitude claimed by Reichardt et al. (2021) disfavours the more extended EoR histories present in the *without kSZ* posterior. While the end of the EoR remains fairly unchanged, constrained by Ly α forest observations (e.g. Choudhury et al. 2021b; Qin et al. 2021b), 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. 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* one. The EoR histories implied by the *with kSZ* posterior are consistent with the Ly 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 preparation). This would leave the CMB τ_e as the only data set preferring a slightly earlier EoR. Such a mild tension between the two CMB data sets could come from calibration or analysis inconsistencies between large- and small-scale data, that is between the *SPT* and Planck data (e.g. Gorce et al. 2022).

On the other hand, constraints on parameters governing the SFRs and stellar-to-halo mass relations (SHMR; i.e. $f_{*,\,10}$, α_*, t_*, $M_{\rm turn}$) are fairly unchanged when including kSZ data. As already shown in Park et al. (2019) and Qin et al. (2021b), observed high-z UV LFs constrain the SHMR ($f_{*,\,10}$, 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.

4.2.2 *Scaling relations of EoR galaxies*

To gain further insight into the implications of our results, we show the corresponding galaxy scaling relations in Fig. 4. In the top panel, we plot the inferred SHMR, defined as the average stellar mass inside a halo of mass $M_{\rm h}$ [including the $f_{\rm duty}$ occupation fraction term from equation (5)]. The redshift-independent median relation from the *with kSZ* posterior is denoted with the red solid line, corresponding to

$$\frac{\bar{M}_*}{M_h} = 0.011^{+0.003}_{-0.002}\left(\frac{M_{\rm h}}{10^{10}\,{\rm M}_\odot}\right)^{0.50^{+0.06}_{-0.06}} \quad (68\ \text{per cent CI}), \qquad (9)$$

while the 95 per cent CI is shown with the purple shading.[12] We also show the corresponding 95 per cent CI 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

[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 & 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.

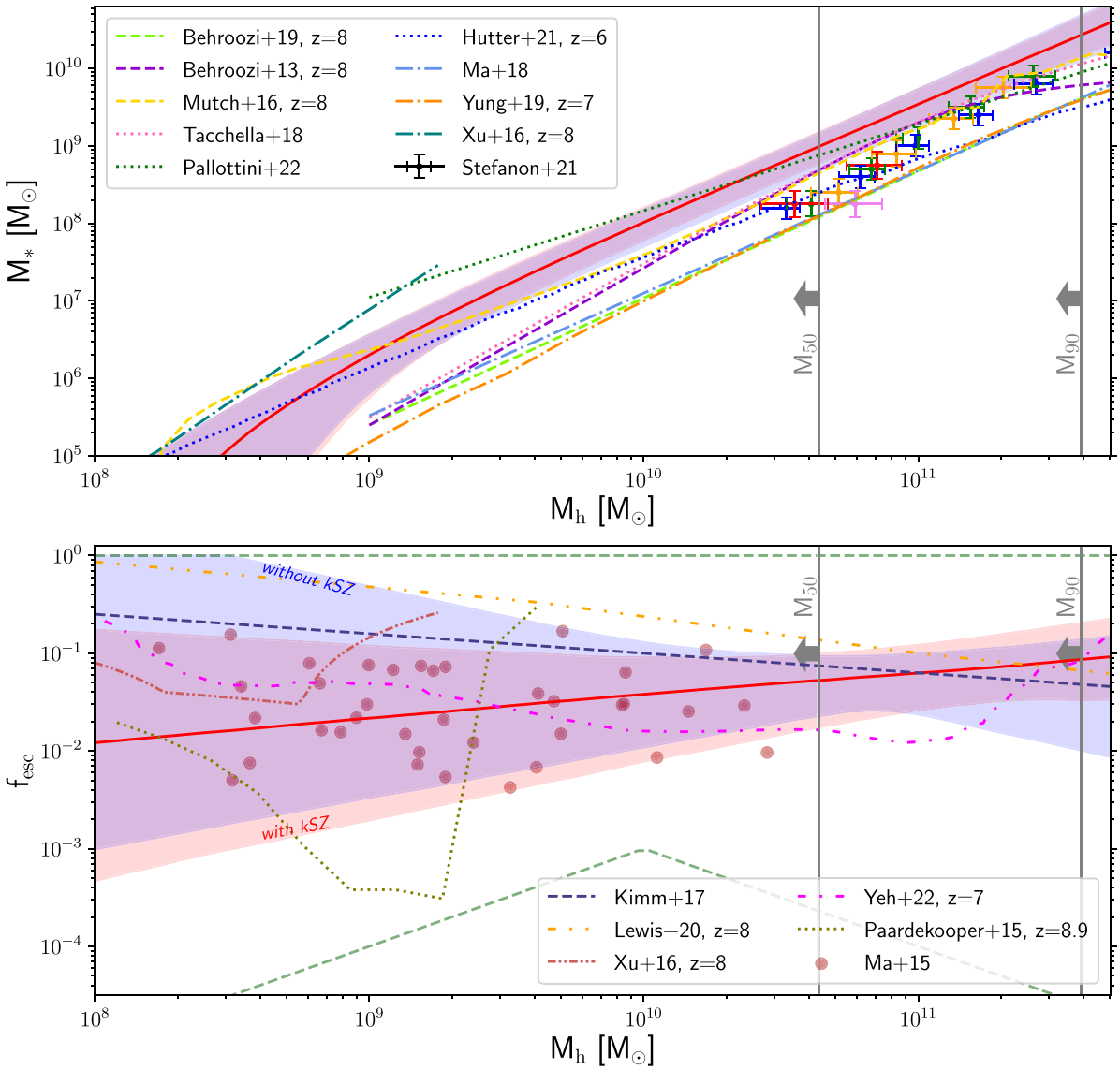

Figure 4. The dependence of average galaxy properties with halo mass. Shaded regions represent 95 per cent CI of *without kSZ* (*blue*) and *with kSZ* (*red*). Vertical lines demarcate the posterior-averaged mean of M_{50} (M_{90}), defined as halo mass upper limit below which galaxies source 50 per cent (90 per cent) 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 (Behroozi, Wechsler & Conroy 2013; Mutch et al. 2016; Tacchella et al. 2018; Behroozi et al. 2019; Yung et al. 2019; Hutter et al. 2021), as well as from cosmological, hydro simulations (Xu et al. 2016; Ma et al. 2018; Pallottini et al. 2022). For Tacchella et al. (2018) and Ma et al. (2018), we use the redshift independent fits from their equation (12) and equation (1), respectively. The Xu 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 Pallottini et al. (2022) curve is a linear fit to their data points. Coloured points with error bars correspond to abundance matching estimates assuming a constant duty cycle (Stefanon 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 (Ma et al. 2015; Paardekooper, Khochfar & Dalla Vecchia 2015; Xu et al. 2016; Kimm et al. 2017; Lewis et al. 2020; Yeh et al. 2023). For Kimm et al. (2017) we show their 'fiducial' model. The Ma et al. (2015) points represents time-averaged escape fractions obtained using their SHMR relation (see their figs 3 and 9.).

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 halo mass function (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]

[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 Section 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 & Stanway 2019; Stanway et al. 2020).

Similarly, in the bottom panel of Fig. 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,

$$f_{\rm esc} = 0.038^{+0.021}_{-0.017}\left(\frac{M_{\rm h}}{10^{10}\,{\rm M}_\odot}\right)^{0.25^{+0.19}_{-0.19}} \quad (68\ \text{per cent CI; } with\ kSZ), \tag{10}$$

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

$$f_{\rm esc} = 0.060^{+0.038}_{-0.028}\left(\frac{M_{\rm h}}{10^{10}\,{\rm M}_\odot}\right)^{0.07^{+0.23}_{-0.33}} \quad (68\ \text{per cent CI; } without\ kSZ). \tag{11}$$

As in the panel above, the corresponding shaded areas demarcate the 95 per cent CI. 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 favours 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 per cent. Interestingly, strong evolution with halo mass is disfavoured.

In Fig. 4 we also demarcate the posterior-averaged mean of M_{50} (M_{90}), defined as halo mass upper limit below which galaxies source 50 per cent (90 per cent) of the ionizing emissivity at $z = 7$, i.e. M_{50} is calculated by solving the equation: $\int_0^{\rm M_{50}} {\rm dM_h\ dn_{ion}/dM_h} = 1/2 \int_0^{\infty} {\rm dM_h\ dn_{ion} dM_h}$. We see that over half of the ionizing photons are sourced by galaxies that are below current detection limits.

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 5–10, depending on the bandpower (Abazajian et al. 2016). Improved foreground modelling should also help reduce the uncertainty on the kSZ amplitude by roughly 30 per cent (Gorce et al. 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 Section 4.2.1, but using a mock future kSZ measurement instead of Reichardt et al. (2021). We assume $\mathcal{D}^{\rm kSZ}_{3000} = 2.0 \pm 0.10\,\mu{\rm K}^2$. 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.

In Fig. 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 rms 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: $\Delta_z \equiv z(\bar{x}_{\rm H} = 0.75) - z(\bar{x}_{\rm H} = 0.25)$. 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.

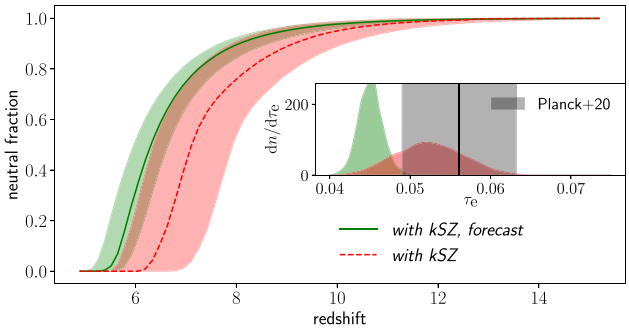

Figure 5. Median EoR histories with 95 per cent CI for the *with kSZ* (*red*) posterior and new posterior with the forecast value of patchy kSZ amplitude: $\mathcal{D}^{\rm kSZ}_{3000} = 2.0 \pm 0.10\,\mu{\rm K}^2$ (*green*). In the inset we also show the corresponding PDFs of $\tau_{\rm e}$.

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 ∼30 per cent CI inferred from *Planck* alone. Interestingly this later EoR history is in very good agreement with the latest independent estimates coming from the Ly α forest opacity fluctuations which imply that EoR finishes at $z \approx 5.3$ (Qin et al., in preparation).

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. Zahn et al. 2012; Battaglia et al. 2013; Reichardt et al.

[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 ∼10 per cent to the EoR (e.g. Qin et al. 2017; Harikane et al. 2023a), 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.

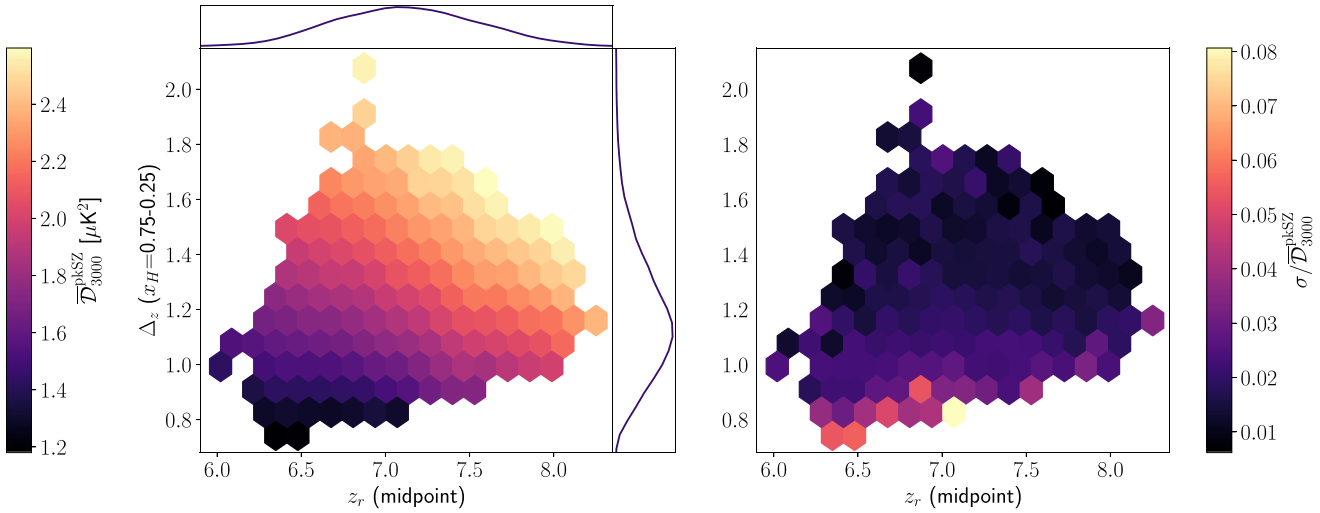

Figure 6. Left panel: Mean value of the patchy kSZ power spectrum amplitude at $l = 3000$ binned as a function of z_r (midpoint of reionization) and Δ_z (duration of reionization; $\Delta_z \equiv z(\bar{x}_{\rm H} = 0.75) - z(\bar{x}_{\rm H} = 0.25)$). The samples are taken from the *with kSZ* run. 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.

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* (Mesinger et al. 2012; Battaglia et al. 2013; Gorce et al. 2020; Paul et al. 2021; Choudhury et al. 2021a; Chen et al. 2023); 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 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 4.2, computing for each sample the midpoint of the EoR, z_r, and its duration, Δ_z. In Fig. 6 we plot the mean (*left panel*) $\bar{\mathcal{D}}_{3000}^{\rm pkSZ}$ and normalized rms (*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 per cent higher compared to some recent estimates, for a given combination of z_r and Δ_z (Chen et al. 2023; Gorce 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 (1) to result in a $\sim 0.2\,\mu{\rm K}^2$ difference in the patchy kSZ amplitude.

[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 & Mesinger 2017; Park et al. 2019; The HERA Collaboration et al. 2022).

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. Ma & Fry 2002; Gorce et al. 2020; Choudhury et al. 2021a; Paul et al. 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 per cent larger than Battaglia et al. (2013) and Chen et al. (2023). 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 per cent CI). 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 Reichardt et al. (2021) who found $\Delta_z = 1.1^{+1.6}_{-0.7}$ (68 per cent CI) and Choudhury et al. (2021a) who found $\Delta_z = 1.30^{+0.19}_{-0.60}$. Our limits are however $\sim$3 times tighter compared to Reichardt et al. (2021) since we use additional complementary observations in the likelihood.

In the right panel of Fig. 6 we quantify the scatter in the $l = 3000$ patchy kSZ power, *at fixed values of* $z_{\rm r}$ *and* Δ_z. We see that the rms scatter in the power is generally at the level of a few per cent. 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 per cent. 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 et al. (2021) find a larger rms scatter, of order $\sim 0.4\mu{\rm K}^2$ 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.

6 CONCLUSIONS

The patchy kSZ signal is an integral probe of the timing and morphology of the EoR. Recently, Reichardt et al. (2021) have claimed a detection of the patchy kSZ signal ($\mathcal{D}_{3000}^{\rm pkSZ} = 1.1^{+1.0}_{-0.7}\,\mu{\rm K}^2$). 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) Ly α 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 et al. [2021b; using data sets (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 favour of faster and later reionization models (Fig. 2). This results in a lower optical depth to the CMB: $\tau_e = 0.052^{+0.009}_{-0.008}$ (68 per cent CI).

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 per cent. We disfavour 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 per cent CI), consistent with other recent results (Choudhury et al. 2021a; Reichardt et al. 2021). We show that the scatter in patchy kSZ power at $l = 3000$, *at a fixed* $z_{\rm r}$ *and* Δ_z, is of order $\sim$ few per cent. 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 error reduced to $0.1\mu K^2$, centred on the MAP model from our inference. Such a futuristic observation can reduce the uncertainties on the recovered EoR history by factors of $\sim$2, 3. However, if the patchy kSZ power is confirmed to be low ($\mathcal{D}_{3000}^{\rm pkSZ} \lesssim 2\,\mu{\rm K}^2$), it would result in a mild tension with the CMB τ_e inferred from primary CMB anisotropies.

ACKNOWLEDGEMENTS

We thank C Mason and A Ferrara for insightful comments on a draft version of this manuscript. We gratefully acknowledge computational resources of the Center for High Performance Computing (CHPC) at Scuola Normale Superiore (SNS). AM acknowledges support from the Ministry of Universities and Research (MUR) through the PNRR project 'Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing' and the PRO3 project 'Data Science methods for Multi-Messenger Astrophysics and Cosmology'. YQ acknowledges that part of this work was supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project #CE170100013. AG's work is supported by the McGill Astrophysics Fellowship funded by the Trottier Chair in Astrophysics, as well as the Canadian Institute for Advanced Research (⟨0:funding-source 3:href="http://dx.doi.org/10.13039/100007631"⟩CIFAR⟨/0:funding-source⟩) Azrieli Global Scholars program and the Canada 150 Programme.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. The code will be made public by merging into the main $21cmFAST$ branch, after the manuscript is accepted for publication.

REFERENCES

Abazajian K. N. et al., 2016, preprint (arXiv:1610.02743)
Abitbol M. H. et al., 2017, preprint (arXiv:1706.02464)
Ade P. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 56
Alvarez M. A., 2016, ApJ, 824, 118
Barlow R., 2004, preprint (arXiv:physics/0406120)
Battaglia N., Natarajan A., Trac H., Cen R., Loeb A., 2013, ApJ, 776, 83
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Bégin J.-M., Liu A., Gorce A., 2022, Phys. Rev. D, 105, 83503
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Bhatawdekar R., Conselice C. J., Margalef-Bentabol B., Duncan K., 2019, MNRAS, 486, 3805
Bolton J. S., Becker G. D., Wyithe J. S. B., Haehnelt M. G., Sargent W. L. W., 2010, MNRAS, 406, 612
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055
Bosman S. E. I. et al., 2022, MNRAS, 514, 55
Bouwens R. J. et al., 2015, ApJ, 803, 34
Bouwens R. J. et al., 2016, ApJ, 830, 67
Bouwens R. J., Illingworth G. D., van Dokkum P. G., Oesch P. A., Stefanon M., Ribeiro B., 2022, ApJ, 927, 81
Bouwens R. J. et al., 2023, MNRAS, 523, 1036
Calabrese E. et al., 2014, J. Cosmol. Astropart. Phys., 2014, 10
Chen N., Trac H., Mukherjee S., Cen R., 2023, ApJ, 943, 138
Choudhury T. R., Ferrara A., 2006, preprint (astro–ph/0603149)
Choudhury T. R., Mukherjee S., Paul S., 2021a, MNRAS, 501, L7
Choudhury T. R., Paranjape A., Bosman S. E. I., 2021b, MNRAS, 501, 5782
Crawford T. M. et al., 2014, ApJ, 784, 143
D'Aloisio A., McQuinn M., Maupin O., Davies F. B., Trac H., Fuller S., Upton Sanderbeck P. R., 2019, ApJ, 874, 154
D'Odorico V. et al., 2023, MNRAS, 523, 1399
Das S. et al., 2011, ApJ, 729, 62
Das S. et al., 2014, J. Cosmol. Astropart. Phys., 2014, 14
Dunkley J. et al., 2013, J. Cosmol. Astropart. Phys., 2013, 25
Efstathiou G., 2021, MNRAS, 505, 3866
Endsley R. et al., 2022, MNRAS, 517, 5642
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Finkelstein S. L. et al., 2015, ApJ, 810, 71
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1

Gaikwad P. et al., 2020, MNRAS, 494, 5091
George E. M. et al., 2015, ApJ, 799, 177
Gorce A., Douspis M., Aghanim N., Langer M., 2018, A&A, 616, A113
Gorce A., Ilić S., Douspis M., Aubert D., Langer M., 2020, A&A, 640, A90
Gorce A., Douspis M., Salvati L., 2022, A&A, 662, A122
Greig B., Mesinger A., 2015, MNRAS, 449, 4246
Greig B., Mesinger A., 2017, MNRAS, 465, 4838
Greig B., Mesinger A., 2018, MNRAS, 477, 3217
Harikane Y. et al., 2023a, preprint (arXiv:2303.11946)
Harikane Y. et al., 2023b, ApJS, 265, 5
Heinrich C., Hu W., 2021, Phys. Rev. D, 104, 63505
Hoag A. et al., 2019, ApJ, 878, 12
Hutter A., Dayal P., Yepes G., Gottlöber S., Legrand L., Ucci G., 2021, MNRAS, 503, 3698
Iliev I. T., Pen U.-L., Bond J. R., Mellema G., Shapiro P. R., 2007, ApJ, 660, 933
Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725
Iršič V. et al., 2017, MNRAS, 466, 4332
Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2018, ApJ, 854, 73
Jin X. et al., 2023, ApJ, 942, 59
Kaur H. D., Gillet N., Mesinger A., 2020, MNRAS, 495, 2354
Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 466, 4826
Konno A. et al., 2018, PASJ, 70, S16
Leonova E. et al., 2022, MNRAS, 515, 5790
Lewis J. S. W. et al., 2020, MNRAS, 496, 4342
Lueker M. et al., 2010, ApJ, 719, 1045
McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499
McQuinn M., Furlanetto S. R., Hernquist L., Zahn O., Zaldarriaga M., 2005, ApJ, 630, 643
Ma C.-P., Fry J. N., 2002, Phys. Rev. Lett., 88, 211301
Ma X., Kasen D., Hopkins P. F., Faucher-Giguère C.-A., Quataert E., Kereš D., Murray N., 2015, MNRAS, 453, 960
Ma X. et al., 2018, MNRAS, 478, 1694
Ma X., Quataert E., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2020, MNRAS, 498, 2001
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Mason C. A. et al., 2019, MNRAS, 485, 3947
Mertens F. G. et al., 2020, MNRAS, 493, 1662
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Mesinger A., McQuinn M., Spergel D. N., 2012, MNRAS, 422, 1403
Murray S., Greig B., Mesinger A., Muñoz J., Qin Y., Park J., Watkinson C., 2020, J. Open Source Softw., 5, 2582
Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 250
Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, ApJ, 855, 105
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544
Pallottini A. et al., 2022, MNRAS, 513, 5621
Park H., Shapiro P. R., Komatsu E., Iliev I. T., Ahn K., Mellema G., 2013, ApJ, 769, 93
Park J., Mesinger A., Greig B., Gillet N., 2019, MNRAS, 484, 933
Park J., Gillet N., Mesinger A., Greig B., 2020, MNRAS, 491, 3891
Park J., Greig B., Mesinger A., 2022, MNRAS, 517, 192
Paul S., Mukherjee S., Choudhury T. R., 2021, MNRAS, 500, 232
Planck Collaboration, 2016, A&A, 596, A108
Planck Collaboration, 2020, A&A, 641, A6
Qin Y. et al., 2017, MNRAS, 472, 2009
Qin Y., Poulin V., Mesinger A., Greig B., Murray S., Park J., 2020, MNRAS, 499, 550
Qin Y., Mesinger A., Greig B., Park J., 2021a, MNRAS, 501, 4748
Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021b, MNRAS, 506, 2390
Reichardt C. L. et al., 2012, ApJ, 755, 70
Reichardt C. L. et al., 2021, ApJ, 908, 199
Scoccimarro R., 1998, MNRAS, 299, 1097
Sehgal N. et al., 2019, in Bulletin of the American Astronomical Society. p. 6, preprint (arXiv:1906.10134)
Seiler J., Hutter A., Sinha M., Croton D., 2019, MNRAS, 487, 5739
Shaw L. D., Rudd D. H., Nagai D., 2012, ApJ, 756, 15
Shirokoff E. et al., 2011, ApJ, 736, 61
Sobacchi E., Mesinger A., 2013, MNRAS, 432, L51
Sobacchi E., Mesinger A., 2014, MNRAS, 440, 1662
Stanway E. R., Chrimes A. A., Eldridge J. J., Stevance H. F., 2020, MNRAS, 495, 4605
Stark D. P., Ellis R. S., Chiu K., Ouchi M., Bunker A., 2010, MNRAS, 408, 1628
Stark D. P. et al., 2017, MNRAS, 464, 469
Stefanon M., Bouwens R. J., Labbé I., Illingworth G. D., Gonzalez V., Oesch P. A., 2021, ApJ, 922, 29
Sun G., Furlanetto S. R., 2016, MNRAS, 460, 417
Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, ApJ, 868, 92
The HERA Collaboration, 2022, ApJ, 924, 51
The HERA Collaboration, 2023, ApJ, 945, 124
Trott C. M. et al., 2020, MNRAS, 493, 4711
Trotta R., 2017, preprint (arXiv:1701.01467)
Wilkins S. M., Lovell C. C., Stanway E. R., 2019, MNRAS, 490, 5359
Wyithe J. S. B., Loeb A., 2013, MNRAS, 428, 2741
Xu H., Wise J. H., Norman M. L., Ahn K., O'Shea B. W., 2016, ApJ, 833, 84
Yeh J. Y. C. et al., 2023, MNRAS, 520, 2757
Yung L. Y. A., Somerville R. S., Popping G., Finkelstein S. L., Ferguson H. C., Davé R., 2019, MNRAS, 490, 2855
Zahn O., Zaldarriaga M., Hernquist L., McQuinn M., 2005, ApJ, 630, 657
Zahn O. et al., 2012, ApJ, 756, 65

APPENDIX A: CALIBRATING SIMULATIONS TO ACCOUNT FOR MISSING LARGE-SCALE KSZ POWER

As discussed in Section 2.1, large boxes are required to accurately simulate the patchy kSZ signal. Shaw et al. (2012) find that a simulation box of side length $100\,h^{-1}$Mpc would miss about 60 per cent 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 (Park et al. 2013; Gorce 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 et al. 2007).

We pick a random sample from the posterior distribution of Qin et al. (2021b), 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. Mesinger et al. 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 $\mathcal{D}_{3000}^{\rm pkSZ}$ from the 500 Mpc simulations is shown in Fig. A1, together with the value from the 1.5 Gpc simulation (blue vertical line). We compute the ratio of the missing power as $f_{0.5\rm Gpc}^{1.5\rm Gpc} = \mathcal{D}_{3000}^{\rm pkSZ-1.5Gpc}/\mathcal{D}_{3000}^{\rm pkSZ-500Mpc}$. Using these 20 realizations, we find $f_{0.5\rm Gpc}^{1.5\rm Gpc} = 1.27 \pm 0.19$. We include this scaling factor and associated uncertainty in the likelihood when

Table A1. Astrophysical parameters used for the scaling test.

	$\log_{10}(f_*)$	α_*	$\log_{10}(f_{\rm esc})$	$\alpha_{\rm esc}$	$\log_{10}(M_{\rm turn}/{\rm M}_\odot)$	t_*	$\mathcal{D}_{3000}^{\rm pkSZ}[\mu{\rm K}^2]$	$f_{0.5{\rm Gpc}}^{1.5{\rm Gpc}}$
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

$f_{0.5{\rm Gpc}}^{1.5{\rm Gpc}}$ is the power spectrum scaling factor (see text). Note that the parameter combination 2 is used in Fig. 1.

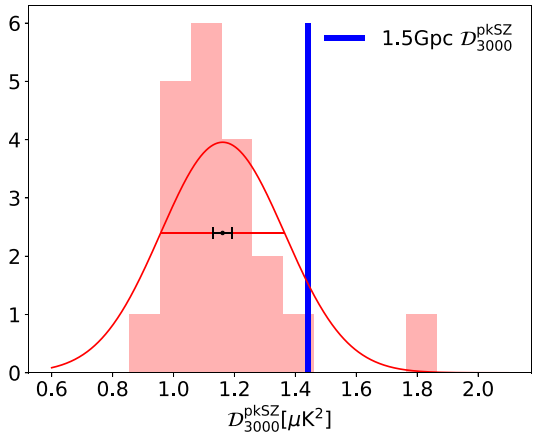

Figure A1. Histogram of the $l = 3000$ patchy kSZ amplitudes generated from 500 Mpc boxes by varying the cosmic initial seed (see text for details). The solid red line corresponds to a Gaussian fit to the histogram, with the standard error σ denoted as a solid line. The vertical blue line denotes the value for the 1.5 Gpc box. We see that the small boxes on average underestimate the kSZ power at $l = 3000$ by ~ 20 per cent, though with sizable scatter. For illustrative purposes, we also demarcate with the black segment the size of the 1σ 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.

performing inference,

$$\ln \mathcal{L}_{\rm kSZ} = -\frac{1}{2}\left(\frac{\mathcal{D}_{3000}^{\rm kSZ,SPT} - \mathcal{D}_{3000}^{\rm kSZ,model}}{\sigma_a + \sigma_b\left(\mathcal{D}_{3000}^{\rm kSZ,SPT} - \mathcal{D}_{3000}^{\rm kSZ,mock}\right)}\right)^2, \tag{A1}$$

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, $\sigma_{\rm u}$ and $\sigma_{\rm l}$ are upper and lower 68 per cent CI limits of the measurement. Since we are adding the scaling factor uncertainty in the quadrature, the final expressions for $\sigma_{\rm u}$ and $\sigma_{\rm l}$ are $\sigma_u = \sqrt{\sigma_{u,SPT}^2 + \sigma_{f_{0.5{\rm Gpc}}}^2}$ and $\sigma_l = \sqrt{\sigma_{l,SPT}^2 + \sigma_{f_{0.5{\rm Gpc}}}^2}$, where the measurement is expressed as $(\mathcal{D}_{3000}^{\rm kSZ,SPT})_{-\sigma_{l,SPT}}^{+\sigma_{u,SPT}} = 1.1_{-0.7}^{+1.0}\mu{\rm K}^2$. Log-likelihood written in equation (A1) 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 $f_{0.5{\rm Gpc}}^{1.5{\rm Gpc}}$ 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. A1, and has a subdominant contribution to the scatter in $\mathcal{D}_{3000}^{\rm pkSZ}$ from the 500 Mpc simulations.

How much does the scaling factor, $f_{0.5{\rm Gpc}}^{1.5{\rm Gpc}}$, 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 $f_{0.5{\rm Gpc}}^{1.5{\rm Gpc}}$ 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 A1. Reassuringly the scaling factors vary by only ~ 3 per cent between the four parameter combinations. This is much smaller than the $\sim 0.7\mu{\rm K}^2$ measurement uncertainty, justifying our assumption of a constant $f_{0.5{\rm Gpc}}^{1.5{\rm Gpc}}$.

APPENDIX B: LIKELIHOODS USED FOR INFERENCE

In Section 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 4.2:

(i) Ly α forest opacity distributions: The log-likelihood for one redshift, z, and effective optical depth, $\tau_{\rm eff}$, bin is assumed to be Gaussian,

$$\ln \mathcal{L}_{\alpha,z,\tau_{\rm eff}}(\theta) = -0.5X^{\rm T}\Sigma^{-1}X, \tag{B1}$$

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 et al. (2021b). The total log-likelihood for the Ly α forest is the sum over redshift and optical depth bins,

$$\ln \mathcal{L}_{\alpha}(\theta) = \sum_z \sum_{\tau_{\rm eff}} \ln \mathcal{L}_{\alpha,z,\tau_{\rm eff}}(\theta) \tag{B2}$$

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

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

$$\ln \mathcal{L}_{\rm DF}(\theta) = \begin{cases} 0 & \text{if } \bar{x}_{{\rm HI},z}(\theta) \leq 0.06 \\ -\frac{1}{2}\frac{(\bar{x}_{{\rm HI},z}(\theta) - 0.06)^2}{\sigma_{\rm DF}^2} & \text{otherwise} \end{cases} \tag{B3}$$

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

(iii) High-redshift galaxy UV LFs: The log-likelihood is the sum over redshifts and magnitudes given by Bouwens et al. (2015, 2016) and Oesch et al. (2018),

$$\ln \mathcal{L}_{\rm LF}(\theta) = -0.5\sum_z \sum_{M_{\rm UV}} \left(\frac{\phi_{\rm LF,model}(\theta, z, M_{\rm UV}) - \phi_{\rm LF,obs}(z, M_{\rm UV})}{\sigma_{\rm LF}(z, M_{\rm UV})}\right)^2. \tag{B4}$$

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

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

$$\ln \mathcal{L}_{\tau_e}(\theta) = -\frac{1}{2}\left(\frac{\tau_{e,\mathrm{model}}(\theta) - \tau_{e,\mathrm{obs}}}{\sigma_{\tau_e}}\right)^2, \tag{B5}$$

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

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