Efficient exploration of reionization parameters for the upcoming 21 cm observations using a photon conserving semi-numerical model SCRIPT

Barun Maity1* and Tirthankar Roy Choudhury1
1National Centre for Radio Astrophysics, TIFR, Pune University Campus, Post Bag 3, Pune 411 007, India

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
One of the most promising probes to constrain the reionization history of the universe is the power spectrum of neutral hydrogen 21 cm emission fluctuations. The corresponding analyses require computationally efficient modelling of reionization, usually achieved through semi-numerical simulations. We investigate the capability of one such semi-numerical code, SCRIPT, to constrain the reionization parameters. Our study involves creating a mock data set corresponding to the upcoming SKA-Low, followed by a Bayesian inference method to constrain the model parameters. In particular, we explore in detail whether the inferred parameters are unbiased with respect to the inputs used for the mock, and also if the inferences are insensitive to the resolution of the simulation. We find that the model is successful on both fronts. We also develop a simple template model of reionization which can mimic the complex physical processes like inhomogeneous recombinations and radiative feedback and show that it can recover the global reionization history reliably with moderate computational cost. However, such simple models are not suitable for constraining the properties of the ionizing sources. Our results are relevant for constraining reionization using high-quality data expected in future telescopes.

Key words: intergalactic medium – cosmology: theory – dark ages, reionization, first stars – large-scale structure of Universe

1 INTRODUCTION
The epoch of reionization is that phase of the history of our universe where the universe consists mainly of hydrogen. At the moment, the Universe is approximately half ionized, a transition that is expected to have occurred by the time the first luminous sources (for reviews, see Barkana & Loeb 2001; Choudhury 2009; Dayal & Ferrara 2018; Gnedin & Madau 2000; Choudhury 2009; Dayal & Ferrara 2018; Gnedin & Madau 2000) which is the epoch when the universe transits from a predominantly neutral to the mostly ionized state via the ionizing photons coming from the very first luminous sources (for reviews, see Barkana & Loeb 2001; Choudhury 2009; Dayal & Ferrara 2018; Gnedin & Madau 2000; Choudhury 2009; Dayal & Ferrara 2018; Gnedin & Madau 2000). Neither the exact timeline of the process of reionization is understood, nor do we have a good understanding of the nature of the first sources that reionized the universe.

The key component of the universe consists mainly of hydrogen. So, the redshifted 21 cm signal originating from the spin flip transition at the ground state of the neutral hydrogen atom is one of the most promising probes of reionization. In particular, the radio interferometric observations that are able to track the fluctuations in neutral hydrogen field should provide detailed information about the state of the intergalactic medium (IGM) during reionization. As the 21 cm signal is very faint, it is very hard to detect it with the sensitivities of the present observatories. Nevertheless, considerable progress has been achieved with the current interferometric observations, which have started to provide upper limits on the power spectrum of the cosmological 21 cm signal. These interferometers include Low Frequency Array (LOFAR; Gehlot et al. 2019; Mertens et al. 2020), Murchison Widefield Array (MWA; Barry et al. 2019; Trott et al. 2020), Giant Metrewave Radio Telescope (GMRT; Paciga et al. 2013) and Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010). It is expected that the future telescopes like Square Kilometre Array (SKA-Low; Koopmans et al. 2015) and Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017; Abdurashidova et al. 2022), will be able to pin down large scale power spectrum within ~ 10% uncertainties in about a few hundred hours of observations.

To correctly interpret these data, we need reliable modelling of the physics of reionization. There exist a variety of models, starting from the simple analytical ones to the very complex radiative simulations. The very early models of reionization would be rather simplistic, where the ionization field could be generated by approximating the ionized regions as spherical bubbles (Furlanetto et al. 2004; Bharadwaj & Ali 2005; Datta et al. 2007; Shin et al. 2008). These models were useful to provide initial insights on the overall characteristics of the 21 cm power spectrum. On the other extreme in terms of complexities and computational requirements, full radiation hydrodynamic simulations have been developed by several groups (Mellema et al. 2006; Ilic et al. 2006; Trac & Cen 2007; Ghara et al. 2015; Ocvirk et al. 2016; Katz et al. 2019; Ocvirk et al. 2020; Kannan et al. 2022; Garaldi et al. 2022; Puchwein et al. 2022) which capture the detailed physics of individual sources and surrounding regions. Although these simulations are useful for capturing the interplay between different physical processes during reionization, one major bottleneck is that they cannot be used to do parameter space exploration because of computational constraints.

The trade-off between accuracy and efficiency can be achieved by the semi numerical approaches. Instead of modelling the full radia-
tive transfer, these models employ some kind of algorithm to count the photons and compare them with the hydrogen distribution to generate the ionization maps (Mesinger & Furlanetto 2007; Geil & Wyithe 2008; Santos et al. 2010; Mesinger et al. 2011; Battaglia et al. 2013). A major advantage of these models is that they can be used to do parameter inference studies using the 21 cm power spectrum as one of the observational probes. In the absence of a real detection of the reionization power spectrum, one usually generates mock data expected from the next generation telescopes and studies the recovery of the model parameters along with the forecast for the errors on them using Bayesian inference techniques (Greig & Mesinger 2015; Hassan et al. 2017; Greig & Mesinger 2017, 2018; Park et al. 2019; Qin et al. 2021a). These calculations require the model to be evaluated numerous times as one samples the parameter space, thus the computational efficiency becomes extremely crucial in such studies. One can bypass the large number of evaluations through machine learning techniques, which also have been employed in forecasting the parameter constraints using the future 21 cm power spectra data (Kern et al. 2017; Shimabukuro & Semelin 2017; Schmit & Pritchard 2018; Doussot et al. 2019; Zhao et al. 2022; Choudhury et al. 2022).

With the recent upper limits on the power spectrum measurements from different telescopes, these semi-numerical models have been exploited extremely to rule out somewhat extreme models of heating and ionization (Pober et al. 2015; Greig & 2016; Ghara et al. 2020; Mondal et al. 2020; Greig et al. 2021a,b; Ghara et al. 2021; Abdurashidova et al. 2022a).

Several of the above models are based on the excursion set method, perhaps the most popular and efficient method of implementing a semi-numerical photon counting algorithm (Mesinger & Furlanetto 2007; Geil & Wyithe 2008; Choudhury et al. 2009; Mesinger et al. 2011). It has been known that the excursion set based models encounter the issue of photon number non-conservation, i.e., the number of ionizing photons produced by the sources is not equal to the number of hydrogen atoms ionized (Zahn et al. 2007, 2011; Paranjape et al. 2016). More recently, Choudhury & Paranjape (2018) have shown that, as a consequence of this photon non-conservation, the large scale 21 cm power spectra depend on the resolution of the simulation box used to generate the ionization fields. This non-convergence of the power spectrum has important consequences for inferring the properties of the IGM from the observational data. In general, one would feel inclined to use coarse resolution simulations while exploring the parameter space, as they require less computational resources. However, if the constraints are found to be dependent on the resolution of the simulations, then one needs to exercise extreme caution in setting up the simulation so that the inferred parameter values are not biased.

A possible solution to the photon non-conservation has been proposed via an explicitly photon conserving algorithm, namely SCRIPT (Choudhury & Paranjape 2018). Although it has been shown that the model naturally produces numerically convergent power spectra at large scales, it still needs to be tested whether the recovered parameters remain unbiased when the resolution is varied. This will be the main aim of the work. Once this is confirmed, we also test if the model is able to efficiently constrain the reionization history when complex processes like the inhomogeneous recombinations and radiative feedback are included. Including these processes makes the code relatively inefficient, hence one needs to devise faster ways to explore the parameter space. To this end, we also develop an approximate template model which can mimic the complex effects through unknown parameters and still provide unbiased estimates of the global ionization fraction. As we will show in the paper, this provides a potential way to constrain the reionization history with relatively moderate computational resources.

The paper is organized as follows: In Section 2, we provide a discussion on the generation of mock power spectra, followed by the formalism for the statistical analysis and parameter space exploration. We then introduce our most basic model of reionization in Section 3 and discuss the recovery of the corresponding model parameters. A more complex model of reionization is discussed in the subsequent Section 4, where we also introduce a simple template reionization model for efficiently sampling the parameter space. Lastly, we summarize our results and discuss the consequences in Section 5. In this paper, the assumed cosmological parameters are $\Omega_m = 0.308, \Omega_\Lambda = 0.691, \Omega_b = 0.0482, h = 0.678, \sigma_8 = 0.829$ and $n_s = 0.961$ (Planck Collaboration et al. 2016).

2 MOCK DATA AND STATISTICAL ANALYSIS

Let us first discuss our method for constructing the mock 21 cm data as appropriate for the upcoming experiments, and the procedure for constraining the model parameters by comparing with the data.

2.1 The mock 21 cm power spectrum and telescope noise

In the absence of actual data which can be analysed to constrain the reionization model parameters, one usually uses the theoretical models to construct the mock 21 cm data. The details of the model used to produce the 21 cm power spectrum will be discussed in the latter sections. In general, any model of reionization would produce the ionized hydrogen fraction $x_{\text{HI}i}$ in grid cells (represented by the index $i$) inside a simulation volume. The differential brightness temperature (assuming spin temperature is very much larger than CMB temperature) is given by (Madau et al. 1997; Ciardi & Madau 2003)

$$\delta T_{b,i} \approx 27 \text{ mK} \left(1 - x_{\text{HI}i}\right) \Delta \left[\frac{1 + z}{10} \frac{0.15}{\Omega_m h^2}\right]^{1/2} \frac{\Omega_b h^2}{0.023},$$

where $\Delta \equiv \bar{\rho}_{m,i}/\bar{\rho}_m$ is the ratio of the matter density $\rho_{m,i}$ in the grid cell and the mean matter density $\bar{\rho}_m$.

The observable we focus on in this paper is the dimensionless 21 cm power spectrum, defined as

$$\Delta^2_{21}(k) = \frac{k^3P_{21}(k)}{2\pi^2},$$

where $P_{21}(k)$ is the power spectrum of the mean-subtracted fluctuation field $\delta T_{b,i} - \left\langle \delta T_{b,i}\right\rangle$.

The dominant contribution to the errors in the 21 cm power spectra comes from the thermal noise of the telescopes. In addition, we also need to account for the cosmic variance arising from surveying only a finite volume of the sky. These uncertainties on the mock 21 cm power spectra are obtained from a modified version of the publicly available package 21cmSense (Pober et al. 2013, 2014)\(^1\). The mathematical details behind the computation of the interferometer sensitivities can be found in Parsons et al. (2012). It can be shown that the dimensionless power spectrum of the thermal noise is given by

$$\Delta^2_N(k) \approx X^2 Y \frac{k^3}{2\pi^2} \frac{\Omega}{2r^3} T^2_{\text{sys}},$$

\(^1\) The modified version of 21cmSense used in this work can be found at https://github.com/paizc001/21cmSense. This version has several new functions and also modules and data files to generate sensitivities for various telescopes.
where \( X^2 - Y \) is the converting factor from bandwidths (frequency) and solid angles to cosmological comoving distances, \( \Omega \) is the primary field of view, \( t \) is the total integration time for Fourier mode \( k \) and \( T_{\text{sys}} \) is the system temperature. After adding the cosmic variance term, the total noise in a bin \( k_{\alpha} \) can be written as

\[
\delta \Delta_{21}^2(k_{\alpha}) = \left( \sum_i \frac{1}{\left[ \Delta_{21}^2(k_i) + \Delta_{21}^2(k_j) \right]^2} \right)^{-\frac{1}{2}},
\]

where \( \Delta_{21}^2(k) \) is the theoretical power spectrum as defined in equation (2), and the sum is over all independent Fourier modes \( j \) that contribute to the particular bin labelled by \( \alpha \).

The next step in creating the mock data is to add the noise and other uncertainties. This requires us to pick a telescope whose specifications would determine the noise. We choose SKA-Low specifications for this study (Braun et al. 2019). The results we obtain would hold qualitatively for other upcoming telescopes, e.g. HERA, as well. For the telescope noise, we take 512 SKA-low stations with sizes of 40 metres in diameter. The system temperature \( T_{\text{sys}} \) is assumed to be 180 K (\( \nu/180 \text{ MHz} \))^{2.5}, where \( \nu \) is the central frequency of observation in MHz units. We compute the noise for observations in a drift scan mode of 6 hours/day for 180 days, which gives a total observing time of around 1080 hours. The bandwidth is taken to be 8 MHz.

We also assume a moderate foreground removal (Pober et al. 2014) where the “foreground wedge” is considered to extend up to wave numbers \( \Delta k_{\parallel} = 0.1h \text{ cMpc}^{-1} \) beyond the horizon limit, \( k_{\parallel} \) being the component of the Fourier mode vector \( k \) along the line of sight. Note that the specifications chosen lead to an extremely low noise for the 21 cm power spectra and correspond to the best quality data we expect in the coming decade or so. Since the aim of our work is to understand the possible biases arising in the recovered parameters because of the assumptions made while modelling, we consider the most optimistic case in terms of error-bars in the model parameters and see if the recovery is within these error-bars. In this sense, this can be considered as possibly the most stringent test for the models.

In this work, we generate the mock data at three redshifts, \( z = 6.2, 7 \) and \( 8 \) which correspond to observational frequencies \( \approx 197, 178 \) and 158 MHz, respectively. These redshifts allow us to probe different representative phases of the reionization history. We use power spectra in \( k \)-bins ranging from 0.102h cMpc\(^{-1} \) to 1.047h cMpc\(^{-1} \), the bins being linearly spaced at intervals of \( \Delta k = 0.035h \text{ cMpc}^{-1} \).

The \( k \)-range considered here are well suited for studying the cosmological signal from reionization, \( k \)-modes smaller than what we consider are dominated by the cosmic variance while the larger ones are dominated by the thermal noise.

To calculate the theoretical power spectrum, we use a simulation box of size 256h\(^{-1}\)cMpc and generate the power spectra using a resolution of \( \Delta x = 2h^{-1}\text{cMpc} \). The box size is sufficiently large to ensure that the effects of missing Fourier modes are not substantial (Iliev et al. 2014; Kaur et al. 2020; Giri et al. 2022) and it also covers the lowest \( k \)-bins we are interested in. The resolution of the simulation is set by a technical limitation of the way we generate the collapsed haloes in the box, see Section 3. However, this does not cause any serious concern in our analysis because the higher \( k \)-modes are heavily dominated by the thermal noise and do not play any significant role in interpreting the cosmological signal. The mock power spectrum used for further analysis is simply the theoretical power spectrum \( \Delta_{21}^2(k_{\alpha}) \) for a fiducial set of parameters plus a random number drawn from a Gaussian distribution having zero mean and standard deviation equal to the thermal noise \( \Delta_{21}^2(k_{\alpha}) \) for the particular \( k \)-bin. This shifting of the data points by adding appropriate random numbers is necessary because we do not explicitly add the thermal noise to the generated 21 cm maps.

In contrast, the scatter due to the cosmic variance is already included in the mock data when we carry out the parameter space exploration. This is because we use a realization of the initial density field for generating the mock power spectra different from the one used for computing the model 21 cm power spectra, see Section 3 for more details.

### 2.2 Likelihood and parameter space exploration

Given the mock data, we employ a Bayesian approach to recover the parameters of our model. The main goal is to compute the conditional probability distribution or the posterior \( P(\theta|D) \) of the model parameters \( \theta \) given the mock data sets \( D \) mentioned in the previous section. This can be computed using the Bayes’ theorem

\[
P(\theta|D) = \frac{P(D|\theta) \pi(\theta)}{P(D)},
\]

where \( P(D|\theta) \) is the conditional probability distribution of data given the parameters or the likelihood, \( \pi(\theta) \) is the prior and \( P(D) \) is the evidence (which can be treated as the normalization parameter and does not play any role in our analysis). The likelihood is assumed to be multidimensional Gaussian which is similar to our earlier study (Maity & Choudhury 2022b)

\[
L(D|\theta) = \exp \left( -\frac{1}{2} \sum_\alpha \frac{\Delta_{21,\alpha}^2(k_{\alpha} - \Delta_{21,\alpha}^2(\theta)^2)}{\delta \Delta_{21}^2(k_{\alpha})} \right)
\]

\[
= \prod_\alpha \exp \left( -\frac{1}{2} \frac{\Delta_{21,\alpha}^2(k_{\alpha} - \Delta_{21,\alpha}^2(\theta)^2)}{\delta \Delta_{21}^2(k_{\alpha})} \right),
\]

where \( \Delta_{21,\alpha}^2(k_{\alpha}; \theta) \) are the model predictions for the parameters \( \theta \), \( \Delta_{21,\alpha}^2(k_{\alpha}) \) are the mock 21 cm data points and \( \delta \Delta_{21}^2(k_{\alpha}) \) are the corresponding error bars on the data computed using equation (4).

The summation index \( \alpha \) runs over all \( k \)-values used in the analysis.

We sample the posterior distribution using the Markov Chain Monte Carlo (MCMC) method, more specifically, the Metropolis-Hastings algorithm (Metropolis et al. 1953). We make use of the publicly available package cobaya (Torrado & Lewis 2021)\(^3\) to run the MCMC chains. The samples are drawn using 20 parallel chains (Lewis & Bridle 2002; Lewis 2013). The chains are assumed to converge when the Gelman-Rubin \( R = 1 \) statistic (Gelman & Rubin 1992) becomes less than a threshold 0.01. This typically needs around 10\(^5\) steps for our model and completes in about 2-3 days for a resolution of \( 2h^{-1}\text{cMpc} \) with a 256h\(^{-1}\text{cMpc} \) box, and in \( 5-6 \) hours when the resolution is coarsened to \( 4h^{-1}\text{cMpc} \). We discard the first 30\% steps in the chains as ‘burn-in’ and work with only the rest.

### 3 ANALYSIS WITH A TWO PARAMETER REIONIZATION MODEL

We now start the discussion on recovering the input model parameters by comparing with the mock data. In this work, we use the explicit

---

\(^2\) The antenna coordinates are taken from the SKA document summarizing the specifications, found at [https://www.skao.int/sites/default/files/documents/d18-SKA-TEL-SKRO-0006422_02_SKA1_LowConfigurationCoordinates-1.pdf](https://www.skao.int/sites/default/files/documents/d18-SKA-TEL-SKRO-0006422_02_SKA1_LowConfigurationCoordinates-1.pdf)

\(^3\) [https://cobaya.readthedocs.io/en/latest/](https://cobaya.readthedocs.io/en/latest/)
photon conserving semi-numerical model SCRIPT (Choudhury & Paranjape 2018) to construct the mock data and for the likelihood analysis. In its simplest and most computationally efficient form, the reionization can be modelled at any given redshift using only two parameters. Generating the ionization field at a given redshift does not require any knowledge of the reionization history at earlier redshifts, hence the requirement of computational resources is minimal; we refer to such models as single-snapshot models. It is possible to extend SCRIPT to include several other physical effects, however that introduces several additional parameters, requires knowledge of the reionization and thermal histories and thus makes the model more computationally expensive. Models which require modelling the full history to compute the ionization maps at a given redshift will be referred to as full-history models.

In this section, we will describe our analysis using the simplest two parameter single-snapshot model and take up more complicated full-history cases in the next section. Let us first describe the main features of the model, for details we refer the reader to Choudhury & Paranjape (2018):

- SCRIPT provides the ionization field as the main output given two input fields at any redshift of interest. These input fields correspond to the dark matter density distribution and collapsed mass fraction in haloes which can produce ionizing photons. In general, these input fields can be got from any N-body dark matter only simulation. In case one is interested only at relatively large scale $\gtrsim 1 h^{-1} \text{cMpc}$ features of the IGM, it is sufficient to use 2LPT formalism to construct the density fields (Hahn & Abel 2011). The collapsed fraction field is computed using subgrid prescription based on conditional ellipsoidal mass function (Sheth & Tormen 2002). The prescription requires mapping the non-linear density field (the Eulerian density) in the simulation volume to the initial linear density field (the Lagrangian density). We use the spherical approximation to carry out the mapping, however, the approximation breaks down for grid cells that are too small. We checked and found that the smallest grid cell we can use corresponds to a length $\Delta x = 2 h^{-1} \text{cMpc}$, which sets the smallest scales we can probe.

- In the most basic version of the code, there are only two free parameters which need to be specified. The first one is the ionizing photon efficiency, $\zeta$, which gives the available number of ionizing photons per unit number of hydrogen atoms. The other one is the minimum threshold mass $M_{\text{min}}$ for haloes that can contribute to the ionizing photon budget. The number density of ionizing photons produced in a grid cell $i$ is given by

$$n_{\text{ion, i}} = \zeta f_{\text{coll}}(M_{\text{min}}) n_{H, i},$$

where $f_{\text{coll}}(M_{\text{min}})$ is the fraction of mass in collapsed haloes with masses $\geq M_{\text{min}}$ and $n_{H, i}$ is the hydrogen number density in the cell. The above expression can be easily generalized to cases where $\zeta$ depends on the halo mass, see Maity & Choudhury (2022a) for such models.

- The photon conserving algorithm consists of two steps. In the first, photons from grid cells containing ionization sources are distributed to neighbouring cells in increasing order of distance until all the photons are exhausted. This process is carried out independently for all source cells, and hence may lead to cells where the ionization fraction exceeds unity. In the second step, which actually consists of a series of substeps, these excess photons in overionized cells are redistributed to nearby cells. The process converges when there are no unphysical overionized cells left in the volume.

The ionization condition of a cell is determined by the number density of ionizing photons available $n_{\text{ion, avail, i}}$ and $n_{H, i}$. The number of available ionizing photons can be written as

$$n_{\text{ion, avail, i}} = \sum_j n_{\text{ion, j} \rightarrow i},$$

where $n_{\text{ion, j} \rightarrow i}$ is the number of photons contributed by the cell $j$ to cell $i$. Clearly, $n_{\text{ion, j} \rightarrow i}$ will be dominated by cell pairs whose distance is small, and will also depend on the density distribution of cells situated between $i$ and $j$. At any step of the algorithm, a cell is assigned fully ionized if

$$n_{\text{ion, avail, i}} \geq n_{H, i},$$

and the excess photons distributed to other cells in the next step. Other cells are assigned an ionized fraction

$$x_{\text{HII, i}} = n_{\text{ion, avail, i}} / n_{H, i}.$$

- Once we obtain the ionization fraction $x_{\text{HII, i}}$ for each cell in the simulation box, the global mass averaged ionization fraction $Q_{\text{HI}} = \langle x_{\text{HII, i}} \rangle$ can be derived, the angular brackets denoting the average over all grid cells in the box. The photon conserving criterion establishes that the total average number of ionizing photons per hydrogen atoms is equal to the global ionization fraction, i.e.,

$$\langle f_{\text{coll}}(M_{\text{min}}) \Delta i \rangle = Q_{\text{HI}}.$$

One important advantage of the photon conserving model is that it produces large scale 21 cm power spectra which are independent of resolutions of the simulation box, a feature we will study next.

### 3.1 Recovery of parameters

We now investigate the recovery of model parameters for the two parameter single-snapshot model. The mock 21 cm power spectra are calculated using fiducial values of the free parameters, namely, $\zeta = 14$ and $M_{\text{min}} = 10^9 M_\odot$. These choices lead to ionization fractions $Q_{\text{HI}} = 0.28, 0.49, 0.75$ at redshifts $z = 8.0, 7.0, 6.2$ respectively.

While sampling the parameter space via the MCMC analysis, we use the same two parameter single-snapshot model, but with an initial density field realization different from the one used to construct the mock data. Let us first take the case where the simulation volume has the same resolution $\Delta x = 2h^{-1} \text{cMpc}$ as the one used to generate the mock data. We study the redshift $z = 7$ first, which corresponds to a $\sim 50\%$ ionized IGM. In Fig. 1, we show the recovered parameter posteriors for this case (blue contours and curves). We find that the input parameters ($\zeta = 14$ and $M_{\text{min}} = 10^9 M_\odot$), shown by dashed lines in the figure, are well within the $1\sigma$ confidence levels of the posterior distribution. We also derive the distribution of $Q_{\text{HI}}$ and find that it peaks around the input value of $\approx 0.5$. The degeneracies between the parameters are straightforward to understand: $\zeta$ and $M_{\text{min}}$ are positively correlated as higher values of $\zeta$ lead to larger number of ionizing photons per halo, and that can be compensated by increasing $M_{\text{min}}$ so that the number of haloes decrease (Greig & Mesinger 2015; Mondal et al. 2022). The degeneracies between $Q_{\text{HI}}$ and the two free parameters too can be understood from similar arguments. In the left-hand panel of Fig. 2, we show the plots of power spectra for 200 random samples from our MCMC chains (blue curves), along with the best-fit and the input models. The points with error-bars are the mock data. The match between the recovered power spectrum and the mock data is quite good. The values of the recovered parameters along with the statistical errors are listed in Table 1. It is clear that the input values are always within the $1\sigma$ of the recovered parameters. For completeness, we also provide the values of $\chi^2 / \nu$ for the best-fit model in the table, $\nu$ being the number of degrees of freedom. The values confirm that the match between the recovered best-fit model and the mock data is excellent.
Let us now check the recovery when the analysis is carried out with a simulation volume of coarser resolution $\Delta x = 4h^{-1}\text{cMpc}$. Note that the mock data remains identical as before, i.e., generated using a finer resolution $\Delta x = 2h^{-1}\text{cMpc}$. This analysis requires less grid cells (for the same simulation volume) and hence is much more efficient than the previous one. The price one has to pay is that the data at $k \geq 0.5h/\text{cMpc}$ cannot be used. This may not be a serious handicap as the mock data at most of the high-$k$ bins have relatively larger error-bars and hence would not contribute significantly to the likelihood.

As can be seen from Fig. 1 (red contours and curves), the recoveries for this coarser resolution are almost identical to those obtained using the finer resolution. The corresponding limits on the parameters can be found Table 1, see the two right-most columns. The constraints for the coarser resolution case are slightly weaker than the finer one as we are not able to use the data points at very small-scales (large $k$ modes), e.g., the errors on $Q_{\text{HI}}^M$ are $\sim 4\%$ for the coarser resolution compared to $\sim 2.5\%$ for the finer resolution.

This analysis confirms the numerical convergence of our photon conserving algorithm with respect to the grid size used. In particular, we confirm that using a coarser resolution does not lead to any bias in the inferred values of the parameters. This result is significant because it allows us to obtain unbiased constraints on the parameters using relatively coarser resolution and thus moderate computing requirements. In an earlier work, Choudhury & Paranjape (2018) have shown that the non-convergence of the large-scale power spectrum with respect to the resolution is directly related to non-conservation of photons in semi-numerical models of reionization. Because the model used above is photon conserving by construction, the convergence is not surprising. However, for photon non-conserving models, e.g., those based on excursion set formalism, this convergence is not guaranteed. We study this in detail in Appendix A and quantify the bias in the recovered parameters for coarser resolution maps in excursion set-based models.

The next obvious step is to extend the analysis to other redshifts with different global ionization fractions. The posterior distributions and the corresponding sampled power spectra for a lower redshift ($z = 6.2$) with a higher ionization fraction ($Q_{\text{HI}}^M \approx 0.75$) are shown in Figs. 3 and 4 respectively, the corresponding parameter values and statistical errors are quoted in Table 1. As seen from the plots, the recovery is quite good, and that too for both the resolutions, thus confirming the generic nature of the numerical convergence of our photon conserving algorithm. Lastly, we show the results for a higher redshift ($z = 8.0$), i.e., a lower ionization state of the universe ($Q_{\text{HI}}^M \approx 0.28$) in Figs. 5 and 6 respectively, see Table 1 for the parameter values and errors. As with the other cases, the recovered parameters match the corresponding input values within the $1\sigma$ for almost all the parameters. The uncertainties on the parameters are relatively larger when the resolution is coarser. This is expected as only the large-scale models are available for the likelihood calculations, and the power spectrum is rather featureless at these scales.

### 4 MODEL WITH RECOMBINATION AND FEEDBACK

The model presented so far is very basic in nature, which does not include several physical processes related to reionization. In this section, we check the possibility of recovering the reionization parameters when the model consists of more complex processes.
As is well known, reionization is associated with other inhomogeneous astrophysical processes like the recombination of free electrons and ionized atoms, and the radiative feedback on star forming haloes due to heating of the medium. Recently, we have extended our model to include these inhomogeneous effects, see Maity & Choudhury (2022a,b). The details of the model can be found in these two papers, we summarize the main features of the model here:

- The temperature of each grid cell in the simulation volume is computed using the appropriate evolution equation. We account for processes like the expansion cooling, adiabatic heating/cooling from evolution of the densities, photoheating and Compton cooling. The calculation requires introduction of a free parameter, namely, the reionization temperature increment $T_{\text{re}}$. This parameter quantifies the increase in temperature when a region transitions from being fully neutral to a fully ionized state, i.e., it is the temperature of the region immediately after the ionization. The calculation of the photoheating rate requires knowledge of $T_{\text{re}}$.
- The number density of recombinations $n_{\text{rec},i}$ in each cell is computed self-consistently by tracking the ionization history of that cell. We account for the subgrid clumping of the IGM by introducing another free parameter $C_{\text{HI}}$, which is the globally averaged clumping factor. In the presence of recombinations, the condition for assigning a cell to be ionized, equation (9), is modified to

$$n_{\text{ion ava}, i} \geq n_{H, i} + n_{\text{rec}, i}.$$ (10)
the rest of the algorithm for generating ionization maps remaining unchanged.

- We implement radiative feedback suppressing star formation in low mass haloes using a Jeans mass-based prescription. The prescription relies on the value of the temperature at each grid cell, and hence the amount of feedback too varies from cell to cell. So, the minimum halo mass for a cell (say \(i\)) is given by \(M_{\min,i} = \max[M_{\text{cool}}, M_{\text{Jeans,i}}]\), where \(M_{\text{cool}}\) is the minimum mass of haloes that can cool via atomic transitions and \(M_{\text{Jeans,i}}\) is the Jeans mass at virial overdensity. The Jeans mass depends upon the temperature of the region \((\sim T^{3/2})\) and is higher than \(M_{\text{cool}}\) in the ionized regions. It is thus clear that the \(M_{\min}\) is determined by the atomic cooling in the neutral regions and by Jeans mass in the feedback affected ionized regions. The ionizing photon production rate of a cell then can be computed by summing over weighted contribution from neutral and ionized regions within a cell. Note that the minimum mass of haloes that can produce ionizing photons is not a free parameter any more, it is given by the atomic cooling condition in neutral regions and by the radiative feedback in the ionized regions.
- It must be emphasized that the ionization map at a given redshift depends on the thermal and ionization history of the IGM. This makes the model very different from its basic form of Section 3. While the maps for the two parameter single-snapshot model at a given redshift can be generated without any knowledge of the history of that cell, that is not possible for this extended full-history model any more. Consequently, the model becomes much slower computationally.

Upon comparing the theoretical model with measurements of the CMB optical depth (Planck Collaboration et al. 2020), the dark pixel fraction at \(z \sim 6\) (McGreer et al. 2015), UV luminosity function of \(z \sim 6\) and 7 galaxies (Bouwens et al. 2015, 2017) and the low-density IGM temperature measurements (Gaikwad et al. 2020), we find that a model with \(\zeta(z) = 8[10/(1+z)]^{2/3}, T_{\text{re}} = 2 \times 10^4\ \text{K}\) and \(C_{\text{HI}} = 3\) is close to the model that best fits the data. The CMB optical depth for this model turns out to be \(\tau_c = 0.053\). These fiducial values are used to construct the mock power spectra data for further analysis in this section.

**4.2 Recovery of parameters using the two parameter single-snapshot model**

Since the mock data has been generated using a rather complex (albeit semi-numerical) model of reionization, it is natural to study the recovery of the parameters using the same model. However, since computing the 21 cm power spectrum at a given redshift requires computing the full history, the model is not computationally efficient, and the parameter space exploration can be quite demanding. We thus take a different approach and try to recover the essential parameters of reionization using a template model which can be used at a given redshift snapshot without solving for the full history. A full parameter space exploration using the detailed model would require some sophisticated interpolation scheme in the parameter space (e.g., using emulators), which we postpone for a future work.

---

**Figure 5.** Same as Fig. 1 but for redshift \(z = 8\).

**Figure 6.** Same as Fig. 2 but for redshift \(z = 8\).

**Figure 7.** Recovery of parameters for the case where the mock 21 cm power spectrum data is generated using the full-history model (consisting of radiative feedback and inhomogeneous recombinations), while the one used for the parameter space exploration is the simple two parameter single-snapshot model. The descriptions of the different panels are the same as in Fig. 1. Although the recovery of the input ionized fraction \(Q_{\text{HI}}^{\text{in}}\) is unbiased when the resolution used for parameter constraints is equal to that used for the mock, the recovery fails for the coarser resolution (red contours and curves).
The first template model we try is our simple two parameter model of Section 3. Clearly, this is a very simplistic reionization model where we have neglected the inhomogeneities in the recombinations and the effect of radiative feedback on the small mass haloes. In particular, the two parameters of the model $\zeta$ and $M_{\text{min}}$ do not have any obvious counterparts in the detailed model. Hence, the constraints obtained on these parameters cannot be compared with any “input” values. What we can check is that whether this two parameter model can recover the value of $Q_{\text{HI}}^{M}$ as implied by power spectra generated using the detailed model.

We start with redshift $z = 7.0$ where the global ionization fraction $Q_{\text{HI}}^{M} \approx 0.5$ for the input model. In Fig. 7, the blue regions and curves show the parameter recoveries using the resolution $\Delta x = 2h^{-1}$ cMpc, same as what was used to generate the mock data. As is clear from the figure, the recovered posterior distribution of $Q_{\text{HI}}^{M}$ matches the input $Q_{\text{HI}}^{M}$ surprisingly well. This indicates that the 21 cm power spectrum can be used for recovering the global ionization fraction even when the physical model uses simplistic descriptions of the underlying physical processes.

We next study the parameter recovery for the coarser resolution $\Delta x = 4h^{-1}$ cMpc, see the regions and curves in red in Fig. 7. Interestingly, we find a bimodality in the posterior distributions of the parameters for this case. In particular, the posterior of $Q_{\text{HI}}^{M}$, our main quantity of interest, cannot be constrained reliably because of the bimodality. One of the peaks of the bimodal distribution is around the input value, however, the other peak is at a much lower value $Q_{\text{HI}}^{M} = 0.2$. The strengths of both the peaks are almost similar, as we have checked from the values of the $\chi^2$. The reason for this bimodality is as follows: since the coarse resolution maps probe only the larger scales and also since the 21 cm power spectrum at large scales is non-monotonic (see Figure 10 of Maity & Choudhury 2022a), the model ends us producing very similar power spectra at two widely different stages of reionization.

We thus conclude that our two parameter single-snapshot model cannot be used reliably to describe the full-history model at different resolutions, as it cannot recover the value of $Q_{\text{HI}}^{M}$ corresponding to the input model. Hence, we do not explore this model any further and move on to a slightly complicated template model which can possibly describe the detailed full-history model.

### 4.3 A four parameter single-snapshot template

It is obvious that the two parameter single-snapshot model cannot recover the global ionization fraction reliably because it is not informed of the effects of inhomogeneous recombination and feedback. To improve on this, we next consider a slightly sophisticated, still single-snapshot, model which can approximate these physical effects without the need to solve for the full ionization history.

To mimic the effect of feedback, we use two characteristic threshold mass instead of a single one. Let the minimum mass of ionizing photon producing haloes be $M_{\text{ne}}$ in the neutral regions, while let it be $M_{\text{io}}$ in the ionized regions. The conservation of photons would then lead to the relation

$$
Q_{\text{HI}}^{M} = \left(1 - Q_{\text{HI}}^{M}\right) \left(\zeta f_{\text{coll},i}(M_{\text{ne}}) \Delta x\right) + Q_{\text{HI}}^{M} \left(\zeta f_{\text{coll},i}(M_{\text{io}}) \Delta x\right) = Q_{\text{HI}}^{M}.
$$

(11)

For given $M_{\text{ne}}, M_{\text{io}}$ and $\zeta$, the above can be solved to obtain $Q_{\text{HI}}^{M}$ without generating the ionization maps. Note that this is applicable only to photon conserving models like ours.

The above equation (11) can be written equivalently as

$$
Q_{\text{HI}}^{M} = \left(1 - Q_{\text{HI}}^{M}\right) \left[\zeta f_{\text{coll},i}(M_{\text{ne}}) \Delta x\right] + \left[\zeta f_{\text{coll},i}(M_{\text{io}}) \Delta x\right].
$$

(12)

If we further write $Q_{\text{HI}}^{M} = \left(\zeta_{\text{eff}} f_{\text{coll},i}(M_{\text{ne}}) \Delta x\right)$, this allows us to identify the effective ionization efficiency as

$$
\zeta_{\text{eff}}(M) = \begin{cases} 
\zeta, & \text{if } M \geq M_{\text{io}}, \\
1 - Q_{\text{HI}}^{M}, & \text{if } M_{\text{ne}} \leq M < M_{\text{io}}, \\
0, & \text{otherwise}. 
\end{cases}
$$

(13)

Thus, the implementation of our feedback prescription is equivalent to introducing a halo mass-dependent efficiency having the form above. This is quite simple to implement in the photon conserving algorithm without compromising on the computing efficiency significantly, e.g., see Maity & Choudhury (2022a) for similar models.

This simple model of the feedback is different from the detailed implementation in several aspects. Firstly, the effect of feedback and hence the effective $M_{\text{io}}$ depends on the thermal history of a cell and thus is not described by a single parameter in the whole box in the detailed model. Further, whether one should use $M_{\text{ne}}$ or $M_{\text{io}}$ in a given cell would depend on whether the cell is already ionized or not. However, in the simple model, we have assumed that a fraction $Q_{\text{HI}}^{M}$ is determined by $M_{\text{io}}$ and rest by $M_{\text{ne}}$ in every cell, irrespective of their ionization state. This needs to be done because we do not know beforehand the ionization state of a cell without generating the ionization map.

The other important effect during reionization is the recombination between ionized atoms and free electrons, which is also intrinsically inhomogeneous in nature. Ideally, one needs to track the density evolution and ionization history of each cell in the simulation box to model this inhomogeneous effect (Sobacchi & Mesinger 2013; Maity & Choudhury 2022a). But, this full evolution can be inefficient for a high resolution box which is needed for 21 cm power spectra computation. So, we follow a mean free path-based approach based on Davies & Furlanetto (2022) where the recombinations can be implemented at a particular redshift without solving for the full ionization evolution.

The implementation requires us to modify equation (8) for computing the available number density of ionizing photons as

$$
n_{\text{ion,avail},i} = \sum_j n_{\text{ion},j-i} \rightarrow \sum_j n_{\text{ion},j-i} \left[\frac{x_{ij}}{\lambda(1 - e^{-x_{ij}/\lambda})}\right]^{-1}
$$

(14)

where $x_{ij}$ is the distance between the cells $i$ and $j$ and $\lambda$ is the mean free path in comoving units. The above modification essentially mimics the loss of photons in a cell from recombination. The effectiveness of the recombinations is characterized by $\lambda$ which is a free parameter in the model. A large $\lambda \rightarrow \infty$ would correspond to the model with no recombinations.

There is one more effect we need to account for, which is the photon absorption inside the source cells where they originate. If we naively apply the above equation, these would correspond to $x_{ij} \rightarrow 0$ and the effect of recombinations would be absent. Instead, we modify the computation of the number of ionizing photons available in the source cells as (Davies & Furlanetto 2016)

$$
n_{\text{ion},j} \rightarrow n_{\text{ion},j} \left[\frac{\epsilon \Delta x}{\lambda(1 - e^{-\epsilon \Delta x/\lambda})}\right]^{-1}
$$

(15)
Mock using full-history, recovery using four parameter single-snapshot, $z=7.0$

Figure 8. Recovery of parameters for the case where the mock 21 cm power spectrum data is generated using the full-history model (consisting of radiative feedback and inhomogeneous recombinations), while the one used for the parameter space exploration is the four parameter single-snapshot model. The off diagonal panels show the joint two-dimensional posterior distributions of each pair of parameters. The contours represent 68% and 95% confidence intervals. The diagonal panels show the marginalized posterior distributions of the parameters. Wherever possible, we show the input values for generating mock data by dashed lines. The blue contours and curves show the results when the resolution used for the MCMC run is $\Delta x = 2\, h^{-1}$ cMpc (same as the one used for the mock data), while the red ones are for the resolution $\Delta x = 4\, h^{-1}$ cMpc. The recovery of the parameters describing the reionization sources are biased with respect to the input, however, the global ionized fraction $Q_{\text{HII}}^M$ is faithfully recovered for both the resolutions.

Figure 9. The 21 cm power spectra for the photon conserving four parameter single-snapshot model at redshift $z = 7$ for 200 random samples drawn from the posterior distributions shown in Fig. 8. The blue lines show the models for a resolution $\Delta x = 2\, h^{-1}$ cMpc while the red ones are for $\Delta x = 4\, h^{-1}$ cMpc. The dashed lines denote the input models.

where $\Delta x$ is the cell size and $\epsilon$ is a fudge factor. The value of $\epsilon$ is tuned so that the results are independent of the resolution used. We find the value to be $\epsilon = 0.3$.

The four-parameter single-snapshot template model thus has four free parameters: $\zeta$, $M_{\text{ne}}$, $M_{\text{io}}$ and $\lambda$. The next step would be to check if the model can be used to recover the reionization history from the 21 cm power spectra.

4.4 Recovery using the four parameter template model

We now present the results when the mock data is generated using the full-history model of Section 4.1 while the Bayesian analysis is carried out using the single-snapshot template introduced above. For the parameters $\zeta$, $M_{\text{ne}}$ and $\lambda$, we choose flat priors over a sufficiently wide range. We also impose a physically motivated condition on $M_{\text{io}}$ that it should be larger than $10^9 M_\odot$ and also that $M_{\text{io}} > M_{\text{ne}}$ (Sobacchi & Mesinger 2013; Hutter et al. 2021).

Let us begin our discussion by showing the results at $z = 7$ (the midpoint of reionization). Fig. 8 shows the posteriors of the model
parameters, the corresponding values are quoted in Table 2. Usually, the quality of parameter recovery is assessed by comparing the recovered values with the inputs. However, in this case, there is no input counterpart of the two parameters \(M_{\text{ne}}\) and \(\lambda\). The input values of the other parameters, including the derived \(Q_{\text{HI}}^M\), are marked by dashed lines in the figure. We can see the recovered values of both \(\zeta\) and \(M_{\text{ne}}\) are lower than the input value for the resolution \(\Delta x = 2h^{-1}\) Mpc corresponding to the one used for generating the mock data (blue curves and regions). This clearly suggests that the simplistic four-parameter single-snapshot model cannot recover the source properties faithfully. The lower recovered \(\zeta\) implies that the recombinations could be underestimated in the simple mean free path approach of the single-snapshot model. A low \(\zeta\) in the model is compensated by a \(M_{\text{ne}}\) higher than the input value used in the mock data. What is interesting is that, in spite of the biased estimates of the source properties, the recovered value of \(Q_{\text{HI}}^M\) is a remarkable match to the input value. This result points towards one utility of the single-snapshot model: it can be used to recover the reionization history faithfully. Fig. 9 shows that the match between the mock power spectra and the recovered ones are also excellent.

The argument can be strengthened by carrying out the analysis with a different resolution \(\Delta x = 4h^{-1}\) Mpc (keeping the mock data same, i.e., generated at a fine resolution of \(\Delta x = 2h^{-1}\) Mpc). The results are shown in Figs. 8 and 9 by the red curves and contours. Firstly, the posteriors of all the parameters are consistent between the two resolutions, thus confirming the numerical convergence of the four parameter single-snapshot model. More importantly, there is no bias in the recovery of \(Q_{\text{HI}}^M\) for this resolution as well.

We can test our postulate that the single-snapshot template model provides a good description of the HI field by comparing the ionization maps visually. The maps are shown in Fig. 10 where we plot the quantity \(\Delta_{\text{HI},i} = (1 - x_{\text{HI},i}) \Delta \lambda_i\). One can see clearly that the maps are very similar. In Fig. 11, we also show the probability distribution function of neutral fraction \(1 - x_{\text{HI}}\) for the full-history model and the four parameter single-snapshot model. It is evident that most of the cells are ionized and rest of the cells are at a low ionized state for both the cases. This is in agreement with the expected trend, i.e., most of the grid cells are at a fully ionized state or at a highly neutral state. However, there also exist a significant fraction of cells that are partially ionized, this is a consequence of our grid cells being relatively coarse. What is important for our work is that the two models produce distributions which agree quite well. This indicates that

| Parameter | Prior | \(\Delta x = 2h^{-1}\) Mpc | \(\Delta x = 4h^{-1}\) Mpc |
|-----------|-------|-----------------|-----------------|
| \(\zeta\) | [2, 100] | 10.19 \(< 18.809\) | 10.420 \(< 32.319\) |
| log \((M_{\text{ne}}/M_\odot)\) | [7, 9] | 8.07 \(< 7.316\) | 7.381 | 7.892 \(< 0.215\) | 7.378 |
| log \((M_{\text{ne}}/M_\odot)\) | [log \((M_{\text{ne}}/M_\odot)\), 11] | - | \(< 10.223\) | 10.970 \(< 9.874\) | 9.247 |
| log \((\lambda/h^{-1}\) Mpc) | [-2.5, 2.5] | 0.394 \(+0.678\) \(-1.261\) | -0.196 | 0.392 \(+0.392\) \(-1.157\) | -0.307 |
| \(Q_{\text{HI}}^M\) | [0.1, 1] | 0.27 | 0.283 \(+0.005\) \(-0.000\) | 0.278 | 0.265 \(+0.010\) \(-0.020\) | 0.259 |
| \(\chi^2/\nu\) | - | - | - | - | - | - |

Table 2. Parameter constraints obtained using the photon conserving four parameter single-snapshot model at redshifts \(z = 7.0\), 8.0 and 6.2. Results are shown for two different resolutions \(\Delta x = 2h^{-1}\) Mpc and \(4h^{-1}\) Mpc used for the MCMC analysis. The mock data is generated using the full-history model which includes radiative feedback and inhomogeneous recombinations, always at the resolution \(\Delta x = 2h^{-1}\) Mpc. For each parameter, we show the prior and the input value used for the mock (wherever possible) along with the obtained mean, 68% confidence limits and best-fit. We also provide the \(\chi^2/\nu\) for the best-fit models, \(\nu\) being the number of degrees of freedom.
the simplistic models can be useful for characterizing the properties of the IGM, even though they are limited in recovering the source properties. Interestingly, using the present limits on the 21 cm power spectrum, it has become possible to put constraints on the properties of the ionized (and hence heated) regions (Ghara et al. 2020); the simple model can be useful in this regard.

Let us next study the recovery of the parameters at other redshifts.

We show the parameter constraints in Table 2 for \( z = 8 \) and 6.2. The recovery of \( Q^M_{\text{H}II} \) can be seen to be reasonable. The worst case seems to be for \( z = 8 \) when the analysis is carried out at the finer resolution. Because of tiny error-bars on the parameter, the input value is outside the 2\( \sigma \) region of the recovered constraints. However, even in this case, the difference in the recovered and input values are within 5 per cent. We can thus hope to recover the reionization history reliably using the single-snapshot template model, the advantage being that the analysis requires very little computational resources. At the same time, we stress that the single-snapshot models are unlikely to be reliable for understanding the properties of the reionization sources, e.g., the ionizing photon efficiency and the mass of the haloes hosting the ionizing sources. This limitation should be kept in mind while using these models to interpret observations.

Although the analysis of this section holds a promise that one can obtain the reionization history from 21 cm maps using a rather simple model of radiative feedback and inhomogeneous recombinations, one should keep in mind that the underlying procedure for generating the ionization maps is almost identical for the mock data and the model used to constrain the parameters. It is thus not clear whether using widely different algorithms for ionization maps would lead to similar conclusions regarding the value of \( Q^M_{\text{H}II} \). In fact, we discuss in Appendix B that two different methods of generating ionization maps, namely excursion set and photon conserving, could lead to very different conclusions on \( Q^M_{\text{H}II} \) even when the source model is exactly the same. At this point, we can only claim that the simple single-snapshot model provides an initial direction on how to obtain the reionization history quickly from the high quality data expected in the future. Another way to appreciate the importance of the analysis is that it is a test case where the theoretical model is only an approximate description of the universe which provided the observational data. Our speculation is that we can recover the reionization history and the IGM properties with the approximate model, but the source
properties require way more detailed modelling and/or comparing with other observations.

5 SUMMARY & CONCLUSIONS

The 21 cm power spectrum of neutral hydrogen from the epoch of reionization is a promising means for constraining the effect of the first stars on the ionization history of the universe. The parameters characterizing the physical processes during this epoch can be constrained by comparing the observational data with theoretical models. In case one attempts to obtain the constraints using MCMC based Bayesian inference techniques, it requires the model to be evaluated many times for different parameter values. As a result, these models need to be computationally efficient. In this work, we check the prospects of constraining the reionization parameters using the photon conserving semi-numerical model SCRIPT. Our main aim is to ensure that the model provides unbiased estimates of the underlying parameters, irrespective of the resolution of the simulation used. For this purpose, we generate and use mock data sets as expected from the upcoming SKA-Low in ~ 1000 hours of observations.

The main results of our work can be summarized as follows:

- For the simplest model where the reionization can be modelled using only two parameters, namely, the ionizing photon efficiency $\xi$ and the minimum threshold halo mass $M_{\text{min}}$ that can contribute to ionizing photons, we find that our model can recover the input parameters extremely well. The conclusion holds for different phases of reionization (and hence different redshifts) and also for different resolutions of the simulation. In particular, the model provides unbiased estimates of the parameters even when the resolution of the simulation used for parameter constraints is different from that used to generate the mock data. We find that the same conclusion does not hold for excursion set based models for generating ionization maps.

- We also explore the capability of our model to recover parameters when the reionization model is taken to be more complex, e.g., by including radiative feedback and inhomogeneous recombinations. This requires us to compute the ionization and thermal histories consistently, and hence the model becomes computationally slower. So, for parameter recovery, we develop a simple template model where these physical processes are approximated by additional parameters. The advantage of this approximate model is that the ionization maps can be computed at a single redshift without any reference to the overall history. In this case, we find that the template single-snapshot model cannot recover the source properties faithfully, which clearly indicates the limitations of such models. However, it manages to recover the global ionization fraction without any bias, over the full reionization history and for different resolutions. This promises that the template can be used to recover the reionization history with moderate computational resources.

The work highlights the necessity of photon number conservation while interpreting the 21 cm data. The large scale convergence with respect to the resolution is important as we do not know about the correct resolution to work with when real observational data will be available. So, it becomes important to take into account the photon number conservation while creating the ionization maps.

There are studies (Park et al. 2019; Qin et al. 2021b) which have shown that the constraints on the reionization parameters improve significantly after including 21 cm data with measurements from other observational probes. So, the next obvious target is to study the joint estimates using the realistic 21 cm mocks and the observational data used in Maity & Choudhury (2022b). These require a more efficient way to compute the models with full ionization history using a high resolution. In a future project, we are planning to build up an emulator which can serve the purpose. These kinds of studies will be very useful before the real observational data become available.

ACKNOWLEDGEMENTS

The authors acknowledge support of the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.02-0700.

DATA AVAILABILITY

A basic version of the code, which does not include the effects of recombinations and feedback on ionization maps, used in the paper is publicly available at https://bitbucket.org/rctirthankar/SCRIPT. The data obtained from the extensions of the code and presented in this article will be shared on reasonable request to the corresponding author (BM).

REFERENCES

Abdurashidova Z., et al., 2022a, ApJ, 924, 51
Abdurashidova Z., et al., 2022b, ApJ, 925, 221
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Barry N., et al., 2019, ApJ, 884, 1
Battaglia N., Trac H., Cen R., Loeb A., 2013, ApJ, 776, 81
Bharadwaj S., Ali S. S., 2005, MNRAS, 356, 1519
Bouwens R. J., et al., 2015, ApJ, 803, 34
Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, ApJ, 843, 129
Braun R., Bonaldi A., Bourke T., Keane E., Wagg J., 2019, arXiv e-prints, p. arXiv:1912.12699
Choudhury T. R., 2009, Current Science, 97, 841
Choudhury T. R., 2022, General Relativity and Gravitation, 54, 102
Choudhury T. R., Paranjape A., 2018, MNRAS, 481, 3821
Choudhury T. R., Haehnelt M. G., Regan J., 2009, MNRAS, 394, 960
Choudhury M., Datta A., Majumdar S., 2022, MNRAS, 512, 5010
Ciardi B., Madau P., 2003, ApJ, 596, 1
Datta K., Bharadwaj S., Choudhury T. R., 2007, MNRAS, 382, 809
Davies F. B., Furlanetto S. R., 2016, MNRAS, 460, 1328
Davies F. B., Furlanetto S. R., 2022, MNRAS, 514, 1302
Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1
DeBoer D. R., et al., 2017, PASP, 129, 045001
Dousset A., Eames E., Semelin B., 2019, MNRAS, 490, 371
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1
Gaikwad P., et al., 2020, MNRAS, 494, 5091
Garaldi E., Kannan R., Smith A., Pakmor R., Vogelsberger M., Hernquist L., 2022, MNRAS, 512, 4909
Gehlot B. K., et al., 2019, MNRAS, 488, 4271
Geil P. M., Wyithe J. S. B., 2008, MNRAS, 386, 1683
Gelman A., Rubin D. B., 1992, Statistical Science, 7, 457
Ghara R., Choudhury T. R., Datta K. K., 2015, MNRAS, 447, 1806
Ghara R., et al., 2020, MNRAS, 493, 4728
Ghara R., Giri S. K., Ciardi B., Mellema G., Zaroubi S., 2021, MNRAS, 503, 4551
Giri S. K., Schneider A., Maion F., Angulo R. E., 2022, MNRAS, 512, 5010
Greig B., Bonaldi A., Bourke T., Keane E., Wagg J., 2019, arXiv e-prints, p. arXiv:1912.12699
Greig B., Mesinger A., 2017, MNRAS, 460, 1328
Greig B., Mesinger A., 2015, MNRAS, 449, 4246
Greig B., Mesinger A., 2016, MNRAS, 455, 4295

Maity & Choudhury

MNRAS 000, 1–15 (2023)
Figure A1. Recovery of parameters using the excursion set two parameter single-snapshot model at redshift \( z = 7 \). The mock data is created by the same single-snapshot model (but with a different realization of the matter density field) using a grid resolution \( \Delta x = 2 h^{-1}\text{cMpc} \). The off diagonal panels show the joint two-dimensional posterior distributions of each pair of parameters. The contours represent 68% and 95% confidence intervals. The diagonal panels show the marginalized posterior distributions of the parameters. The dashed lines represent the input values for generating mock data. The blue contours and curves show the results when the resolution used for the MCMC run is \( \Delta x = 2h^{-1}\text{cMpc} \) (same as the one used for the mock data), while the red ones are for the resolution \( \Delta x = 4h^{-1}\text{cMpc} \). The recovery of the input parameters is biased when the resolution used is different from that used to generate the mock.

Figure A2. The 21 cm power spectra for the excursion set two parameter single-snapshot model at redshift \( z = 7 \) for 200 random samples drawn from the posterior distributions shown in Fig. A1. The blue lines show the models for a resolution \( \Delta x = 2h^{-1}\text{cMpc} \) while the red ones are for \( \Delta x = 4h^{-1}\text{cMpc} \). The dashed lines denote the input models.

APPENDIX A: RECOVERY USING EXCURSION SET BASED APPROACH

In Section 3, we discussed the recovery of model parameters for a simple two parameter model using our photon conserving code SCRIPT. In this appendix, we discuss an identical analysis on parameter recoveries but for the excursion set based model. It is known

\[
\Delta x = 4h^{-1}\text{cMpc}
\]

Best fit
mock data (SKA noise, \( z = 7.0 \))

\[
\Delta x = 2h^{-1}\text{cMpc}
\]

Best fit
mock data (SKA noise, \( z = 7.0 \))
that the excursion set based models do not conserve the number of photons (Zahn et al. 2007, 2011; Paranjape et al. 2016). It was shown in an earlier work by Choudhury & Paranjape (2018) that an important consequence of this non-conservation is the non-convergence of the large scale 21 cm power spectra for different resolutions. The aim of the analysis here is to check how the parameter recoveries are affected by the non-convergent power spectra.

Similar to the previous case, we generate the mock dataset for the finest resolution ($\Delta x = 2h^{-1}$ cMpc) using the excursion set based model. The density and collapsed halo fields used for this purpose are identical to those used for the photon conserving case. The details of our implementation of the excursion set model can be found in earlier papers (Choudhury et al. 2009; Majumdar et al. 2014), in particular see Section 3.1 of Choudhury & Paranjape (2018). We calculate the uncertainties due to the thermal noise of the telescope and the cosmic variance to the mock power spectrum as in equation (4). We then shift the power spectrum data points by adding a random number having a Gaussian distribution with zero mean and a standard deviation equal to the associated thermal noise.

Let us discuss the case which corresponds to the middle stages of the reionization, i.e., $z = 7$. We use parameters $\xi = 14$ and $M_{\text{min}} = 10^9 M_\odot$ as the input model to generate the mock data. The global ionization fraction for this input model is $Q_{\text{HI}}^M = 0.45$. We first study the recovery when the resolution is the same $\Delta x = 2h^{-1}$ cMpc as the one used for generating the mock data. The posterior distributions are shown in blue in Fig. A1, the corresponding values can be found in Table A1. It is clear that the recovery of all the parameters is excellent when the resolution used for the analysis is the same as the mock data.

However, the posterior distributions for the coarser resolution $\Delta x = 4h^{-1}$ cMpc (red) show significant deviation from the finer ones. The input values (shown by dashed lines) deviate significantly from the two-dimensional joint posterior distributions. The same can be concluded from Table A1 too. If we concentrate on $Q_{\text{HI}}^M$, we find that the best-fit values for the two resolutions differ by $\sim 20\%$ (the corresponding difference was $\sim 2.5\%$ for the photon conserving model, see Table 1). The recovered mean values of $Q_{\text{HI}}^M$ are inconsistent at a level $\geq 3\sigma$ (they were well within $1\sigma$ for the photon conserving case). Thus, using a coarser resolution map for parameter estimation could lead to a significantly biased value of $Q_{\text{HI}}^M$ for the excursion set models.

We see interesting trends for other redshifts as well, see Table A1 for the constraints and compare with the input values. At $z = 8$, we find that the mean values of $Q_{\text{HI}}^M$ for the two resolutions are consistent with each other. This result is consistent with the findings in our earlier work (Choudhury & Paranjape 2018) that the effect of photon non-conservation is less at early stages of reionization when the bubble sizes are smaller than the grid resolution used. At late stages of reionization $z = 6.2$, the mean values of $Q_{\text{HI}}^M$ for the two resolutions are within $\sim 1.5\sigma$, which shows that the effect is less at late stages too. The effect of photon non-conservation, leading to non-convergence of the power spectra, is maximum at the middle stages of reionization where the large-scale signal is maximum.

We should mention here that the results presented here are valid only for the excursion set model as implemented by us. For other implementations of the algorithm, which could vary from ours regarding how the haloes are identified, smoothing of the density field, type of filters used to identify self-ionized regions, the results could be different. However, the main lesson from our analysis is that regardless of the implementation, it is important to check for the numerical convergence of the results for any excursion set algorithm.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{FigureB1.png}
\caption{Recovery of parameters using the excursion set two parameter single-snapshot model at redshift $z = 7$. The mock data is created by the photon conserving single-snapshot model (with a different realization of the matter density field) using a grid resolution $\Delta x = 2h^{-1}$ cMpc. The off-diagonal panels show the joint two-dimensional posterior distributions of each pair of parameters. The contours represent 68% and 95% confidence intervals. The diagonal panels show the marginalized posterior distributions of the parameters. The dashed lines represent the input values for generating mock data. The blue contours and curves show the results when the resolution used for the MCMC run is $\Delta x = 2h^{-1}$ cMpc (same as the one used for the mock data), while the red ones are for the resolution $\Delta x = 4h^{-1}$ cMpc.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{FigureB2.png}
\caption{The 21 cm power spectra for the excursion set two parameter single-snapshot model at redshift $z = 7$ for 200 random samples drawn from the posterior distributions shown in Fig. A1. The blue lines show the models for a resolution $\Delta x = 2h^{-1}$ cMpc while the red ones are for $\Delta x = 4h^{-1}$ cMpc. The dashed lines denote the input models.}
\end{figure}

\section*{APPENDIX B: RECOVERY USING ES MODEL WITH PC MOCKS}

It has been shown by Choudhury & Paranjape (2018) that the ionization maps and power spectra predicted by the photon conserving and excursion set models, for the same input parameters, differ from each other. To study the consequence of this difference on recovery of parameters, we carry out an exercise where we generate the mock data using the photon conserving model and run the MCMC chains.
to constrain the parameters using the excursion set model. Differences between the input and recovered parameters would indicate the mismatch between the two algorithms.

As before, we focus on $z = 7$. The mock data is generated using the photon conserving model at a resolution of $\Delta x = 2 h^{-1}$ cMpc, i.e., the data is identical to that used in Section 3. We first check the parameter recovery for the excursion set model at the same resolution, the results are shown in Fig. B1. As is obvious, the recovered parameters are significantly different from the input ones. In particular, the recovered $Q_{\text{HII}}^M$ is different by $\sim 10\%$ from the input value, as can be seen in Table B1. This would then be the typical level of discrepancy between the two algorithms at the mid-stages of reionization.

It was also shown by Choudhury & Paranjape (2018) that the amount of photon non-conservation becomes less when the resolution is coarser. To check the implications, we also run a MCMC with the coarser resolution of $\Delta x = 4 h^{-1}$ cMpc. The corresponding results (red) in Fig. B1 indicate that the recovery of $Q_{\text{HII}}^M$ is much better in this case. The same conclusion can be drawn from Table B1 which shows that the recovered $Q_{\text{HII}}^M$ is within $\sim 2.5\%$ of the input value. Thus, the match between excursion set and photon conserving models improves as we coarsen the resolution.

We have also run the chains for $z = 8$ and $z = 6.2$ for completeness. The recoveries of $Q_{\text{HII}}^M$ at $z = 8.0$ is reasonably well (input value is with $1\sigma$ uncertainties). This is due to that the photon number conservation is less dominant at higher redshifts when the global ionization fraction is small. On the other hand, the recoveries in the case of $z = 6.2$ are worse.

This analysis confirms that the recovery of the reionization history using 21 cm power spectra depends on the semi-numerical algorithm used for generating the ionization maps. It thus becomes important to develop some kind of consensus within the community of researchers working on semi-numerical models of reionization as to what would be the best way of identifying a physically meaningful model which remains computationally efficient.

**APPENDIX C: NUMERICAL CONVERGENCE OF THE 21 CM POWER SPECTRA FOR PC MODELS**

We study the numerical convergence of the 21 cm power spectra obtained using our photon conserving models of reionization. In Fig. C1, we show the power spectra at $z = 7$ for three different grid sizes, namely, $\Delta x = 2$, 4 & 8 $h^{-1}$ cMpc for the different fiducial models used in the study. It is evident that the power spectra at large scales convergence to within 5\% with respect to the resolutions used for generating the maps. The numerical convergence for the two-parameter single snapshot model was already shown by Choudhury & Paranjape (2018). It is interesting that the results hold also for the full-history models.

This paper has been typeset from a TeX/LaTeX file prepared by the author.
Table B1. Parameter constraints obtained using the excursion set two parameter single-snapshot model at redshifts $z = 7.0$, 8.0 and 6.2. Results are shown for two different resolutions $\Delta x = 2h^{-1}\text{cMpc}$ and $4h^{-1}\text{cMpc}$ used for the MCMC analysis. The mock data is generated using the two parameter photon conserving model, always at the resolution $\Delta x = 2h^{-1}\text{cMpc}$. For each parameter, we show the prior and the input value used for the mock along with the obtained mean, 68% confidence limits and best-fit. We also provide the $\chi^2/\nu$ for the best-fit models, $\nu$ being the number of degrees of freedom.

Figure C1. The 21 cm power spectra at $z = 7$ for three different grid sizes $\Delta x = 2$, 4 & $8 h^{-1}\text{cMpc}$ used for generating the ionized maps. We show the results for the different fiducial models used in the study.