Comparison of reionization models: radiative transfer simulations and approximate, seminumeric models

Oliver Zahn, Andrei Mesinger, Matthew McQuinn, Hy Trac, Renyue Cen and Lars E. Hernquist

1 Berkeley Center for Cosmological Physics, Physics Department and Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94708, USA
2 Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA
3 Harvard–Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
4 Berkeley Astronomy Department, University of California, Berkeley, CA 94720, USA
5 Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Accepted 2011 January 28. Received 2011 January 26; in original form 2010 April 29

ABSTRACT
We compare the predictions of four different algorithms for the distribution of ionized gas during the Epoch of Reionization. These algorithms are all used to run a 100 Mpc $h^{-1}$ simulation of reionization with the same initial conditions. Two of the algorithms are state-of-the-art ray-tracing radiative transfer codes that use disparate methods to calculate the ionization history. The other two algorithms are fast but more approximate schemes based on iterative applications of a smoothing filter on the underlying source and density fields. We compare these algorithms’ resulting ionization and 21-cm fields using several different statistical measures. The two radiative transfer schemes are in excellent agreement judging by the power spectra of both the ionization fields and the 21-cm emission fields (agreeing to better than 10 per cent) and are in good agreement with the analytic schemes (better than 50 per cent) over the range of ionized fractions and wavevectors we compare ($0.2 < \bar{x}_i < 0.7; 0.1 < k < 10 h \text{ Mpc}^{-1}$).

This agreement suggests that the different approximations involved in the ray-tracing algorithms are sensible and that seminumerical schemes provide a numerically inexpensive, yet fairly accurate, description of the reionization process.

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

1 INTRODUCTION
At $z \gtrsim 20$, structures massive enough for the gas to cool and form the first galaxies first broke away from the Hubble flow. Stars and black holes formed in these structures and ultimately ionized and heated the intergalactic gas. However, exactly when and how this process occurred are questions that are still unresolved. There are currently two prominent constraints on the reionization era. First, Ly$\alpha$ forest absorption spectra towards high-redshift quasars show a rapidly increasing opacity to Ly$\alpha$ photons by the intergalactic medium (IGM) at $z \gtrsim 6$ (e.g. Fan et al. 2006). Several studies have interpreted this increase as evidence for the end of reionization. At the very least, the amount of transmission in the Ly$\alpha$ forest at $z \lesssim 6$ indicates that the bulk of reionization occurred at higher redshifts. Secondly, measurements of the large-angle cosmic microwave background (CMB) polarization anisotropies suggest that the median redshift of the reionization process was roughly $z \simeq 10.4 \pm 1.4$ (Komatsu et al. 2009).

Taken at face value, these observations favour an extended reionization epoch (Cen 2003; Fan et al. 2006; Bolton & Haehnelt 2007; Wyithe & Cen 2007). However, the interpretation of quasar absorption spectra is hampered by the large cross-section for Ly$\alpha$ absorption, which can lead to complete absorption even if hydrogen is highly ionized. This makes the high-redshift Ly$\alpha$ forest spectra difficult to interpret (Lidz, Oh & Furlanetto 2006; Becker, Rauch & Sargent 2007), and the data may even be consistent with reionization completing at redshifts $z < 6$ (Lidz et al. 2007; Mesinger 2010; McGreer, Mesinger & Fan 2011). Furthermore, the CMB polarization measurements only offer an integral constraint on the ionization history (Kogut et al. 2003; Page et al. 2007).

Additional model-dependent constraints on reionization have been derived from other astrophysical probes such as (1) the size of the proximity zone around quasars (Wyithe, Loeb & Carilli 2005; Fan et al. 2006, but see Mesinger, Haiman & Cen 2004; Bolton...
Redshifted 21-cm emission from the hyperfine transition of neutral hydrogen has the potential to provide detailed 3D information about the evolution and morphology of the reionization process (e.g. Zaldarriaga et al. 2004). Several large interferometers will devote significant integration times during the next few years towards detecting this signal. These efforts include the Mileura Widefield Array (MWA; Bowman, Morales & Hewitt 2005), 1 the Low Frequency Array (LOFAR; Harker et al. 2010), 2 the Giant Metrewave Radio Telescope (GMRT; Pen et al. 2009), the Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010) and eventually the Square Kilometer Array (SKA). 3

Our ability to infer information about the properties of the first sources from the 21-cm signal, as well as other observations, hinges on the accuracy of the theoretical modelling of the ionization structure during reionization. Therefore, a number of groups have developed 3D radiative transfer (RT) codes (e.g. Gnedin 2000; Sokasian, Abel & Hernquist 2001; Razoumov et al. 2002; Ciardi, Ferrara & White 2003; Mellem et al. 2006; McQuinn et al. 2007a; Semelin, Combes & Baek 2007; Trac & Cen 2007; Altay, Croft & Pelupessy 2008; Aubert & Teyssier 2008; Finlator, Özel & Davé 2009; Petkova & Springel 2009; see Trac & Gnedin 2009 for a recent review). Furthermore, given the large uncertainties in the rate ionizing photons were produced and escaped from high-redshift galaxies (and also in how they were absorbed by the dense systems), a large region of parameter space needs to be modelled in order to interpret observations. These concerns have prompted several groups to develop more approximate, but much faster schemes (e.g. Zahn et al. 2005; Mesinger & Furlanetto 2007; Geil & Wyithe 2008; Alvarez et al. 2009; Choudhury, Haehnelt & Regan 2009; Thomas et al. 2009).

Accurate models of the Epoch of Reionization (EoR) must include the evolution of the dark matter, gas, radiation, galaxies as well as a plethora of feedback processes. Although the morphology of reionization may be robust to many of these modelling uncertainties (McQuinn et al. 2007a), the scope of the parameter space is daunting. Simulations must resolve the small mass galaxies that are expected to dominate the ionizing photon budget (probably corresponding to the atomic cooling threshold, with halo masses of \( \sim 10^9 \, M_\odot \) at \( z \sim 7-10 \)), as well as the photon sinks. On the other hand, reionization simulations must also be large enough to statistically sample the distribution of \( H\alpha \) regions, which can span tens of comoving megaparsecs in size towards the end of reionization (Furlanetto & Oh 2005; Zahn et al. 2005; Mesinger & Furlanetto 2007; Zahn et al. 2007; Shin-S., Trac & Cen 2008). A few groups have recently come close to achieving this dynamic range (Iliev et al. 2006a; McQuinn et al. 2007a; Trac & Cen 2007; Shin-S., Trac & Cen 2008; Trac, Cen & Loeb 2008; see the recent review in Trac & Gnedin 2009). 4 In addition, promising analytic alternatives have been suggested (e.g. Furlanetto, Zaldarriaga & Hernquist 2004). The Monte Carlo implementation of these alternatives in large-volume simulations allows for a description of the non-spherical \( H\alpha \) region morphology that can be compared side by side with simulations (e.g. Zahn et al. 2005, 2007; Mesinger & Furlanetto 2007).

In this paper, we investigate the convergence of different algorithms for modelling the epoch of reionization (EoR). We compare two different cosmological RT codes, which follow ionizing radiation from sources identified as haloes in an N-body simulation. Our comparison also includes ionization fields 5 that are generated with two different fast, seminumerical algorithms. All of these calculations are performed on the same density fields and N-body halo fields. We test the codes for the same 100 Mpc h\(^{-1}\) simulation of reionization in order to compare them on the ‘full’ problem as well as study the numerical convergence of the predicted 21-cm signal. Our approach is different but complimentary to Iliev et al. (2006b, 2009), which compared several RT codes on a number of simpler tests problems.

The structure of this paper is as follows. In Section 2, we present the N-body simulation we employ. Section 3 discusses the two different RT schemes that are run in post-processing on top of the N-body simulation, and Section 4 describes the two seminumerical schemes. Section 5 investigates the spatial agreement between the ionization fields of these algorithms. Section 6 compares the predictions of these schemes for the 21-cm power spectrum. In Appendix A, we present further discussion and tests of the seminumerical schemes.

The simulations and analytic calculations presented here are based on a Λ cold dark matter (ΛCDM) cosmology with the following parameter values: \( \sigma_8 = 0.82, \ h = 0.7, \ \Omega_m = 0.28, \ \Omega_b = 0.046, \ n_s = 0.96, \) consistent with latest constraints from the Wilkinson Microwave Anisotropy Probe (WMAP; Komatsu et al. 2009).

2 TEST PROBLEM

In this paper, the RT and seminumerical calculations were performed by post-processing input density fields from Trac et al. (2008). A high-resolution N-body simulation with 3072 6 dark matter particles was used to evolve the matter distribution in a periodic box of comoving length 100 Mpc h\(^{-1}\). We chose to work with 82 snapshots of the matter density field (every 10 Myr from redshift \( z = 27 \) to 6), gridded on to a 256\(^3\) Cartesian grid for the reionization algorithms. The baryons were assumed to trace the dark matter down to the grid cell spacing, \( \Delta x = 390 \, \text{kpc} \, h^{-1} \), which is close to the anticipated physical smoothing scale for the intergalactic gas (the Jeans length for 10\(^5\) K gas is equal to 370 kpc h\(^{-1}\) at the mean density at \( z = 6 \)).

Radiation sources embedded in the large-scale structure of the matter distribution were modelled using haloes catalogued in the N-body simulation. Dark matter haloes were located on the fly using a friends-of-friends algorithm, with a linking length of \( b = 0.2 \) times the mean interparticle spacing. Haloes were located down to

\[^{4}\] Note, however, that Lyman limit systems (LLSs), which can dominate the absorption of ionizing photons (see for example the appendix of Furlanetto \& Oh 2005), are still too small to be resolved by state-of-the-art reionization simulations and therefore must be included via some analytic prescription (e.g. Choudhury et al. 2009; Alvarez & Abel 2010; Crociani et al. 2010).

\[^{5}\] By ‘ionization fields’, we mean the 3D distributions of the hydrogen ionized fraction, \( x_i \).
a minimum group size of \( \sim 40 \) particles \((M = 1.0 \times 10^8 \, M_\odot)\). This minimum mass corresponds roughly to the mass threshold that has a sufficient circular velocity to cool by atomic transitions \((T \gtrsim 10^6 \, K)\). The first source appeared at \( z \sim 27 \) in this simulation. By \( z = 6 \), there were \( >6 \) million sources and, when binned on a \( 256^3 \) Cartesian grid, the clustered sources occupied \( \approx 17 \) per cent of the grid.

### 3 Radiative Transfer Algorithms

This section describes the two RT codes that are performed on the test problem described in Section 2: the McQuinn et al. code (Section 3.1) and the Trac & Cen code (Section 3.2).

#### 3.1 McQuinn et al.

The algorithm presented in McQuinn et al. (2007a) is an improvement of the Sokasian et al. (2001) ray-tracing code. It is a more approximate ray-tracing scheme than Trac & Cen (2007), but has more modest CPU and memory requirements. This code was specifically designed to perform ray tracing on the millions of sources required to study the EoR and was the first ray-tracing code to achieve this feat (McQuinn et al. 2007a). In addition, this code has enabled the first parameter space studies of this epoch (McQuinn et al. 2007a).

This code sends out 768 rays from every source between every time-step, and rays travel within that time-step until their photons are expended (i.e. an infinite speed of light). It uses the HEALP-based adaptive ray-splitting scheme of Abel & Wandelt (2002), such that rays split as they travel to ensure that a minimum number of rays reach each cell. For the simulations in this paper, each cell receives at least 2.1 rays from each source when averaged over HEALP orientations. The numbers 768 (corresponding to HEALP level 3) and 2.1 were calibrated with convergence tests in the appendix of McQuinn et al. (2007a) and were the values that this code has been run with for large-scale simulations of cosmological reionization. The order that rays are cast in this code is randomized over both the sources and directions for each time-step. Finally, in the test problem in this paper, each time-step is taken to be \( \Delta t = 10 \, \text{Myr} \), which corresponds to the time between stored snapshots of the density field. Comparable values for \( \Delta t \) have been used in published studies with this code (McQuinn et al. 2007a,b).

The McQuinn et al. RT algorithm for hydrogen reionization works as follows. At the beginning of the time-step, the ionization fraction in each cell is corrected for recombinations that occur during that time-step. Next, rays are cast. Each ray carries a set number of photons and when a ray reaches a cell that contains neutral gas, it deposits all of the photons into ionizations. The approximation that all the photons are absorbed at the ionization front is motivated by the short mean free path of ionizing photons in a neutral medium for a physically motivated source spectrum during hydrogen reionization (i.e. a mean free path of tens of kpc is expected, which is much smaller than the cell size). When the path to the edge of an ionized region becomes comparable to the box size, this algorithm slows significantly because it has to follow each ray to the front edge of each time-step. The solution to the ionization field that this algorithm obtains technically depends on the order rays are cast from each cell, but convergence studies have demonstrated that in practice a nearly identical solution is reached with different ray orderings (McQuinn et al. 2007a). After rays have ionized a cell, it is assumed to be in photoionization equilibrium with the photoionizing background.

Finally, as we mentioned previously, this algorithm assumes an infinite speed of light. This approximation is the least valid at the end of reionization when the \( H \) regions are the largest. Because of this approximation, the McQuinn et al. algorithm produces several per cent larger ionization fractions close to the end of reionization, compared with the other RT algorithm in this comparison by Trac & Cen (2007). This issue could be alleviated in higher resolution simulations than performed here where Lyman limit absorption systems are resolved, which will limit the mean free path.

#### 3.2 Trac & Cen

The Trac & Cen (2007) RT algorithm follows the propagation of ionizing photons using an adaptive ray-tracing technique, featuring several novel approaches that distinguish it from other previous ray-tracing algorithms. The major characteristics include conservation of photons, a time-dependent solution that is causal and computational scaling that is independent of the number of sources. Here we provide a basic description of the algorithm and highlight differences with other approaches.

As a basis, the adaptive ray-splitting scheme of Abel & Wandelt (2002) is used to efficiently improve spatial resolution. For a given source, a preset number of base rays are cast, and they travel a short distance before each split into four daughter rays. Successive generations of splitting continue downstream such that the angular resolution continually increases farther away from sources. This procedure ensures that a minimum number of rays intersect each grid cell element of the density field within the ionized region. For this comparison study, the number of base rays was set at 192, and the minimum number of rays intersecting each grid cell was set at two.

One major feature of the algorithm is that new source rays are cast for every RT time-step, which is set equal to the light-crossing time, defined as the length of a grid cell divided by the speed of light. Correspondingly, all rays are advanced a maximum distance (when the medium is optically thin) equal to a grid cell length in each time-step. This synchronization ensures a causal solution, a characteristic usually not shared by other ray-tracing algorithms where rays from multiple sources are treated independently. Generally, this approach is costly because the light-crossing time at high resolution can be quite small compared to the duration of reionization. However, Pawlik & Schaye (2008) have pointed out that when the radiation filling factor is close to unity near the end of reionization, the light-crossing time-stepping can actually be computationally faster than the infinite speed of light approach.

Another major feature of the algorithm is an adaptive ray-merging scheme designed to circumvent the \( \mathcal{O}(N^2) \) scaling in the multi-source problem. In the brute-force approach, every source must emit roughly as many rays as there are density grid cells when the radiation filling factor is close to unity. For a \( 256^3 \) grid with close to 3 million source cells, a computationally infeasible number of 50 trillion rays would be required in total. Trac & Cen (2007) adaptively restrict the maximum number of rays entering and exiting a given cell by merging near-parallel rays from multiple sources. For this comparison study, approximately 100 rays per cell and as many as 2 billion rays to propagate within a time-step were required by the end of reionization. One disadvantage is that a large amount of memory is required to store the necessary rays. In the current version of the implementation, each ray requires 44 bytes to store.

---

\(^6\) Admittedly, \( H_\text{i} \) cooling mass haloes were unresolved in the \( N \)-body simulation used in this study and were instead included in post-processing prior to ray tracing.
intrinsic variables plus another 8 bytes per frequency bin to keep track of the frequency and photon count.

For each RT time-step, the photoionization rates are calculated on the same 256$^3$ grid using the incident radiation flux from the rays. The equations for the ionization evolution are then solved using a non-equilibrium solver where the time integration involves a stable implicit scheme. For this comparison study, only photoionization at one frequency is considered. However, the RT algorithm can also handle photoheating and multiple frequencies in a cosmological hydrodynamic simulation (Trac et al. 2008).

4 SEMINUMERICAL MODELLING

Our comparisons also include two seminumerical algorithms for generating the ionization field. These techniques are motivated by the intuition that sources ionize the regions immediately around them and that the clustering of sources drives the structure of reionization. The RT algorithms described previously can require large computer clusters and can take days to process a single redshift output. However, for many cosmological reionization applications, it might be useful to sacrifice some accuracy for an increase in speed and dynamic range. An important example of such an application is simulating the signal for the upcoming 21-cm surveys. The effective resolution of all planned instruments, including the SKA, is much worse than the resolution of the typical RT scheme. Additionally, the field of view (FoV) of some of these instruments will be enormous (e.g. $5000 \times 5000$ Mpc$^2$ at $z \sim 8$ for MWA), far out of reach of conventional RT algorithms. Also, very little is presently known about the properties of the underlying sources, translating to an enormous parameter space in need of exploration. Thus, a ‘relatively’ accurate yet fast ionization algorithm can be an invaluable tool in reionization studies. Below, we detail the seminumerical algorithms we use: one introduced in Mesinger & Furlanetto (2007) and a new one, somewhat refined scheme, dubbed fast Fourier radiative transfer (FFRT). These are capable of generating large-scale 3D ionization fields in a matter of minutes on a single processor (see Zahn et al. 2005, 2007; Geil & Wyithe 2008; Alvarez et al. 2009; Choudhury et al. 2009; Thomas et al. 2009 for related/alternate seminumerical ionization algorithms).

Both our seminumerical algorithms are based on the excursion-set formalism for modelling reionization developed by Furlanetto et al. (2004). The foundation of this approach is to require that the number of ionizing photons inside a region be larger than the number of hydrogen atoms that it contains.

In particular, we require that an ionized simulation cell at coordinate $x$ satisfy the criterion (cf. Zahn et al. 2005; Mesinger & Furlanetto 2007).

$$f_{\text{coll}}(x, z, R) \geq \zeta^{-1},$$

where $\zeta$ is the efficiency factor by which the mass in collapsed systems can ionize the mass around it, and $f_{\text{coll}}$ is the fraction of mass residing in collapsed haloes with masses $>10^8$ M$_\odot$ inside a sphere of mass $M = 4/3\pi R^3\rho[1 + \langle\delta_{\text{NL}}\rangle],\frac{\rho}{\langle\delta_{\text{NL}}\rangle}$, with mean physical overdensity $\langle\delta_{\text{NL}}\rangle$, centred on Eulerian coordinate $x$, at redshift $z$.

Following the excursion-set approach (Bond et al. 1991; Lacey & Cole 1993), our algorithm decreases the filtering scale $R$, starting from some maximum value $R_{\text{max}}$, and progressing down to the cell size, $R_{\text{cell}}$, with logarithmic filter spacing of $R_{\text{next}} = 0.9 \times R_{\text{prev}}$. If at any filter scale the criterion in equation (1) is met around a cell, this cell is flagged as ionized.

Our algorithms also take into account bubbles that are smaller than the cell size by setting the cell’s ionized fraction to $\zeta$ times the collapse fraction in the cell at the last filtering step for those cells that are not fully ionized. A two-phased medium (containing only fully ionized and fully neutral cells) is a good approximation for stellar sources and simulation resolution we study (see Appendix A). Such partial ionizations are important at the early stages of reionization, and this correction becomes more significant as the cell size is increased.

We include two variations of the said seminumerical ionization algorithm. These variations are distinguished by their method of computing $f_{\text{coll}}$, their choice of filter and their H II region-flagging scheme, namely the following.

(i) Mesinger & Furlanetto 2007 (MF07): this scheme computes $f_{\text{coll}}$ using the same $N$-body source field and evolved density fields as the RT algorithms. MF07 also uses a top-hat filter to compute $f_{\text{coll}}$, and it flags the entire sphere with radius $R$ as ionized. The only difference with the algorithm published in MF07 is the inclusion of partial ionizations, as mentioned above.

(ii) Fast Fourier Radiative Transfer: this variant computes $f_{\text{coll}}$ using the conditional Press–Schecter (PS) formalism (Lacey & Cole 1993; Somerville & Kolatt 1999) directly on the evolved (i.e. non-linear) density field. FFRT also uses a sharp $k$-space filter to compute $f_{\text{coll}}$ instead of spherical top-hat. Finally, it only flags the...
central filter cell at \( x \) as ionized (as in Zahn et al. 2005) instead of ‘painting’ the entire sphere with radius \( R \) as ionized (as in MF07), as the former approach is faster.

In setting the partial ionizations in the FFRT scheme, we also account for Poisson fluctuations in the halo number when the mean collapse fraction obtained through conditional PS formalism\(^{12} \) becomes small: \( f_{\text{cell}}(x, z, R_{\text{cell}}) [1 + \delta_{\text{cell}}(x, z, R_{\text{cell}})] M_{\text{cell}} < 50 \, M_{\odot} \), where \( M_{\text{cell}} \) is the mean, total (DM+baryonic) mass corresponding to a cell, \( \delta_{\text{cell}} \) is the fractional overdensity in the cell’s mass, and our faintest ionizing sources correspond to halo masses of \( M_{\text{min}} \approx 10^8 \, M_{\odot} \) in this application. We make the simplifying assumption that the subgrid sources are all of the same mass, \( M_{\text{min}} \) (likely a decent approximation for partially ionized cells, since the mass function is fairly steep in this regime). We sample a Poisson distribution with mean \( f_{\text{cell}}(1 + \delta_{\text{cell}})/M_{\text{min}} \) to obtain the number of subgrid sources, \( N_{\text{min}} \). We then compute that cell’s collapse fraction, \( f_{\text{cell}} \text{sampled} = N_{\text{min}} M_{\text{min}} /[M_{\text{cell}} (1 + \delta_{\text{cell}})] \). Poisson fluctuations are found to be important when the cell size increases to \( > 1 \) Mpc (see Fig. A1 and associated discussion). Note, however, that this feature should be left as an option, depending on the particular application. Although improving the statistics of the ionization fields at a fixed redshift, such stochastic fluctuations in the collapsed mass and ionization fields make it impossible to deterministically track the redshift evolution of a particular realization.

These seminumerical schemes can be extended to contain additional physics such as spatially inhomogeneous recombinations (Choudhury et al. 2009), a mass-dependent ionization efficiency (Furlanetto, McQuinn & Hernquist 2006a) and radiative feedback (Geil & Wyithe 2008). Since these processes are also poorly understood and the RT schemes we compare against do not take them into account, we will limit our refinements in this work to the simplest case that ignores feedback and inhomogeneous recombinations.

5 COMPARISON OF IONIZATION FIELDS

This section compares the ionization fields generated from the four schemes: McQuinn et al. (2007a), Trac & Cen (2007), Mesinger & Furlanetto (2007) and FFRT. Fig. 1 shows ionization maps for three different volume-weighted ionization fractions, \( x_{\odot} = 0.25, 0.51 \) and 0.72. The maps are from the same slice, spanning 100 Mpc \( h^{-1} \times 100 \) Mpc \( h^{-1} \) with a depth of 0.4 Mpc \( h^{-1} \), through the simulation box. In all of the simulations, the reionization process appears qualitatively similar. In the early stages, \( \text{H} \, \text{II} \) regions typically contain relatively few galaxies and are spatially clustered similar to the sources. The \( \text{H} \, \text{II} \) regions expand and merge with one another as reionization progresses in these calculations. By the time the process is half completed, the ionization structures appear much more connected, and the largest \( \text{H} \, \text{II} \) regions contain tens of thousands of galaxies. In the final stages, the ionization fronts expand into the remaining neutral regions until there is complete overlap.

The maps from the two RT simulations show the closest agreement, with very similar ionization structures at all times. The ionization fronts are similarly resolved such that even small neutral regions between converging fronts are well preserved. Of the two seminumerical methods, the MF07 scheme produces maps that are more similar to the RT ones, although the \( \text{H} \, \text{II} \) regions are more spherical due to the spherical bubble-flagging algorithm. Conversely, FFRT produces \( \text{H} \, \text{II} \) regions that appear to be not as spherical as the RT ones and that are more connected. Such differences are understandable, as the seminumeric models do not include the physics needed to exactly capture the position and structure of the ionization fronts. In the rest of this section, we compare several statistics to quantify the reionization morphology and to check the degree of convergence between the different schemes.

5.1 Bubble size distribution

Fig. 2 shows the ionized bubble size distribution, defined as the probability that a randomly chosen ionized cell from the simulation volume is part of an ionized region of radius \( R \). An ionized cell is assigned to be in a bubble of radius \( R \), where \( R \) is set equal to the largest sphere surrounding the cell whose mean enclosed ionized fraction is 0.9 (see Zahn et al. 2007, for more discussion). This definition is convenient because it is similar to that which is used in excursion set reionization models (Furlanetto et al. 2004). In addition, it has the advantage of being quickly computable with fast Fourier transforms.

The distributions from the four schemes are generally in good agreement, as expected from the visual inspection of the ionization maps. The bubble size distributions have a characteristic peak size that shifts towards larger scales as \( \text{H} \, \text{II} \) regions grow and reionization progresses. The two RT simulations show excellent agreement, with the disagreement being largest on small scales early in reionization. The overconnectedness of FFRT’s ionization fields translates to a shift to larger bubble scales in the early and middle stages of reionization. The MF07 scheme also produces larger bubbles in the early stages. This could be due in part to the fact that MF07 more spherical \( \text{H} \, \text{II} \) regions are more easily flagged in their entirety by our definition of ‘bubble size’. The fact that both seminumerical codes seem to produce larger \( \text{H} \, \text{II} \) regions at \( \bar{\xi}_{i} = 0.25 \) could also be due in part to their underestimate of partially ionized cells (see Appendix A). By normalizing at fixed \( \bar{\xi}_{i} \), we require more fully ionized cells to compensate for the lack of partially ionized ones, thereby resulting in larger \( \text{H} \, \text{II} \) regions. Nevertheless, by \( \bar{\xi}_{i} = 0.5 \), the MF07 bubble size distribution agrees better with those of the RT algorithms than the FFRT bubble size distribution. At \( \bar{\xi}_{i} = 0.7 \), the seminumeric schemes are in best agreement with the RT.

5.2 Power spectra and cross-correlations

Another measure of reionization morphology is the auto power spectrum of ionized regions \( P_{XY} \). Aside from its intrinsic value in characterizing the ionization field, the power spectrum is important because it is directly related to the 21-cm power spectrum (see equation 3). Fig. 3 shows the evolution of the ionization power spectra. As previously noted (e.g. McQuinn et al. 2007a; Zahn et al. 2007), the power shifts to larger scales as the bubbles grow and combine. Furthermore, the power spectrum is relatively flat.
Figure 1. Comparison of ionization fields generated from four schemes: McQuinn et al., Trac & Cen, MF07 and FFRT. The maps are from the same slice (100 Mpc $h^{-1} \times 100$ Mpc $h^{-1}$ with depth of 0.4 Mpc $h^{-1}$) through the simulation box.
Comparison of reionization models

Figure 2. The ionized bubble size distribution for the four schemes: Trac & Cen (red solid line), McQuinn et al. (green dashed line), FFRT (blue short-dashed line) and MF07 (magenta dotted line). Plotted is the likelihood that a given ionized patch is part of a region with size $R$ (see text).

Figure 3. Ionization field autopower spectra for the four schemes. As reionization progresses and bubbles grow and merge, the power shifts to larger scales. The small-scale upturn in power at $k \gtrsim 8 \, h \, \text{Mpc}^{-1}$ is due to discreteness noise.

$P_{X_1X_2} / \sqrt{P_{X_1X_1} P_{X_2X_2}}$, where $P_{X_1X_2}$ is cross power spectrum of ionization fields $X_1$ and $X_2$. While the schemes may agree on $P_{XX}$ by having similar distributions of bubble sizes, $r_{X_1X_2}$ specifically tests whether the ionized cells are in the correct locations by comparing the phases of the Fourier modes of the transformed ionization fields.

To better interpret $r_{X_1X_2}$, consider the case where $X_2$ is a random translation of $X_1$ by the distance $r_0$. In this case, the cross-correlation coefficient is equal to $\cos(k \cdot r_0)$ for each $k$ and $\sin (k r_0) / (k r_0)$ when binning by $k \equiv |k|$. If instead each bubble is translated by a distance $r_0$, but in random directions, the cross-correlation will also go to zero at $k r_0 \gg 1$.

Fig. 4 shows the cross-correlation between the Trac & Cen (2007) ionization fields and those of the other schemes. For reference, the dot–dashed cyan curve in the middle panel corresponds to the algorithm presented in Zahn et al. (2007). On all scales, the ionization structure in the RT simulations agrees well, with $r_{X_1X_2} > 0.8$ at $k < 10 \, h \, \text{Mpc}^{-1}$. The slightly weaker correlations on scales near the grid spacing are due to small differences in the locations and structure of the ionization fronts. The ionization field of the seminumerical schemes have a smaller correlation with that of Trac & Cen (2007). The MF07 cross-correlation coefficients are only weakly dependent on $\bar{x}_i$, with $r_{X_1X_2} > 0.8$ at $k < 1 \, h \, \text{Mpc}^{-1}$ throughout the EoR. Also according to this statistic, FFRT performs better in the early stages and worse in the late stages of reionization.

as a function of scale during the final half of reionization, though seminumerical studies predict a drop in power on scales larger than these simulation boxes (Mesinger & Furlanetto 2007). While all of our schemes show these trends, they are not as obviously present in the RT simulations of Iliev et al. (2006b).

The small differences in $P_{XX}$ between the four schemes are consistent with the differences seen previously in the bubble size distributions. The two RT simulations agree extremely well. The seminumerical schemes have $\sim 30$ per cent more large-scale power in the early stages, probably due to the aforementioned underprediction of the partial ionizations. FFRT performs slightly better than MF07 on moderate scales early in reionization, but the overconnectedness of its ionization field overproduces the large-scale power throughout. Nevertheless, the power spectra of both seminumerical schemes agree fairly well with the RT ones (to within $\sim 30$ per cent) throughout the EoR on all scales probed by the simulations. We note that some of the excess large scale power in both the FFRT and MF07 schemes compared to the radiative transfer schemes is due to their under-estimation of the number of partially ionized cells (see Fig. 8 in the Appendix). This can be alleviated by using higher resolution simulations (in which fewer cells are partially ionized), as we have verified by running FFRT with 9 times higher resolution.

As another test of the agreement between the ionization fields, we consider the cross-correlation coefficient $r_{X_1X_2} \equiv P_{X_1X_2} / \sqrt{P_{X_1X_1} P_{X_2X_2}}$, where $P_{X_1X_2}$ is cross power spectrum of ionization fields $X_1$ and $X_2$. While the schemes may agree on $P_{XX}$ by having similar distributions of bubble sizes, $r_{X_1X_2}$ specifically tests whether the ionized cells are in the correct locations by comparing the phases of the Fourier modes of the transformed ionization fields.

To better interpret $r_{X_1X_2}$, consider the case where $X_2$ is a random translation of $X_1$ by the distance $r_0$. In this case, the cross-correlation coefficient is equal to $\cos(k \cdot r_0)$ for each $k$ and $\sin (k r_0) / (k r_0)$ when binning by $k \equiv |k|$. If instead each bubble is translated by a distance $r_0$, but in random directions, the cross-correlation will also go to zero at $k r_0 \gg 1$.

Fig. 4 shows the cross-correlation between the Trac & Cen (2007) ionization fields and those of the other schemes. For reference, the dot–dashed cyan curve in the middle panel corresponds to the algorithm presented in Zahn et al. (2007). On all scales, the ionization structure in the RT simulations agrees well, with $r_{X_1X_2} > 0.8$ over the $k$ captured in the simulation ($k < 10 \, h \, \text{Mpc}^{-1}$). The slightly weaker correlations on scales near the grid spacing are due to small differences in the locations and structure of the ionization fronts. The ionization field of the seminumerical schemes have a smaller correlation with that of Trac & Cen (2007). The MF07 cross-correlation coefficients are only weakly dependent on $\bar{x}_i$, with $r_{X_1X_2} > 0.8$ at $k < 1 \, h \, \text{Mpc}^{-1}$ throughout the EoR. Also according to this statistic, FFRT performs better in the early stages and worse in the late stages of reionization.

© 2011 The Authors, MNRAS 414, 727–738
Month Notices of the Royal Astronomical Society © 2011 RAS
stages of reionization, with the cross-correlation coefficient at $k = 1 \, h \, \text{Mpc}^{-1}$ changing from 0.9 to 0.6 as reionization progresses from $x_i = 0.25$ to 0.72. Both seminumerical schemes perform better than the one in Zahn et al. (2005; see the middle panel), mostly owing to the latter’s use of linear (unevolved) density field.

We also investigate how the bubbles are distributed with respect to the large-scale overdensities of dark matter and gas. As suggested already by many studies, large-scale overdense regions tend to be ionized earlier than large-scale underdense regions because they harbour more sources. The seminumeric filtering techniques are motivated by this general, ‘inside-out’ description of the reionization process. In order to quantify the validity of this assumption, we calculate the cross-power spectrum $P_{XD}$ between the ionization and density fields, which is also directly related to the 21-cm power spectrum (see equation 3). Fig. 5 shows $P_{XD}$, and Fig. 6 shows the cross-correlation coefficient $r_{XD} \equiv P_{XD}/\sqrt{P_{XX}P_{DD}}$. In general, the ionization and density fields are perfectly correlated on large scales, but the correlation gets progressively weaker on smaller scales. This is expected as the sources (and sinks) are not perfectly correlated with the density field on small scales either.

The two RT simulations show good agreement overall in both $P_{XD}$ and $r_{XD}$. The Trac & Cen (2007) simulation appears to track the density field more closely, particularly for the $x_i = 0.25$ case, with $r_{XD}$ differing by $\sim 0.1$ on scales near the grid spacing. Not surprisingly, the FFRT scheme overpredicts the correlation of ionization and density and yields the highest values of $P_{XD}$ and $r_{XD}$ (particularly at $k > 1 \, h \, \text{Mpc}^{-1}$). This is because in FFRT the source and ionization fields are determined directly from the evolved density field. The MF07 field underpredicts this correlation.

6 COMPARISON OF THE 21-CM SIGNAL

One of the main applications of reionization models will be to aid in the interpretation of high-redshift 21-cm observations. Hence, we briefly explore the differences in the 21-cm signal predicted by our various schemes. We only study the power spectrum as this is a simple, often-studied statistic, which can encode information about the reionization state of the Universe (e.g. Lidz et al. 2008). In this section, we make the simplifying assumption that the spin temperature of the gas is much larger than the CMB temperature (which should be accurate for all but the very early stages of reionization; e.g. Furlanetto 2006). We also neglect redshift-space distortions, which mainly affect scales smaller than the bubble scales during the reionization epochs studied here (McQuinn et al. 2006; Mesinger & Furlanetto 2007; Mesinger, Furlanetto & Cen 2010).

With these simplifications, the excess 21-cm brightness temperature over that of the CMB from emission at a redshift of $z$ is
Comparison of reionization models

Figure 6. Cross-correlation coefficients between the ionization and density fields for the four schemes. The correlation is perfect on large scales, but it weakens with increasing $k$ because the sources (and sinks) are not perfectly correlated with the density field on small scales.

\[
\delta T \approx 26 (1 + \delta) \bar{x}_{\text{HI}} \left( \frac{\Omega_m h^2}{0.022} \right) (1 + \frac{z}{10})^{1/2} \text{mK},
\]

where $\delta$ is the density contrast of gas and $\bar{x}_{\text{HI}} \equiv 1 - x_{\text{LV}}$ is the average neutral fraction. The first-order expansion in ionization and density fluctuations for the 21-cm power spectrum is

\[
P_{21\text{cm}}(k) = T_b^2 \left[ P_{XX} - 2 \bar{x}_{\text{HI}} P_{XD} + \bar{x}_{\text{HI}} P_{DD} \right],
\]

where $T_b$ is the average brightness temperature in regions with $\bar{x}_{\text{HI}} = 1$, and $P$ is the power spectrum of the ionization (X) and density (D) fields.

In Fig. 7, we show the 21-cm power spectra computed using the four algorithms. As anticipated from $P_{XX}$ and $P_{XD}$ shown in the previous section, the 21-cm power spectra of the RT schemes agree quite well, differing by $\lesssim 10$ per cent on all scales. The MF07 scheme results in $\sim 50$ per cent more 21-cm power, due to the combined effects of its overprediction of $P_{XX}$ and underprediction of $P_{XD}$ (the cross-power enters the 21-cm expression above with a negative sign, as 21-cm signal stems from the neutral regions). Except for very large scales, the FFRT algorithm yields better overall agreement. This appears to be just a coincidence and owes partially to the fact that its overprediction of $P_{XX}$ is mitigated by its overprediction of $P_{XD}$.

The error bars in the middle panel of Fig. 7 correspond to forecasted precisions for the MWA and the SKA, assuming the configurations described in McQuinn et al. (2006). These errors assume a 1000-h integration time and a bandwidth of 6 MHz. The differences between the four schemes are smaller (comparable for MF07) than the MWA $1 \sigma$ errors at $k > 0.5 \text{ h Mpc}^{-1}$. On larger scales, however, uncertainty in the seminumerical modelling might bias reionization constraints with the MWA and especially with the SKA. Note that the power spectrum predictions are uncertain near the box scale owing to sample variance (and possibly at smaller scales owing to correlations with box-length modes). We expect the relative differences between these methods to be better converged given that the ionization field from all of the methods is computed from the same density field.

7 CONCLUSIONS

We have compared the results of different algorithms for computing the morphology of reionization. Two of these algorithms were RT simulations, which made disparate approximations for propagating H\textsc{ii} fronts from the sources, and the other two are less computationally expensive seminumeric schemes, which use the excursion-set formalism to identify ionized regions. The algorithms were all...
compared on equal footing, using the same realization of the large-scale matter distribution on which to generate the ionization field.

The two RT schemes are in excellent agreement with each other, with both $P_{\text{XX}}$ and $P_{\text{21cm}}$ agreeing to better than 10 percent over the range of ionization fractions and wavevectors that we have compared ($0.2 < x_i < 0.7; 0.1 < k < 10 \text{ Mpc}^{-1}$). The seminumerical schemes are also in good agreement with the RT, with $P_{\text{XX}}$ and $P_{\text{21cm}}$ differing by less than 50 percent during most of the EoR. In general, MF07 results in more accurate ionization morphology than FFRT (except in the early stages of reionization). However, FFRT yields more accurate 21-cm power spectra and is a faster algorithm allowing for larger dynamic range.

The differences between these schemes are smaller than those arising from our current astrophysical uncertainties (i.e. our inability to accurately model the sources and sinks of ionizing photons). This level of agreement suggests that the different approximations involved in the ray-tracing algorithms are sensible and that seminumerical schemes provide a numerically inexpensive, yet fairly accurate, realization of the reionization process. The speed and accuracy of these schemes can aid in efficient parameter space studies and in the interpretation of future observations.

ACKNOWLEDGMENTS

OZ wishes to express special gratitude to the co-authors for being supportive in completing this work after he was injured in the fall of 2009. The authors wish to thank Adam Lidz, Steven Furlanetto and Matias Zaldarriaga for helpful discussions. OZ is supported by an Inaugural Fellowship by the Berkeley Center for Cosmological Physics. AM is supported by NASA through Hubble Fellowship grant HST-HF-51245.01-A, awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. MM is supported by NASA through an Einstein Fellowship. HT is supported by an Institute for Theory and Computation Fellowship. RC acknowledges support from NASA grant NNG06GI09G.

REFERENCES

Abel T., Wandelt B. D., 2002, MNRAS, 330, L53
Altay G., Croft R. A. C., Peeples I., 2008, MNRAS, 386, 1931
Alvarez M. A., Abel T., 2010, ArXiv e-prints (1003.6132)
Alvarez M. A., Busha M., Abel T., Wechsler R. H., 2009, ApJ, 703, L167
Aubert D., Teyssier R., 2008, MNRAS, 387, 295
Barkana R., Loeb A., 2007a, MNRAS, 377, 1043
Bolton J. S., Haiman Z., 2007, ApJ, 669, L15
Bolton J. S., Haehnelt M. G., 2007a, MNRAS, 371, L35
Bolton J. S., Haehnelt M. G., 2007b, MNRAS, 382, 325
Bond J. R., Cole S., Easthrough G., Kaiser N., 1991, ApJ, 379, 440
Bowman J. D., Morales M. F., 2010, MNRAS, 344, L7
Bovon J. D., Morales M. F., Hewitt J. N., 2005, ApJS, 148, 161
Cen R., 2003, ApJ, 670, 39
Choudhury T. R., Haehnelt M. G., Regan J., 2009, MNRAS, 394, 960
Ciardi B., Ferrara A., White S. D. M., 2003, MNRAS, 344, L7
Crociain D., Mesinger A., Moscardini L., Furlanetto S., 2010, MNRAS, 1655

13 So as to provide an equal footing, all schemes in this paper used the evolved density field obtained with a cosmological N-body code. We note, however, that when used in an 'end-to-end' seminumerical simulation such as 21cmFAST (http://www.astro.princeton.edu/~mesinger/Sim.html), which generates the evolved density field with perturbation theory, the resulting 21-cm power spectra are comparable in accuracy to those generated with MF07 (Mesinger et al. 2010).
APPENDIX A: FURTHER TESTS OF THE FFRT SCHEME

A1 Distributions of the ionization fraction

Here we investigate how well the analytic schemes perform at computing the level of partial ionization in simulation cells. This is where we expect the most disagreement between the RT and the FFRT, as these cells contain the edges of H\textsc{ii} regions/ionization fronts, where detailed RT effects are the most important.

We remind the reader that the FFRT scheme assigns partial ionization fractions based on the collapse fraction inside each cell. Therefore, it takes into account the ionizing photons from subgrid sources, which were not luminous enough to ionize the entire cell. This is done in anticipation of the use of FFRT on very large scales, with fairly large cell sizes (e.g. Mesinger 2010). As the cell size increases, fluctuations in the source number among neighbouring cells decreases, and our approximation becomes increasingly accurate. However, as the cell size becomes small (\(M_{\text{cell}} \sim M_{\text{min}}\)), partially ionized cells increasingly correspond to the edges of ionization fronts driven by external sources instead of the subgrid sources churning away at their local H\textsc{ii} as assumed by the FFRT scheme. Here we wish to quantify these limits, checking how well the FFRT reproduces the distribution of partially ionized cells, given a certain cell size.

In Fig. A1, we show the fraction of simulation cells with a given ionization fraction, at \((z, x_i) = (8.49, 0.25), (7.56, 0.51), (7.11, 0.72)\), top to bottom. Red circles, green triangles, blue squares and magenta pentagons correspond to the ionization algorithms of Trac & Cen (2007), McQuinn et al. (2007b), FFRT and FFRT without Poisson fluctuations in the halo number (see Section 4), respectively. The left-hand panel of Fig. A1 shows the ionization fraction distributions for the 256\(^3\) grid used throughout this paper. The two RT schemes agree extremely well, signalling that these schemes have converged. Note that on this grid scale, and assuming a relatively soft stellar ionization spectrum, the ionization field can indeed be well represented by a binary field. The FFRT algorithm tends to underpredict the number of mostly ionized cells. This underprediction is less severe in the early stages of reionization when the surface area of ionization fronts is smaller.

Another interesting thing to note from the left-hand panel is that our fiducial FFRT model, which includes Poisson scatter in the subgrid source number, has no partially ionized cells with ionization fractions less than \(\sim 0.1\). This is because a single source with our minimum halo mass \(M_{\text{min}}\) is luminous enough to ionize its host cell to a degree \(\geq 0.1\). On the other hand, if one ignores that sources are discrete and instead uses the mean conditional collapse fraction (magenta pentagons), the distribution of partial ionizations agrees fairly well with the RT in this regime. This indicates that partial ionizations in the 256\(^3\) grid with its \(\Delta x = 0.39\) \(\text{Mpc}^{-1}\) cells is dominated by unresolved ionization fronts driven by sources outside of each cell, as discussed above. Hence, these cells are too small to justify the assumption of partial ionizations driven by subgrid sources, and the good performance of the FFRT without Poisson scatter at \(x_i \lesssim 0.1\) is coincidental. Fortunately, the ionization field on

\begin{align*}
\Delta x = 0.56 \text{ Mpc} & \quad \Delta x = 1.1 \text{ Mpc} \\
\text{z}=8.49, x_i=0.25 & \quad \text{z}=8.49, x_i=0.25 \\
\text{z}=7.66, x_i=0.51 & \quad \text{z}=7.56, x_i=0.51 \\
\text{z}=7.11, x_i=0.72 & \quad \text{z}=7.11, x_i=0.72
\end{align*}

Figure A1. The fraction of simulation cells with a given ionization fraction, at \((z, x_i) = (8.49, 0.25), (7.56, 0.51), (7.11, 0.72)\), top to bottom. Red circles, green triangles, blue squares and magenta pentagons correspond to the ionization algorithms of Trac & Cen (2007), McQuinn et al. (2007b), FFRT and FFRT without Poisson fluctuations in the halo number (see Section 4), respectively. The left-hand panel was created from the 256\(^3\) grid (cell size \(\Delta x = 0.39\) \(\text{Mpc}^{-1}\)) used throughout this paper, while the right-hand panel was created from a 128\(^3\) grid (cell size \(\Delta x = 1.1\) \(\text{Mpc}\)). Note that the distributions corresponding to the RT simulations in the right-hand panel were computed from just the (boxcar filter) smoothed 256\(^3\) grid. However, the FFRT ionization fields were generated directly from the low-resolution 128\(^3\) root grid of the low-resolution density field (i.e. the 128\(^3\) FFRT ionization fields are not just smoothed versions of the 256\(^3\) fields).

© 2011 The Authors, MNRAS 414, 727–738
Monthly Notices of the Royal Astronomical Society © 2011 RAS
these scales is well approximated by a binary (fully neutral or fully ionized) field, and so correctly computing the partial ionization fraction is not very important for observables such as the 21-cm power spectrum.

As discussed above, we expect our partial ionization scheme to be most effective when cell sizes are large, $M_{\text{cell}} \gg M_{\text{min}}$. Unfortunately, we do not have RT fields for very large scales and cell sizes with which to check this assumption. Instead, we approximate larger cell sizes by smoothing the $256^3$ density field down to $128^3$ and by using this lower resolution grid as the root grid for the FFRT algorithm. We compare the resulting distributions to the same RT fields (i.e. with the $256^3$ root grid), but smoothed down to $128^3$. The resulting distributions are shown in the right-hand panel of Fig. A1. These distributions are not as smooth as the $256^3$ ones, due to smaller number statistics. However, we can note that the FFRT scheme does in fact perform better on these $\Delta x = 1.1$ Mpc cells. This may or may not be a physical effect, due to the increasing relevance of subgrid sources in setting the partial ionization fraction of larger cells. We certainly expect agreement to improve with larger cell sizes.

Another interesting point evident from the right-hand panel of Fig. A1 is that not including the discreteness of sources on this cell size results in a notable lack of fully neutral pixels. This is to be expected since the mean value of the collapse fraction (which is the quantity directly computed in PS formalism) is never zero for $M_{\text{cell}} > M_{\text{min}}$. Therefore, including the discrete nature of sources, such as with our Poisson scatter approach, is important in identifying pristine, fully neutral regions on this scale.

This paper has been typeset from a \TeX/\LaTeX file prepared by the author.