



**HAL**  
open science

## Simulating cosmic reionization: how large a volume is large enough?

Ilian T. Iliev, Garrelt Mellema, Kyungjin Ahn, Paul R. Shapiro, Yi Mao, Ue-Li Pen

► **To cite this version:**

Ilian T. Iliev, Garrelt Mellema, Kyungjin Ahn, Paul R. Shapiro, Yi Mao, et al.. Simulating cosmic reionization: how large a volume is large enough?. Monthly Notices of the Royal Astronomical Society, 2014, 439, pp.725-743. 10.1093/mnras/stt2497 . insu-03645685

**HAL Id: insu-03645685**

**<https://insu.hal.science/insu-03645685>**

Submitted on 22 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Simulating cosmic reionization: how large a volume is large enough?

Ilian T. Iliev,<sup>1</sup>★ Garrelt Mellema,<sup>2</sup> Kyungjin Ahn,<sup>3</sup> Paul R. Shapiro,<sup>4</sup> Yi Mao<sup>5</sup>  
and Ue-Li Pen<sup>6</sup>

<sup>1</sup>*Astronomy Centre, Department of Physics & Astronomy, Pevensey II Building, University of Sussex, Falmer, Brighton BN1 9QH, UK*

<sup>2</sup>*Department of Astronomy & Oskar Klein Centre, Stockholm University, AlbaNova, SE-10691 Stockholm, Sweden*

<sup>3</sup>*Department of Earth Sciences, Chosun University, Gwangju 501-759, Korea*

<sup>4</sup>*Department of Astronomy, University of Texas, Austin, TX 78712-1083, USA*

<sup>5</sup>*UPMC Univ Paris 06, CNRS, Institut Lagrange de Paris, Institut d'Astrophysique de Paris, UMR7095, 98 bis, boulevard Arago, F-75014, Paris, France*

<sup>6</sup>*Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada*

Accepted 2013 December 30. Received 2013 December 30; in original form 2013 October 25

## ABSTRACT

We present the largest-volume ( $425 \text{ Mpc } h^{-1} = 607 \text{ Mpc}$  on a side) full radiative transfer simulation of cosmic reionization to date. We show that there is significant additional power in density fluctuations at very large scales. We systematically investigate the effects this additional power has on the progress, duration and features of reionization and on selected reionization observables. We find that comoving volume of  $\sim 100 \text{ Mpc } h^{-1}$  per side is sufficient for deriving a convergent mean reionization history, but that the reionization patchiness is significantly underestimated. We use jackknife splitting to quantify the convergence of reionization properties with simulation volume. We find that sub-volumes of  $\sim 100 \text{ Mpc } h^{-1}$  per side or larger yield convergent reionization histories, except for the earliest times, but smaller volumes of  $\sim 50 \text{ Mpc } h^{-1}$  or less are not well converged at any redshift. Reionization history milestones show significant scatter between the sub-volumes, as high as  $\Delta z \sim 1$  for  $\sim 50 \text{ Mpc } h^{-1}$  volumes. If we only consider mean-density sub-regions the scatter decreases, but remains at  $\Delta z \sim 0.1\text{--}0.2$  for the different size sub-volumes. Consequently, many potential reionization observables like 21-cm rms, 21-cm PDF skewness and kurtosis all show good convergence for volumes of  $\sim 200 \text{ Mpc } h^{-1}$ , but retain considerable scatter for smaller volumes. In contrast, the three-dimensional 21-cm power spectra at large scales ( $k < 0.25 \text{ h Mpc}^{-1}$ ) do not fully converge for any sub-volume size. These additional large-scale fluctuations significantly enhance the 21-cm fluctuations, which should improve the prospects of detection considerably, given the lower foregrounds and greater interferometer sensitivity at higher frequencies.

**Key words:** radiative transfer – methods: numerical – H II regions – galaxies: high-redshift – intergalactic medium – dark ages, reionization, first stars.

## 1 INTRODUCTION

The first billion years of cosmic evolution remain the only period in the history of the universe still largely unconstrained by direct observations. While we now have fairly detailed data on the cosmic microwave background (CMB) originating from the last scattering surface at redshift  $z \sim 1100$  and a wealth of multiwavelength observations at later times,  $z < 6$ , the intermediate period remains largely uncharted. A number of ongoing observational programs aim to provide observations of this epoch in e.g. high- $z$  Ly $\alpha$  (e.g. Krug et al. 2012), CMB secondary anisotropies (e.g. Zahn & et al. 2012) and redshifted 21-cm (e.g. Lonsdale & et al. 2009; Harker et al.

2010; Parsons & et al. 2010). Improved observational constraints could provide a wealth of information about the nature of the first stars and galaxies, their properties, abundances and clustering, the timing and duration of the reionization transition and the complex physics driving in this process.

However, most observations will yield only statistical measures, e.g. power spectra and auto- and cross-correlations, whose interpretation requires detailed simulations and modelling in order to extract the quantities of interest. In recent years there has been significant improvement in terms of the dynamic range of such simulations and their sophistication in implementing the variety of relevant physical processes (see e.g. Trac & Gnedin 2011, for a recent review).

One question that has not received much attention is what simulation volume is required for obtaining reliable results on the different aspects of cosmic reionization – for example the mean reionization

★E-mail: I.T.Iliev@sussex.ac.uk

history, the scales and distribution of the patchiness and the various corresponding observational features. Due to limited dynamic range and the necessity to resolve the low-mass galaxies driving reionization, the earliest numerical simulations of this period inevitably followed relatively small volumes, typically just a few comoving Mpc per side (Gnedin & Ostriker 1997; Ciardi et al. 2000; Gnedin 2000; Ricotti, Gnedin & Shull 2002). Such small-volume simulations yield a rather late, sharp and fast reionization occurring over a short interval in redshift. Subsequent simulations managed to increase the studied volumes to  $10\text{--}20 h^{-1}$  Mpc per side (Ciardi, Ferrara & White 2003; Sokasian et al. 2003). This resulted in more extended reionization histories, in better agreement with the then current constraints on the CMB optical depth for electron scattering (for a review see Spergel & et al. 2003).

Barkana & Loeb (2004) used theoretical calculations based on analytical halo mass functions and the extended Press–Schechter conditional mass function to show that local abundances of galaxies vary significantly on up to tens of Mpc scales. Consequently, periodic cosmological volumes of sizes less than this significantly underestimate the abundance of rare haloes (which at high redshifts are most or all of the star-forming haloes), with corresponding effects on the reionization observables. In particular, the scatter in the local reionization is greatly reduced in the latter case, resulting in quicker reionization than larger volumes would predict. The global mean halo collapsed fraction was shown to be largely converged for  $\sim 100$  Mpc volumes. However, the question what simulation volume is required for a particular reionization feature to be derived reliably was left open, to be tested and quantified by future simulations and observations.

The first cosmological reionization simulations at large scales (specifically,  $100 h^{-1}$  Mpc = 143 Mpc per side) were presented in Iliev et al. (2006). By sub-dividing this large volume into sub-volumes of different sizes and calculating their reionization history, they demonstrated that a mean-density volume of a few tens of Mpc per side is sufficient to derive the mean reionization history with a reasonable accuracy. However, the same work also showed that there is considerable scatter in the reionization histories for all (i.e. not just mean-density) volumes. This scatter becomes very large for  $5\text{--}10 h^{-1}$  Mpc volumes, to the point of making even the mean reionization history unreliable. The convergence of other properties with volume, e.g. the size distributions of the ionized regions, remained an open question since individual ionized regions can reach tens of Mpc across, similar to the size of the sub-divided volumes, and thus even such  $\sim 100 h^{-1}$  Mpc sized volumes are insufficient for such a study. Furthermore, the answer to this question likely depends on the property being studied. In the currently favoured  $\Lambda$  cold dark matter universe density fluctuations occur on all scales, albeit with diminishing amplitude for larger sizes, so any result should be quantified based on the level of convergence required.

Previously, reionization at large scales has only been approached by semi-numerical models (Kohler, Gnedin & Hamilton 2007; Santos et al. 2010). Kohler et al. (2007) used small-scale reionization simulations as sub-grid prescription in a very large, 1 Gpc, volume to study the regions around luminous quasi-stellar objects (QSOs) at the end of reionization. More recently, the results in Santos et al. (2010), Mesinger, Furlanetto & Cen (2011) indicated that the additional large-scale density power does result in larger ionized patches and increased 21-cm fluctuations. The semi-numerical approach, whereby the reionization transition is modelled based on analytical prescriptions for the local halo abundances and clustering combined with smoothed density fields generated as initial conditions for  $N$ -body simulations, has the advantage that it is computationally

much cheaper than full simulations. This allows for variation of the unknown parameters and thus testing a variety of models. However, the available semi-numerical and semi-analytical models generally do not include many important reionization features, such as the suppression of low-mass sources due to Jeans-mass filtering, recombinations or non-linear halo clustering, which limits their applicability. Furthermore, to date the semi-numerical models have only been tested against simulations at smaller scales, below  $\sim 100$  Mpc, for lack of larger-scale simulations. Such larger-scale simulations would therefore be also useful for verification and further development of semi-numerical modelling.

In addition to being able to probe very large scales, simulations of at least a few hundred comoving Mpc per side also have the advantage of matching better wide-field surveys of high-redshift structures such as the one currently conducted by LOFAR, with field-of-view (FOV) of a few tens of square degrees. The first detections are likely to be statistical, measuring the redshifted 21-cm power spectra and variance (Harker et al. 2009; Iliev et al. 2012), although a limited form of imaging might be possible in some cases (Datta et al. 2012; Zaroubi et al. 2012). Another important advantage of large simulation volumes is that they provide a wider range of environments (e.g. larger, deeper voids, higher density peaks, large-scale bulk motions) and much richer statistics of objects, particularly very rare ones which might not be found at all in a smaller volume (e.g. Watson et al. 2014), for example luminous high-redshift QSOs as found by e.g. Sloan Digital Sky Survey (Fan & et al. 2001). In fact, the simulation we present here has already been used to study the impact and detectability of a rare QSO (Datta et al. 2012).

In this work we present the first direct, radiative transfer (RT) simulation of reionization at very large scales, in a comoving volume of  $425 h^{-1}$  Mpc = 607 Mpc per side, and investigate the effects of the previously ignored long wavelength fluctuations on the formation of early structures and on the progress and features of cosmic reionization. The structure of this paper is as follows: in Section 2 we present our simulations; in Section 3.1 we discuss the formation of early structures; the effects on the reionization patchiness is presented in Section 3.2; the convergence with volume of the reionization history, its milestones and some of the resulting observables are discussed in Section 3.3; finally, our summary and conclusions are given in Section 4.

Results of the detailed, large-volume RT +  $N$ -body simulation of reionization described here have already been applied elsewhere to predict some additional observational probes of reionization, for which the large box size had distinct advantages. Park et al. (2013) used these simulations, along with those in boxes of smaller size,  $114 h^{-1}$  Mpc = 163 Mpc on a side, to predict the contribution of patchy reionization to the secondary CMB temperature fluctuations from the kinetic Sunyaev–Zel’dovich (kSZ) effect. We demonstrated there that a correction based upon linear perturbation theory for the growth of the underlying matter density and velocity fluctuations are necessary in order to account for the missing power on large scales when predicting the effect of patchy reionization on the kSZ using the smaller-box simulations, while finding that this correction is small for the larger box size. Shapiro et al. (2013) and Jensen et al. (2013) applied these big-box simulation results to model the anisotropy of the 3D power spectrum of 21-cm brightness temperature fluctuations from the epoch of reionization caused by the redshift-space distortion introduced by the Doppler shift from velocity fluctuations along the line of sight, to determine if measurements could be used to separate out the effects of cosmology from astrophysics. A large box was required to be able to make predictions

at the small wavenumbers that will be probed by upcoming 21-cm radio surveys, with enough modes in the box to make the sampling variance small enough. Finally, Datta et al. (2012) used this large-box reionization simulation to predict the signature of luminous QSOs in the 21-cm background maps, to show that a matched filter technique could detect the ionized patches surrounding the QSOs as a gap in the signal. This required a box large enough to find a rare galactic halo massive enough to host a bright QSO, while also much larger than the size of the  $H\text{II}$  region that might surround such a source.

Here, we will describe the underlying, large-box RT +  $N$ -body simulation, itself, in some detail, and address a range of measures of the effect of box size on the results for patchy reionization and its observable consequences, with the latter focused on statistical predictions of the 21-cm background from the epoch of reionization.

The simulations presented in this work use the following set of cosmological parameters  $\Omega_\Lambda = 0.73$ ,  $\Omega_M = 0.27$ ,  $\Omega_b = 0.044$ ,  $h = 0.7$ ,  $\sigma_8 = 0.8$ ,  $n_s = 0.96$  where  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ , consistent with the *Wilkinson Microwave Anisotropy Probe* (WMAP) 5-year data (Komatsu et al. 2009) and the recent Planck results (Planck Collaboration et al. 2013).

## 2 SIMULATIONS

We simulate the formation of high-redshift cosmic structures using a set of high-resolution  $N$ -body simulations, summarized in Table 1. Our main simulation has  $5488^3$  (165 billion) particles in a volume of  $425 h^{-1} \text{ Mpc}$  per side. For comparison, we also consider two smaller-volume simulations with a slightly higher resolution, each following  $3072^3$  particles in a volume of  $114 h^{-1} \text{ Mpc}$  per side and each with a different random realization of the initial Gaussian-random field to control for cosmic variance. The simulations were performed with the  $P^3M$   $N$ -body code Cube $P^3M$  (Harnois-Déraps et al. 2013) on the (now retired) *Ranger* computer at the Texas Advanced Computing Center (TACC; SunBlade  $\times 6420$  with AMD  $\times 86$  64 Opteron Quad Core, 2.3 GHz, 9.2 GFlops/core Barcelona processors and Infiniband networking). The smaller-volume,  $114 h^{-1} \text{ Mpc}$  simulations were previously presented in Iliev et al. (2012), while the large-volume simulation was part of the suite of simulations used in Watson et al. (2013) for deriving the halo mass function and its evolution with redshift. The large-volume,  $425 h^{-1} \text{ Mpc}$ -box,  $N$ -body simulation required 2 million core-hours to reach redshift  $z = 6$  and was ran using 2744 MPI processes. Each of MPI nodes initially contained four OpenMP-threaded computing cores and 8 GB shared memory (for a total of 10 976 cores). During the later stages of the simulation, strong clustering meant that the size of an MPI node had to be increased to eight cores and 16 GB memory (for a total of 21 956 cores) to accommodate the larger memory requirement. The Cube $P^3M$  code shows an excellent (weak) scaling up to 4096 cores and very good scaling up to 22 000 computing cores. For full details on the Cube $P^3M$  code and its performance see Harnois-Déraps et al. (2013).

All simulations were started at redshift  $z = 300$  and with initial conditions generated using the Zel'dovich approximation and power

spectrum of the linear fluctuations given by the CAMB code (Lewis, Challinor & Lasenby 2000). This starting redshift is sufficiently early to guarantee linearity of the density field even at such high resolution and thus avoiding numerical artefacts (Crocce, Pueblas & Scoccimarro 2006), see also Watson et al. (2013) for discussion of this point regarding the current simulations. The smaller-volume simulations were ran until redshift  $z = 6$ , and the larger-volume to  $z = 2.6$ . We have identified collapsed haloes at 76 redshifts equally spaced in time (every 11.5 Myr) between  $z = 30$  and  $z = 6$  using an on-the-fly halo finder based on the spherical overdensity method with overdensity of  $\Delta = 178$  with respect to the mean density. We also calculated the density and velocity distributions on several regular grids (from  $252^3$  up to  $2016^3$  for the  $425 h^{-1} \text{ Mpc}$  box and  $256^3$  and  $512^3$  for the  $114 h^{-1} \text{ Mpc}$  boxes) using an SPH-like smoothing over the particle distributions, which produces smoother distributions than other commonly used interpolation schemes [e.g. cloud-in-cell (CIC), nearest grid point (NGP)] and minimizes noise in low-density regions. These gridded data were then used as an input to the RT simulations of reionization discussed below, as well as for subsequent analysis of the results from those simulations.

The halo catalogues are used to construct the sources of ionizing radiation, largely using the approach introduced in Iliev et al. (2006, 2007), with some important modifications. The smaller-volume, higher-resolution simulations resolve all haloes with mass of  $10^8 M_\odot$  or more, i.e. roughly all atomically cooling haloes (ACHs), which we split into low-mass atomically cooling haloes (LMACHs), which are affected by radiative feedback and high-mass atomically cooling haloes (HMACHs) which are not. On the other hand, the larger-volume simulation only resolves haloes of mass  $10^9 M_\odot$  and above, corresponding to the HMACHs only. Consequently, for the reionization simulation we model the LMACHs as a sub-grid population using a recently developed extended Press–Schechter type model (Ahn et al., in preparation). We do this by calculating a cell-by-cell local LMACH population based on a fit to the data from higher-resolution simulations. This yields the local LMACH collapsed fraction as function of the cell density. Our procedure reproduces the mean halo mass function for the whole box and reflects the halo clustering down to the grid scale, but at present it does not account for the (modest) scatter in the local halo number–overdensity relation observed numerically. We then combine the local LMACH collapsed fraction yielded by our model with the directly resolved HMACH population binned on to the same grid cells to create the complete ionizing source lists. We do not include minihaloes (MH; haloes with masses  $10^5 \lesssim M/M_\odot \lesssim 10^8$ ), even though they are expected to play an important role in the earliest stage of reionization, as previously shown in Ahn et al. (2012). The reason for this is that star formation in MHs is more uniformly distributed on large scales and is strongly regulated by the long-range Lyman–Werner feedback, which further smooths out their distribution and thus we expect their effect to be less correlated with large-scale modes in our  $425 h^{-1} \text{ Mpc}$  volume. We defer the study of these effects to future work.

Another new element in the current RT simulations is that we also include an approximate treatment of the absorber systems which

**Table 1.** Simulation parameters. Background cosmology is based on the WMAP 5-year results.

Simulation	Boxsize	$N_{\text{part}}$	Mesh	Force softening	$m_{\text{particle}}$	$M_{\text{halo, min}}$	RT grid	$g_\gamma$ , HMACH	$g_\gamma$ , LMACH
XL2	$425 h^{-1} \text{ Mpc}$	$5488^3$	$10976^3$	$3.88 h^{-1} \text{ kpc}$	$4.97 \times 10^7 M_\odot$	$9.93 \times 10^8 M_\odot$	$504^3$	1.7	7.1
L2.1	$114 h^{-1} \text{ Mpc}$	$3072^3$	$6144^3$	$1.86 h^{-1} \text{ kpc}$	$5.47 \times 10^6 M_\odot$	$1.09 \times 10^8 M_\odot$	$256^3$	1.7	7.1
L2.2	$114 h^{-1} \text{ Mpc}$	$3072^3$	$6144^3$	$1.86 h^{-1} \text{ kpc}$	$5.47 \times 10^6 M_\odot$	$1.09 \times 10^8 M_\odot$	$256^3$	1.7	7.1

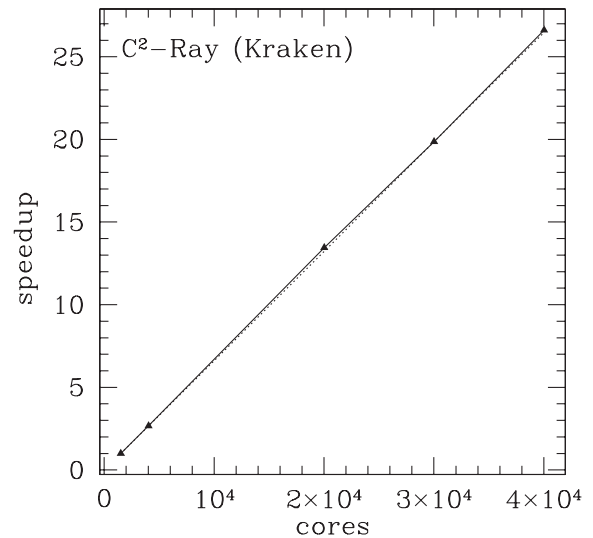
limit the mean-free path of the ionizing photons during the late stages of reionization (Songaila & Cowie 2010). Numerically, this is done by limiting the photon propagation to a box of  $80 h^{-1}$  Mpc comoving centred on each source. This size roughly corresponds to the observed mean-free path due to Lyman Limit Systems at  $z \sim 6$ . Any photons reaching that boundary are presumed lost, absorbed by the Lyman-limit systems.

Once the source lists are constructed, we assign to each an ionizing photon production rate per unit time,  $\dot{N}_\gamma$ , proportional to the total mass in haloes within that cell,  $M$  as previously discussed in Iliev et al. (2006, 2007, 2012):

$$\dot{N}_\gamma = \frac{g_\gamma M \Omega_b}{\Omega_0 m_p} \left( \frac{10 \text{ Myr}}{\Delta t} \right), \quad (1)$$

where  $m_p$  is the proton mass and  $f_\gamma = f_{\text{esc}} f_\star N_\star$  is an ionizing photon production efficiency parameter which includes the efficiency of converting gas into stars,  $f_\star$ , the ionizing photon escape fraction from the halo into the IGM,  $f_{\text{esc}}$ , and the number of ionizing photons produced per stellar atom,  $N_\star$ . The latter parameter depends on the assumed IMF for the stellar population and could vary significantly, between  $\sim 4000$  (Pop II, Salpeter IMF) and  $\sim 100\,000$  (Pop III, top-heavy IMF). HMACH and LMACH haloes are assigned different luminosities, based on the reasonable assumption that the latter would typically be less metal-enriched and thus their stellar content should be more dominated by massive stars, which are efficient producers of ionizing photons.  $\Delta t = 11.46$  Myr is the time between two snapshots from the  $N$ -body simulation. For all RT simulations in this paper we adopt  $g_{\gamma, \text{LMACHs}} = 7.1$  and  $g_{\gamma, \text{HMACHs}} = 1.7$ , which values yield a reionization history consistent with the current observational constraints (Iliev et al. 2012), with electron-scattering optical depth  $\tau = 0.0566$  and overlap reached at redshift  $z = 6.5$ . LMACHs are assumed to be suppressed within ionized regions (for ionization fraction higher than 10 per cent), through Jeans-mass filtering, as discussed in Iliev et al. (2007).

Due to the nature of its algorithm, our  $\text{C}^2\text{-Ray}$  RT code is quite flexible and can run on serial, shared-memory (using OpenMP) or distributed parallel systems (using MPI), as well as on hybrid systems (OpenMP+MPI), with the latter being the most common mode of usage. The code can run on any number of computational cores, as long as there is sufficient memory per processor/MPI process. The  $425 h^{-1}$  Mpc-box RT simulation presented here was performed on several different clusters, utilizing our available allocations. The initial phases were ran on the Apollo cluster at The University of Sussex (168 cores at the time) and the Sciama SEPNet computer (1008 cores) at University of Portsmouth. The later stages of the evolution, when many more sources needed to be handled, were ran on larger clusters, the TACC *Ranger* and *Lonestar* (22 656 cores, 44.3 TB RAM, 302 TFlops peak) clusters, the *Kraken* Cray XT-5 (112 896 compute cores, 147 TB RAM, 1.17 PetaFlops) at the National Institute for Computational Sciences/University of Tennessee (NICS) and the SciNet computer at University of Toronto (30 240 cores, 60 TB RAM, 306 TFlops peak). The simulation was performed over 10 months and used up a total of approximately 8 million *Lonestar*-equivalent (3.3 GHz, 13.3 GFlops) core-hours. At different stages of the computation we used between 32 and 40 032 cores simultaneously. For large enough problems and on machines with fast communication network like the NICS *Kraken* the code scaling is essentially perfect as far as we have tested it, here on up to 40 032 computing cores (see Fig. 1).



**Figure 1.** Computational speedup (normalized to the smallest run) versus number of computing cores (solid line) and the ideal linear weak scaling (dotted line) for the  $\text{C}^2\text{-Ray}$  code on *Kraken* (Cray XT5) at University of Tennessee.

## 3 RESULTS

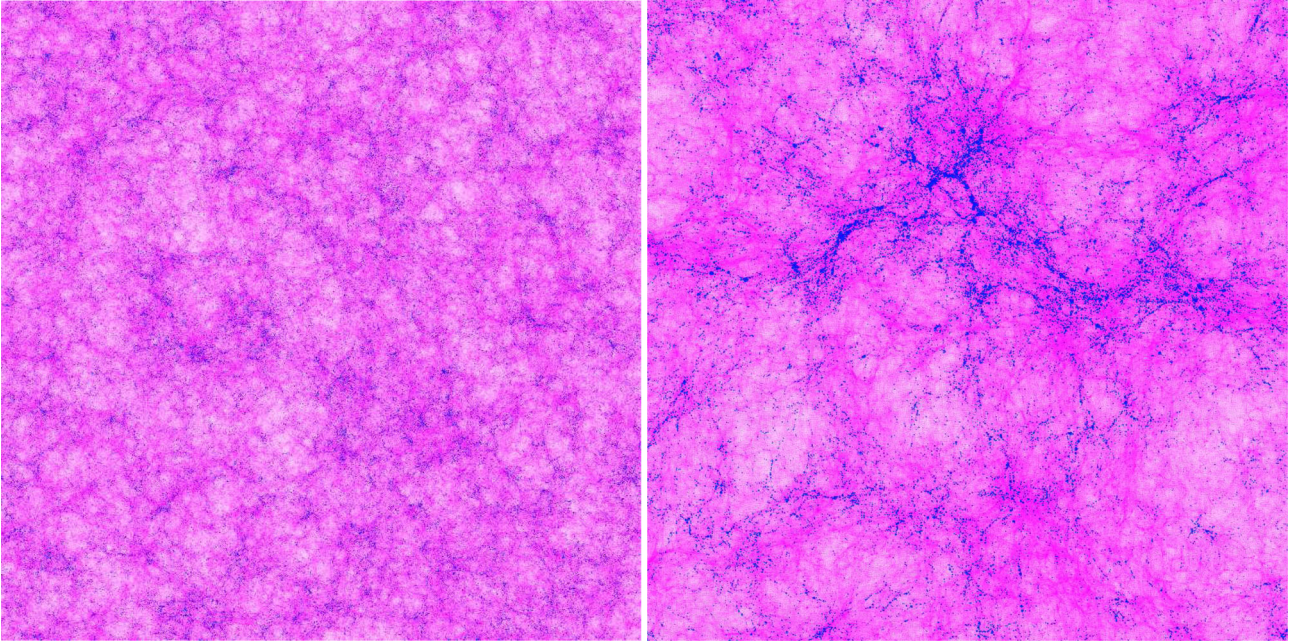
### 3.1 Early cosmic structure formation

In Fig. 2 we show an illustrative image of the cosmic web at redshift  $z = 6$  extracted from our  $425 h^{-1}$  Mpc volume. It shows a  $15 h^{-1}$  Mpc thick slice of the density distribution with the haloes (in their actual sizes) overlaid. It is clear that while the universe is fairly homogeneous at such large scales, we can clearly observe that the long-wavelength density modes result in significant density fluctuations even at scales as large as tens to 100 Mpc. As we will see below, these large-scale variations have important implications for the reionization patchiness on large scales. The  $N$ -body haloes are strongly clustered around the high-density peaks, again at scales of up to tens of Mpc, and along the filaments and there are very large, with sizes of up to 50 Mpc or more, underdense regions mostly devoid of haloes.

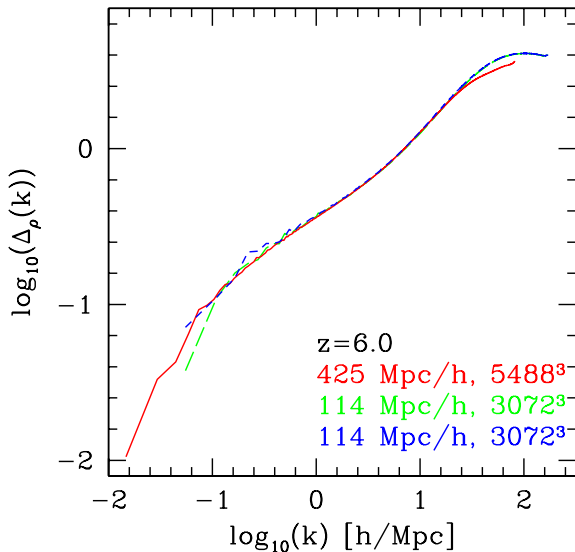
#### 3.1.1 Density power spectra and halo bias

In Fig. 3 we show the dimensionless power spectra of the density field,  $\Delta_\rho = (k^3 P(k)/2\pi^2)^{1/2}$ , for all three simulations. The density fields were obtained by interpolating the  $N$ -body particles on a regular grid ( $10\,976^3$  for the  $425 h^{-1}$  Mpc volume and  $6144^3$  for the  $114 h^{-1}$  Mpc volumes) using CIC interpolation. All three power spectra agree closely over the wavenumber range which they all resolve well,  $-0.5 < \log_{10}(kh/\text{Mpc}) < 1.4$ . At smaller scales (higher  $k$ ) there is less power in the large box due to its slightly lower resolution, while large scales, near the respective box sizes, are affected by cosmic variance. Apart from these expected differences, in the  $425 h^{-1}$  Mpc box there is clearly significant additional power at large scales ( $k \lesssim 0.06 h/\text{Mpc}$ , corresponding to scales of  $\gtrsim 100 h^{-1}$  Mpc), which fluctuations are completely missing in the smaller volumes.

The halo bias derived from our simulations, defined as  $b_{\text{hh}} = \Delta_{\text{hh}}/\Delta_\rho$ , where  $\Delta_{\text{hh}}$  is the halo-halo auto-correlation power spectrum calculated on the same grid as the matter power spectrum

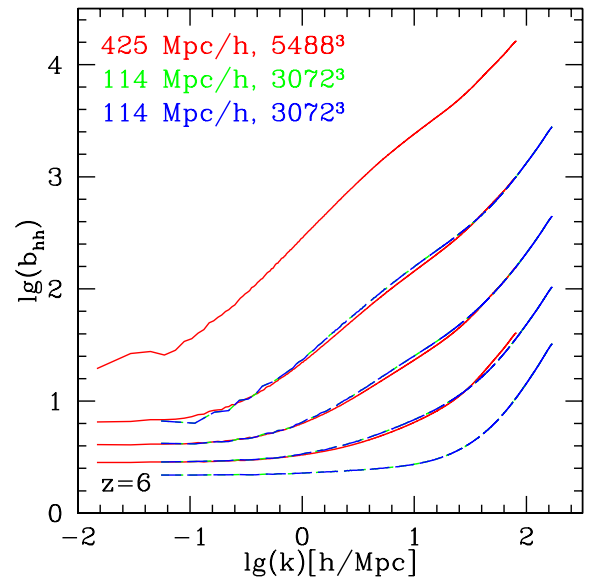


**Figure 2.** Large-scale structures and cosmic web at redshift  $z = 6$  from our CubeP<sup>3</sup>M simulation with  $5488^3$  particles (165 billion) on a  $10976^3$  fine grid in a comoving volume of  $425/h = 607$  Mpc on a side: (left) full simulation volume, and (right) zoomed-in  $41.5 \times 41.5$  Mpc region of the same image. Shown are the density field (violet) and haloes (blue). Slices are  $15 h^{-1}$  Mpc thick.



**Figure 3.** Power spectrum of the density field at redshift  $z = 6$  for the  $425 h^{-1}$  Mpc volume (solid, red) and two realizations of  $114 h^{-1}$  Mpc volume (dashed, green and blue).

$\Delta_\rho$ , again using CIC interpolation, is shown in Fig. 4. We plot the halo bias in five mass bins, one per each decade in mass, starting from  $10^8 M_\odot$ . Due to resolution the lowest mass bin,  $10^8 M_\odot \leq M_{\text{halo}} < 10^9 M_\odot$  (LMACHs), is only present in the smaller,  $114 h^{-1}$  Mpc boxes, while the largest-mass bin,  $M_{\text{halo}} \geq 10^{12} M_\odot$ , is absent for the same smaller boxes due to lack of statistics, there are only a few such large haloes present in those volumes. The results from all three boxes show an excellent agreement at all scales and masses present in both simulation volumes, with the two  $114 h^{-1}$  Mpc box simulations yielding essentially identical results.



**Figure 4.** The halo bias,  $b_{\text{hh}} = \Delta_{\text{hh}}/\Delta_\rho$ , at redshift  $z = 6$  for the  $425 h^{-1}$  Mpc volume (solid, red) and two realizations of  $114 h^{-1}$  Mpc volume (dashed, green and blue). Lines are for haloes binned by decades of mass (bottom to top curve)  $10^8 M_\odot \leq M_{\text{halo}} < 10^9 M_\odot$  (LMACHs),  $10^9 M_\odot \leq M_{\text{halo}} < 10^{10} M_\odot$ ,  $10^{10} M_\odot \leq M_{\text{halo}} < 10^{11} M_\odot$ ,  $10^{11} M_\odot \leq M_{\text{halo}} < 10^{12} M_\odot$  and  $10^{12} M_\odot \leq M_{\text{halo}}$  (HMACHs all).

At small scales the bias is highly non-linear and can reach quite high values. At large scale all the curves asymptote to a constant value, which for the lower mass bins roughly corresponds to the large-scale bias derived in Sheth, Mo & Tormen (2001). However, for higher masses (i.e. very rare haloes) the numerically calculated bias is well above the analytical prediction, indicating the latter is not valid for such rare haloes.

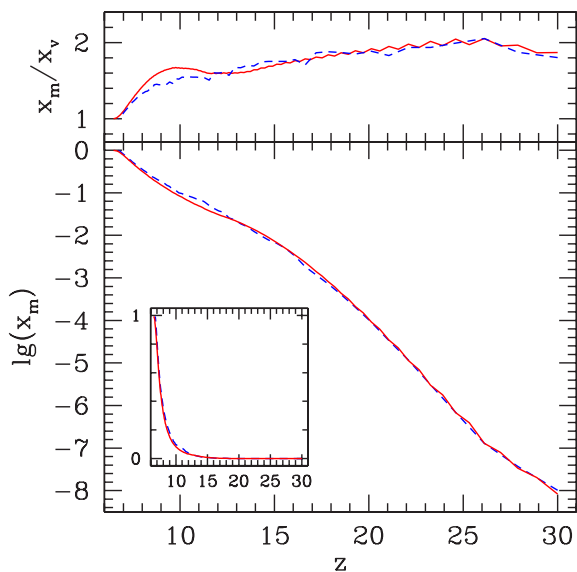
### 3.1.2 Dark matter halo mass function

In a recent paper (Watson et al. 2013), we presented detailed results and fits of the dark matter halo mass functions based on a large ensemble of  $N$ -body simulations which included the ones used here. This was done based on three different methods for halo finding and included precision fits to the mass function, including one specifically targeted at high-redshift data. The results confirmed the trend observed previously by us and others (Iliev et al. 2006, 2012; Reed et al. 2007; Lukić et al. 2007), namely that at the high redshifts of interest here the halo mass functions differ significantly from both the Press–Schechter (Press & Schechter 1974) and Sheth–Tormen (Sheth & Tormen 2002) analytical fits. This departure is particularly important for very rare haloes (corresponding to the massive haloes at the lower end of the redshift range we consider,  $z \sim 6$ –7, and to all ACHs at higher redshifts), whose abundance is very strongly underestimated by Press–Schechter and over-estimated by Sheth–Tormen (both by up to an order of magnitude at e.g.  $z = 17$ ). Therefore, such analytical approximations should not be used in analytical and semi-numerical modelling involving high-redshift haloes and instead more precise fits, such as the ones we presented in Watson et al. (2013), should be applied.

## 3.2 Large-scale cosmic reionization

### 3.2.1 Reionization history

The reionization history resulting from our RT simulations is shown in Fig. 5. In all simulations the first sources appear at  $z \sim 30$  (in the large-box case those sources are sub-grid) and the end-of-reionization (defined by ionized fraction by mass of  $x_m = 0.99$ ) is reached at similar redshifts of  $z = 6.5$  and  $6.6$  for XL2 and L2.1, respectively. The evolution of the ionized fractions is roughly exponential at early times, tracking the exponential rise in the number of (mostly LMACH) ionizing sources. The reionization history thereafter flattens starting at  $\sim 15$ , due to the LMACHs suppression becoming more wide-spread, and finally again becomes exponential



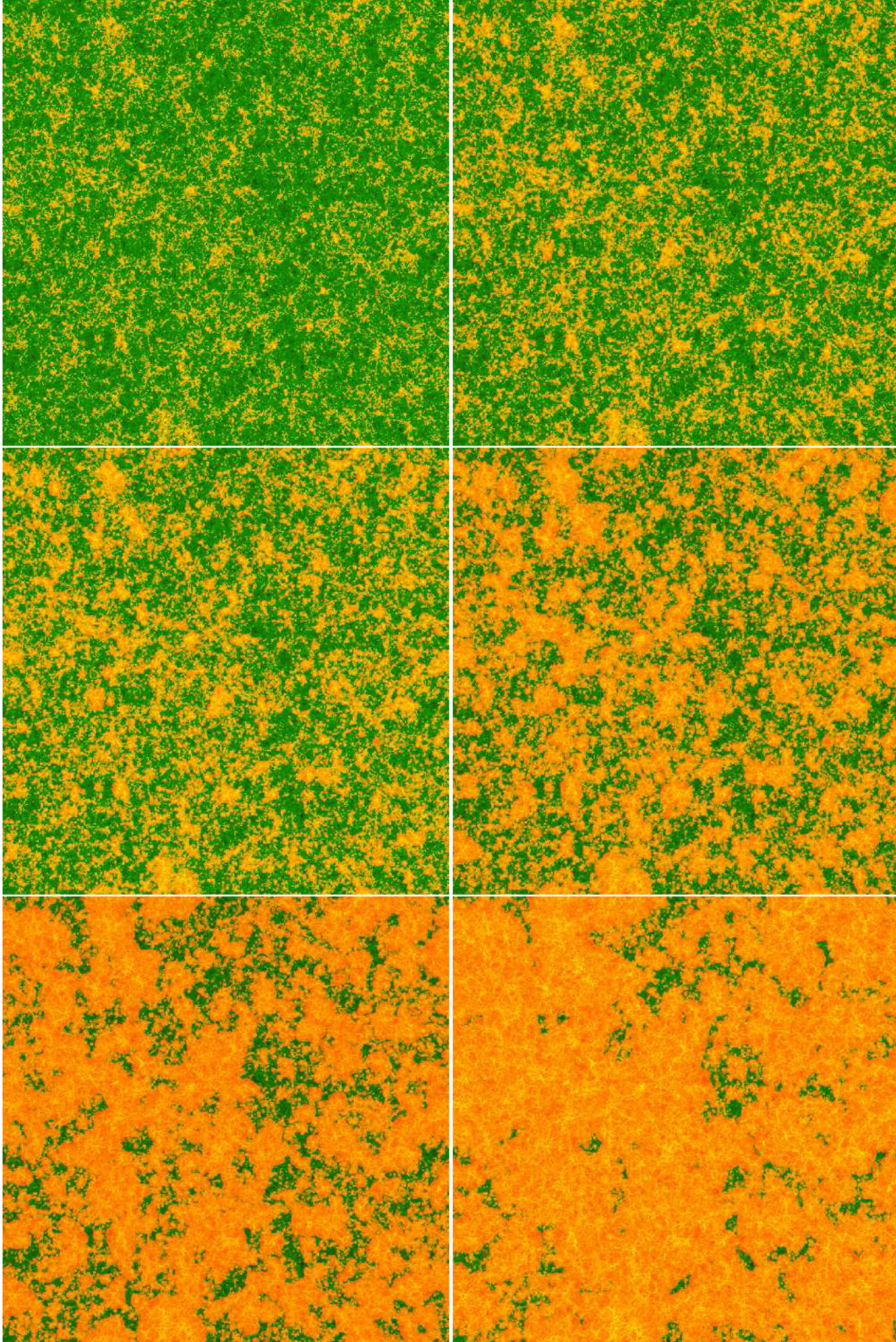
**Figure 5.** Bottom: mass-weighted reionization histories,  $x_m$ , for XL2 (red, solid) and L2.1 (blue, dashed). Top: ratio of the mass- and volume-weighted mean ionized fractions ( $x_m$  and  $x_v$ , respectively)

starting from at  $z \sim 8$ , at which point the unsuppressible HMACH sources come to dominate the evolution. The overall reionization history is near-identical for the two volumes due to the same assumed ionizing photon efficiencies  $g_\gamma$  for both types of sources, LMACHs and HMACHs. The only notable differences between the two runs are observed for  $z \sim 10$ –12, where the reionization of the larger volume slightly lags the smaller one. The reason for this lag is the different modelling of the LMACH sources, which is based on the directly resolved haloes (L2.1) or as a sub-grid population (XL2). As we discussed above, the latter is solely based on the average relation between collapsed fraction in haloes and the local density and at present does not model the scatter around that mean relation observed in simulations. This results in slightly higher suppression of LMACHs due to their increased local clustering. The inside-out character of reionization, evidenced by the ratio of mass-weighted to volume-ionized fraction,  $x_m/x_v$ , being always above one (Iliev et al. 2006) is largely unaffected by the size of the simulation volume and the difference in grid resolution.

### 3.2.2 Reionization geometry and $H\text{II}$ region size distributions

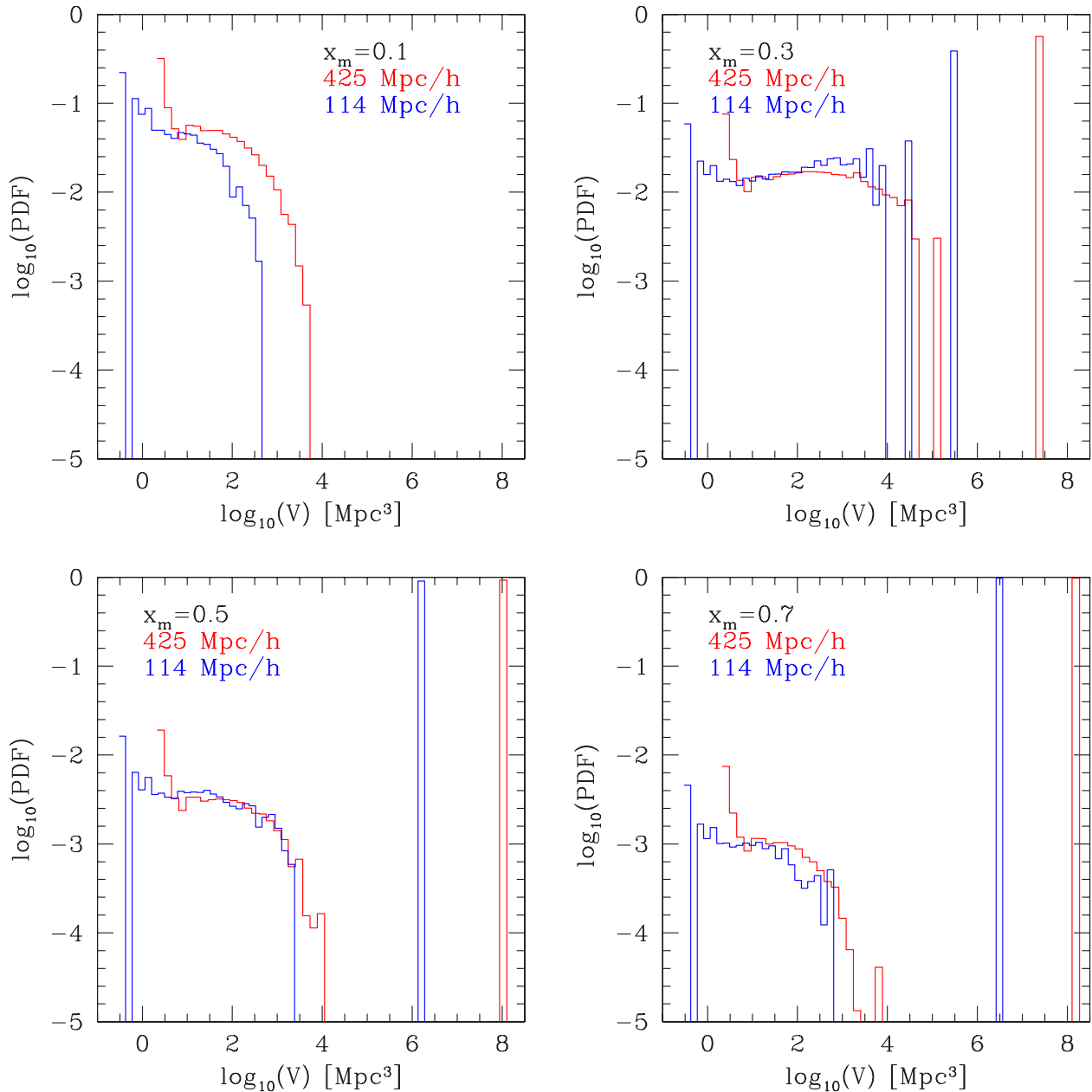
The significant density fluctuations at very large scales, from tens to hundreds of comoving Mpc, and the strong clustering of high- $z$  galaxies discussed on the previous section, could be expected to result in enhanced large-scale patchiness of reionization. In Fig. 6 we illustrate the reionization geometry at several key stages of the evolution, corresponding to mass-weighted ionized fraction of  $x_m = 0.1, 0.2, 0.3, 0.5, 0.7$  and  $0.9$ , or a redshift range of  $z = 9.6$  to  $z = 6.8$ . Initially, a large number of fairly small, Mpc-sized  $H\text{II}$  regions form. They are strongly clustered on small scales, following the clustering of the sources, but are relatively uniformly distributed on larger scales, where the long-wavelength modes modulating the density fluctuations are still very low. Locally these small  $H\text{II}$  regions quickly start merging into larger ones, with sizes between a few and  $\sim 10$  Mpc across. We note that of course these are 2D cuts of the ionization field and  $H\text{II}$  regions can, and do, have different sizes depending on the direction, as quantified e.g. in Iliev et al. (2008). Significant large-scale percolation of the  $H\text{II}$  regions only occurs when the universe reaches  $\sim 50$  per cent ionization by mass, at which point many ionized regions reach sizes of tens of Mpc and become connected by bridges to other nearby large ionized regions of similar sizes. At the same time there are also similarly sized regions which remain largely neutral. The  $H\text{II}$  regions continue percolating up to still larger scales and by  $x_m = 0.7$  some reach hundreds of Mpc across, with large, while at the same time large, tens of Mpc across, neutral regions remain between them, both reflecting the large-scale fluctuations of the underlying density field. Finally, when the volume is 90 per cent ionized all  $H\text{II}$  regions have percolated into one, although fairly large neutral regions still exist even in this late phase.

These qualitative observations are further supported by more quantitative measures of the size distributions of the ionized regions obtained based on different approaches: the friends-of-friends (FOF) method (Iliev et al. 2006; Friedrich et al. 2011), the spherical average method (SPA; McQuinn et al. 2007; Zahn et al. 2007) and the three-dimensional power spectra of the ionization fraction field (Iliev et al. 2006; Friedrich et al. 2011). The results from the FOF method, which links together topologically connected ionized regions, at several representative stages of reionization and our two different simulation volumes are shown in Fig. 7. During the early stages of reionization (ionized fraction by mass of  $x_m = 0.1$ ) the two distributions have similar shape, with the



**Figure 6.** Left to right and top to bottom: spatial slices of the ionized and neutral gas density from our XL2 RT simulation with boxsize  $425/h$  Mpc at box-averaged ionized fractions by mass of  $x_m = 0.1$  ( $z = 9.6$ ),  $x_m = 0.2$  ( $z = 8.6$ ),  $x_m = 0.3$  ( $z = 8.1$ ),  $x_m = 0.5$  ( $z = 7.5$ ),  $x_m = 0.7$  ( $z = 7.1$ ),  $x_m = 0.9$  ( $z = 6.8$ ). Shown are the density field (green) overlaid with the ionized fraction (orange/yellow).



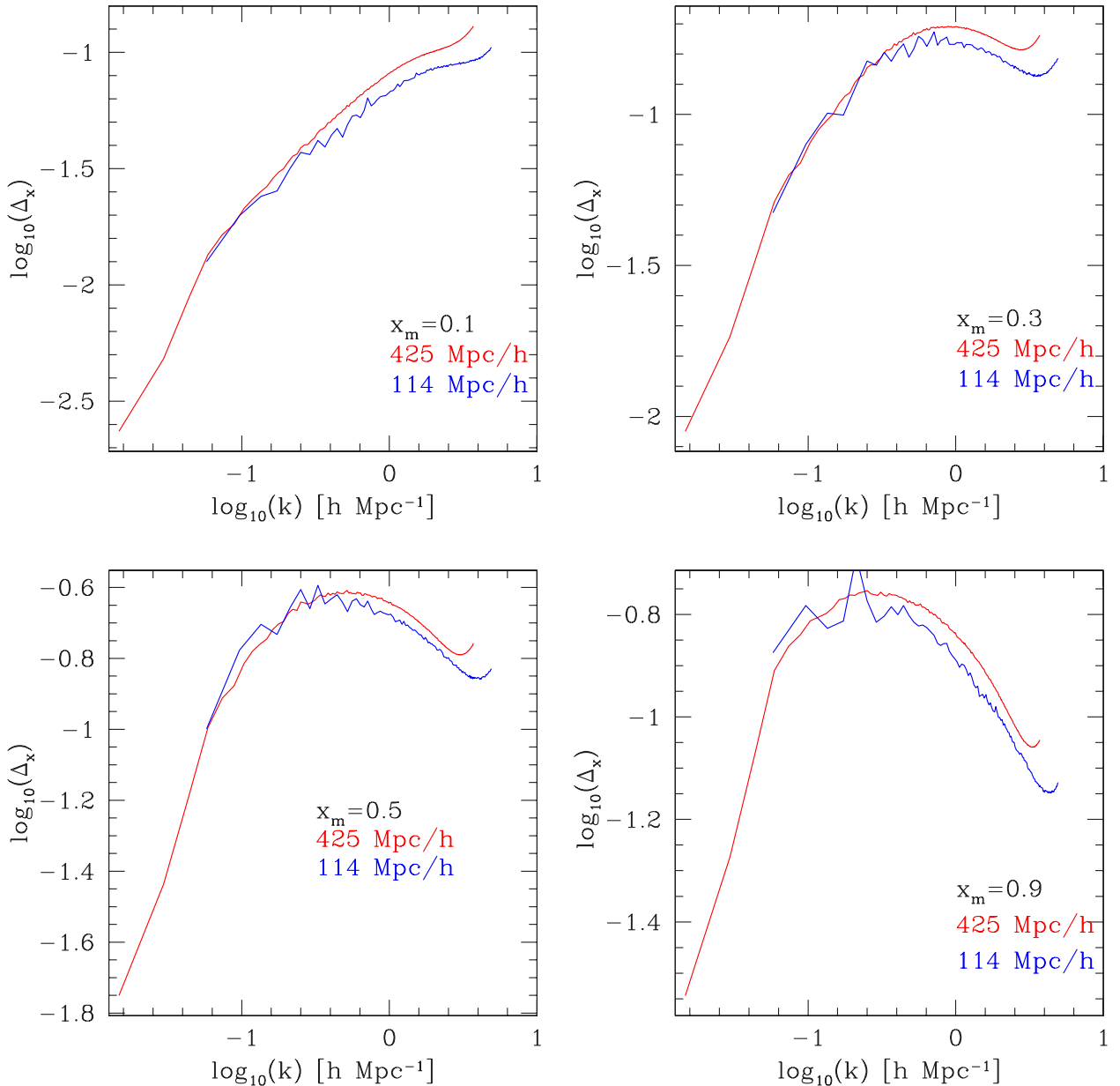


**Figure 7.** Volume-weighted probability distribution function,  $V(dp/dV)$ , for H II regions extracted from our simulations at different stages of the reionization process, as labelled. Shown are the results based on XL2 (red) and L2.1 (blue).

majority of the ionized volume found in small, but numerous H II regions with typical sizes ranging from less than a Mpc up to a few Mpc. The fraction of the simulation volume occupied by the smallest regions which are resolved in both simulations ( $V \sim 10 \text{ Mpc}^3$ ) is very similar. However, in XL2 there are considerably larger ionized regions ( $V > 30 \text{ Mpc}^3$ ). In L2.1 there are no H II regions with volume larger than about  $500 \text{ Mpc}^3$ , while the largest ones in XL2 reach volumes in excess of  $10^3 \text{ Mpc}^3$  even at this early stage, indicating a longer characteristic scale for local percolation in the larger volume.

The ionized regions continue to grow over time and at  $x_m = 0.3$  we already observe considerable local percolation, whereby multiple H II regions merge into a single very large one, with a volume filling a significant fraction of the simulation box. Similar behaviour

was previously observed in Iliev et al. (2006) and Friedrich et al. (2011). For both XL2 and L2.1 this percolated region fills roughly 10 per cent of the total volume at this stage of the evolution. Again, there are significant differences between the two boxes in terms of the abundances of the smaller H II regions. The resolved regions in the two simulations are in excellent agreement up to  $V \sim 10^4 \text{ Mpc}^3$ . However, in the smaller volume there are no ionized regions with volume above  $V \sim 3 \times 10^4 \text{ Mpc}^3$ , while such regions still take a significant fraction of the large volume. Similar trends are observed during the later stages of the evolution ( $x_m > 0.5$ ). The percolation of the ionized regions proceeds much further, with one large, connected region filling most of the volume in either case (with its volume limited only by the total simulation volume), while the smaller regions decrease in both abundance and size. There is agreement in

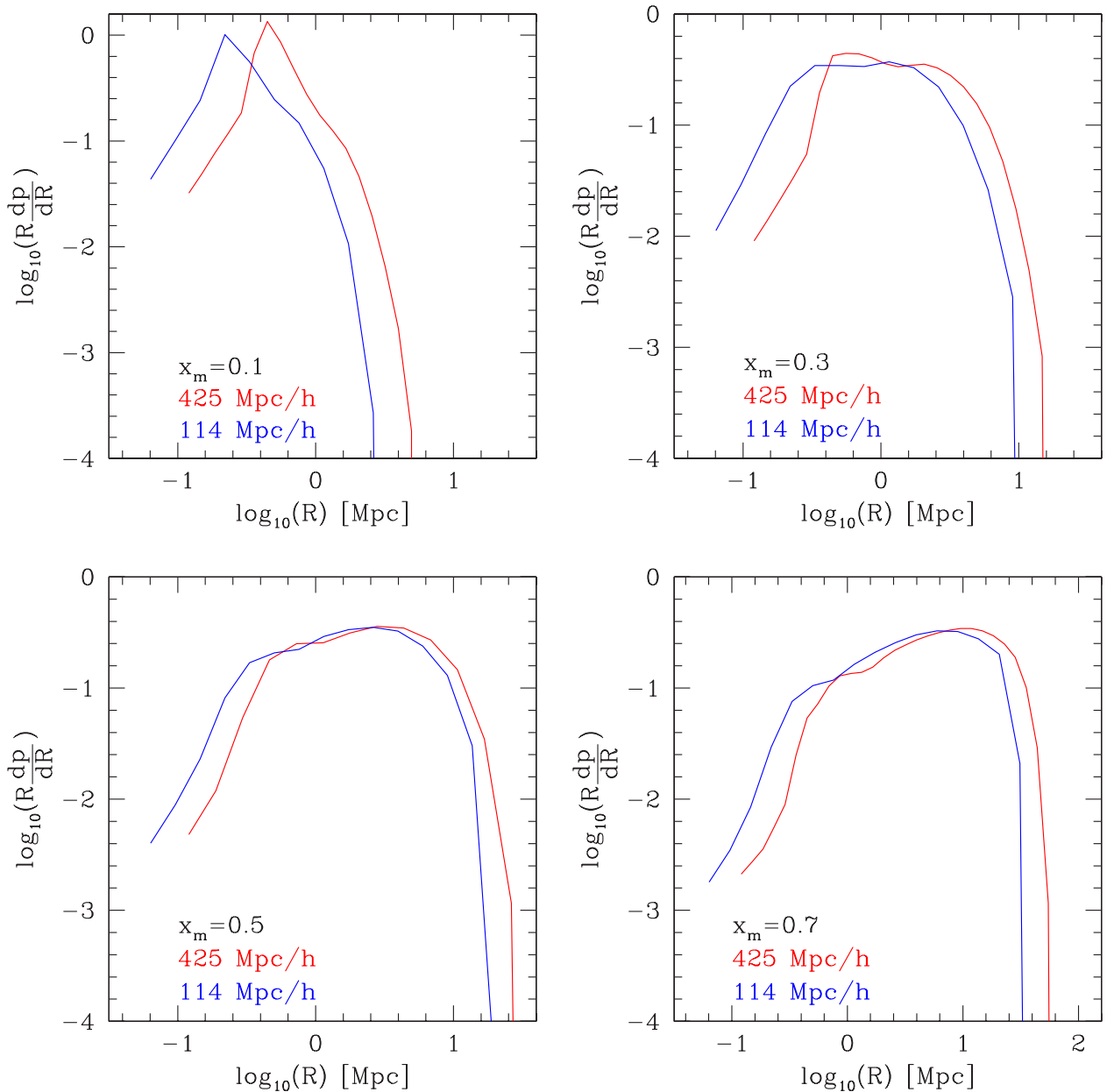


**Figure 8.** Dimensionless power spectra,  $\Delta_x$ , of the volume-weighted ionized fraction,  $x_v$ , for H II regions extracted from our simulations at different stages of the reionization process, as labelled. Shown are the results for XL2 (red) and L2.1 (blue).

the abundances of smallest resolved H II regions ( $V < 10^3 \text{ Mpc}^3$  at  $x_m = 0.5$ ,  $V < 10^2 \text{ Mpc}^3$  at  $x_m = 0.7$ ), but in the larger simulation volume there are still many more mid-sized regions, some reaching up to  $V \sim 10^4 \text{ Mpc}^3$  in volume. We therefore find that, in agreement with our previous results, a bimodal H II region distribution develops once the local percolation starts becoming wide-spread, around  $x_m \sim 0.3$ . Ionized patches cannot surpass certain size before they merge into one very large H II region which encompasses significant fraction of the total volume. While this quantitative behaviour is always observed, independent of the simulation volume, the maximum volume which isolated ionized patches can reach before percolation is dependent on the box size. They can reach up to  $V \sim 10^4\text{--}10^5 \text{ Mpc}^3$  in the larger volume, but rarely surpass  $V \sim 10^3 \text{ Mpc}^3$  in the smaller one (except around  $x_m \sim 0.3$ , where some reach  $V \sim 3 \times 10^4 \text{ Mpc}^3$ , while in the larger volume they reach

up to  $V \gtrsim 10^5 \text{ Mpc}^3$  at the same epoch). The reason for this is that in the larger volume the voids (which due to the inside-out nature of reionization remain largely neutral for longer than do high-density regions) are considerably larger, thus allowing the H II regions to grow further before merging. This has important implications for the redshifted 21-cm line fluctuations, which we will discuss below. We note that in neither case the value of this maximum volume of isolated regions reaches more than a fraction of a per cent of the total volume and is thus unlikely to be affected by straightforward finite box effects.

We next turn our attention to the other, complementary measures of the H II region size distribution we previously mentioned, namely the 3D power spectra of the ionization fraction field, shown in Fig. 8, and the SPA size distribution, shown in Figs 9 and 10. During the initial phases of reionization ( $x_m = 0.1$ ) the 3D ionization field

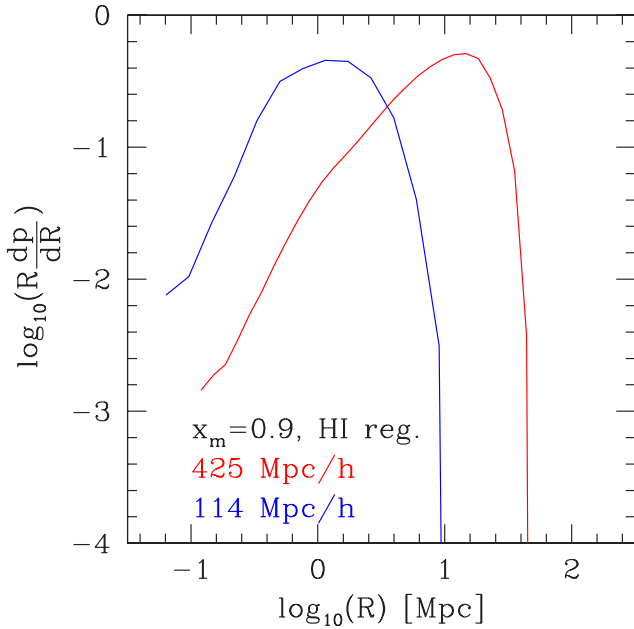


**Figure 9.** Probability distribution function per logarithmic radial bin,  $R \frac{dp}{dR}$ , spherical regions with radius  $R$  as given by the SPA method, based on the ionized distribution given by our simulations. Shown are different stages of the reionization process, for ionized fraction by mass  $x_m = 0.1, 0.3, 0.5$  and  $0.7$ , as labelled. Shown are the results based on XL2 (red) and L2.1 (blue).

power spectra have similar shapes, uniformly rising towards smaller scales without a clearly identified characteristic scale. However, the larger volume yields up to 50 per cent more power than the smaller one at the (relatively small) scales corresponding to the  $\text{H II}$  regions at that time, as well as some additional power at very large scales which are not present in the smaller box. Some of differences at the smallest scales are due to the different grid resolutions; however these scales fall below the resolution of most current observational experiments. As reionization proceeds, the local percolation of  $\text{H II}$  regions starts and a characteristic peak scale for the power spectra develops in both volumes, albeit at slightly different scales and overall magnitude. By the time ionized fraction reaches 50 per cent the power spectra become essentially the same in shape over the common scales of the two simulations. However, the power

spectrum derived from XL2 contains noticeably more power at all scales (and appears smoother) compared to L2.1. In both cases the peak scale gradually moves towards larger scales, tracking the growth of the ionized regions.

The SPA method yields similar results. At very small scales there is a sharp cut-off in the size distribution which is related to the grid resolution in each case and is thus a numerical effect, rather than a physical one. At all times there are larger  $\text{H II}$  regions in XL2, which reach tens of Mpc in radius at late times. On the other hand, the number of intermediate-sized ionized regions (radii from few Mpc to below one Mpc) is quite similar in the two boxes, except during the early evolution, before a characteristic scale develops. This again indicates consistency in the results at scales that are well below the size of the computational box and thus not affected by it.



**Figure 10.** Same as in Fig. 9, but for the *neutral* H I patches at late times, for ionized fraction by mass  $x_m = 0.9$ .

In contrast, the larger ionized patches are affected by the volume size and large-scale density fluctuations, and show to be considerably larger in the larger box.

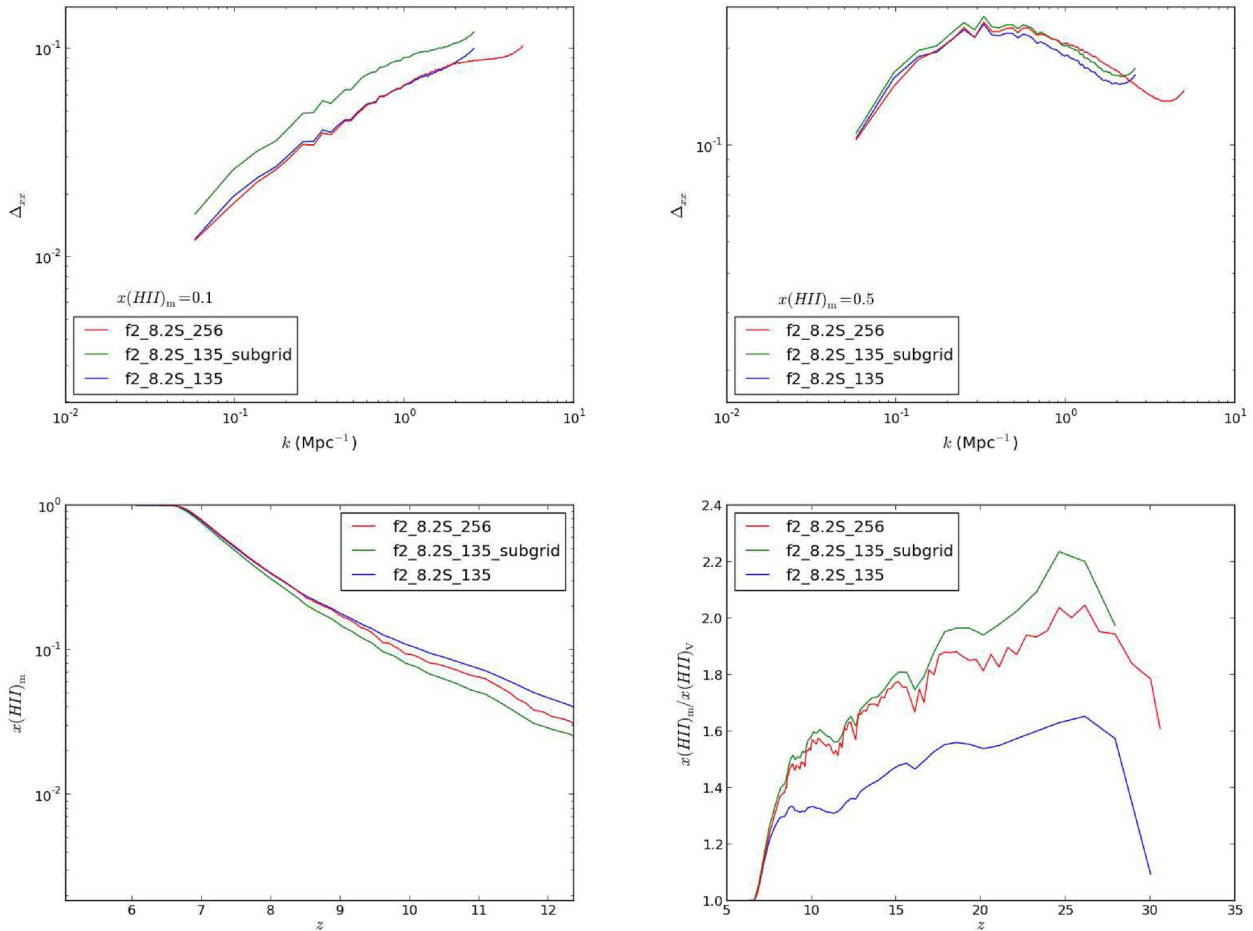
As we noted above, at late times all H II regions percolate into one large, connected region, thus at that epoch from point of view of the SPA method it makes better sense to study the *neutral* patch distribution instead, as those are the patchy regions which still keep their clear identity. Results for the epoch when the ionized fraction by mass is  $x_m = 0.9$  are shown in Fig. 10. Once again, there are many more large neutral patches in the larger volume. Those also reach much larger sizes, up to  $R \sim 45$  Mpc, while in the smaller volume they do not surpass  $R = 10$  Mpc. In both cases there is a clear peak of the distribution, which is at  $R \sim 1\text{--}2$  Mpc for L2.1, and at  $R \sim 15$  Mpc for XL2. In the latter case the peak is less broad and thus the peak scale is better-defined. The grid resolution-related cutoff at small scales is still present, but is now at scales well below the peak scale and is therefore not the reason for it, i.e. the peak scale we observe is physical, not numerical. The presence of such large neutral patches at late times, not seen in smaller-volume simulations, is important because it means better detectability at the redshifted 21-cm and other probes. Most such probes rely on fluctuations between sky patches, which are clearly enhanced here. Furthermore, radio sensitivity is significantly better at lower redshifts/higher frequencies, and the low-frequency foregrounds are much weaker, making the late stages of reionization an important target for the first, likely statistical, detection, though some limited imaging might also be possible (Zaroubi et al. 2012).

### 3.2.3 Effect of resolution and sub-grid modelling

As discussed in Section 2, there are differences in the numerical modelling in the two cases discussed above, which could potentially affect our results. First, the large-volume,  $425 h^{-1}$  Mpc RT simulation has lower grid resolution than the  $114 h^{-1}$  Mpc simulation, by almost a factor of 2 per dimension. Furthermore, the low-mass, suppressible LMACH sources in the larger volume are

modelled using sub-grid prescription, as their host haloes are not directly resolved in the corresponding  $N$ -body simulation. In order to evaluate the effect of these differences, we performed two additional RT simulations, both using the  $114 h^{-1}$  Mpc volume and grid resolution reduced to  $135^3$ , corresponding to the spatial resolution of the large RT volume. The first of these simulations (which will be referred to as L2lr1) used the LMACHs directly based on the resolved haloes, as done in the higher-resolution production simulation, while the second one (L2lr2) used the sub-grid prescription to model the LMACH host haloes. Therefore, by directly comparing them to simulation L2 (f2\_8.2S\_256 in our previous, more expanded notation) and each other we can evaluate the effect of resolution (L2lr1, same as f2\_8.2S\_135 in our previous notation) and the sub-grid model (L2lr2, same as f2\_8.2S\_135\_subgrid). When the resolution is decreased (L2 versus L2lr1), one obvious effect is the under-resolving of the smaller H II regions, resulting in smoothing of the corresponding fluctuations and in less power on small scales. However, since the resolution length in both cases is sub-Mpc, this only affects patches smaller than a few Mpc, or roughly arcminute scales (Fig. 11, top panels), which will not be resolved in the current generation of reionization experiments due to insufficient sensitivity. An additional, more subtle resolution effect is related to the way the LMACHs suppression is implemented – based on an ionization threshold. Averaging the ionized fraction over larger cells means that a cell could remain under the threshold even if sub-volumes of it would be above it if resolved. This yields a lower level of suppression, particularly at the edges of ionized regions in the more coarsely resolved simulation. The result is a somewhat faster-proceeding reionization (Fig. 11, bottom left panel). It also becomes less inside-out in character, with the mean mass-weighted ionized fraction closer to the volume one (Fig. 11, bottom right panel). However, this has no significant effect on the reionization patchiness and the observable signatures. This is demonstrated by all available measures – FOF, SPA and power spectra, which at large scales are essentially the same in the two cases throughout the reionization history. The only exception is the FOF PDF distribution at late times ( $x_m = 0.7$ ), where the coarser-grained simulation yields a small fraction ( $\sim 10^{-3}$  of the total volume) of larger H II regions which become topologically connected together due to the blurred boundaries between them due to the lower resolution.

The LMACHs sub-grid modelling (L2lr2) is currently done deterministically – following the mean density–halo number relation derived from higher-resolution simulations – and does not take into account the scatter in that relation. This has the effect of concentrating the model LMACHs towards the higher density regions more than is the case for the directly resolved LMACHs. Due to the inside-out nature of reionization, where the higher-density regions are preferentially ionized earlier, the model LMACHs are more strongly suppressed, which delays reionization slightly (Fig. 11, bottom left panel). The result is that any given mean ionized fraction is reached at a later time and therefore the H II regions manage to become slightly larger. This is especially noticeable during the early ( $x_m = 0.1\text{--}0.3$ ) and late ( $x_m = 0.7$ ) stages of the reionization process in the power spectra and SPA patchiness measures, while during the middle stages ( $x_m = 0.5$ ) this effect largely disappears (Fig. 11, top panels). In contrast, the FOF PDF distribution is only noticeably affected early on ( $x_m = 0.1$ ) and very little after that time. Interestingly, the separate effects of lower resolution and sub-grid model on the mass-weighted to volume-weighted ionized fraction evolution largely cancel out (Fig. 11, bottom right panel), indicating that when both effects are present the inside-out ionization properties are very close to the fully resolved simulation.



**Figure 11.** Effect of resolution and sub-grid source modelling: power spectra of the ionized field,  $\Delta_{xx}$ , at mass-weighted ionized fraction of  $x_m = 0.1$  (top left) and  $x_m = 0.5$  (top right); late-time ( $z < 12$ ) mass-weighted reionization history,  $x_m(z)$  (bottom left) and ratio of the mass- and volume-weighted ionized fractions versus redshift (bottom right) for the simulations L2 (f2\_8.2S\_256), L2lr1 (f2\_8.2S\_135) and L2lr2 (f2\_8.2S\_135\_subgrid), as labelled.

Overall, compared to the large-volume effects discussed in Section 3.2.2, the consequences of the lower resolution and sub-grid model on the reionization patchiness are much smaller and therefore do not invalidate any of our results.

### 3.3 Convergence of the reionization history and observables with simulation volume

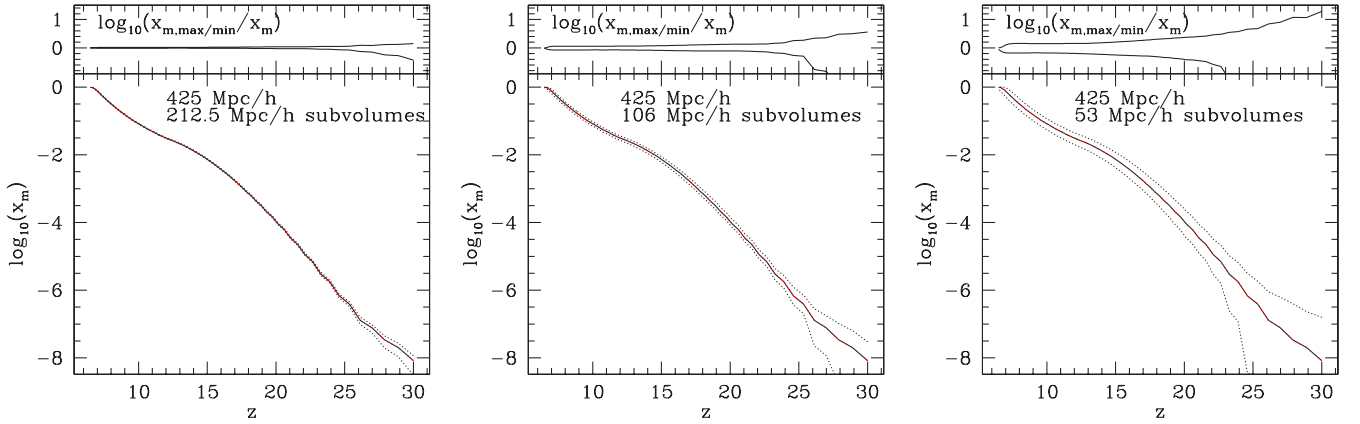
#### 3.3.1 Reionization history

As we showed above, the mean reionization history is already well converged for  $\sim 100$  Mpc volumes (or even for ones of few tens of Mpc across; see Iliev et al. 2006). However, locally for smaller patches it does vary significantly from patch to patch. In order to better quantify the convergence of the reionization history with the sample volume size, we split our XL2 simulation volume into a number of (non-overlapping) volumes of different sizes and then calculate their individual reionization histories, i.e. mean mass-weighted ionized fraction,  $x_m$  versus redshift. In Fig. 12 we show the reionization history averaged over all sub-volumes, as well as the maximum and minimum sub-volume ionized fraction at each redshift. Neither the maximum nor the minimum value necessarily belong to the same sub-volume at different times, thus the minimum and maximum lines are in fact enveloping the full variety of reionization histories which occur in the sub-volumes. For reference

and easier direct comparison we also show the mean reionization history of our full volume discussed in the previous section.

We first note that for all sub-volume sizes considered here the mean reionization history of the sub-volumes is essentially identical to the average one for the full volume. This statement would of course be identically true for the volume-based reionization histories; however it is easy to show that in general this is not the case for the averaged mass-weighted histories, particularly when the mean densities of the sub-volumes vary significantly.

On the other hand, there is a clear variation in the reionization histories of different sub-volumes. Naturally, the spread is strongest at early times and smaller sub-volumes. For sub-volumes of  $212.5 h^{-1}$  Mpc on a side we observe clear convergence to the mean evolution, apart from modest variation at the earliest times,  $z > 25$ . The convergence worsens somewhat, but still remains reasonable for sub-region size of  $106.25 h^{-1}$  Mpc, which size is close to the usually assumed convergence value of  $\sim 100 h^{-1}$  Mpc (cf. Barkana & Loeb 2004); however there are now clear, albeit relatively small deviations from the mean reionization history. The full variation range in ionized fraction among the sub-regions is larger than 10 at the earliest times,  $z > 27$ , when the cosmic variance due to Poisson noise is strongest. As the reionization progresses this ratio drops to a value of the order of 2 and finally slowly approaches the mean reionization history at late times, although the variations remain present at low level all the way to the end of reionization.



**Figure 12.** Bottom panels: Reionization history variation in (non-overlapping) subvolumes of a different size: (left)  $212.5 h^{-1}$  Mpc; (middle)  $106.25 h^{-1}$  Mpc; and (right)  $53.125 h^{-1}$  Mpc. Shown are the mean, minimum and maximum of all sub-regions (black) and the mean mass-weighted reionization history of the full volume (red). Top panels: ratios of the largest and smallest sub-region mass-weighted ionized fraction and the mean one over the full volume.

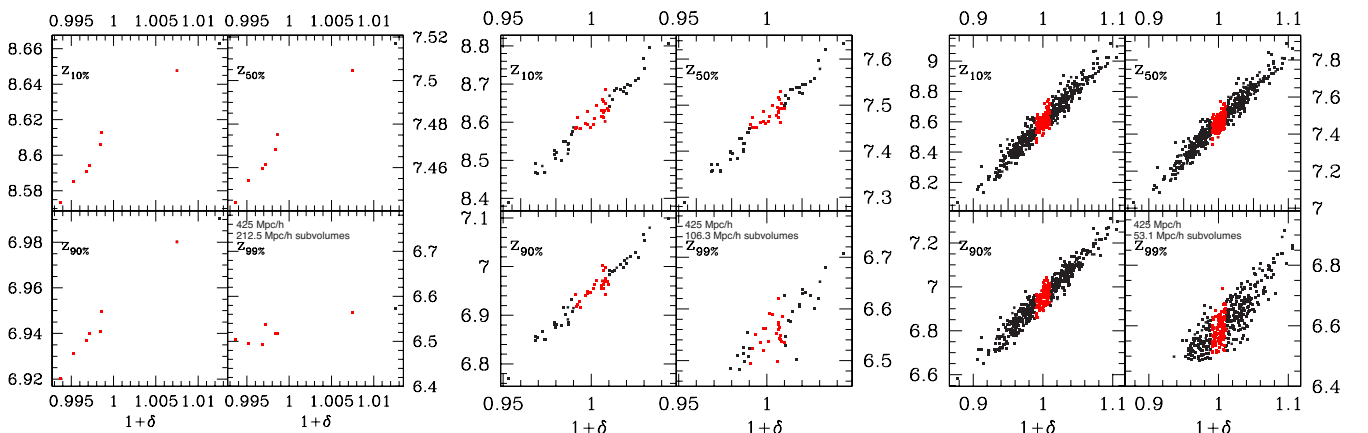
For sub-region size of  $53.125 h^{-1}$  Mpc (which is a volume similar to or even larger than the one adopted by many reionization simulations) the reionization history variations become very considerable. This is a consequence of the significant differences in the mean densities of the sub-regions, which surpasses 20 per cent at late times, which is then reflected exponentially in the local abundance of ionizing sources, and to a lesser extent in the local recombinations. At the earliest epochs,  $z > 25$ , some regions have ionized fractions as much as 10 times larger than the mean, while others still remain fully neutral. These variations decrease over time, but nonetheless stay significant, with a ratio between the extreme cases of the order of a few at all redshifts all the way up to the end of reionization. The completion of reionization in each sub-region also varies accordingly.

The integrated electron scattering optical depth corresponding to each sub-volume's reionization history is tightly correlated to the mean density of that volume. For the  $53 h^{-1}$  Mpc it varies significantly around the mean value corresponding to the full simulation volume, ranging between  $\tau = 0.0518$  and  $0.0611$ . For the  $106 h^{-1}$  Mpc sub-volumes this variation decreases and is between  $\tau = 0.0544$  and  $0.0583$ . For the  $212 h^{-1}$  Mpc sub-regions the optical depth essentially converges to the mean, ranging between

$\tau = 0.0562$  and  $0.0570$ . The optical depth being an integrated, line-of-sight quantity, these values are mostly of theoretical interest. Of more practical importance are the variations of the optical depth for mean-density sub-volumes, as these are analogous to small-box reionization simulations which typically take their average density to be the mean one for the universe. For mean-density sub-volumes the optical depth variations are small, with a range of  $\pm 0.001$  for the  $53 h^{-1}$  Mpc sub-volumes and a negligibly small one for the larger sub-volumes. This indicates that even relatively small-box simulations could be reliably used for calculating the optical depth for a given reionization scenario. We, however, note that at still smaller scales, below 10 Mpc (a few arcmin) the local electron scattering optical depth still varies significantly. It is highly correlated with the ionization field and anticorrelated with the neutral one (Holder, Iliev & Mellema 2007) and can contribute strongly to the CMB polarization at arcmin scales (Doré et al. 2007).

### 3.3.2 Ionization fraction–density correlation

In order to further quantify the variations of the reionization history between the sub-regions and its dependence on the local density, in Fig. 13 we show a scatter plot (each point represents a



**Figure 13.** Redshift at which certain reionization stages are reached:  $x_m = 0.1$  ( $z_{10\text{per cent}}$ ),  $x_m = 0.5$  ( $z_{50\text{per cent}}$ ),  $x_m = 0.9$  ( $z_{90\text{per cent}}$ ), and  $x_m = 0.99$  ( $z_{99\text{per cent}} = z_{\text{ov}}$ ) versus  $1 + \delta = \rho_{\text{sub-region}} / \rho_{\text{global mean}}$ , the average density of that sub-region in units of the global mean for different size sub-regions: (left)  $212.5 h^{-1}$  Mpc; (middle)  $106.25 h^{-1}$  Mpc; and (right)  $53.125 h^{-1}$  Mpc. The regions with roughly mean density (within 1 per cent of the mean) are shown in red.

non-overlapping sub-region) of the correlation between the redshift at which certain reionization milestone is reached versus the average density (with respect to the global mean) of that sub-region at the end of reionization. We have chosen as milestones the epochs when ionized fractions  $x_m = 0.1, 0.5, 0.9$  and  $0.99$  are reached, corresponding to early, middle, late and final reionization stages, respectively, and the same three sub-volume sizes as in the previous section.

The  $212.5 h^{-1}$  Mpc sub-regions (left panels) correspond to a small range of densities, all within 1.5 per cent of the mean and the maximum density variation is a little above 1 per cent. This yields a similarly small variation of  $\sim 1$  per cent in the redshifts at which each reionization stage is reached. The correlation between redshift and  $1 + \delta$  is roughly linear. Only one of the eight sub-regions deviates from the mean by more than 1 per cent and just two of them are overdense. Accordingly, those two regions reach each of the reionization stages noticeably earlier than the rest, by  $\Delta z \sim 0.04$ .

Similar qualitative trends are observed for the smaller-sized sub-regions. The correlation is still roughly linear and fairly tight in all cases, i.e. higher-density regions generally ionize earlier than low-density ones, as expected. However, the smaller the sub-regions are, the larger the variations among them. The sub-region densities vary between  $1 + \delta = 0.95$  and  $1.05$  for  $106.25 h^{-1}$  Mpc sub-regions and  $1 + \delta = 0.9-1.1$  for  $53.125 h^{-1}$  Mpc ones. Consequently, the redshifts at which each reionization stage is reached vary considerably, by as much as  $\Delta z = \pm 0.2$  and  $\Delta z = \pm 0.5$ , respectively. Interestingly, there is a significant scatter even among sub-regions with the same or similar densities. For sub-regions with density close to the mean one and size of  $106.5$  Mpc we have a maximum variation of  $\Delta z \sim \pm 0.1$ , increasing to  $\Delta z \sim \pm 0.2$  for  $53.125$  Mpc regions. Therefore, the redshift of end of reionization (or any of the other EoR milestones) cannot be determined with a good precision based on a small-box simulation even for fixed photon production efficiencies. This uncertainty becomes much worse for very small volumes, which are therefore quite inappropriate for modelling even of the mean ionization history, let alone the EoR patchiness.

### 3.3.3 21-cm emission: power spectra, rms and non-Gaussianity

We now turn our attention to some of the observable signatures of cosmic reionization based on the redshifted 21-cm emission from neutral hydrogen. The differential brightness temperature of the redshifted 21-cm emission with respect to the CMB is determined by the density of neutral hydrogen,  $\rho_{\text{HI}}$ , and its spin temperature,  $T_S$  and is given by

$$\begin{aligned} \delta T_b &= \frac{T_S - T_{\text{CMB}}}{1+z} (1 - e^{-\tau}) \\ &\approx \frac{T_S - T_{\text{CMB}}}{1+z} \frac{3\lambda_0^3 A_{10} T_* n_{\text{HI}}(z)}{32\pi T_S H(z)} \\ &= 28.5 \text{ mK} \left( \frac{1+z}{10} \right)^{1/2} (1+\delta) x_{\text{HI}} \left( \frac{\Omega_b}{0.042} \frac{h}{0.73} \right) \left( \frac{0.24}{\Omega_m} \right)^{1/2} \end{aligned} \quad (2)$$

(Field 1959), where  $z$  is the redshift,  $T_{\text{CMB}}$  is the temperature of the CMB radiation at that redshift,  $\tau$  is the corresponding 21-cm optical depth, assumed to be small when writing equation (2),  $\lambda_0 = 21.16$  cm is the rest-frame wavelength of the line,  $A_{10} = 2.85 \times 10^{-15} \text{ s}^{-1}$  is the Einstein A-coefficient,  $T_* = 0.068$  K corresponds to the energy difference between the two hyperfine levels,  $x_{\text{HI}}(1+\delta) = n_{\text{HI}}/\langle n_{\text{H}} \rangle$  is the mean number density of neutral

hydrogen in units of the mean number density of hydrogen at redshift  $z$ ,

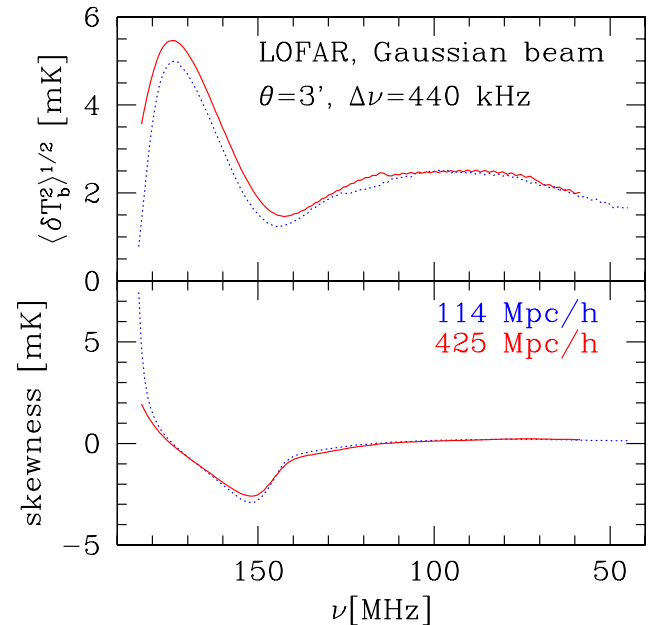
$$\begin{aligned} \langle n_{\text{H}} \rangle(z) &= \frac{\Omega_b \rho_{\text{crit},0} (1+z)^3}{\mu_{\text{H}} m_p} \\ &= 1.909 \times 10^{-7} \text{ cm}^{-3} \left( \frac{\Omega_b}{0.042} \right) (1+z)^3, \end{aligned} \quad (3)$$

with  $\mu_{\text{H}} = 1.32$  the corresponding mean molecular weight for neutral gas of primordial composition (assuming 24 per cent He abundance), and  $H(z)$  is the redshift-dependent Hubble constant,

$$\begin{aligned} H(z) &= H_0 [\Omega_m (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda]^{1/2} \\ &= H_0 E(z) \approx H_0 \Omega_m^{1/2} (1+z)^{3/2}, \end{aligned} \quad (4)$$

where  $H_0$  is its value at present, and the last approximation is valid for  $z > 1$ . Throughout this work we assume that  $T_S > T_{\text{CMB}}$ , i.e. that there is sufficient Ly $\alpha$  background to completely decouple the 21-cm transition from the CMB and that the neutral gas is heated well above the CMB temperature (due to e.g. a small amount of X-ray heating). Under these conditions the 21-cm line is seen in emission. These assumptions are generally well-justified, except possibly during the very early times (see e.g. Furlanetto, Oh & Briggs 2006, and references therein).

In Fig. 14 we show the 21-cm differential brightness temperature fluctuations rms and skewness of the PDF distribution smoothed with a Gaussian beam of size 3 arcmin and bandwidth of  $\Delta\nu = 440$  kHz. Comparing the results from our two simulation volumes, XL2 and L2.1, we see that during the early evolution ( $\nu_{\text{obs}} < 120$  MHz) the ionized regions are still few and small and thus the 21-cm fluctuations are mostly due to the underlying density fluctuations and track each other closely. However, once the patchiness becomes significant due to the growth of the H II regions ( $\nu_{\text{obs}} > 120$  MHz), the XL2 simulation yields noticeably higher fluctuations. The largest rms fluctuations occur at the same epoch in the two cases, at  $\sim 127$  MHz, but with a peak amplitude larger by



**Figure 14.** Top: the evolution of the rms of the 21-cm fluctuations smoothed with beam and bandwidth roughly as expected for the LOFAR EoR experiment; (bottom) evolution of the skewness of the 21-cm PDFs using the same smoothing. Both quantities are shown for XL2 (red) and L2.1 (blue).

about 10 per cent in the XL2 case. At late times the difference grows larger, reaching a factor of  $\sim 2.5$  at overlap. This is important since such stronger signal at late times/higher frequencies significantly improves the chances of detection, given the lower foregrounds at these higher frequencies and the better sensitivity of the radio interferometers there.

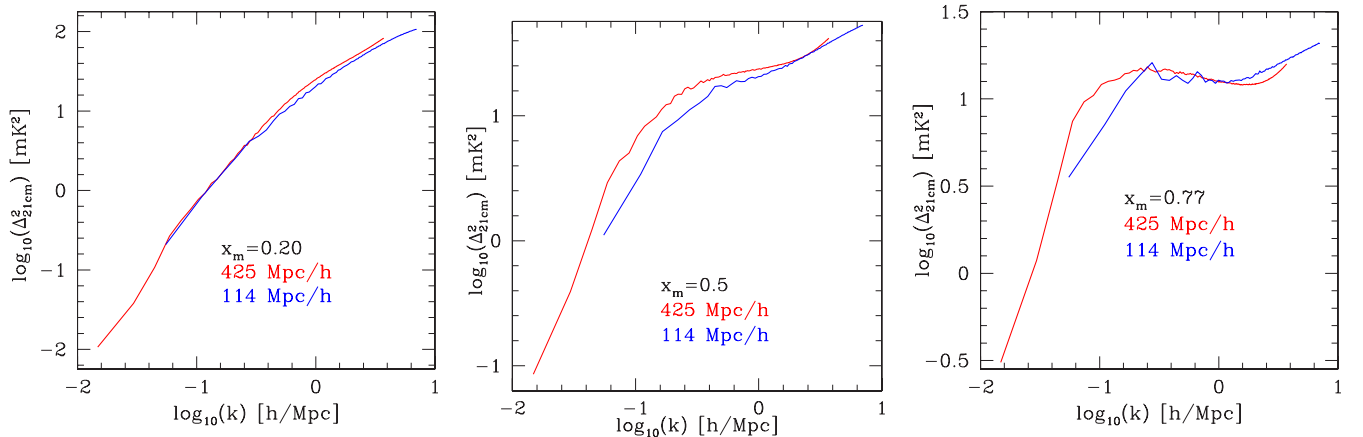
On the other hand, the skewness – a measure of the departure of the 21-cm beam- and bandwidth-smoothed PDF distribution from a Gaussian one – is largely identical in the two cases, indicating the insensitivity of this measure to the simulation volume (Fig. 14, bottom panel). In both cases the skewness starts close to zero during the initial, density fluctuation-dominated phase, and then exhibit a notable dip in the skewness to negative values reaching  $-3$  mK and coinciding with the initial stages of the 21-cm rms steep rise due to EoR patchiness. The numerical grid resolution does not matter in either of the two cases since resolution elements are much smaller than the beam/bandwidth smoothing scale.

The power spectra of the 21-cm fluctuation fields (including the redshift-space distortions, for the calculation methodology, see Mao et al. 2012) are shown in Fig. 15. Initially (shown is the epoch at which the ionized fraction by mass is  $x_m = 0.2$ ), they closely track each other within their scales of overlap. Interestingly, this is in contrast to the significant boost to the dimensionless power spectra of

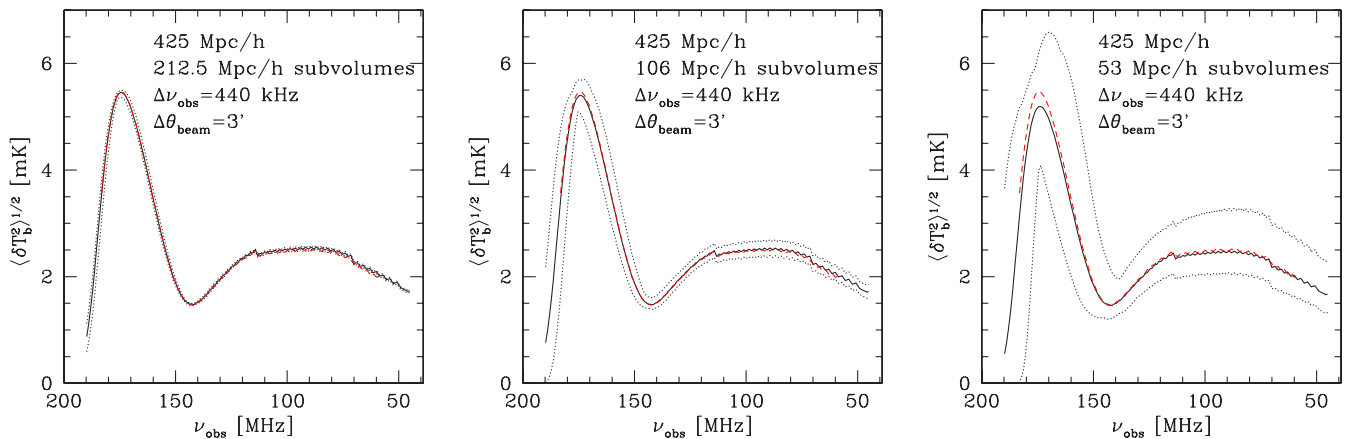
the ionization field in Fig. 8, which is seen at small scales even during the early evolution. This apparent discrepancy is explained by the strong dominance of the early 21-cm fluctuations by the density auto-correlation and the density-ionized fraction cross-correlation terms as compared to the ionized fraction auto-correlation term  $\Delta_x$ . Because these dominant terms are similar in the two simulations, the resulting 21-cm power spectra are also similar. However, again in agreement with the 21-cm rms results above, during the middle ( $x_m = 0.5$ ) and late ( $x_m = 0.77$ ) stages of reionization in simulation XL2 there is considerably more power at large scales ( $k < 1 h\text{Mpc}^{-1}$  for  $x_m = 0.5$ , with the scale of departure between the two increasing to  $k < 0.3 h\text{Mpc}^{-1}$  for  $x_m = 0.77$  as the H II regions grow in size). Therefore, although the ionized regions are larger in XL2 at all times, as was shown in Section 3.2.2, this is noticeable in the 21-cm rms and power spectra only when the patchiness term starts to dominate the total power.

Finally, we consider the variations of the 21-cm observables between sub-volumes of a given size extracted from the XL2 simulation. Results are shown in Figs 16–18 for 21-cm rms, skewness and kurtosis, and in Fig. 19 for 21-cm power spectra.

For 21-cm rms, skewness and kurtosis, the mean over all sub-volumes closely tracks the full volume evolution of the same quantities, regardless of the size of the sub-volumes. All three quantities

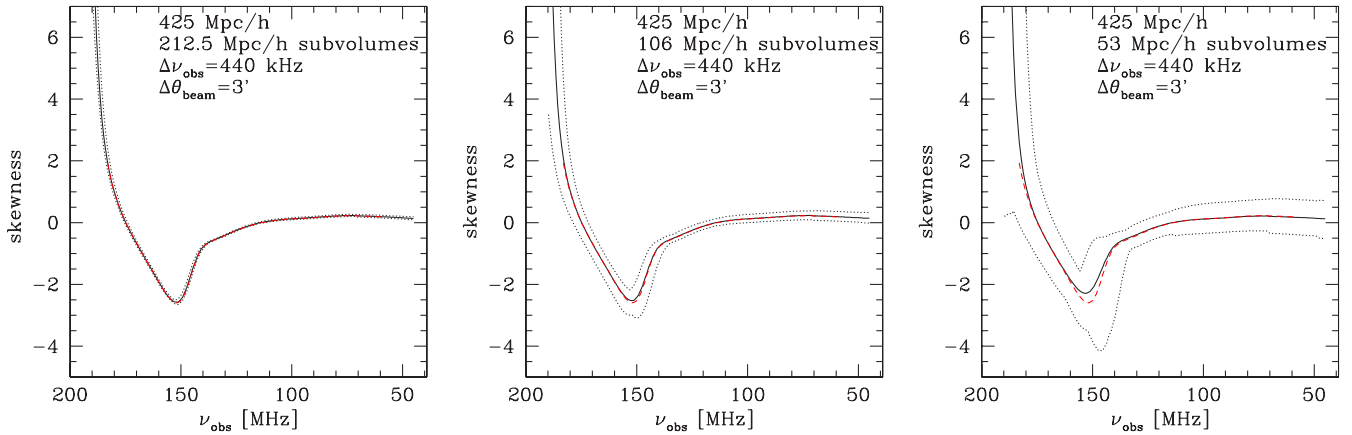


**Figure 15.** 21-cm differential brightness temperature fluctuation power spectra for the XL2 (red) and L2.1 (blue) at different stages of reionization: (left)  $x_m = 0.2$ , (middle)  $x_m = 0.5$  and (right)  $x_m = 0.77$ .

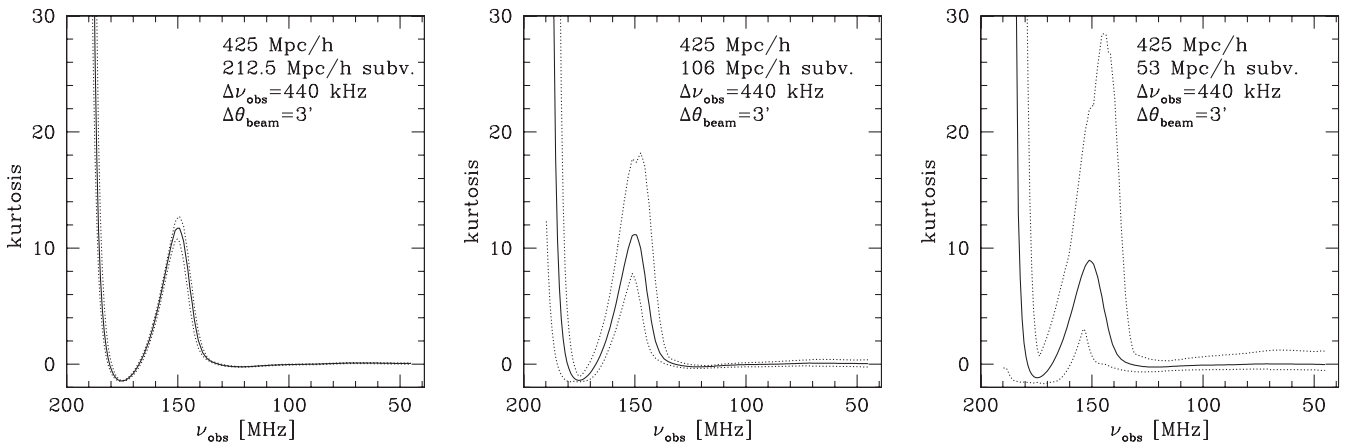


**Figure 16.** 21-cm differential brightness temperature rms variations, smoothed with LOFAR-like beam and bandwidth, between sub-regions of XL2 with a given size, as labelled. Shown are the mean of all sub-volumes (solid, black) and the minimum and maximum lines enveloping the variations between all sub-regions of that size. The average over the full volume is also indicated (red).





**Figure 17.** 21-cm differential brightness temperature field skewness variations in mK, smoothed with LOFAR-like beam and bandwidth, between sub-regions of XL2 with a given size, as labelled. Shown are the mean of all sub-volumes (solid, black) and the minimum and maximum lines enveloping the variations between all sub-regions of that size. The average over the full volume is also indicated (red).



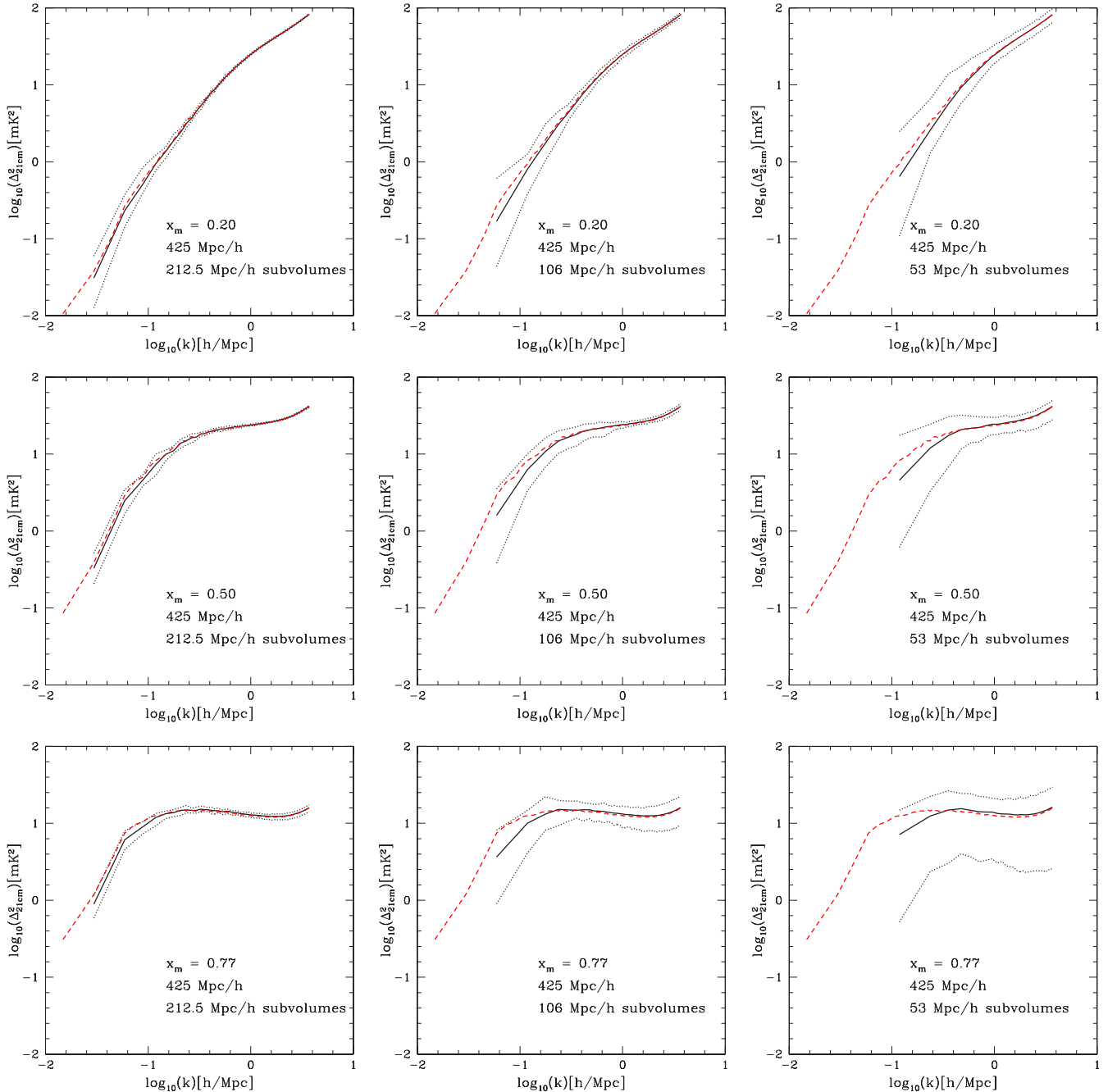
**Figure 18.** 21-cm differential brightness temperature field kurtosis variations, smoothed with LOFAR-like beam and bandwidth, between sub-regions of XL2 with a given size, as labelled. Shown are the mean of all sub-volumes (solid, black) and the minimum and maximum lines enveloping the variations between all sub-regions of that size.

are essentially converged for  $212.5 h^{-1}$  Mpc sub-volumes, but there are notable variations for smaller-sized sub-volumes. For example, in terms of the peak differential brightness temperature rms value, these variations could be as high as 20 per cent for the  $106.25 h^{-1}$  Mpc sub-volumes and reach over 60 per cent for the  $53.125 h^{-1}$  Mpc sub-volumes. These variations grow even larger at late times. For example, at 190 MHz some sub-volumes are already ionized (thus have zero rms), while the mean for all sub-regions is still above 0.5 mK and the maximum rms value is as large as  $\sim 2$  mK ( $\sim 4$  mK) for the  $106.25 h^{-1}$  Mpc ( $53.125 h^{-1}$  Mpc) sub-volumes, respectively. Even the lower, pre-reionization peak at early times ( $\nu \sim 80$  MHz) can differ by as much as 75 per cent for the smallest-sized sub-regions, ranging between 2 mK and over 3 mK.

The higher moments of the 21-cm brightness PDF distributions, skewness and kurtosis also show significant variations among sub-regions, as shown in Figs 17 and 18. The average skewness of all sub-regions is also very close to the full-volume result, for all sub-region sizes. However, once again only the large,  $212.5 h^{-1}$  Mpc volumes are converged and virtually indistinguishable from the full-box results. Even for  $106 h^{-1}$  Mpc volumes there is a difference of up to 50 per cent in the depth of the skewness feature, which varies between  $\sim 2$  and  $\sim 3$  mK. For the smaller sub-volumes this variation becomes as large as a factor of 3, varying between  $\sim 1.5$  mK (i.e.

very weak dip feature) and  $\sim 4.5$  mK. The frequency at which this feature occurs also varies, by as much as  $\sim 10$  MHz for the smallest sub-regions case. This behaviour is mirrored also in the kurtosis peak, found at the same frequencies as the skewness minimum, as shown in Fig. 18. The kurtosis is essentially zero throughout most of the evolution except for this broad peak feature, which in this particular simulation occurs around 150 MHz. The height of that peak varies by about a factor of 2 for  $106 h^{-1}$  Mpc sub-volumes and by up to factor of  $\sim 7$  for the smaller volumes.

However, for 21-cm power spectra, the mean over sub-volumes appears to lose powers on large scales but converge on small scales, relative to the power spectra calculated from the full volume, as shown in Fig. 19. The loss depends on the stage of reionization: the larger the global ionization fraction is, the worse the mean over sub-volumes tracks the values from the full volume. This is simply because of the growth of the ionization bubbles as the reionization proceeds. When the typical bubble size is comparable to the size of the sub-volume, the information of the correlation of the fluctuations across sub-volumes is lost when only fluctuations within sub-volumes are counted for power spectrum calculation. On the other hand, the variations of power spectra for the  $106 h^{-1}$  Mpc and  $53 h^{-1}$  Mpc sub-volumes grow very large at late times ( $x_m \geq 0.50$ ). For example, when  $x_m = 0.50$ , the variation for



**Figure 19.** 21-cm differential brightness temperature fluctuation power spectra, in sub-regions of XL2 with a given size and at different stages of reionization, as labelled. Shown are the mean of all sub-volumes (solid, black) and the minimum and maximum lines enveloping the variations between all sub-regions of that size. The average over the full volume is also indicated (dashed, red).

$106 h^{-1}$  Mpc sub-volumes can be  $\sim 60$  per cent at  $k = 0.1 h \text{ Mpc}^{-1}$ . These variations can be even larger for smaller  $k$ , smaller sub-volumes and later times.

#### 4 SUMMARY AND CONCLUSIONS

We presented the first very large-scale, full RT simulation of Cosmic Reionization. We compared the results from this to a more typical-sized,  $\sim 100 h^{-1}$  Mpc simulation with otherwise the same parameters. The large-scale density fluctuations at scales of tens to hundreds of comoving Mpc, which are damped or not present at all in

the smaller volume, strongly modulate the local number of ionizing sources and introduce significant additional patchiness in the ionization distribution. By all different-sized distribution measures – FOF, spherical average and power spectra – the ionized patches reach considerably larger sizes in the larger simulation volume. This in turn introduces increased rms fluctuations in the redshifted 21-cm emission, with the peak up by  $\sim 10$  per cent for our reionization model and LOFAR-like beamsizes and bandwidth. The 21-cm rms signal boost is larger still, reaching more than a factor of 2, at late times, when extensive neutral patches still remain in the large volume, in contrast to the smaller one. Therefore, taking into

account the long-wavelength perturbations should considerably facilitate the detection of the late-time 21-cm signal.

We used jackknife approach to evaluate the variation of the reionization history in sub-volumes of sizes ranging from  $212.5 h^{-1}$  Mpc to  $53 h^{-1}$  Mpc. The reionization history is largely converged for the larger sub-volumes, but for  $106 h^{-1}$  Mpc it shows significant variations at very high redshift,  $z > 20$ , which persist, albeit at a lower level, all the way to the end of reionization. Finally, the  $53 h^{-1}$  Mpc sub-volumes vary in their degree of ionization by factors of a few or more, indicating that studying such volumes does not yield a reliable estimate of the mean global reionization history. The sub-volumes of a certain size also reach each stage of the reionization history at a range of redshifts, correlated with the mean density of that region. However, there is a significant scatter in that relation, which increases considerably for the smaller-sized sub-volumes. Consequently, even regions of mean density can reionize at quite different times, varying by up to  $\Delta z \sim 1$  for  $\sim 50 h^{-1}$  Mpc volumes.

Recently, Battaglia et al. (2013) proposed a new semi-analytical approach for modelling reionization at large scales based on the correlation between the density and ionized fraction fields observed in simulations. Using volumes of  $100 h^{-1}$  Mpc per side, they showed that this correlation is reasonably tight for regions of size  $\sim 10$  Mpc and larger ( $k < 0.6 h \text{Mpc}^{-1}$ ), but rapidly worsens for smaller sizes. While this trend is in qualitative agreement with our results, we find quite large scatter in the reionization milestones even for regions as large as  $50\text{--}100 h^{-1}$  Mpc, indicating that the close density-ionized fraction correlation found by Battaglia et al. (2013) is likely overestimated due to their relatively small simulation volume.

The same sub-volume variations of reionization result also in corresponding variation in the 21-cm rms, skewness and kurtosis, which for the smaller sub-volumes can vary at peak by factors of  $\sim 2$ ,  $\sim 2.5$  and  $\sim 10$ , respectively for the three quantities. The variations are smaller, but still noticeable for  $106 h^{-1}$  Mpc sub-volumes, at  $\sim 15$  per cent,  $\sim 25$  per cent and  $\sim 2.5$ , respectively for rms, skewness and kurtosis extreme values, and only converge for sub-volumes as large as  $212.5 h^{-1}$  Mpc. On the other hand, the 21-cm 3D power spectra prove more sensitive to the simulation volume, generally with larger variations between sub-volumes than other 21-cm observables, and at large scales ( $k \lesssim 0.25 h \text{Mpc}^{-1}$ ) do not converge even for volumes  $\gtrsim 200 h^{-1}$  Mpc per side. For smaller,  $\sim 50 h^{-1}$  Mpc sub-volumes the 21-cm power spectra do not converge well at any scale probed here, varying by a factor of  $\sim 2\text{--}10$  depending on the reionization stage. However, the average power of the sub-volumes does converge to the full-volume result at small scales.

This work is just a proof of concept and an initial evaluation of the effects of large scales for a particular reionization scenario. Clearly, the specific numbers derived here depend on the assumptions made about the reionization parameters like source efficiencies. Nonetheless, our conclusions about the convergence with simulation volume are general, at least as long as the reionization process proceeds along the lines assumed here, namely it is dominated by soft-UV stellar sources, which is currently the reionization scenario best supported by the observational data. The source efficiencies assumed in this work were chosen so as to satisfy the current observational constraints like the EoR duration, overlap epoch and IGM photoionization rate at late times. However, a full evaluation in terms of observations at redshifted 21-cm or other wavelengths requires running suites of simulations exploring the available parameter space. This large-scale simulation data have already been used for deriving the kSZ signal from reionization (Park et al. 2013), signatures of luminous QSOs (Datta et al. 2012) and 21-cm redshift-space distortions (Jensen et al. 2013; Shapiro et al. 2013).

## ACKNOWLEDGEMENTS

This work was supported by the Science and Technology Facilities Council (grant numbers ST/F002858/1 and ST/I000976/1); and The Southeast Physics Network (SEP-Net). PRS was supported in part by U.S. NSF grants AST-0708176 and AST-1009799, NASA grants NNX07AH09G, NNG04G177G and NNX11AE09G, and Chandra grant SAO TM8-9009X. GM was supported in part by Swedish Research Council grant 60336701. YM was supported by French state funds managed by the ANR within the Investissements d’Avenir programme under reference ANR-11-IDEX-0004-02. KA is supported by NRF grant funded by the Korean government MEST No. 2012R1A1A1014646 and by research funds from Chosun University, 2010. The authors acknowledge the TACC at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported within this paper. This research was supported in part by an allocation of advanced computing resources provided by the National Science Foundation through TACC and the National Institute for Computational Sciences (NICS), with part of the computations performed on Lonestar at TACC (<http://www.tacc.utexas.edu>) and Kraken at NICS (<http://www.nics.tennessee.edu/>). Some of the numerical computations were done on the Apollo cluster at The University of Sussex and the Sciama High Performance Compute (HPC) cluster which is supported by the ICG, SEPNet and the University of Portsmouth. Part of the computations were performed on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund – Research Excellence; and the University of Toronto.

## REFERENCES

- Ahn K., Iliev I. T., Shapiro P. R., Mellema G., Koda J., Mao Y., 2012, *ApJ*, 756, L16
- Barkana R., Loeb A., 2004, *ApJ*, 609, 474
- Battaglia N., Trac H., Cen R., Loeb A., 2013, *ApJ*, 776, 81
- Ciardì B., Ferrara A., Governato F., Jenkins A., 2000, *MNRAS*, 314, 611
- Ciardì B., Ferrara A., White S. D. M., 2003, *MNRAS*, 344, L7
- Crocce M., Pueblas S., Scoccimarro R., 2006, *MNRAS*, 373, 369
- Datta K. K., Friedrich M. M., Mellema G., Iliev I. T., Shapiro P. R., 2012, *MNRAS*, 424, 762
- Doré O., Holder G., Alvarez M. A., Iliev I. T., Mellema G., Pen U.-L., Shapiro P. R., 2007, *Phys. Rev. D*, 76, 043002
- Fan X. et al., 2001, *AJ*, 122, 2833
- Field G. B., 1959, *ApJ*, 129, 536
- Friedrich M. M., Mellema G., Alvarez M. A., Shapiro P. R., Iliev I. T., 2011, *MNRAS*, 413, 1353
- Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, 433, 181
- Gnedin N. Y., 2000, *ApJ*, 535, 530
- Gnedin N. Y., Ostriker J. P., 1997, *ApJ*, 486, 581
- Harker G. J. A. et al., 2009, *MNRAS*, 393, 1449
- Harker G. et al., 2010, *MNRAS*, 405, 2492
- Harnois-Déraps J., Pen U.-L., Iliev I. T., Merz H., Emberson J. D., Desjacques V., 2013, *MNRAS*, 436, 540
- Holder G. P., Iliev I. T., Mellema G., 2007, *ApJ*, 663, L1
- Iliev I. T., Mellema G., Pen U.-L., Merz H., Shapiro P. R., Alvarez M. A., 2006, *MNRAS*, 369, 1625
- Iliev I. T., Mellema G., Shapiro P. R., Pen U.-L., 2007, *MNRAS*, 376, 534
- Iliev I. T., Shapiro P. R., McDonald P., Mellema G., Pen U.-L., 2008, *MNRAS*, 391, 63
- Iliev I. T., Mellema G., Shapiro P. R., Pen U.-L., Mao Y., Koda J., Ahn K., 2012, *MNRAS*, 423, 2222

- Jensen H. et al., 2013, MNRAS, 435, 460  
 Kohler K., Gnedin N. Y., Hamilton A. J. S., 2007, ApJ, 657, 15  
 Komatsu E. et al., 2009, ApJS, 180, 330  
 Krug H. B. et al., 2012, ApJ, 745, 122  
 Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473  
 Lonsdale C. J. et al., 2009, IEEE Proc., 97, 1497  
 Lukić Z., Heitmann K., Habib S., Bashinsky S., Ricker P. M., 2007, ApJ, 671, 1160  
 McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043  
 Mao Y., Shapiro P. R., Mellema G., Iliev I. T., Koda J., Ahn K., 2012, MNRAS, 422, 926  
 Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955  
 Park H., Shapiro P. R., Komatsu E., Iliev I. T., Ahn K., Mellema G., 2013, ApJ, 769, 93  
 Parsons A. R. et al., 2010, AJ, 139, 1468  
 Planck Collaboration et al., 2013, preprint ([arXiv:1303.5076](https://arxiv.org/abs/1303.5076))  
 Press W. H., Schechter P., 1974, ApJ, 187, 425  
 Reed D. S., Bower R., Frenk C. S., Jenkins A., Theuns T., 2007, MNRAS, 374, 2  
 Ricotti M., Gnedin N. Y., Shull J. M., 2002, ApJ, 575, 33  
 Santos M. G., Ferramacho L., Silva M. B., Amblard A., Cooray A., 2010, MNRAS, 406, 2421  
 Shapiro P. R., Mao Y., Iliev I. T., Mellema G., Datta K. K., Ahn K., Koda J., 2013, Phys. Rev. Lett., 110, 151301  
 Sheth R. K., Tormen G., 2002, MNRAS, 329, 61  
 Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1  
 Sokasian A., Abel T., Hernquist L., Springel V., 2003, MNRAS, 344, 607  
 Songaila A., Cowie L. L., 2010, ApJ, 721, 1448  
 Spergel D. N. et al., 2003, ApJS, 148, 175  
 Trac H. Y., Gnedin N. Y., 2011, Adv. Sci. Lett., 4, 228  
 Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230  
 Watson W. A., Iliev I. T., Diego J. M., Gottlöber S., Knebe A., Martínez-González E., Yepes G., 2014, MNRAS, 437, 3776  
 Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, ApJ, 654, 12  
 Zahn O. et al., 2012, ApJ, 756, 65  
 Zaroubi S. et al., 2012, MNRAS, 425, 2964

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