# Analysis of realistic HI 21-cm maps from Epoch of Reionization using Largest Cluster Statistics (LCS)

S. Dasgupta*, S. Bag†, S. Majumdar*, S. Dutta*
*Department of Astronomy, Astrophysics
and Space Engineering, Indian Institute of Technology Indore, Khandwa Rd., Simrol, MP 453552, India
†Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Yuseong-gu, 305-348 Daejeon, Korea
saswataextra@gmail.com

***Abstract*—Using the topological and morphological evolution of neutral hydrogen distribution during the different stages of the Epoch of Reionization (EoR), we try to reveal the nature of the ionizing sources and the astrophysical processes in the Intergalactic Medium (IGM) during this period by studying the evolution of the largest ionized region (via the Largest Cluster Statistics - LCS). We show that It can identify the difference between two distinct reionization scenarios – inside-out and outside-in, even when realistic observational conditions are taken into account i.e. the simulated HI 21cm maps are initially translated to a lower resolution, then corrupted with a simulated Gaussian noise having an RMS value of $\sim 3mK$ in line with the future SKA1-low observations. Afterwards, the effect of the beam was simulated and incorporated into the maps, and then the images are analyzed via LCS. Additionally, we use Shapefinders (SFs) to investigate the morphology of ionized volumes as they change in different reionization models. We observed that the greatest ionized volume abruptly expands solely in terms of its third SF after percolation while the first two SFs stay almost constant. Ergo, when approaching percolation, the ionized hydrogen in these circumstances becomes very filamentary, with a 'characteristic cross-section' that differs between source models.**

## I. Introduction

The Cosmic Dawn (CD) and the subsequent Epoch of Reionization (EoR) are the two least known periods of our cosmic history. After the end of cosmic Dark Ages, the first luminous sources started to form which marked the beginning of the Cosmic Dawn (CD). The high energy photons that radiated outward from these sources started ionizing the baryons in the Intergalactic Medium (IGM) which consisted mostly of neutral hydrogen (HI). This era is termed as the Epoch of Reionization (EoR). The lack of observations of this epoch hinders our understanding of the "phase transition" of the universe from the epoch of last scattering to the current state [1].

In principle, the best way to study these epochs is to image them directly with different telescopes. The very recent James Webb Space Telescope (JWST) actually targets to observe the EoR as one of its major science goals. But, unless these observations are not made in large numbers, the result will be biased towards the brightest objects. So, different complimentary probes using the most abundant baryonic element, hydrogen are used to study this epoch. By studying the Lymann-$\alpha$ emitting high redshift quasars and Thompson scattering of CMBR photons it is constrained that the end of reionization happened at a redshift $z \sim 6$. However, a lot of questions remain unanswered regarding the EoR and the type of the sources that led this event. The 21cm line of neutral hydrogen caused by the spin flip of an electron in the ground state from parallel to anti-parallel state has the potential to answer a lot of questions related to the EoR as it directly probes neutral hydrogen [**?**], [**?**].

However, the 21cm signal is very weak and it is easily concealed by the galactic and extra-galactic foregrounds which are $\sim 10^6$ times brighter than the actual signal. Along with this, various other factors like instrument noise, telescope beam effect and ionospheric effects add up and creates further constraints to detecting the signal [2]. Present generation radio telescopes like the upgraded Giant Meterwave Radio Telescope (uGMRT), Low frequency array (LOFAR), Murchison Widefield Array (MWA), Precision Array for Probing the Epoch of Reionization (PAPER) and Hydrogen Epoch of Reionization Array (HERA) are trying to detect the statistical fluctuations in the signal in Fourier domain and constrain its spherically averaged power spectrum. The upcoming Square Kilometer Array (SKA), however, has the potential to make tomographic images of the redshifted EoR-21cm signal over a large redshift range. Studying these images can help us understand the way reionization has progressed.

There are various ways one can analyse these images in real domain that would be supposedly produced by SKA. Minkowski Tensors and Minkowski functionals (MFs) are widely used to constrain the reionization histoty. Methods that are based on percolation theory, granulometry and persistence theory are also used to quantify topological properties of the HI regions during the EoR. Method of local variance and Betti numbers are a few other methods to probe the reionization morphology. However, our study which is based on percolation theory focuses on the largest ionized region and uses a novel statistic called the Largest Cluster Statistics (LCS) along with the Minkowski functionals using a sophisticated called called SURFGEN2 [3]. The cosmological parameters that are used throughout our analysis are the WMAP five year data release $h = 0.7$, $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $\Omega_b h^2 = 0.0226$ [4].

## II. Simulating the realistic HI 21cm maps

The HI 21cm signal is observed in contrast to the Cosmic Microwave Background Radiation (CMBR) and is measured in terms of the differential brightness temperature. The expression can be written as:

$$\delta T_b \approx 27 x_{HI}(1+\delta)\left(\frac{1+z}{10}\right)^{\frac{1}{2}} ... \\ ...\left(1-\frac{T_{CMB}(z)}{T_S}\right)\left(\frac{\Omega_b}{0.044}\frac{h}{0.7}\right)\left(\frac{\Omega_m}{0.27}\right)^{-\frac{1}{2}}\left(\frac{1-Y_P}{1-0.248}\right) mK \quad (1)$$

$x_{HI}$ and $\delta$ denotes the mass averaged neutral hydrogen fraction and the density fluctuation respectively. $T_{CMB}(z)$ and $T_S$ are the CMB temperature at a redshift z and the spin temperature of the two states of hydrogen respectively. $Y_P$ in this equation denotes the primordial helium abundance. According to the equation above, when $T_S$ is completely coupled to the CMB temperature, the 21-cm signal will not be seen. The Wouthuysen Field effect predicts that the spin temperature will approach the gas temperature during EoR, allowing the 21-cm signal to be seen against the backdrop of the CMB. The signal is seen in emission when the gas temperature is above $T_{CMB}$ and in absorption when it is below. The signal is seen in emission when the gas temperature is above $T_{CMB}$ and in absorption when it is below. Assuming a high spin temperature limit, we generate the HI-21cm maps using a semi-numerical technique.

As a trade-off between a full Numerical simulation and a 3D radiative transfer simulation Majumdar et. al. 2014 [5] use a semi-numerical N-Body code ([6];[5];[7]) to generate dark matter distribution for any given redshift. Then, using Friends of Friends (FoF) or Spherical smoothing collapsed dark mater halos are identified within the simulated matter distribution. Lastly, assuming that these halos are the most probable hosts for the ionizing photons, an ionization field is generated using an excursion set formalism to finally generate the HI 21cm field.

The generated signal cube has a volume of $\mathbf{714 cMPc^3}$. 69123 *particles of mass* $4.0 \times 10^7 M_\odot$ on a 138243 mesh is considered. This is then down-sampled to a $600^3$ grid for reionization modeling. The resulting resolution for the simulation is 1.19 cMpc along each side. For each redshift output of the N-body, halos were identified using a spherical over-density scheme. Minimum halo mass that is being considered here is $2.02 \times 10^9 M_\odot$.

After obtaining these maps, a random sampling of the 8 neighboring cells in the maps are done and only one out of those points are chosen. This way we coarsen the map to twice the initial resolution, making it 2.38 cMpc along each length. The average values of all the pixels were subtracted from each the pixel value to make the data *interferometric* as zero length baselines are impossible to design which will result in mean subtracted maps. Then these maps are coarsened with a 3D Gaussian filter having an FWHM which corresponds to what is expected from SKA1-Low [8]. After this smoothing, the maps are corrupted with a random noise having an RMS value of $\sim 3.10 mK using the$[9].

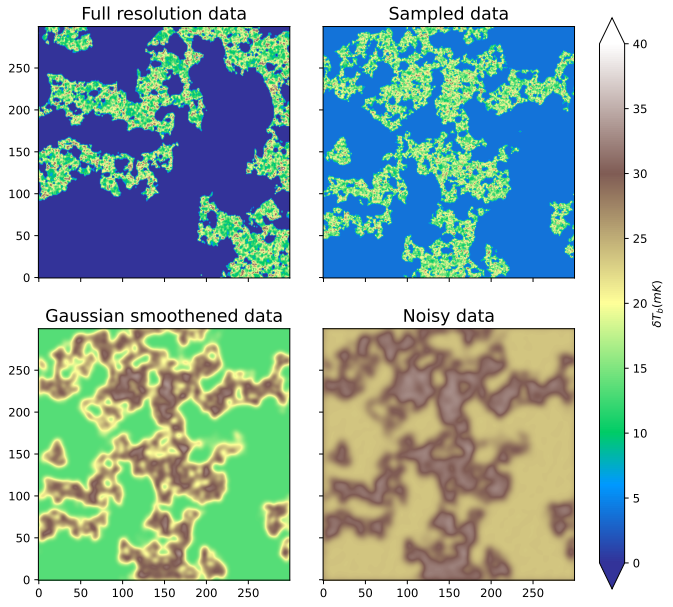


Fig. 1. Slices of reionization simulation maps for z = 7.221. *Top left*: A slice of the full resolution map of $600^3$ grid size. Image has been cropped because it is plotted in 300 grids. left bottom of the other slices correspond to this part of the map. *Top right*: The same slice when it is randomly sampled to a coarser grid of $300^3$ cells. *Bottom left*: The sampled data when it is smoothened with a 3D Gaussian filter of FWHM=9 corresponding to SKA1-Low. *Bottom right*: The final map obtained by adding a 3.10mK noise with the smoothened map.

## III. Methodology

### A. Percolation and LCS

When the first luminous objects began to appear at the time of Cosmic Dawn (CD), the high energy photons released from them started to ionize the Intergalactic Medium (IGM) which was dominated with hydrogen. With the progress of reionization the small ionized pockets of hydrogen started to grow in size and numbers but also started to merge with one another. At some cosmic time, most of these ionized regions suddenly merge together and form a single connected large ionized region. This transition is called the **percolation transition**. In our simulated case, we say that the percolation transition has occurred when the largest ionized volume stretches around the entire simulation box [10].

To track the progress of percolation transition we employ a novel statistic called the Largest Cluster Statistics (LCS). LCS, for the case of reionization is defined as follows:

$$LCS = \frac{\textbf{Volume of the largest ionized region}}{\textbf{Volume of all individual ionized regions}} \quad (2)$$

[11] It is expected that the LCS will show a sharp change at the time of percolation transition.

### B. SURFGEN2

We employ a sophisticated code called SURFGEN2 to find the LCS and Shapefinders for the 21cm differential brightness temperature field. The first version of SURFGEN was developed by Sheth et al. [3] to study large scale structures of the universe. SURFGEN2 uses the *Marching Cube33* algorithm to triangulate an iso-density surface and finally calculates the Minkowski functionals and subsequently the Shapefinders for each ionized pixel-cluster. The entire algorithm only works for a binarized field. So an appropriate theshold needs to be chosen to employ SURFGEN on a particlar field [12].

### C. Threshold determination

After the simulated mean subtracted HI-21cm maps were coarsened down, smoothened and corrupted with noise, the determination of the threshold to binarize the field becomes a non-trivial task. This was done with the help of the Tools21cm package [9].

We use a gradient descent method to find the first minima of the typical bimodal histogram (Figure 2) of the 21cm field where the two peaks quantify the ionized pixels and the neutral pixels. For high redshift maps, where there are almost no ionized regions, this scheme fails to work. So, the result could be only obtained for a certain range of redshift.

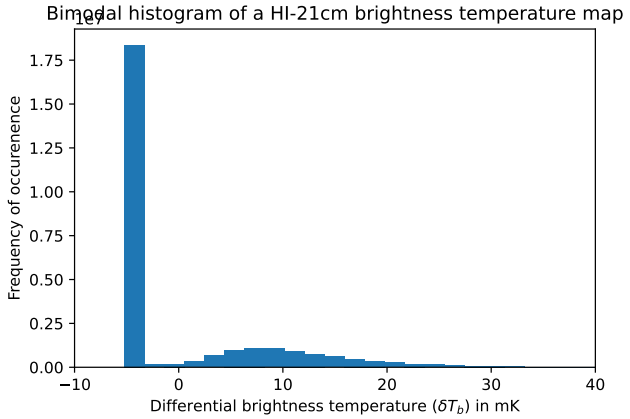


Fig. 2. A histogram plot shown for a coarse gridded reionization simulation map of z = 7.221. It is assumed that the left hump contains all the ionized pixels within the volume and the right hump contains mostly with the neutral pixels. As at z= 7.221, most of the universe is ionized, the left peak is $\sim 7$ times sharper than the right peak.

## IV. Results and Discussion

We plot the LCS with changing redshift in Figure 3 to see the percolation transition for different datasets. Firstly, it is observed that for the full resolution data the percolation happens at a redshift $\sim 8.515$. As we coarsen the data to a resolution of 2.38 Mpc along each side of the simulation volume, we observe that the percolation transition occurs at an earlier redshift of $\sim 8.283$. After smoothing the dataset with a 3D Gaussian filter of FWHM = 9 that corresponds to the SKA1-Low observation for a baseline of 2km, we see a significant change in the LCS. Firstly, as the data had been smoothened out, the smaller ionized regions for the case of higher redshifts were smoothened out and could not be resolved anymore. Alongside, it is noted that the larger ionized bubbles were merged because of this smoothing effect and SURFGEN2 considered them to be a single ionized region. For higher redshift values the LCS could not be computed because of the lack of ionized regions in the smoothened data. Hence, we could only obtain LCS for a redshift range starting from 8.283 to 7.221. Finally, when a Gaussian noise of 3.10mK was added to the smoothened signal map, it further the determination of threshold to binarize the 21cm field. But, it is worth noting that for all of the LCS curves, the abrupt change in topological phase quantified by the LCS is quite visible and percolation transition can be observed in all of the cases.

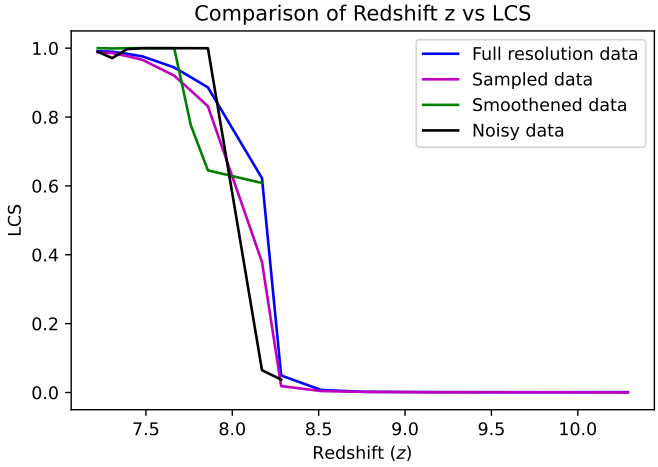


Fig. 3. Evolution LCS of ionized hydrogen with changing redshift in different scenarios are shown.

## V. Future Scope

We plan to employ real SKA-1 Low observational effects on the HI-21cm brightness temperature maps to constrain the robustness of LCS as a tool to directly apply on images. We would also like to incorporate the ionospheric effect on the signal and validate LCS on those maps. Alongside, we would like to constrain the viability of LCS analysis on other telescope models like LOFAR, MWA and HERA.

## References


1 Ryden, B., *Introduction to cosmology*, 2003.

2 Datta, A., Bowman, J., and Carilli, C., "Bright source subtraction requirements for redshifted 21 cm measurements," *The Astrophysical Journal*, vol. 724, no. 1, p. 526, 2010.

3 Sheth, J. V., Sahni, V., Shandarin, S. F., and Sathyaprakash, B. S., "Measuring the geometry and topology of large-scale structure using surfgen: methodology and preliminary results," *Monthly Notices of the Royal Astronomical Society*, vol. 343, no. 1, pp. 22–46, 2003.

4 Kapahtia, A., Chingangbam, P., and Appleby, S., "Morphology of 21cm brightness temperature during the epoch of reionization using contour minkowski tensor," *Journal of Cosmology and Astroparticle Physics*, vol. 2019, no. 09, p. 053, 2019.

5 Majumdar, S., Mellema, G., Datta, K. K., Jensen, H., Choudhury, T. R., Bharadwaj, S., and Friedrich, M. M., "On the use of seminumerical simulations in predicting the 21-cm signal from the epoch of reionization," *Monthly Notices of the Royal Astronomical Society*, vol. 443, no. 4, pp. 2843–2861, oct 2014. [Online]. Available: https://academic.oup.com/mnras/article/443/4/2843/1010602
6 Choudhury, T. R., Haehnelt, M. G., and Regan, J., "Inside-out or outside-in: The topology of reionization in the photon-starved regime suggested by Ly$\alpha$ forest data," *Monthly Notices of the Royal Astronomical Society*, vol. 394, no. 2, pp. 960–977, apr 2009. [Online]. Available: https://academic.oup.com/mnras/article/394/2/960/1072369
7 Mondal, R., Bharadwaj, S., Majumdar, S., Bera, A., and Acharyya, A., "The effect of non-Gaussianity on error predictions for the Epoch of Reionization (EoR) 21-cm power spectrum," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 449, no. 1, pp. L41–L45, may 2015. [Online]. Available: https://academic.oup.com/mnrasl/article/449/1/L41/1028566
8 Mellema, G., Koopmans, L., Shukla, H., Datta, K. K., Mesinger, A., and Majumdar, S., "Hi tomographic imaging of the cosmic dawn and epoch of reionization with ska," *arXiv preprint arXiv:1501.04203*, 2015.
9 Giri, S., Mellema, G., and Jensen, H., "Tools21cm: A python package to analyse the large-scale 21-cm signal from the epoch of reionization and cosmic dawn," *Journal of Open Source Software*, vol. 5, no. 52, p. 2363, 2020.
10 Bag, S., Mondal, R., Sarkar, P., Bharadwaj, S., and Sahni, V., "The shape and size distribution of h ii regions near the percolation transition," *Monthly Notices of the Royal Astronomical Society*, vol. 477, no. 2, pp. 1984–1992, 2018.
11 da Silva, C., Lyra, M., and Viswanathan, G., "Largest and second largest cluster statistics at the percolation threshold of hypercubic lattices," *Physical Review E*, vol. 66, no. 5, p. 056107, 2002.
12 Bag, S., Mondal, R., Sarkar, P., Bharadwaj, S., Choudhury, T. R., and Sahni, V., "Studying the morphology of h i isodensity surfaces during reionization using shapefinders and percolation analysis," *Monthly Notices of the Royal Astronomical Society*, vol. 485, no. 2, pp. 2235–2251, 2019.