

# **Analysing HI 21-cm Images From the Epoch of Reionization using Largest Cluster Statistics (LCS)**

**M.S.(R) Thesis**

By

**Saswata Dasgupta**



**Department of Astronomy, Astrophysics and Space Engineering  
INDIAN INSTITUTE OF TECHNOLOGY INDORE**

**June 22, 2023**

# **Analysing HI 21-cm Images From the Epoch of Reionization using Largest Cluster Statistics (LCS)**

**M.S.(R) Thesis**

*Submitted in partial fulfillment of the  
requirements for the awards of the degree  
of  
Master of Science (Research)*

by

**Saswata Dasgupta**



**Department of Astronomy, Astrophysics and Space Engineering**

**INDIAN INSTITUTE OF TECHNOLOGY INDORE**

**June 22, 2023**



# INDIAN INSTITUTE OF TECHNOLOGY INDORE

## CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Analysing HI 21-cm Images From the Epoch of Reionization** in the fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE (RESEARCH)** and submitted in the **Department of Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from August, 2021 to May, 2023 under the supervision of Dr. Suman Majumdar, Associate Professor, IIT Indore and Prof. Abhirup Datta, Professor, IIT Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

*Saswata Dasgupta 22/06/23*  
Signature of the student with date  
**SASWATA DASGUPTA**

-----  
This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

*Suman Majumdar 22/06/2023*  
Signature of the Supervisor of  
MS (Research) thesis #1 (with date)  
**SUMAN MAJUMDAR**

*Abhirup Datta 22/6/23*  
Signature of the Supervisor of  
MS (Research) thesis #2 (with date)  
**ABHIRUP DATTA**

-----  
**Saswata Dasgupta** has successfully given his MS (Research) Oral Examination held on **22/06/2023**.

*Susmita Paul 22/06/23*  
Signature of Chairperson (OEB) with date

*Suman Majumdar 22/06/2023* *Abhirup Datta 22/6/23*  
Signature(s) of Thesis Supervisor(s) with date

*Manoneeta Chakraborty 22/06/2023*  
Signature of MS coordinator with date

*S.D. Das 22/6/23*  
Signature of Convener, DPGC with date

*Abhirup Datta 22/6/23*  
Signature of Head of Discipline with date  
-----

## **ACKNOWLEDGEMENTS**

I would like to acknowledge Dr. Suman Majumdar and Prof. Abhirup Datta, my supervisors, for giving me the opportunity to work in this project and guiding me thoroughly, alongside being excellent motivators in my difficult times and believing in me. I acknowledge the support from Mr. Samit Kumar Pal on the mock SKA observations using the 21cmE2E-pipeline. Dr. Satadru Bag, Mr. Aadarsh Pathak and Dr. Mohd Kamran have been a constant support in this project and my learning journey. Dr. Kanan K Datta and Dr. Sayiad Ali from Jadavpur university have given their valuable suggestions to improve upon this project which has helped me a lot. The valuable and enriching discussions that I had with my peers like Chandra Shekhar, Leon, Priyatam and Vednarayan has helped me learn a lot. I would also like to thank the entire department of Astronomy, Astrophysics and Space Engineering for giving me the opportunity to contribute to this field. I cannot thank my friends and colleagues like Sohini, Rashmi, Samit and Bhuvnesh enough in the department for making me feel at home here. Finally, thanks to Baba, Maa and Bhai, without their encouragement and support, I would have never had the courage to step my foot in this entirely new field for me.

## Abstract

Two of the most enigmatic eras in the universe's history are the Cosmic Dawn and Epoch of Reionization (CD-EoR). We can learn more about the early luminous sources and astrophysical processes in the Intergalactic Medium (IGM) at those times by analyzing the changes in the distribution of neutral and ionized hydrogen during these epochs. Utilizing the topological and morphological evolution of the redshifted 21-cm signal originating from the neutral hydrogen distribution throughout various CD-EoR stages, we examine these changes. The evolution of the largest ionized region (LIR) during these stages is examined using the Largest Cluster Statistics (LCS). We show that LCS is a reliable statistic that can distinguish between inside-out and outside-in reionization models by comparing their corresponding percolation transition period, the cosmic period during which the majority of the individual ionized regions become connected and form a single, infinitely long ionized region. We also use Shapefinders to examine the shape and morphology of ionized regions under various reionization models as a complement to the percolation analysis. In all the models, we observe that the largest ionized region's volume abruptly increases in length during percolation, but its thickness and breadth essentially hold constant. We scale down the simulated 21-cm maps to a lower resolution to be consistent with the upcoming SKA1-Low. The scaled-down H<sub>I</sub> 21-cm maps from EoR are then contaminated with simulated Gaussian noise with varying RMS values to match the upcoming square Kilometer Array (SKA) in its low frequency regime observations, specifically the SKA1-Low. Then, the maps were combined with a 3D Gaussian kernel with a range of length scales to simulate the impact of the telescope beam, and the images were examined with LCS. Further research was done to determine how the array synthesized beam impacts the 21-cm signal's LCS analysis, particularly in light of upcoming SKA1-Low observations. We used an accurate OSKAR simulation based on the 21cmE2E-pipeline to achieve this. According to our research, the array beam introduces bias into the LCS estimation of synthetic observations, which causes the apparent percolation transition point to shift toward the later stages of reionization. Interpreting any 21-cm signal images from telescopes like the SKA requires careful consideration of the impact of the telescope's synthesized beam and lower resolution. In order to better deconvolve the array synthesized beam from the 21-cm images and obtain a more accurate estimate of LCS, we suggest using denser  $uv$ -coverage at longer baselines.

# Contents

<b>List of Figures</b>	<b>i</b>
<b>List of Tables</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Our Universe’s timeline . . . . .	3
1.2 Observation of Cosmic Dawn (CD) and Epoch of Reionization (EoR) . . . . .	5
1.2.1 Lyman-alpha forest . . . . .	6
1.2.2 Thompson scattering of the CMB photons . . . . .	7
1.2.3 The 21-cm Line of neutral hydrogen . . . . .	8
1.3 21-cm signal from H <sub>I</sub> . . . . .	9
1.3.1 Applications of the 21-cm line in Astrophysics and Cosmology . . . . .	10
1.4 Analysis of the cosmological 21-cm signal . . . . .	10
1.4.1 Importance of image analysis of cosmological 21-cm maps . . . . .	12
<b>2 Topological analysis of 21-cm maps From the EoR</b>	<b>14</b>
2.1 How to Quantify Topology? . . . . .	15
2.1.1 Different topological methods to analyse the 21-cm field . . . . .	17
2.2 Topology in the context of H <sub>I</sub> 21-cm signal . . . . .	18
2.3 Minkowski functionals and Minkowski tensors . . . . .	18
2.3.1 Hadwiger’s theorem . . . . .	18
2.3.2 Granulometry . . . . .	21
2.3.2.1 Theoretical idea . . . . .	21
2.3.2.2 Application on the 21-cm field . . . . .	21
2.3.3 Persistence theory & Betti numbers . . . . .	23
2.3.3.1 Homology groups and Betti Numbers . . . . .	23
2.3.4 Percolation theory . . . . .	25
2.3.5 Largest Cluster Statistics . . . . .	26

<b>3</b>	<b>Simulation of the 21-cm field</b>	<b>27</b>
3.1	Components of generating a reionization model . . . . .	27
3.2	Principles of EoR 21-cm signal modelling . . . . .	29
3.2.1	Analytical models . . . . .	30
3.2.2	Numerical models . . . . .	31
3.2.3	Semi-numerical models . . . . .	31
3.2.4	Simulations used in the current project . . . . .	33
3.3	Source Models . . . . .	34
3.3.1	Ultraviolet photons (UV photons) . . . . .	34
3.3.2	Uniform Ionizing Background (UIB Photons) . . . . .	35
3.3.3	Soft X-Ray photons (SXR) . . . . .	35
3.3.4	Power Law mass dependent efficiency (PL) . . . . .	36
3.4	Reionization scenarios . . . . .	36
<b>4</b>	<b>Topology of the 21-cm field</b>	<b>40</b>
4.1	Overview on the previous work . . . . .	40
4.2	Discussion on the results of Pathak et al. 2022 . . . . .	41
4.2.1	Considerations for identification of ionized regions . . . . .	41
4.2.2	Analysis methods and results . . . . .	42
4.2.3	Analysis of LCS with percolation analysis . . . . .	43
4.2.4	Shapefinders . . . . .	46
4.2.5	SURFGEN2 code and Marching Cube 33 algorithm . . . . .	46
4.2.6	Determination of Shapefinders . . . . .	47
4.2.7	Variation in number of ionized regions ( $N_C$ ) with different parameters . . . . .	49
4.2.8	Results obtained from Shapefinders . . . . .	50
4.2.8.1	Genus . . . . .	50
4.2.8.2	Planarity and Filamentarity . . . . .	51
4.2.8.3	T, B and L . . . . .	51
4.3	Analysis on the mean subtracted field . . . . .	51
4.3.1	Issues in analysing mean subtracted fields . . . . .	52
4.3.2	Summary of this work . . . . .	53
4.3.3	Prospects . . . . .	54
<b>5</b>	<b>Analysis of LCS on sampled maps of lower resolution</b>	<b>67</b>
<b>6</b>	<b>Testing LCS on realistic H<sub>I</sub> 21-cm maps</b>	<b>70</b>

6.1	Addition of noise	71
6.2	Determination of threshold	72
6.3	Results after addition of noise and smoothing	73
<b>7</b>	<b>Creating observational maps and determination of LCS</b>	<b>77</b>
7.1	The End-to-End pipeline at hand	77
7.1.1	Telescope model	79
7.1.2	OSKAR	80
7.1.2.1	OSKAR simulation	82
7.1.2.2	Sky model used	83
7.1.3	CASA	84
7.1.4	Results obtained	89
<b>8</b>	<b>Summary and future work plan</b>	<b>94</b>
8.1	Summary	94
8.2	Future work-plan	95

# List of Figures

1.1	Earth, 1.44 billion km away in this image, appears as a blue dot at centre right. (Image: International Cassini Spacecraft ) . . . . .	1
1.2	Lyman- $\alpha$ forest is a series of absorption lines in the spectra of distant galaxies and quasars that result from the neutral hydrogen atom's Lyman-alpha electron transition. [Murphy, 2016] . . . . .	6
1.3	Energy levels of the hydrogen atom. Different transition lines in the hydrogen atom is also shown in the image [Bradt, 2008]. . . . .	9
2.1	An image of a sphere. This is rendered using an opensource tool called GeoGebra. . . . .	16
2.2	An image of an ellipsoid. This is rendered using an opensource tool called GeoGebra. . . . .	17
3.1	Variation of mass averaged neutral fraction ( $\bar{x}_{\text{HI}}$ ) with redshift ( $z$ ) . . . . .	38
3.2	Simulated 21-cm image field for different models are shown for $\bar{x}_{\text{HI}} \approx 0.5$ . . . . .	39
4.1	Variation of FF with $\bar{x}_{\text{HI}}$ as seen in [Pathak et al., 2022]. . . . .	42
4.2	Same plot reproduced as a part of this work. . . . .	43
4.3	Variation of LCS with FF and $\bar{x}_{\text{HI}}$ as seen in [Pathak et al., 2022]. . . . .	44
4.4	LCS vs FF and $\bar{x}_{\text{HI}}$ as reproduced in this work. . . . .	56
4.5	Variation of NC with FF and $\bar{x}_{\text{HI}}$ as seen in [Pathak et al., 2022]. . . . .	57
4.6	$N_C$ vs FF and $\bar{x}_{\text{HI}}$ as reproduced in this work. . . . .	58
4.7	Variation of genus with $\bar{x}_{\text{HI}}$ as seen in [Pathak et al., 2022]. . . . .	59
4.8	Variation of Genus as reproduced in this work. . . . .	59
4.9	Variation of Planarity and Filamentarity with $\bar{x}_{\text{HI}}$ as seen in [Pathak et al., 2022]. . . . .	60
4.10	Planarity and Filamentarity vs $\bar{x}_{\text{HI}}$ as reproduced in this work. . . . .	61
4.11	Variation of Length and $T \times B$ with $\bar{x}_{\text{HI}}$ as seen in [Pathak et al., 2022]. . . . .	62
4.12	Variation of Length and $T \times B$ with $\bar{x}_{\text{HI}}$ as reproduced in this work. . . . .	63
4.13	Threshold for non mean subtracted hydrogen map is plotted against changing neutral fraction ( $x_{\text{HI}}$ ) for different step sizes for Fiducial model. . . . .	64
4.14	Threshold for non mean subtracted hydrogen map is plotted against changing filling factor for different step sizes for Fiducial model. . . . .	64

4.15	Threshold for non mean subtracted hydrogen map is plotted against changing filling factor for different step sizes for Fiducial model. . . . .	65
4.16	Threshold for mean subtracted hydrogen map is plotted against changing filling factor for different step sizes for Fiducial model. . . . .	66
5.1	For three reionization source models, the LCS results from our original resolution maps (solid/dashed curves) are compared with those from the low-resolution 21-cm maps (dotted curves). The critical $\bar{x}_{\text{HI}}$ at the start of percolation transitions in all the models is unchanged, and the LCS results are consistent for the different resolutions.	68
5.2	In three reionization situations, the growth of the largest ionized region’s “cross-section” has been analysed for low-resolution (dotted curves) and the original resolution fields (solid/dashed curves). Again, the Shapefinders results in the two resolutions are consistent. . . . .	69
6.1	Slices of the brightness temperature maps for $z = 7.221$ ( $\bar{x}_{\text{HI}} = 0.2$ ). <i>Left:</i> The original down-sampled map having a grid size of 300 cells along each side. <i>Center:</i> Map after adding a Gaussian random noise of rms 3.10mK. Random fluctuations in the pixel intensities can be seen. <i>Right:</i> Representation of the map after smoothing the noisy map with a Gaussian smoothing kernel of smoothing length 7.14 cMpc to mimic the SKA-I Low behaviour. . . . .	71
6.2	An image of a typical histogram of an ideal HI 21-cm image cube. It can be seen that there are sharp bimodal features. The peak on the left correspond to the ionized pixels ( $\rho_{\text{HI}} = 0$ ) and the peak on the right correspond to the neutral pixels. . . . .	73
6.3	Variation of LCS with neutral fraction ( $\bar{x}_{\text{HI}}$ ) is shown. It is seen that as we go to higher noise levels, the sharp bimodal feature of the 21-cm maps are lost and LCS cannot be computed for higher $\bar{x}_{\text{HI}}$ values. . . . .	74
6.4	LCS vs neutral fraction ( $\bar{x}_{\text{HI}}$ ) for changing smoothing lengthscales. Due to the disappearance of small ionized regions at higher neutral fractions, the behaviour of LCS do not follow the expected trend. . . . .	75
6.5	LCS vs $\bar{x}_{\text{HI}}$ plotted for noise and Gaussian smoothing combined. It is seen that due to the incorrect threshold determination and error in the cluster-number detection, the percolation transition point is shifted to a later stage of reionization which is different from the value determined from simulation. . . . .	76
7.1	Schematic diagram of the 21cmE2E-pipeline used for analysing the HI 21-cm maps from the future SKA observations. . . . .	78

7.2	The telescope layout that has been used in this work. It comprises of stations around 2km from the central station. . . . .	80
7.3	$uv$ -coverage for 30 minutes observation. . . . .	81
7.4	$uv$ -density for the observation that has been taken into consideration. . . . .	81
7.5	The OSKAR simulator's general architecture. The front end and back end are separate components that communicate with one another using a file interface. [Mort et al., 2010]	82
7.6	LCS vs $\bar{x}_{\text{HI}}$ is plotted for maps of $300^3$ grid volume and $256^3$ grid volume. It can be seen that the percolation transition threshold remains the same even though the simulation volumes are different. . . . .	85
7.7	Iterative Image Reconstruction - Major and Minor Cycles. Image from CASA tclean	86
7.8	The schematics representation of the CASA tclean weighting and tapering summary. Image from CASA tclean . . . . .	87
7.9	Slice of the original HI 21-cm signal map using ReionYuga at a neutral fraction value of $\bar{x}_{\text{HI}} \approx 0.71$ from an image cube of size $300^3$ grid units. The image below shows artefacts due to incorrect weighing in the extended source pixels in Hogbom CLEAN method with natural weighting. . . . .	88
7.10	Comparison between the dirty image vs cleaned image slice of the simulated HI 21-cm map at a neutral fraction $\bar{x}_{\text{HI}} = 0.2$ . . . . .	89
7.11	Comparison between the slice of the original HI 21-cm signal map using ReionYuga and the simulated image after Multiscale CLEANing at a neutral fraction value of $\bar{x}_{\text{HI}} \approx 0.71$ from an image cube of size $300^3$ grid units. It can be compared from the simulated image shown above that the features of the contrast difference could be recovered using Multiscale CLEAN. . . . .	90
7.12	Comparison between the slice of the original HI 21-cm signal map and the simulated image after Multiscale CLEANing with the Briggs weighting scheme. It is important to note that these images have a grid volume of $256^3$ grid units. . . . .	91
7.13	Effects of artefacts as seen in LCS vs $\bar{x}_{\text{HI}}$ . Due to improper cleaning we see a trend in the LCS which is not expected and very different from our original case. . . . .	91
7.14	Estimated LCS for different neutral fractions using two Multiscale CLEAN techniques: natural weighted and Multiscale weighted with Briggs weighting (employing the robust parameter = 0.5). The LCS estimated using Briggs weighted CLEAN is superior to the natural weighted CLEAN in two ways. Firstly, we were able to calculate LCS values up to almost 0.7, which was not feasible with the natural weighted CLEAN. Secondly, the percolation transition was not shifted as significantly as in the case of natural weighted CLEAN. . . . .	93

# List of Tables

2.1	Minkowski functionals shown for 3 – $D$ Euclidean spaces . . . . .	20
4.1	Summary of the reproduced work of [Pathak et al., 2022]. Critical values indicate the values where percolation transition has occurred. . . . .	53
7.1	Observational parameters used in the simulations . . . . .	79

# Chapter 1

## Introduction



Figure 1.1: Earth, 1.44 billion km away in this image, appears as a blue dot at centre right. (Image: International Cassini Spacecraft )

“Look again at that dot. That’s here. That’s home. That’s us. On it everyone you love, everyone you know, everyone you ever heard of, every human being who ever was, lived out their lives. The aggregate of our joy and suffering, thousands of confident religions, ideologies, and economic doctrines, every hunter and forager, every hero and coward, every creator and destroyer of civilization, every king and peasant, every young couple in love,

every mother and father, hopeful child, inventor and explorer, every teacher of morals, every corrupt politician, every “superstar”, every “supreme leader”, every saint and sinner in the history of our species lived there—on a mote of dust suspended in a sunbeam.

The Earth is a very small stage in a vast cosmic arena. Think of the rivers of blood spilled by all those generals and emperors so that, in glory and triumph, they could become the momentary masters of a fraction of a dot. Think of the endless cruelties visited by the inhabitants of one corner of this pixel on the scarcely distinguishable inhabitants of some other corner, how frequent their misunderstandings, how eager they are to kill one another, how fervent their hatreds.

Our posturings, our imagined self-importance, the delusion that we have some privileged position in the Universe, are challenged by this point of pale light. Our planet is a lonely speck in the great enveloping cosmic dark. In our obscurity, in all this vastness, there is no hint that help will come from elsewhere to save us from ourselves.

The Earth is the only world known so far to harbor life. There is nowhere else, at least in the near future, to which our species could migrate. Visit, yes. Settle, not yet. Like it or not, for the moment the Earth is where we make our stand.

It has been said that astronomy is a humbling and character-building experience. There is perhaps no better demonstration of the folly of human conceits than this distant image of our tiny world. To me, it underscores our responsibility to deal more kindly with one another, and to preserve and cherish the pale blue dot, the only home we’ve ever known.”

— Carl Sagan [[Sagan, 1997](#)]

Cosmology is an empirical science that tells us the tale of our genesis. It is a branch of Astronomy that encompasses the origin and evolution of the entire universe and gives an outlook to the large scale properties of the universe.

## 1.1 Our Universe's timeline

The best model so far, we have about our universe's origin is called the Big Bang model of the universe. Crucially, the model is consistent with the Hubble–Lemaître rule, which states that the further a galaxy is from Earth, the quicker it moves away. The idea predicts an increasingly condensed universe preceded by a singularity in which space and time lose meaning when the universe is extrapolated backwards in time using existing physics equations from Einstein's General Relativity and The standard model of particle physics. The Big Bang singularity was roughly 13.8 billion years ago, according to detailed calculations of the universe's expansion rate, and is consequently considered the universe's birth [Ryden, 2003].

Right after the big bang, it is theorised that the size of the universe expanded by more than 60 “e-folds”, or a factor of  $10^{26}$  in  $10^{-30}$  seconds. For the next 18,000 years, the universe was a hot, uniform ‘soup’ of subatomic particles and photons. The cosmos was filled with a sea of plasma - charged particles – that formed a thick, virtually uniform fluid instead of stars and galaxies. There were very small density fluctuations of the order of  $10^{-5}$ . As the temperature was too high, the particles could not stick together, rather they started bouncing off of each other. Waves of pressure – sound – travelled through the plasma as the gravitational pull and the repelling force alternated. As the cosmos cooled with time, particles merged to form neutral atoms. Though the bouncing and the oscillations stopped, their imprints, are still visible throughout the universe. These oscillations are known as the Baryon Acoustic Oscillation [Guth, 1997].

As universe cooled down for further 400,000 years, decoupling of matter and radiation started to happen and radiation started leaking out all around the universe in all directions uniformly. It is called the Cosmic Microwave Background (CMB) and this point in time is also called the epoch of last scattering. Very tiny fluctuations in matter density can be found in this primordial, uniform, smooth and featureless ideal blackbody radiation [Loeb, 2010].

As time progressed and universe started to cool down and expanded,

baryons started recombining and the very first element in the universe, hydrogen was formed. Radiation finally decoupled from matter and this recombination of matter led to the start of structure formation with the help of collapsing cold dark matter halos. This epoch is called the Dark Ages [Loeb, 2010].

About 375,000 years after the Dark Ages the first stars and galaxies started to form in the universe. The tiny density fluctuations that were there, started to grow in volume. Huge amount of gas coalesced with the help of gravitational pull and led to the formation of the first stars and galaxies. This epoch is called the Cosmic Dawn [Loeb and Furlanetto, 2013].

The outburst of ultraviolet photons from these early stars, galaxies and other sources started ionizing the surrounding intergalactic medium filled with neutral hydrogen present from the time of Dark Ages creating an ionized bubble around these sources. As time progressed the hydrogen in the entire universe started getting ionized all over again, hence this era is known as the Epoch of Reionization. The cosmos became ‘transparent’ once the neutral hydrogen was nearly fully ionized as a result of this reionization [Loeb and Furlanetto, 2013].

To summarize the history and evolution of the universe, it started as a uniform, smooth CMBR and is continuing to expand in today’s non uniform universe with very high matter density fluctuation. To understand this “phase transition” of the universe where it became so non uniform from that uniform background radiation, we need to look at the time in between. Observations show that nothing much happened for about 400 million years after the CMBR where there was no luminous source when the universe was in neutral state in its Dark Ages. But afterwards, as the first stars started to form and they started to heat and ionize their surrounding IGM at the time of Cosmic Dawn (CD) and the subsequent Epoch of Reionization (EoR) [Furlanetto et al., 2006]. These two epochs are a crucial point in the phase transition of the universe. They can essentially tell us about the first luminous sources that created the ionizing photons which helped in the phase transition of the universe and brought us to our current state. The epoch of reionization can be probed using some particular transition lines. Unfortunately, these

lines coming from far off distance are extremely faint and are easily perturbed by the foreground sources. Till now, though no direct observation has been done for this Epoch of Reionization due to these constraints, many next-gen radio telescopes like the Square Kilometer Array (SKA) might as well have the potential to directly image the EoR [[Wel, 2020](#)].

## **1.2 Observation of Cosmic Dawn (CD) and Epoch of Reionization (EoR)**

The formation of the first compact objects like the stars and the galaxies during the time of Cosmic Dawn led to the ionization of the surrounding neutral baryonic elements predominantly the hydrogen and Helium and started the Epoch of Reionization. The understanding for these two epochs in the science community is still very feeble as mentioned in the previous section. The crucial questions that arise can be summarized as follows:

- a) When exactly did reionization start?
- b) What was the nature of reionization? Was it a gradual process or was it abrupt? Did ionization happened in a homogeneous manner or not?
- c) What were the dominant sources that caused reionization?

As we know that because light has a finite velocity, the farther we look in the sky, the older version of the universe is visible to us. So, if we want to look at the evolution of the cosmic history in these two epochs, in principle we can target our telescopes and observe. The very recently launched James Webb Space Telescope (JWST) has a major science goal for doing the same [[Gardner et al., 2006](#)]. Otherwise, we can use extremely large ground based optical telescopes like the 30m telescope [[Sanders, 2013](#)]. But, there is a caveat to this, i.e the results for these will be biased towards the brightest objects because they will be the easiest to detect, unless a large number of observations are made. So, this gives us a necessity to have complementary probes; especially the ones devoid of this bias [[Kalirai, 2018](#)].

hydrogen is the most abundant baryonic element in the universe, followed by helium. It covers almost three quarters of the baryonic mass budget, so they can be essentially excellent complimentary probes for the CD and EoR

[Choudhury et al., 2009]. In this report, we will be focusing on neutral hydrogen only. If we try to analyse the distribution of neutral hydrogen, it will give us an idea on the first galaxies that went on to ionize the surrounding IGM and reionizing the universe as more and more neutral hydrogen will become ionized as time progresses. So, different observables from hydrogen can be used to probe the EoR [Pritchard and Loeb, 2012].

### 1.2.1 Lyman-alpha forest

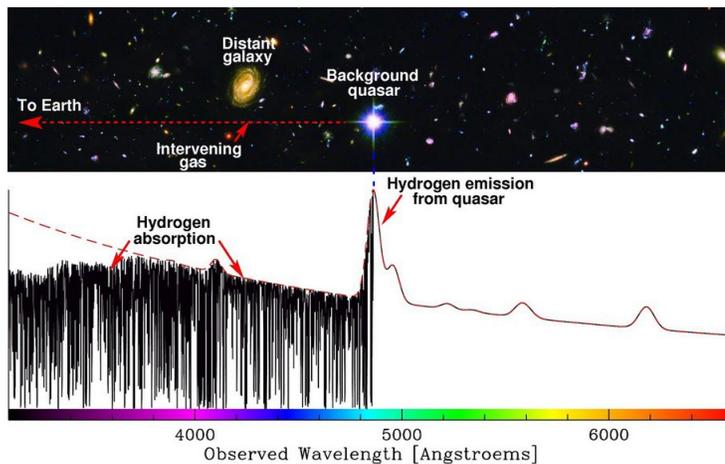


Figure 1.2: Lyman- $\alpha$  forest is a series of absorption lines in the spectra of distant galaxies and quasars that result from the neutral hydrogen atom's Lyman- $\alpha$  electron transition. [Murphy, 2016]

In Figure 1.2, a spectrum of a quasar is being shown at some particular redshift. The emission peak arises due to Lyman- $\alpha$  transition in hydrogen atoms in the surroundings of the quasar [Fan, 2003]. The Lyman- $\alpha$  photons are emitted when an electron transits from the ground ( $n=1$ ) to the second excited state ( $n=2$ ) and vice versa for absorption. Very dense absorption features are seen as we look further low in wavelength. They arise due to the hydrogen present in the path of photons in between the observer and the quasar. These absorption lines altogether seem like a forest, hence the name [Rauch, 1998].

For absorption only, the Cosmological radiative transfer equation [Gnedin and Ostriker, 1997] yields-

$$\nu = I_{\nu_Q}(t_Q) \left( \frac{1}{1+Z_Q} \right)^3 e^{-\tau_\nu} \quad (1.1)$$

Here,  $I_\nu$  is the specific intensity that is observed,  $I_{\nu_Q}(t_Q)$  is the specific intensity emitted from the quasar.

$\left( \frac{1}{1+Z_Q} \right)^3$  is the redshift dilution factor and  $\tau$  is the optical depth. The optical depth  $\tau$  is given by the following equation:

$$\tau_\nu = \int_0^s ds' n_{abs}(s', t') \sigma_{\nu'} \quad (1.2)$$

$n_{abs}$  is the number of absorbers and  $\sigma_{\nu'}$  is the cross section of absorption. So, if we calculate the optical depth in terms of redshift assuming the absorption profile to be a delta function at  $z \sim 3$ , [Rauch, 1998] we would get-

$$\tau_\nu \approx 10^5 \left( \frac{n_{HI}}{n_H} \right) \quad (1.3)$$

From here, we can conclude that the observed flux is the emitted flux with a factor of  $\exp(-10^5 x_{HI})$  where  $\bar{x}_{HI}$  or the neutral fraction is the ratio of mass density of neutral hydrogen to the density of total hydrogen.

This implies that even if the amount of neutral hydrogen is very low, the observed flux will be very less. But, in the figure we see that there are some emission features present in the Lyman- $\alpha$  forest implying that the amount of neutral hydrogen fraction is lesser than  $10^{-5}$  indicating a highly ionized universe. This effect in the Lyman- $\alpha$  forest is known as the Gunn-Peterson effect [Gunn and Peterson, 1965]. Hence, Lyman- $\alpha$  can be a probe for EoR studies. But as we go to higher redshifts (i.e  $z \geq 5.5$ ), this feature starts to fall off and at even higher redshifts, this feature goes completely missing. So, we require other probes.

### 1.2.2 Thompson scattering of the CMB photons

After the surface of last scattering, the emitted CMB photons that was streaming freely in the universe in all directions. But, when reionization happened, the free electrons from reionization started scattering these

CMB photons. So, reionization left a permanent mark on these CMB photons, which in theory can help us to probe EoR [Komatsu et al., 2011, Qin et al., 2020]. Now, the scattering mechanism that was involved is Thomson scattering and it has an angular dependence; so, it redistributes the angular pattern of the CMB photons and hence creates polarization which can be seen in the CMB temperature anisotropy [Komatsu et al., 2011].

To find the reionization effects on the CMB, we study the optical depth from Thomson scattering-

$$\tau = \sigma_T c \int_t^t dt \bar{n}_e (1+z)^3 \quad (1.4)$$

The optical depth is proportional to the number of free electrons i.e.  $\bar{n}_e$  so, it can complement the Lyman- $\alpha$  probe and provide us with valuable information about the EoR [Komatsu et al., 2011, Qin et al., 2020].

### 1.2.3 The 21-cm Line of neutral hydrogen

The 21-cm line is generated When an electron in the parallel state in neutral hydrogen flips into the anti parallel state, the energy change corresponds to a rest frame frequency of 1420 MHz or a corresponding rest frame wavelength of 21-cm. As hydrogen is the most abundant baryonic element in the universe, almost covering 75% of the entire mass budget, the 21-cm line can be an excellent complementary measure to look at the evolutionary history of our universe [Furlanetto et al., 2004, Loeb and Furlanetto, 2013]. What we can in principle do is that we can create a map of the evolving hydrogen density distribution over a large redshift range. So if we are able to track the evolution of hydrogen distribution we can tell about the sources and their properties that might have reionized the surrounding IGM and it can essentially give us a time-lapse video of our universe's history [Bradt, 2008]. Many next-gen radio telescopes like the Square Kilometer Array (SKA) [Datta et al., 2016] might as well have the potential to directly image the EoR. But there are certainly a lot of challenges that we face when we try to measure the 21-cm signal. The ability of 21-cm interferometric studies to differentiate Epoch of Reionization (EoR) signal models is generally restricted by the simplicity of data models, which commonly simplify or ignore foreground

signals [Watkinson et al., 2021], instrument features such as the instrumental effect on the beam [Iheanetu et al., 2019] and the effect of ionosphere [Sokolowski et al., 2015]. The focus of this report will be on this particular probe for the EoR.

### 1.3 21-cm signal from H<sub>I</sub>

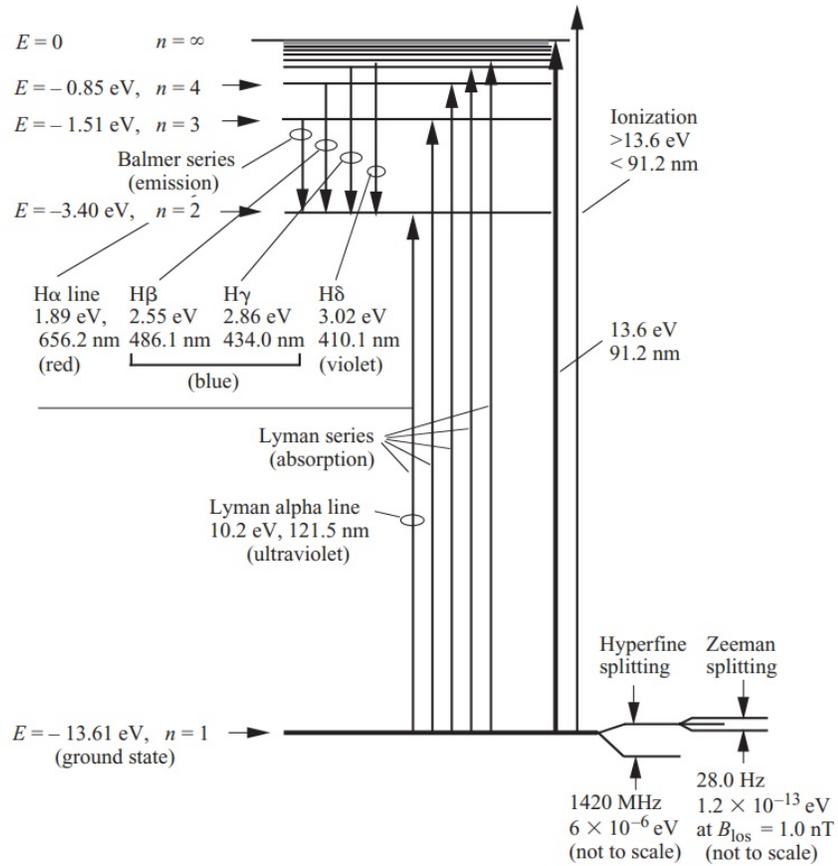


Figure 1.3: Energy levels of the hydrogen atom. Different transition lines in the hydrogen atom is also shown in the image [Bradt, 2008].

The hydrogen atom consists of an electron and a proton. Both the proton and the electron have angular momentum (or spin) and a magnetic moment associated with them. The electron spin in the ground ( $n = 1$ ) state of hydrogen can be either parallel or anti-parallel to the proton spin. The interaction of magnetic moments cause an energy difference generation between these two states that is  $\Delta E = 6 \times 10^{-6}$  which corresponds to a rest

frame wavelength of  $\lambda = 21\text{cm}$  or a rest frame frequency  $\nu = 1420\text{ MHz}$ . This is known as the 21-cm hyperfine spin-flip transition of neutral hydrogen.

### 1.3.1 Applications of the 21-cm line in Astrophysics and Cosmology

Hydrogen being the most abundant baryonic element in the universe, the 21-cm line from neutral hydrogen can be an excellent complimentary probe for the CD and EoR. The probability per unit time for a neutral hydrogen atom for a spin-flip transition is extremely low i.e.  $2.85 \times 10^{-15}\text{ s}^{-1}$  [Bradt, 2008]. It means in principle it would take approximately 10 million years for the spin-flip transition to occur naturally in a hydrogen atom indicating that the transition is extremely rare. But, In diffuse locations, the number density of neutral atomic hydrogen (  $\text{H I}$  ) can reach from  $10^9\text{ atoms/m}^3$  to  $\geq 10^{13}\text{ atoms/m}^3$  in the  $\text{H I}$  clouds which is quite high. At the time of EoR the number density was even higher, which essentially counterbalances the rarity of the transition. Even though the transition is extremely rare, the abundance of neutral hydrogen present at the time of EoR gives rise to a significant amount of radiation to be detected.

As a result of the Heisenberg Uncertainty relationship, we can expect a highly sharp emission line with a minimal  $\delta E$  energy dispersion (or line width) in frequency [Liu, 2008]. This feature enables for very precise  $\text{H I}$  source velocity estimations using simple measurements of the 21-cm line's Doppler shift. Along with this, precise Doppler spectroscopy of the 21-cm line allows for the exploration of the galactic structures and their dynamics with velocity curve predictions [Liu, 2008].

## 1.4 Analysis of the cosmological 21-cm signal

Many first generation radio interferometers like the Giant Meterwave Radio Telescope (GMRT)[Swarup, 1991], LOw Frequency ARray (LOFAR)[van Haarlem et al., 2013a, Yatawatta et al., 2013, van Haarlem et al., 2013b, Jelić et al., 2014, Mertens et al., 2020], Murchison Widefield Array (MWA) [Tin, 2013, Bowman et al., 2013] are trying to do interferometric detection of the 21-cm signal from the IGM over a vast range of cosmic time. However, due to their low

sensitivity it will not be possible to make tomographic images of this signal. Therefore, they adopt the approach of statistical detection of the signal. As the basic observable in these radio interferometers is visibility [SNELL, 2021], which is nothing but the Fourier transform of the sky brightness convolved with the primary beam, thus it is natural to use Fourier statistics, e.g. power spectrum as the target statistic for the detection of the signal [Furlanetto et al., 2006, Shi, 2015, Choudhuri et al., 2020]. As the 21-cm field at the time of EoR is highly non-Gaussian [Bharadwaj and Pandey, 2005, Yamauchi, 2022, Mellema et al., 2006] higher order Fourier statistics need to be done. Bispectrum [Majumdar et al., 2018, Hutter et al., 2020, Majumdar et al., 2020, Kamran et al., 2021, Watkinson et al., 2021, Tiwari et al., 2022] and Trispectrum [Cooray et al., 2008, Shaw et al., 2019, Mondal et al., 2016, Lewis, 2011] are the different higher order Fourier statistics that are carried out to analyse the signal further. Upcoming telescopes like the Square Kilometer Array (SKA) [Datta et al., 2016] might as well have the required sensitivity to directly image the EoR and create high resolution tomographic images from the data received. But there are certainly a lot of challenges that we face when we try to measure the 21-cm signal. The ability of 21-cm interferometric studies to differentiate Epoch of Reionization (EoR) signal models is generally restricted by the simplicity of data models, which commonly simplify or ignore foreground signals [Di Matteo et al., 2002, Jelić et al., 2008, Datta et al., 2010, Watkinson et al., 2021], instrument features such as the telescope thermal noise [Morales, 2005, McQuinn et al., 2006], instrumental effect on the beam [Iheanetu et al., 2019] and the effect of ionosphere [Sokolowski et al., 2015]. However, the Fourier statistics methods do not contain the phase information of the 21-cm signal and thus complementary analysis are required to be done.

There are several methods that are being employed directly on the simulated tomographic image of the 21-cm field in real space. Minkowski Functionals (MFs) [Friedrich et al., 2011, Bag et al., 2019] is one such example that is used to track the morphological and topological evolution of the 21-cm field during the EoR. The generalized tensorial

form for MFs are the Minkowski tensors [Kapahtia et al., 2019] which contains the directional information of the MFs. Other methods like the percolation theory [Iliev et al., 2006], granulometry [Kakiichi et al., 2017] and persistence theory [Elbers and van de Weygaert, 2019] are employed to find the topological phases of the ionized field at the time of EoR. Another method based on the Betti numbers [Giri and Mellema, 2021] can provide us with information related to the topological state of the IGM like its connectivity, number of cavities and tunnels [Giri et al., 2018a]. The method of local variance [Gorce et al., 2021] tracks the morphological evolution of the reionization environment. However, most of these methods essentially require the detection of large number of ionized hydrogen regions and for the analysis to be done. On the contrary, the method of Largest Cluster Statistics (LCS) [Bag et al., 2018, Bag et al., 2019] tracks the largest ionized region and applies the percolation theory based on it. Alongside, a method of “Shapefinders” [Sahni et al., 1998] are employed to get a complementary analysis.

#### **1.4.1 Importance of image analysis of cosmological 21-cm maps**

As mentioned earlier, the Fourier statistics of the 21-cm field do not contain the information related to its phase. This is one major reason that direct analysis in real space on the 21-cm field should be done. The image analysis of the field would theoretically involve topological analysis of the field which we will discuss later. This analysis has a potential to be a descriptor of the evolving geometrical features of the ionized regions during the time of EoR. In principle, the analysis would provide us with great detail on the nature of evolution of the IGM which can tell us about the physics related to IGM and also about the nature of ionizing sources and their properties at the time of EoR. Broadly speaking, the possible information that can be acquired by image analysis are the following:

Firstly, it can tell about the reionizing sources where we can possibly tell about their

- (a) Distribution
- (b) Mass

(c) Photon production rate

(d) Spectra

Secondly, We can have an idea on the physics of the IGM which involve

(a) Lyman- $\alpha$  coupling

(b) X-Ray heating

Thirdly, the distribution of matter at the time of reionization can be known. In principle we can track the entire reionization history from the data which essentially tells us the evolutionary pattern of how uniformity of the CMBR evolved to today's high matter density fluctuating universe.

This project aims to demonstrate that the information regarding the first sources of light and IGM physics during the EoR can be extracted using the analysis of the tomographic maps of H<sub>I</sub> 21-cm signal under realistic observational conditions. Our work mainly focuses on using the simulated 21-cm maps to analyse the topological and morphological evolution of the ionized or neutral hydrogen from the time of EoR using a novel statistical method called LCS as a follow up of [Pathak et al., 2022] along with a method called Shapefinders [Bag et al., 2019, Pathak et al., 2022]. To create a more realistic environment which mimics the actual observational scenario, the mean is subtracted from the 21-cm field in order to mimic interferometric data. We aim to apply LCS and Shapefinders on this data to verify if the statistics yield similar results and verify if LCS is still a strong enough probe.

## Chapter 2

# Topological analysis of 21-cm maps From the EoR

Topology in general deals with the mathematical properties of an object's geometrical features [Dominus, 2010]. Coarsely, topology can be of four types:

- (a) **General Topology or Point Set Topology:** General topology usually deals with the local properties of spaces. It generalizes the concept of continuity to define topological spaces, in which limits of sequences can be considered.
- (b) **Combinatorial Topology:** The oldest branch of topology called the Combinatorial topology tells us about the global properties of spaces which can be built up from a network of vertices, edges, and faces.
- (c) **Algebraic Topology:** Algebraic topology uses algebraic objects such as groups and rings to answer topological questions related to the global properties of spaces.
- (d) **Differential Topology:** Differential topology deals with spaces with some kind of smoothness associated to each point. Differential topology is mostly used for understanding the properties of vector fields, such as a magnetic or electric fields.

The analysis of the topology of the 21-cm field is extremely important

because it can potentially tell us a lot about the geometrical evolution of our universe during the EoR. Also, it is to be noted that in this report the topology that is used is the *Combinatorial Topology* [Friedrich et al., 2011]

## 2.1 How to Quantify Topology?

The topology of a structure is quantified by its topological properties. A topology is defined by metrics like the metric of Euclidean spaces, and broadly, topological spaces are the generalized form of any metric space [Dominus, 2010]. Formally, the definition of a topological space is given below.

There are two types of deformations that are taken into account in topology which are homeomorphisms and homotopies [Kuronya, 2010]. The properties that are invariant under such deformations are called topological properties. [Kuronya, 2010]: A topological space is an ordered pair  $(X, \tau)$ , where  $X$  is a set,  $\tau$  a collection of subsets of  $X$  satisfying the following properties:

1.  $\Phi, X \in \tau$ ,
2.  $U, V \in \tau$  implies  $U \cap V$
3.  $U_\alpha \mid \alpha \in I$  implies  $U_{\alpha \in I}$  implies  $U_\alpha \in \tau$

The topological properties can be understood with the help of a few examples:

- i **Dimension:** This property helps to distinguish between a line and a surface.
- ii **Compactness:** Compactness allows to differentiate between a line and a circle.
- iii **Connectedness:** This property helps distinguishing a circle from two non-intersecting circles.

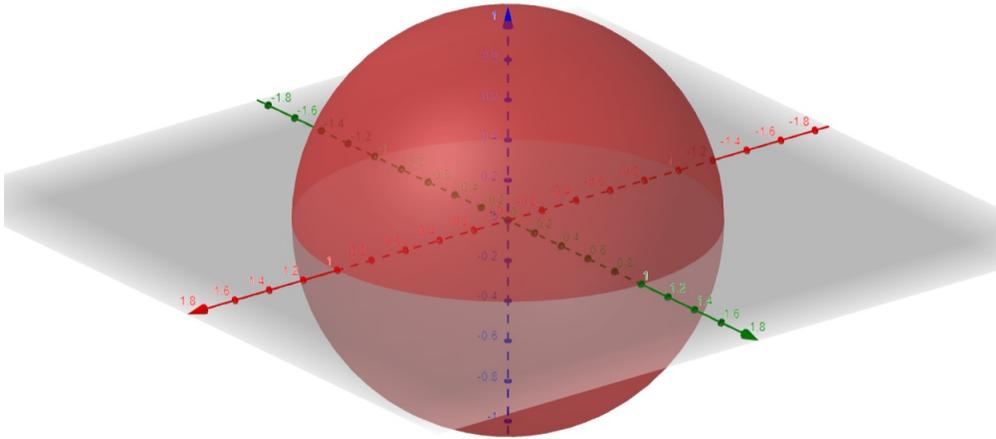


Figure 2.1: An image of a sphere. This is rendered using an opensource tool called GeoGebra.

Topological spaces can be easier understood with visualization using different geometrical objects. For ease of understanding, let us take a sphere that is made of play dough. We can convert this sphere into some other shape like an ellipsoid. But, there are two conditions that we must need to satisfy:

- (a) We cannot create any hole in the dough
- (b) We can not concatenate two points on the dough to merge them (like a doughnut)

If some shape can be transformed into another shape with those conditions being met, then they are topologically equivalent and the property that they are following is called homeomorphism. Another famous example of topological homeomorphism is between a coffee mug and a doughnut.

The application of topology extends to different branches of Physics including Astrophysics and Cosmology. Calabi-Yau manifolds [Yau, 2008] hold important relevance in String Theory. In Cosmology, specifically called the spacetime topology, topology is used to find the shape of the universe.

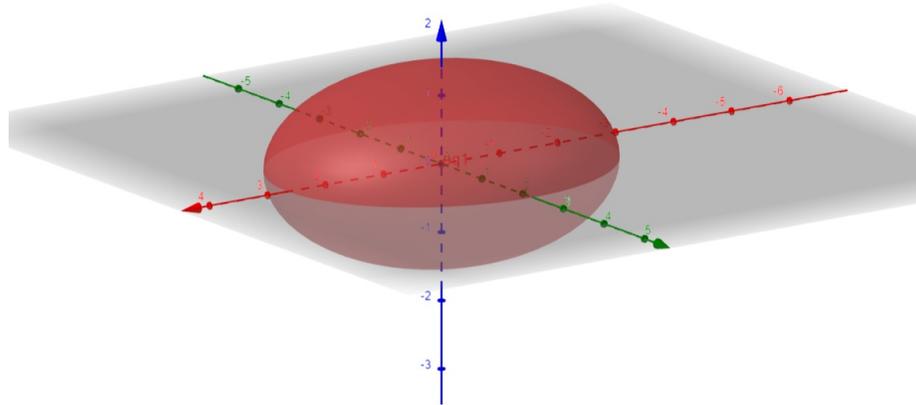


Figure 2.2: An image of an ellipsoid. This is rendered using an opensource tool called GeoGebra.

Along with this, in our context, the evolution of the morphology and topology of neutral hydrogen during the EoR can provide us with rich information related to the ionizing sources of that era. In the context of image analysis certain morphological and topological methods are implemented and this report will only contain the relevant ones for our need.

### 2.1.1 Different topological methods to analyse the 21-cm field

There are numerous widely used methods to analyse the 21-cm field in real space to understand the topological and morphological evolution of neutral hydrogen during the EoR [Friedrich et al., 2011]. A few of them are

-

- (a) Minkowski Functionals and Minkowski Tensors [Friedrich et al., 2011, Kapahtia et al., 2019]
- (b) Granulometry [Kakiichi et al., 2017]
- (c) Persistence Theory [Elbers and van de Weygaert, 2019]
- (d) Percolation Theory [Iliev et al., 2006, Bag et al., 2018, Bag et al., 2019]
- (e) Betti Numbers [Giri and Mellema, 2021]

(f) Largest Cluster Statistics [Bag et al., 2018, Bag et al., 2019, Pathak et al., 2022]

A brief discussion on these will be given in the next section.

## 2.2 Topology in the context of HI 21-cm signal

As mentioned in the previous section, the topological analysis of the 21-cm can be done with various methods as mentioned in the previous section. So, let us discuss them in brief.

## 2.3 Minkowski functionals and Minkowski tensors

One of the most important mathematical methods to describe complex patterns in space is integral geometry using a branch of morphological measures called the Minkowski functionals [Mecke et al., 1993, Buchert, 1994].

Let us assume, there are a set of points  $x_i/i = 1, \dots, N$  that represent the positions of a sample of galaxies in 3-D Euclidean space  $\mathbb{E}^3$ . We cover each of these points and create a neighborhood with a sphere of radius  $r$   $B_r = \{x \in \mathbb{E}^3 / \|x - x_i\| \leq r\}$ . The union of these spheres represents a collection of galaxies. Two spheres will be the part of same cluster if they are connected by a series of intersecting spheres. Minkowski functionals can be used to quantify the content, shape and connectivity of the unions  $B_r = \bigcup_{i=1}^N B_r(x_i)$  considering different  $r$  values. To understand integral geometry, especially in the context of Minkowski functionals, one needs to understand the Hadwiger's theorem [Chen, 2004].

### 2.3.1 Hadwiger's theorem

Let us assume there is a manifold with a group of transformations  $\mathcal{G}$ . For any  $d$  dimensional Euclidean space let us consider there is the set  $\mathcal{K}$  of convex bodies embedded and there exists a convex ring  $\mathcal{R}$  of all the finite unions of the convex bodies. To characterize a body  $\mathcal{B}$  from  $\mathcal{R}$  it is essential to find scalar functionals  $\mathcal{M}$  which satisfy the following requirements:

1. **Motion Invariance** : The functional  $\mathcal{M}$  should be independent of the body's position and orientation in space,

$$\mathcal{M}(g\mathcal{B}) = \mathcal{M}(\mathcal{B}) \text{ for any } g \in \mathcal{G}, \mathcal{B} \in \mathcal{R}. \quad (2.1)$$

2. **Additivity** : The union of two bodies result in addition in the functionals minus the functional of the intersection,

$$\mathcal{M}(B_1 \cup B_2) = \mathcal{M}(B_1) + \mathcal{M}(B_2) - \mathcal{M}(B_1 \cap B_2) \text{ for any } B_1, B_2 \in \mathcal{R}. \quad (2.2)$$

3. **Conditional Continuity** : The functionals of convex approximations to a convex body converge to the functionals of the body,

$$\mathcal{M}(K_i) \rightarrow \mathcal{M}(K) \text{ as } K_i \rightarrow K \text{ for } K, K_i \in \mathcal{K}. \quad (2.3)$$

For a  $d$  dimensional space, Hadwiger's theorem states that there are  $d+1$  independent functionals that satisfy the above requirements. The formal statement is given as follows:

Let  $\mathcal{R}$  be the convex ring embedded in  $d$  dimensional space. Then there exist  $d+1$  functionals  $\mathcal{M}_\mu, \mu = 0, \dots, d$  on  $\mathcal{R}$  such that any functional  $M$  on  $\mathcal{R}$  that is motion invariant, additive and conditionally continuous can be expressed as a linear combination of them:

$$M = \sum_{\mu=0}^d c_\mu \mathcal{M}_\mu \quad (2.4)$$

with numbers  $c_\mu$ .

In general, if  $A$  is a compact domain in  $\mathbb{R}^d$  with regular boundary  $\partial A \in \mathbb{C}^2$  and  $d-1$  principal radii of curvature  $R_i (i = 1, \dots, d-1)$ . The functionals  $W_\mu(A)$ , with  $\mu \geq 1$ , can be defined using the following,

$$W_{\mu+1}(A) = \frac{1}{(\mu+1) \binom{d}{\mu+1}} \int_{\partial A} S_\mu \left( \frac{1}{R_1}, \dots, \frac{1}{R_{d-1}} \right) d\mathbf{S} \quad (2.5)$$

Geometric Quantity	$\mu$	$M_\mu$	$V_\mu$	$W_\mu$	$\overline{V}_{3-\mu}$	$\omega_\mu$
V volume	0	V	V	V	V	1
A surface	1	A/8	A/6	A/3	A/2	2
H mean curvature	2	H/2 $\pi^2$	H/3 $\pi$	h/3	H/ $\pi$	$\pi$
$\chi$ Euler characteristic	3	3 $\chi$ /4 $\pi$	$\chi$	4 $\pi\chi$ /3	$\chi$	4 $\pi$ /3

Table 2.1: Minkowski functionals shown for 3 – D Euclidean spaces

Here,  $S_\mu$  denotes the  $\mu$ -th elementary symmetric function and  $dS$  is the (d-1) surface element. In 3-d the formula yields the following,

$$W_1 = \frac{1}{3} \int dS, W_2 = \frac{1}{3} \int H dS, W_3 = \frac{1}{3} \int G dS \quad (2.6)$$

Where  $G = \frac{1}{R_1 R_2}$  is called the Gaussian and  $H = \frac{1}{2} \left( \frac{1}{R_1} + \frac{1}{R_2} \right)$  is called the mean curvature.

In other words, if  $\omega_\mu$  is the volume of a  $\mu$  dimensional unit sphere, then four important notations are defined which are -  $M_\mu$ ,  $V_\mu$ ,  $\overline{V}_\mu$  and  $W_\mu$ . They are defined as follows:

$$V_\mu := \frac{\omega_{d-\mu}}{\omega_d} M_\mu, \overline{V}_{d-\mu} := \frac{\omega_{d-\mu}}{\omega_d} \binom{d}{\mu} M_\mu,$$

$$W_\mu := \frac{\omega_\mu \omega_d}{\omega_{d-\mu}} M_\mu, \text{ with } \omega_\mu := \frac{\pi^{\mu/2}}{\Gamma(1+\mu/2)}.$$

The notations for the Minkowski functionals in 3d space is shown in Table 4.1. The first functional denoted by V represents the volume of the body, second one is the surface area A. The third functional which is the Integral mean curvature H provides with the information related to the body's shape. The fourth Minkowski functional is called the Euler characteristic  $\chi$  and it is a pure topological quantity that provides information related to a body's genus  $g = 1 - \chi$  and it can be calculated by the simple formula

$$\chi = \text{number of components} - \text{number of tunnels} + \text{number of cavities}.$$

It is important to state that the analysis that is being done in this report has significant usage of the Minkowski functionals.

### 2.3.2 Granulometry

In mathematical morphology and image analysis, granulometry is a technique that helps to calculate a size distribution of grains in binary images [Soille et al., 1999]. In our context, this technique is used to obtain a measure of the size distribution of ionized hydrogen ( HII ) regions and 21-cm cold spots, the latter one is defined as the regions where the differential brightness temperature is lesser than a certain threshold value.

The idea of granulometry can be easier understood with an analogy of sifting of rocks in a gravel heap which are sieved through sievers of increasing size resulting in the small rocks to pass through. Similarly, the opening of an image using a fixed size of structuring element can be understood were the residue after each opening is quantified in terms of a granulometric curve which gives important features in the image such as object sizes in it.

#### 2.3.2.1 Theoretical idea

A granulometry is defined using a transformation that has a size parameter  $\lambda$  and satisfies the following three axioms:

- a Anti-extensivity: The particles that cannot pass through the sieve are a subset of the initial particles.
- b Increasingness: The particles that remain in the sieve after any subset of particles has been sifted through already are a subset of those remaining particles from the first subset.
- c Absorption: If there are two sifting transformations of two different sized  $\lambda$  and  $\nu$ , called  $\phi_\lambda$  and  $\phi_\nu$ , then the order of sifting will not change the result of overall sifting. That means sifting with  $\phi_\lambda$  and then sifting with  $\phi_\nu$  would yield the same result if done in the reverse order. In mathematical terms-

$$\phi_\lambda \phi_\nu = \phi_\nu \phi_\lambda = \phi_{\max(\lambda, \nu)} \quad (2.7)$$

#### 2.3.2.2 Application on the 21-cm field

In [Kakiichi et al., 2017] a sphere of radius R (or a disc of radius R for 2D nalysis) is chosen as the structuring element, defined by,

$$S_R = \prod (\mathbf{r} - \mathbf{r}_0 | R). \quad (2.8)$$

where,  $\mathbf{r}_0$  is the coordinate of the centre of the sphere. Afterwards, some operations of mathematical morphology is employed to quantify the shapes.

For a binary image,  $X$ , that is probed using a symmetric structuring element,  $S$ , there are four elemental operations-

- a Minkowski addition: It is denoted by the symbol  $\oplus$ , and defined as the union of a binary field,  $X$  and a structuring element,  $S$  as the element centre is moved inside the binary field.
- b Minkowski subtraction: It is denoted by the symbol  $\ominus$ , is defined as the intersection of a binary field,  $X$  and a structuring element  $S$  as the element centre is moved inside the binary field.
- c Morphological opening: It is denoted by the symbol  $\circ$ , is defined as a Minkowski subtraction followed by a Minkowski addition. For a binary field,  $X$  and structural element  $S$ , the morphological opening is expressed by,

$$X \circ S \equiv (X \ominus S) \oplus S. \quad (2.9)$$

- d Morphological closing: It is denoted by the symbol  $\bullet$ , and it is expressed as the following,

$$X \bullet S \equiv (X \oplus S) \ominus S \quad (2.10)$$

The sieving of a binary image  $X$  through a sieve of radius  $R$  is essentially a morphological opening and the new binary field can be expressed as-

$$X'(\mathbf{r}) = X \circ S_R \quad (2.11)$$

The sieved image contains the HII regions or the low density regions larger than the radius of the structuring element. By this method the size distributions of HII regions and cold spots are formulated mathematically.

### 2.3.3 Persistence theory & Betti numbers

This is a subsection of algebraic topology. Before understanding the theory we need to define and compute homology groups [Institute, 2012]. In order to quantify and express ‘holes’ in a geometrical object homology groups are formed. Simplicial complexes are introduced to compute homology groups. Chain, cycle and boundary groups are defined on these simplicial complexes. Betti numbers and persistence diagrams are topologies of a simplicial complex. A few important definitions need to be introduced to understand this concept properly.

1. **Simplices:** The d-dimensional simplex (d-simplex) is a spanning of d vertices,  $v_{i1}, v_{i2}, \dots, v_{id}$ . It is generally called  $\sigma$  and is denoted by  $\sigma = v_{i1}v_{i2}\dots v_{id}$ , with  $i_n \neq i_m \forall n, m$  with  $n \neq m$ . A 0-dimensional simplex is a point, 1-D is a line, 2-D is a face spanned by 3 vertices. A simplicial complex K is the collection of all simplices in a given structure.
2. **Chains:** Given a simplicial complex K, a p-chain or a p-dimensional chain is a collection of simplices given as  $c = \sum a_i \sigma_i$ , where  $\sigma_i$  are all p-simplices  $\in K$ , and  $a_i \in \mathbb{Z}/2\mathbb{Z}$ . The p-dimensional chain group  $C_p = \{c = \sum a_i \sigma_i \mid \sigma_i \in K, a_i \in \mathbb{Z}/2\mathbb{Z} \text{ and } \dim(\sigma_i) = p, \forall i\}$ .
3. **Boundary Map:** Given a p-simplex  $\sigma = v_1 \dots v_p \in K$ , the boundary map  $\partial_p : C_p \rightarrow C_{p-1}$ , is defined as  $\partial_p \sigma = \sum_{i=0}^p [v_1, \dots, \hat{v}_i, \dots, v_p]$ , the vertex with the hat is omitted,  $[v_1, \dots, \hat{v}_i, \dots, v_p] = v_1 \dots v_{i-1} v_{i+1} \dots v_p$ .
4. **Cycle group:** Given a simplicial complex K, the group of p-cycles is  $Z_p = \{c \in C_p \mid \partial_p c = 0\}$
5. **Boundary group:** Given a simplicial complex K, the group of p-boundaries is  $B_p = \{c \in C_p \mid \exists d \in C_{p+1} \text{ such that } c = \partial_{p+1} d\}$ .

#### 2.3.3.1 Homology groups and Betti Numbers

**Homology Groups:** Given the cycles  $Z_p$  and boundaries  $B_p$  of a simplicial complex in p-dimensions, the p-th homology group is defined as  $H_p = Z_p/B_p$ .

**Betti number:** The p-th Betti number is defined as the rank of the p-th homology group. Hence,  $\beta_p = \text{rank}H_p$ .

Intuitively speaking, the topological features in 0,1 and 2-dimensions are these:

0 represent the number of “objects” or connected components

1 represent the number of holes or tunnels

2 represent the number of voids or cavities

However, the proper way of calculating the Betti numbers is by using the definition  $\beta_p = \text{rank}H_p$ . Otherwise, an algorithm called incremental algorithm is used to quantify Betti numbers.

For example the letter ‘O’ has the Betti numbers  $\beta_0 = 1, \beta_1 = 1 \& \beta_2 = 0$  and a ball has the Betti numbers as  $\beta_0 = 1, \beta_1 = 0 \& \beta_2 = 1$

**Persistence:** The idea of persistence is based on the intricacies of the spatial connections between the numerous topological spaces, holes or borders underlying the global homology qualities [Rote and Vegter, 2006, Edelsbrunner and Harer, 2010]. Persistence establishes topology as a hierarchical notion, allowing for a far more detailed explanation of the topological structure of the cosmic mass distribution than is possible with traditional descriptions based on genus and even Betti numbers. It is founded on the notion that a systematic investigation of a field’s singularity structure may yield a wealth of topological information.

Morse theory, a branch of mathematics that analyses the singularity structure of a field, i.e. the positions of minima, maxima, and saddle points, as well as their mutual connections, plays a key role. The mathematical premise that there is a close link between the topology of the space and the critical points of any smooth function on the topological space is crucial in this regard [Milnor, 1969, Edelsbrunner and Harer, 2010]. Following this insight, Morse theory studies the critical points of a matching Morse function, i.e. a smooth scalar function defined on the topological space, to characterise the topology of the space. When the interval between the two defining threshold values does not contain any critical point, submanifolds defined as areas where the Morse function is in excess of

a certain functional threshold value are topologically identical or, more accurately, diffeomorphic. This has the essential implication that all changes in a space's topology occur only at key locations [Pranav et al., 2016].

[Elbers and van de Weygaert, 2019] suggests that persistent homology is a very ideal tool to study reionization by analysing the topological features of the ionization field where the connected components of the field are the  $H_{II}$  regions. For example, these topological features can be the ionization bubbles, neutral filaments that punch through the ionising bubbles creating tunnels or the neutral patches that are enclosed by ionized material. The betti numbers can also be related to the Euler characteristic as  $\chi = \beta_0 - \beta_1 + \beta_2$ . However, if they are considered separately, complementary measures can be quantified [Elbers and van de Weygaert, 2019].

### 2.3.4 Percolation theory

Percolation theory is the most basic, though imperfectly solved, description of a phase transition. Knowledge the percolation theory problem often leads to a better understanding of a variety of different physical systems. Furthermore, the notion of fractals, which is closely connected to the percolation theory problem, is of wide interest since it appears in Nature in many forms. These concepts are rigorously used in physics and cosmology, especially in the context of reionization. Before understanding percolation, let us have a brief introduction to scaling and renormalisation group theory [Kim et al., 2016].

- a Cluster: A cluster is a group of the closest neighboring sites. Percolation theory quantifies clusters and their properties.
- b Cluster number: Cluster number denoted by  $n_s(p)$  is the number of clusters per lattice size.
- c Percolation threshold: It is the concentration or the probability of occupancy  $p$  at which an infinite cluster appears for the first time in an infinite lattice. It is denoted by  $p_c$ .

Reionization of the neutral hydrogen can be understood very well with percolation theory [Furlanetto and Oh, 2016] as percolation is a very simple

example of phase transition and can contain rich information related to the morphology and topology of the regions we are dealing with. The ionized regions created by sources of soft ionizing spectra have sharp edges, hence, they have very well-defined boundaries where they can be modelled as clusters, mathematically. The work presented in this report also uses this theory in a robust manner to identify the transition where the small ionized bubbles suddenly merge to form a very big region that spans the entire simulation box, which will be discussed in later subsections.

### 2.3.5 Largest Cluster Statistics

In percolation analysis [[Klypin and Shandarin, 1993](#)] introduces the Largest Cluster Statistics or LCS as follows:

$$LCS = \frac{\text{Volume of the largest neutral or ionized region}}{\text{Total volume of all the neutral or ionized regions}} \quad (2.12)$$

With the progress of reionization, the largest bubble of ionized hydrogen grows in size, stretching the entire simulation volume at one point. Details on the simulation volume will be given in the next section. The point where this transition occurs can also be termed as the percolation transition. By analysing the change in the LCS, insights related to the geometry and topology of ionisation history can be gained [[Bag et al., 2018](#), [Bag et al., 2019](#), [Pathak et al., 2022](#)]. This report primarily deals with LCS which will be discussed in subsequent sections. In this work, we apply LCS on the mean subtracted hydrogen density maps and see the impact of the array synthesized PSF from SKA1-Low, subsequently.

# Chapter 3

## Simulation of the 21-cm field

Because of the large dynamic range requirements, simulating the reionization environment is a difficult process. The main challenge in modelling the reionization scenario from basic principles is to represent both big and small scale cosmology and astrophysics [Choudhury et al., 2016]. This necessitates simulating reionization in a big enough cosmic volume which is almost  $\sim 1 \text{ Gpc}^3$  to account for the effects of large scale matter density variations. At the same time, it is necessary to determine the reionizing sources which are typically galaxies nearly 10 Kpc in size in order to accurately imitate their characteristics.

### 3.1 Components of generating a reionization model

In order to put constraints on the reionization history some important ingredients are to be taken into account:

- a The primary step is to form the sources of ionization. This can be done by considering some analytical models where the standard forms of abundant dark matter haloes are given by the halo mass function  $dn(M, z)/dM$  in [Press and Schechter, 1974, Bond et al., 1991, Sheth and Tormen, 1999]. The masses and locations of these haloes can be found by  $N$ -Body simulations and subsequently group finder algorithms like the Friends-of-friend algorithm [Davis et al., 1985].

b After the dark matter haloes are obtained, the astrophysical processes related to galaxy formation, radiation emitted by the stars and the escape of the ionizing photons should be considered. Due to the high uncertainty of these processes at  $z \geq 6$ , some assumptions are taken into account which gives a relation between the number of ionizing photons to the halo mass. The simplest of these is given by the following formula:

$$N_\gamma = \zeta \left( \frac{\Omega_H}{\Omega_m} \right) \frac{M}{m_H}, \quad (3.1)$$

where  $\zeta$  is related to the star formation efficiency, number of ionizing photons produced and the fraction of escaped photons from the host galaxy to the IGM. The alternate expression for the given formula is:

$$\dot{n}_\gamma = \zeta n_H \frac{df_{coll}}{dt}, \quad (3.2)$$

where  $\dot{n}_\gamma$  and  $n_H$  are the number of photons in the IGM per unit time per unit comoving volume and the comoving number density of hydrogen and  $f_{coll}$  is the collapsed fraction that denotes the mass fraction in collapsed dark matter haloes which form stars.

$$f_{coll} = \frac{1}{\bar{\rho}_m} \int_{M_{min}}^{\infty} dM M \frac{dn(M, z)}{dM}, \quad (3.3)$$

Here,  $M_{min}$  is the minimum halo mass to form stars which is determined by the cooling efficiency of the gas in collapsed haloes.

c Now, as the baryonic density field is inhomogeneous, a descriptor for it is required. A suitable descriptor may be the P.D.F  $P(\delta_B)$  of the baryonic overdensity  $\delta_B$  which is being smoothed over the Jeans scale. The lognormal distribution [Bi and Davidsen, 1996, Choudhury et al., 2000, Choudhury and Ferrara, 2006] or any other fitting function obtained by hydrodynamic simulations [Becker and Bolton, 2013] can be used. It is important to be able to resolve small scales so as to identify the dense optically thick systems. As these regions have high recombination rate, they act as

sinks of ionizing photons and can cause changes in the distribution of the ionized regions.

d With the inhomogeneous baryonic density field and the underlying ionizing sources, the cosmological radiative transfer equation has to be solved through the IGM accounting for all astrophysical processes going on. The equation is as follows:

$$\frac{\partial I_\nu}{\partial t} + \frac{c}{a(t)} \hat{n} \cdot \nabla_x I_\nu - H(t) \nu \frac{\partial I_\nu}{\partial \nu} + 3H(t) I_\nu = -c\kappa_\nu I_\nu + \frac{c}{4\pi} \epsilon_\nu. \quad (3.4)$$

Here,  $I_\nu \equiv I(t, x, n, \nu)$  is the monochromatic specific intensity of the radiation field.  $\hat{n}$  is the unit vector along the propagation of radiation,  $\kappa_\nu$  is the absorption coefficient and  $\epsilon_\nu$  is the emissivity.

In order to create the simulation environment, the equation has to be solved in seven dimensions of the  $(t, x, n, \nu)$  space; but, this is very much computationally challenging. Hence, different numerical schemes and approximations are used [Iliev et al., 2006].

The radiative transfer equation can be simplified by taking a global average considering the assumption that the mean free path of the photons are way to small in comparison with the horizon size. In terms of the volume filling factor, the equation is given as:

$$\frac{dQ_{HII}}{dt} = \frac{\dot{n}_\gamma}{n_H} - Q_{HII} C \alpha(T) n_H, \quad (3.5)$$

where,  $C$  is the clumping factor and  $\alpha(T)$  is the recombination coefficient at temperature  $T$ .

### 3.2 Principles of EoR 21-cm signal modelling

In order to model the redshifted 21-cm signal from EoR the differential brightness temperature is used which is given by:

$$\delta T_b = \bar{T}_{xHI} \delta_B \left( \frac{T_s - T_{CMB}}{T_S} \right). \quad (3.6)$$

Here,  $T_S$  is the spin temperature of the gas and  $\bar{T}$  is given by:

$$\bar{T} = 27mK \left( \frac{\Omega_b h^2}{0.023} \right) \left( \frac{0.15}{\Omega_m} \frac{1+z}{10} \right)^{1/2}. \quad (3.7)$$

$x_{HI}$  is known as the neutral fraction of hydrogen and  $\delta_B$  is the baryonic overdensity.

Other than the early stages of reionization, it can be considered that IGM is sufficiently heated by X-rays emitted from the sources of that time and  $T_S$  couples with the gas temperature by  $Ly\alpha$  coupling (Wouthuysen, 1952; Field, 1959). Here,  $T_S \gg T_{CMB}$ , so, the simple relation  $\delta T_b \propto x_{HI} \Delta_B$  holds.

### 3.2.1 Analytical models

Analytical reionization models are based on simulating the size distribution of ionized bubbles surrounding galaxies, and then extrapolating the power spectrum. The power spectrum is defined as follows:

$$\langle \delta T_b(\hat{k}) \delta \hat{T}_b^*(K') \rangle = (2\pi)^3 \delta_D(k - k') P(k), \quad (3.8)$$

where,  $\delta T_b(\hat{k})$  is the Fourier transform of  $\delta T_b$ .

The most basic approach would be to represent the ionisation field as a collection of non-overlapping spheres with a radius of  $R$ . [Bharadwaj and Ali, 2005, Zahn et al., 2007]. To produce a certain ionized fraction  $Q_{HII}(z) = n_{bub} \frac{4}{3} \pi R^3$  at a given epoch, the values of  $R$  and the numerical density of such bubbles  $n_{bub}$  can be selected. Because these simplistic ionized bubble models do not account for overlap, they are only viable when the ionized fraction is very tiny. The excursion formalism technique developed by [Furlanetto et al., 2004], which we refer to as FZH04, can be used to account for such overlaps. A spherical region of radius  $R$  having a density contrast  $\delta$  will be fully ionized if the following condition is met:

$$\langle f_{coll} \rangle_{\delta, R} \geq \zeta^{-1}. \quad (3.9)$$

A modified excursion set peak (ESP) model can account for the fact that the haloes would prefer forming near the high density regions which predicts larger ionized bubbles.

### 3.2.2 Numerical models

Only the numerical solution of the cosmological radiative transfer equation can account for the full complexity of radiative transfer via the clumpy IGM. Radiative transfer simulations, on the other hand, are still computationally expensive, hence the equation is generally solved using appropriate approximations calculating the ionized hydrogen fraction's temporal evolution [Mellema et al., 2006, Iliev et al., 2006]. The radiative transfer code “Conservative Causal Ray-tracing technique” ( $C^2$ -RAY) uses one such approach, which works by tracing rays from all sources and solving for the temporal evolution of the ionized hydrogen percentage repeatedly [Mellema et al., 2006, Iliev et al., 2006]. It turns out that the  $C^2$ -RAY's ionisation fields have features that are quite close to those of the semi-numerical computations reported in [Majumdar et al., 2014].

A quicker way is to using the density field and haloes from a dark matter only  $N$ -Body simulation and post-processing with a spherically symmetric one-dimensional radiative transfer technique [Thomas and Zaroubi, 2008, Thomas et al., 2009]. The basic concept is to create spherically symmetric 21-cm patterns around individual sources (galaxies) and then account for the overlap of these patterns adequately in such areas. This has been implemented in [Ghara et al., 2015].

### 3.2.3 Semi-numerical models

Although analytical calculations based on excursion sets provide a reasonable description of the growth of HII regions during reionization, it is difficult to incorporate various complexities into the analytical framework, such as the overlapping bubbles can be non-spherical, the effects of line of sight peculiar velocities, and quantifying cosmic variance. Radiative transfer simulations may be used to create realistic radio maps by realization of the ionizing field, although they aren't always suitable for exploring the parameter space. A 3D radiative transfer simulation like  $C^2$ -RAY may be used to check for different physical processes occurring during the EoR. It operates on the notion of ray tracing by tracking the ionisation fronts in the IGM [Ricotti et al., 2002, Thomas et al., 2009, Iliev et al., 2014]. However, utilising these simulations to investigate

the multi-dimensional reionization parameter space is nearly impossible computationally. A realistic trade-off to account for these constraints can be using semi-numerical simulations. The semi-numerical approach is based on the excursion set formalism and is computationally efficient [Mesinger and Furlanetto, 2007, Poole et al., 2016]. One of these methods involves using perturbation theory to generate the dark matter density field, then using the analytical expression for the conditional mass function to calculate the collapsed fraction within each grid cell (of size  $R$  and density fluctuation  $\delta$ ), and finally using the excursion set formalism to generate the ionisation field [Mesinger et al., 2011]. A somewhat different strategy is to conduct a complete dark matter only  $N$ -Body simulation and use an appropriate group-finder technique to detect the haloes [Zahn et al., 2007, Choudhury et al., 2009]. The excursion set formalism that is used to generate the ionisation field, stays unchanged.

In order to understand the working of the semi-numerical models, the density field with the collapsed fraction at each grid point of the simulation box need to be realised. The spherically averaged collapsed fraction  $\langle f_{coll} \rangle_R$  for each grid point for a large range of  $R$  values is calculated. If the condition,  $\langle f_{coll} \rangle_{\delta,R} \geq \zeta^{-1}$  is satisfied, then the region is flagged as ionized. The points that do not satisfy the condition are assigned with a neutral fraction  $\zeta \langle f_{coll} \rangle_{R_{cell}}$ , where  $R_{cell}$  is the size of the grid cell set by the map resolution. So, finally we get a realisation of the ionization field for any given  $\zeta$ .

Recombination can also be taken into account as a spatially homogeneous effect and the ionization condition is modified as follows:

$$\langle f_{coll} \rangle_{\delta,R} \geq \zeta^{-1}(1 + N_{rec}^-) \equiv \zeta_{eff}^{-1}, \quad (3.10)$$

where,  $N_{rec}^-$  is the average number of recombinations per hydrogen atom.

However, as recombinations are inhomogeneous, high density regions will recombine faster and start self-shielding from the ionizing radiation and will also act as sinks of ionizing photons. This is accounted for by introducing an additional condition [Choudhury et al., 2009]:

$$\langle n_\gamma \rangle_R \geq n_H \frac{\epsilon t_H L}{t_{rec} R}, \quad (3.11)$$

where,  $\langle n_\gamma \rangle_R$  is the number density of ionizing photons averaged over radius  $R$ ,  $t_{rec}$  is the recombination time,  $\epsilon t_H$  is the time in which recombination will be significant where  $t_H$  is the Hubble time and  $L$ , the comoving size of the absorbing region considered to be the local Jeans scale.

### 3.2.4 Simulations used in the current project

Our work uses the semi-numerical approach as used in [Majumdar et al., 2016]. [Majumdar et al., 2016] employed a semi-numerical strategy that consists of three phases:

- a It uses  $N$ -Body dark matter gravity alone simulations to build the dark matter distribution at any chosen redshift. The  $N$ -Body simulations were performed with the *CUBEP<sup>3</sup>M* code (Harnois-Deraps et al.'2013), which is based on the earlier algorithm PMFAST, as part of the PRACE4LOFAR project (PRACE projects 2012061089 and 2014102339) [Merz et al., 2005]. At low distances, gravitational forces are calculated on a particle–particle basis, whereas at longer distances, they are calculated on a mesh. The size of the simulation volume was  $500/h = 714 Mpc$  (comoving) along each side.  $6912^3$  particles of mass  $4.0 \times 10^7 M_\odot$  on a  $13824^3$  mesh is considered. This is then downsampled to a  $600^3$  grid for reionization modeling. For each redshift output of the  $N$ -Body, haloes were identified using a spherical overdensity scheme. Minimum halo mass that is being considered here is  $2.02 \times 10^9 M_\odot$ .
- b Second, it locates collapsed dark matter halos within the simulated matter distribution, which may be done using methods such as Friends of Friend (FoF) or spherical smoothing.
- c Finally, it considers halos to be the most likely hosts of ionising photon sources, and it generates the ionisation field using excursion set formalism, which is then transformed into the 21-cm field.

The average number of photons in a given volume is compared to the average number of neutral hydrogen atoms in the same volume using this approach. The neutral hydrogen is assumed to follow the dark matter distribution in this case. After the dark matter density field at a fixed redshift is generated using the  $N$ -Body simulation, an instantaneous ionising photon field at the same redshift in a grid the same size as the dark matter density field is generated, depending on the source model of the reionization scenario under consideration that will be discussed in subsequent sections. These photon and neutral hydrogen density fields are usually built on a grid that is coarser than the real  $N$ -Body resolution. This grid has a resolution of 1.19 Mpc, or  $600^3$  cells in our case. It is worth mentioning that the cosmological parameters that we have used are according to the WMAP five year data release  $h = 0.7$ ,  $\Omega_m = 0.27$ ,  $\Omega_\Lambda = 0.73$ ,  $\Omega_b h^2 = 0.0226$  [Komatsu et al., 2009]

### 3.3 Source Models

Each of our simulated reionization situations is made up of a unique combination of ionising photon sources. There are four different source models that we have considered and with those four models we have simulated six different reionization scenarios. The numerous source types are described here, and the reionization possibilities are described in this section. In these simulated reionization scenarios, the source models differ in two essential ways: a) how the amount of ionising photons released by the sources is linked to their host halo mass, and b) how the ionising photons' rest frame energy is distributed. The four different models are described as follows:

#### 3.3.1 Ultraviolet photons (UV photons)

Galaxies with collapsing dark matter haloes are expected to be the primary producers of ionising photons in most reionization models. There is still a lot that we don't know about these high-redshift galaxies and their radiation. In most simulations, the total amount of ionising photons ( $N_\gamma(M_h)$ ) provided by a halo of mass  $M_h$  that hosts such galaxies is simply assumed as follows:

$$N_\gamma(M_h) = N_{ion} \frac{M_h \Omega_b}{m_p \Omega_m}. \quad (3.12)$$

Where,  $N_{ion}$  essentially quantifies the number of photons entering in the IGM per baryon,  $m_p$  is the proton mass. Radiative transfer or semi-numerical simulations of reionization consider this kind to be the most dominant kind of source model producing an “Inside-out” type of reionization scenario which is driven by Ultraviolet photons [Majumdar et al., 2016].

### 3.3.2 Uniform Ionizing Background (UIB Photons)

Unless there is a large rise in the escape fraction of ionising photons from them with rising redshift, or galaxies below the detection limits of current surveys contribute a significant fraction of the total ionising photons, the observed population of galaxies at high redshifts appears unable to maintain the universe ionized [Kuhlen and Faucher-Giguère, 2012, Mitra et al., 2015]. Another theory is that in the early Universe, sources of hard X-ray photons (such as active galactic nuclei or X-ray binaries) were prevalent, resulting in a nearly homogeneous ionising background. Hard X-rays would be able to depart their home galaxies and traverse great distances and get redshifted to the UV part of the electromagnetic spectrum and start ionising hydrogen [McQuinn, 2012, Mesinger et al., 2011]. If 100% of the photons comes from this type of background then that scenario is termed as an “Outside-in” reionization scenario that has the same number of ionizing photons at every location.

### 3.3.3 Soft X-Ray photons (SXR)

The proportion of extreme ultraviolet photons that escape from their home galaxies is a contentious topic. Soft X-ray photons, on the contrary, would not have much trouble escaping into the IGM from their host sources. There’s also a chance that X-ray generation was more common in high-redshift galaxies than in low-redshift galaxies [Kaaret et al., 2011, Fragos et al., 2013]. It is considered that the photons are uniformly distributed around the host haloes within a radius equal to their mean free path at that particular redshift. The mean free path is calculated using the same equation as used for UV photons. It is to be noted that at

high redshifts, the mean free path of ionising photons is a mostly unknown quantity. For the UV sources, a constant maximum smoothing radius of 70 comoving Mpc at all redshifts were employed, which is consistent with [Songaila and Cowie, 2010] findings at  $z \sim 6$ .

### 3.3.4 Power Law mass dependent efficiency (PL)

In this case, the number of ionising UV photons emitted by sources in collapsed dark matter halos is proportional to the halo mass to the  $n^{th}$  power, as follows:

$$N_\gamma(M_h) \propto M_h^n. \quad (3.13)$$

The scenario that we have considered in this report is of the index 3. Higher-mass haloes create more ionising photons in this source model, resulting in fewer but bigger ionized zones. This model is a rough approximation of a situation where reionization is fueled by rare strong sources like quasars, resulting in extreme type of topology.

## 3.4 Reionization scenarios

For this investigation, we use [Majumdar et al., 2016]’s 21-cm maps of six alternative reionization scenarios. These scenarios are made up of various combinations of the above-mentioned source models.

The **fiducial** scenario assumes that galaxies in the halos of mass  $2.02 \times 10^9 M_\odot$  provide 100% of the ionising UV photons. The **clumping** and **PL (n = 3)** scenarios are similar in that they involve 100% UV photon input from halos, but they differ somewhat from the fiducial. The only situation in which non-uniform density dependent recombination has been considered is clumping. High mass halos, on the other hand, have a larger weightage in UV photon output than low mass halos in the PL (n = 3) scenario because they follow the power law equation instead of being proportional to the halo mass. The **UIB dominated**, **SXR dominated**, and **UV+SXR+UIB** have a mixed contribution of different kinds of ionising photons. For instance, UIB

dominated scenario assumes 20% contribution from UV photon whereas 80% contribution from Uniform ionizing background photons. SXR dominated case on the other hand considers 20% contribution from UV photons but 80% contribution comes from Soft X-Rays. The UV+SXR+UIB scenario has 50% contribution from Ultraviolet photons, 10% from UIB and rest 40% from Soft X-Ray photons. All of these three scenarios are considered using a uniform recombination type.

The proportionality constant  $N_{ion}$  was calibrated in the reionization scenarios discussed here so that they all have the same mass averaged neutral fraction ( $\bar{x}_{H_I}(z)$ ) at a particular time/redshift, i.e. they have the same reionization history [Majumdar et al., 2016, Pathak et al., 2022]. The mass averaged neutral fraction is defined as follows:

$$\bar{x}_{H_I}(z) = \frac{\text{Average mass density of neutral hydrogen}}{\text{Average mass density of total hydrogen}}. \quad (3.14)$$

A complementary parameter filling factor (FF) is also defined:

$$FF = \frac{\text{Total volume of all ionized regions}}{\text{Total simulation volume}}. \quad (3.15)$$

Figure 3.1 depicts the evolution of the neutral proportion as a function of redshift for all of the scenarios. We discuss our results as functions of  $\bar{x}_{HI}$  rather than redshift ( $z$ ) since different scenarios follow the same  $\bar{x}_{HI}(z)$  profile. In figure 3.2, 21-cm field images are shown for different reionization scenarios.

In the latter part of the study where we study the instrumental effect on the 21-cm images, we created new H I 21-cm maps for the EoR by using  $1024^3$  particles with a mass of  $1.089 \times 10^8 M_\odot$  on a  $2048^3$  mesh grid, and then coarsened the grid with a factor of 8 to produce a  $256^3$  volume. The smallest halo considered in this simulation had a mass of  $1.089 \times 10^9 M_\odot$ , and the simulation cube had a length of 143.36 cMpc on each side. We kept the mean free path constant at  $R_{mfp} = 20.0$  cMpc and adjusted the  $N_{ion}$  values for each redshift to maintain consistency with the Fiducial model used in [Majumdar et al., 2016] and ensure that the reionization history (measured

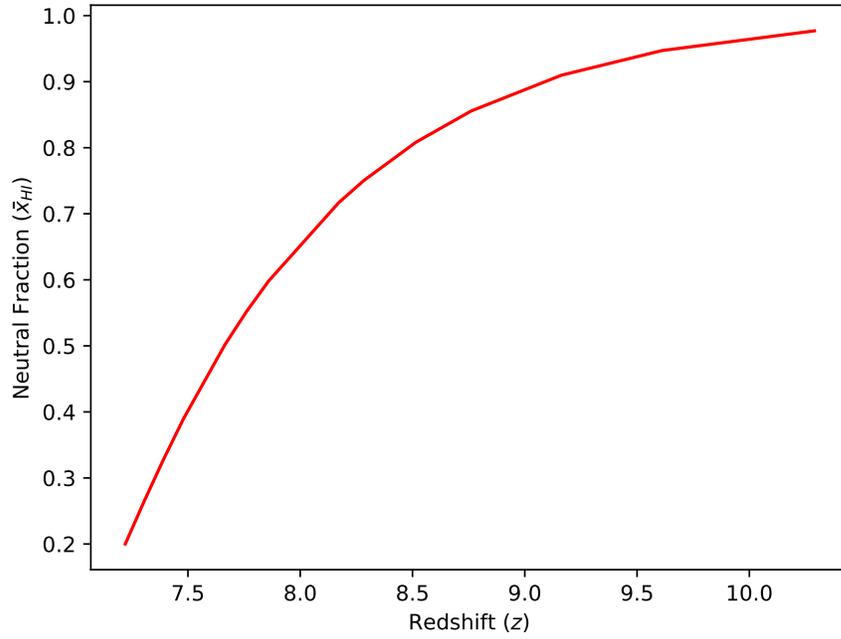


Figure 3.1: Variation of mass averaged neutral fraction ( $\bar{x}_{\text{HI}}$ ) with redshift ( $z$ )

by the neutral fraction  $\bar{x}_{\text{HI}}$ ) remained unchanged. The reasons for doing so is justified thoroughly in Section [7.1.4](#).

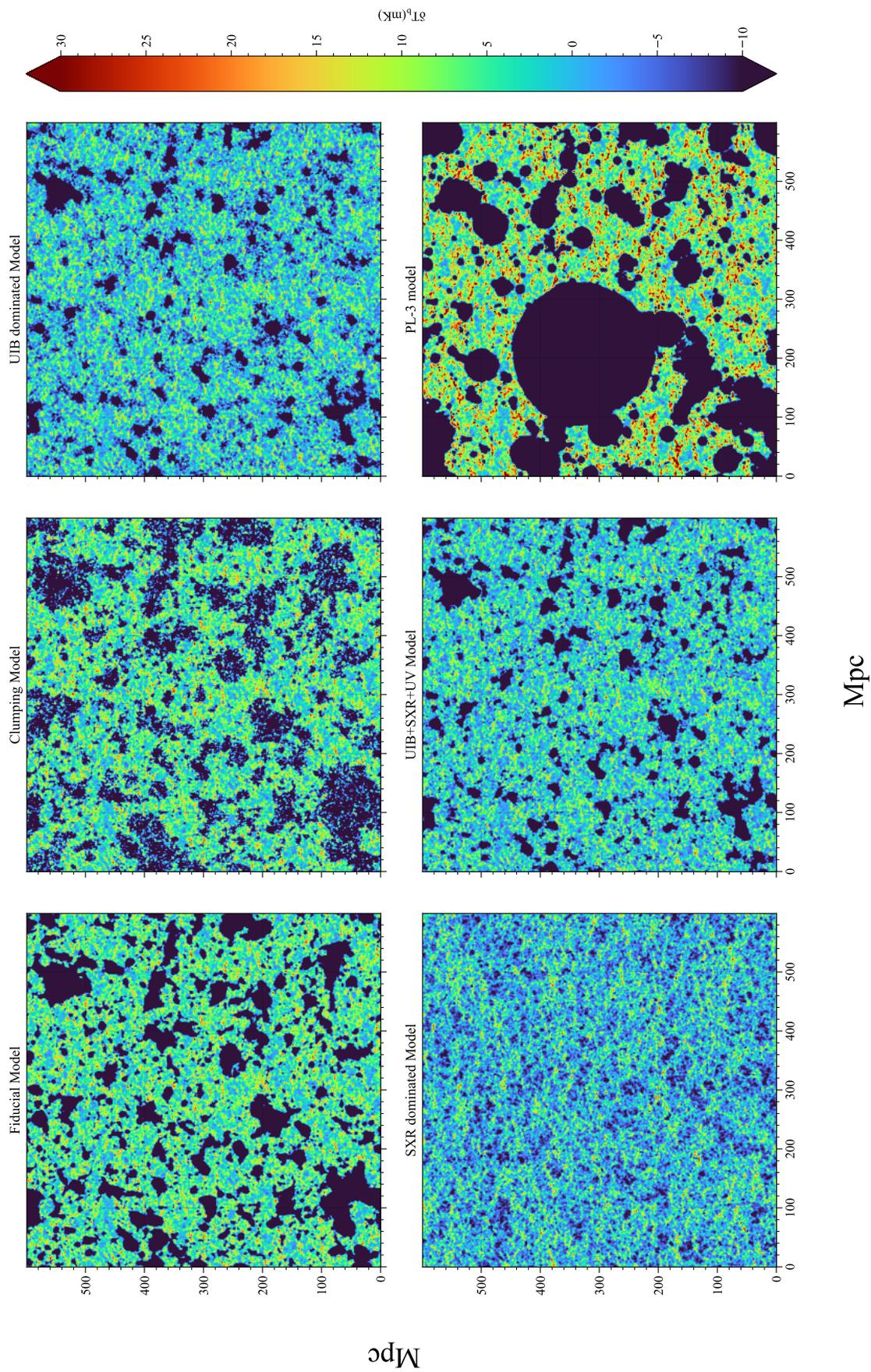


Figure 3.2: Simulated 21-cm image field for different models are shown for  $\bar{x}_{\text{HI}} \approx 0.5$ .

# Chapter 4

## Topology of the 21-cm field

### 4.1 Overview on the previous work

The topology of the 21-cm field have been checked using the methods in chapter 2 in literature. Our work focuses on the technique that [Bag et al., 2018, Bag et al., 2019] followed, which is using the Largest Cluster Statistics (LCS). [Bag et al., 2018] focused on the largest ionized region only for the fiducial model of reionization and along with LCS, they complemented their analysis with Minkowski functionals (MF) using a code called SURFGEN2.

As an add-on to this work [Pathak et al., 2022] did the same analysis but on different reionization scenarios as mentioned in [Majumdar et al., 2016]. In brief, they have shown that the LCS when applied to different scenarios can help to distinguish between the two types of reionization process, the “Inside-Out” and the “Outside-In” scenario. They have also concluded that at percolation, the global neutral fraction is significantly higher for the case of inside-out scenario when compared with the outside-in case. [Pathak et al., 2022] also did a complementary analysis to quantify the evolving topology and morphology of the ionized region for different reionization environments using the ratios of Minkowski functionals called Shapefinders (SFs), about which, we will discuss as well. It is seen that for non-uniform reionization processes (e.g. Clumping model) happening in the inter-galactic medium (IGM), the process of reionization may be discerned by the non-uniformity of the recombination process. It is also

observed that LCS has an abrupt growth with an SF called the ‘length’ keeping others almost steady indicating a highly filamentary structure in the hydrogen distribution on and after percolation. Finally, it was concluded that by doing different topological analysis, the reionization processes can be understood in a better manner.

## 4.2 Discussion on the results of Pathak et al. 2022

The work that is done in [Pathak et al., 2022] is discussed in this section. It is to be noted that the results of [Pathak et al., 2022] have been reproduced and similar results have been found which will also be discussed here.

### 4.2.1 Considerations for identification of ionized regions

Pathak et al. considers the regions that have neutral hydrogen density  $\rho_{HI} > 0$  are completely neutral. It is to be kept in mind that partially ionized regions are considered neutral as well. Our work on the other hand moves to a more realistic scenario where the images that are analysed are converted in a manner such that they mimic the image that will be received by low-frequency radio interferometers [SNELL, 2021]. The images received by the interferometers will have its  $0^{th}$  K-mode in the Fourier plane to be missing; which implies the signal received from the interferometers will have their mean subtracted. So the consideration for identification of ionized regions will be slightly modified and is given as follows:

$$\rho'_{HI}(x) = (\rho_{HI}(x) - \bar{\rho}_{HI}) > 0 \implies \text{Neutral region}$$

$$\rho'_{HI}(x) \leq 0 \implies \text{Ionized region}$$

However, when run in SURFGEN2, a very small value of the density threshold has been considered. It is done because 0 value will result in a singularity in the SURFGEN2 code while doing the mapping process. In [Pathak et al., 2022], the value is taken to be 0.01 wherein the value considered in our case is 0.1 about which we will discuss in later subsections. When considering the instrumental effects on these images, the choice of threshold becomes non-trivial. For these cases we employ the method of Gradient descent which we have described elaborately in Section 6.2.

## 4.2.2 Analysis methods and results

It is already mentioned that the different reionization scenarios despite having different evolution pattern, they are forced to follow the same reionization history. This eases the understanding of the topological evolution and its relation to the neutral fraction for different reionization scenarios.

It is important to note that though the different scenarios follow the same reionization history, as the mass averaged neutral fraction ( $\bar{x}_{HI}$ ) is same for any given redshift, the total volume of ionized regions are different for different scenarios. This implies that the Filling Factor (FF) is different for different ( $\bar{x}_{HI}$ ) values. This is shown in 4.1 and 4.2.

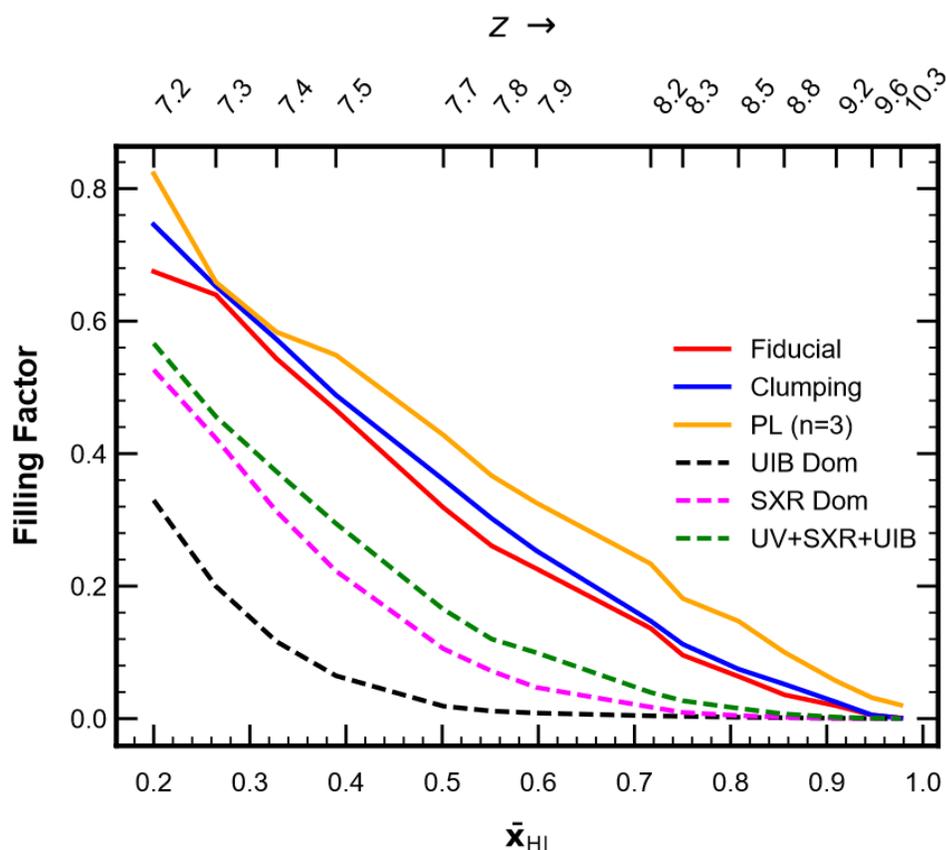


Figure 4.1: Variation of FF with  $\bar{x}_{HI}$  as seen in [Pathak et al., 2022].

The UIB dominated scenario provides the lowest FF for a given  $\bar{x}_{HI}$  or redshift ( $z$ ), whereas  $PL = 3$  produces the largest FF. As hard X-ray photons may escape to longer distances in the UIB dominated scenario and

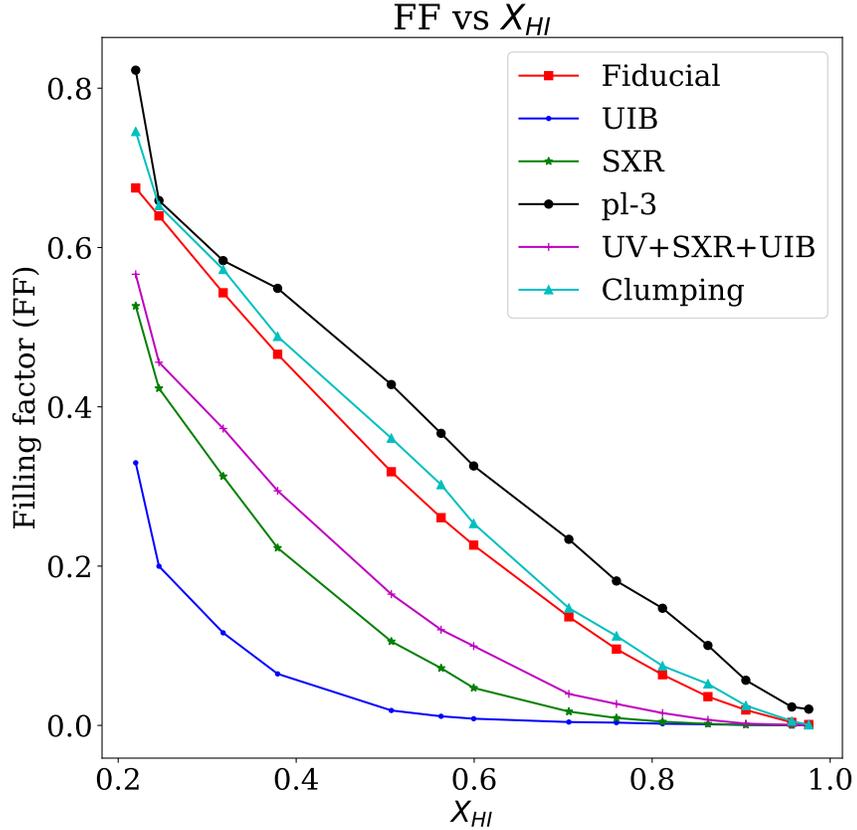


Figure 4.2: Same plot reproduced as a part of this work.

effectively provide a uniform ionising background, this results in a ‘outside-in’ reionization, in which the low density areas ionize first and the high density regions follow afterwards. As a result, in contrast to PL or Fiducial situations, a significant volume will be partly ionized in this scenario. The reionization will be ‘inside-out’ in nature under the Fiducial, Clumping, or PL scenarios, with UV radiations from collapsing halos escaping to further distances only after they have considerably ionized their local IGM. The UV+SXR+UIB and the SXR are in the middle of the ‘inside-out’ and ‘outside-in’ scenarios.

### 4.2.3 Analysis of LCS with percolation analysis

When the first sources formed at CD, they began to ionize the neutral hydrogen field around them. As reionization occurs, the size and amount

of these little ionized hydrogen areas grows, and they begin to overlap. However, depending on the reionization scenario, these bubbles will join at some point in the future to form a big linked single ionized zone. We term this rapid topological shift in the ionisation field as percolation transition [Klypin and Shandarin, 1993, Yess et al., 1997].

The percolation transition occurs when the greatest ionized zone spreads from one face of the simulation box to the opposite face. Due to the periodic boundary condition of our simulation volume, the greatest ionized zone is thus indefinitely expanded. The greatest ionized region, as defined statistically using LCS, aids in the study of ionized region expansion and the understanding of the percolation process. LCS is projected to grow dramatically with filling factor (FF) or  $\bar{x}_{\text{HI}}$  during percolation. However, when and how percolation occurs in an ionized hydrogen field is determined by the reionization process. In figure 4.3 and 4.4, the evolution of LCS with FF and neutral fraction is shown, which can be used to identify the type of source models used in different reionization environments.

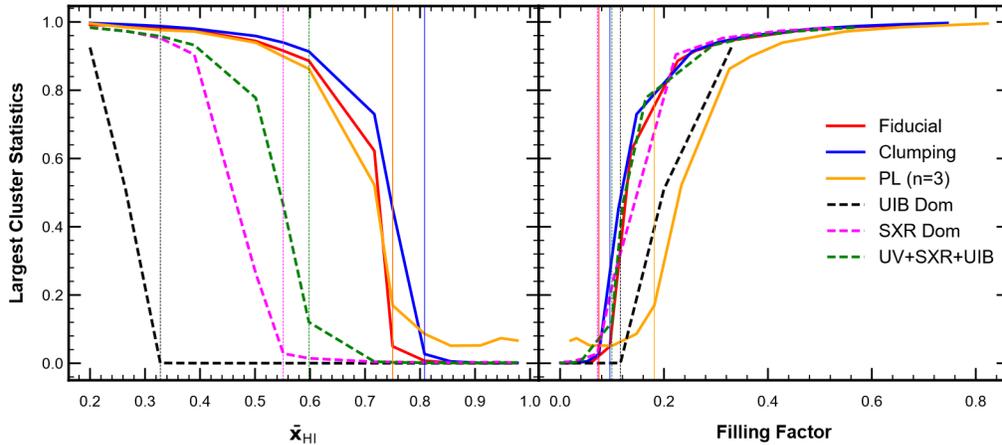


Figure 4.3: Variation of LCS with FF and  $\bar{x}_{\text{HI}}$  as seen in [Pathak et al., 2022].

When the largest ionized region abruptly develops in all directions and extends throughout the simulation volume, the vertical dashed lines depict the commencement of percolation. A little change in  $\bar{x}_{\text{HI}}$  or FF during the percolation transition causes a dramatic spike in LCS, as seen in the relevant panels. This formally discontinuous growth in LCS may be

used to characterise the percolation transition [Klypin and Shandarin, 1993]. Percolation transition occurs for the fiducial model at  $\bar{x}_{\text{HI}} \approx 0.75$  and  $FF \approx 9.6\%$  corresponding to a redshift value  $z \approx 8.1$ , which is in line with previous findings in the literature [Iliev et al., 2006, Chardin et al., 2012, Furlanetto and Oh, 2016, Bag et al., 2018] where the considered models were a little different. After percolation, a big, single ionized region is formed by the merger of the smaller ionized regions and LCS value saturates to unity. Similar results were discovered in the case of other reionization scenarios. A lot of information about the source models can be extracted from these curves. For example-

- (i) The value of the neutral fractions at percolation for each reionization scenario are different indicating a connection between percolation with the topological properties of the evolving IGM.
- (ii) The shape of the LCS curve is different from one another for different reionization scenarios. It can be inferred that the PL3 model, the large ionized regions are connected easily and hence percolates at an earlier stage of reionization. On the contrary, for the UIB model, transition occurs at a late stage as ionized regions are smaller and the FF value is significantly lower. In gist, for inside-out models, percolation occurs at higher neutral fraction wherein for outside-in models, it occurs at a lower neutral fraction value.
- (iii) For the PL3 model, the starting value of LCS is relatively higher indicating that this scenario has larger ionized regions in the beginning itself.
- (iv) Clumping model has the earliest percolation stage as it is the only non-uniform recombination scenario.

The reproduced result of LCS vs FF in the case of UIB model vary from the [Pathak et al., 2022] result due to a slightly different consideration in threshold. With very tiny deviation in threshold, the UIB model is varied significantly as in all the cases partially ionized regions are considered to be neutral as well, so for a uniform ionizing case, many number of such regions will occur.

#### 4.2.4 Shapefinders

Shapefinders is a complementary analysis done along with the percolation method to study the geometry of the ionized regions. This analysis is done on individual ionized regions. Shapefinders are essentially ratios of the Minkowski functionals (MFs) [Sahni et al., 1998] which are explained in section 2.3. In 3-d the three Shapefinders are as follows:

1. Thickness:  $T = \frac{3V}{A}$ ,
2. Breadth:  $B = \frac{S}{H}$ ,
3. Length:  $L = \frac{H}{4\pi}$ .

The Shapefinders have the dimension of length and can be used extensively to measure the extension of an object. It is worth mentioning that the Shapefinders are defined in such a way that they are spherically normalized which means the first Minkowski functional  $V = \frac{4\pi}{3}T.B.L$ .

The ratio of the Shapefinders, further, yield two more quantities that help in understanding the morphology of the ionized regions or any object, in general. They are defined as follows:

1. Planarity :  $P = \frac{B-T}{B+T}$ ,
2. Filamentarity:  $F = \frac{L-B}{L+B}$ .

P and F denote the 'planarity' and 'filamentarity' of an object, respectively, and  $0 \leq P, F \leq 1$ , by definition. To understand in a more intuitive way, a sphere has  $P \simeq F \simeq 0$ , but a ribbon has  $P \sim F \sim 1$ . A planar object for example a plane sheet, has  $P \gg F$ , whereas a filament is the opposite i.e  $F \gg P$ . Studying 'Shapefinders' of ionized volumes at various phases of reionization, in conjunction with percolation analyses, would give information on the evolution of the ionization field's geometry, morphology, and topology.

#### 4.2.5 SURFGEN2 code and Marching Cube 33 algorithm

To find different topological and morphological parameters like the LCS, Shapefinders etc. the SURFGEN2 algorithm is used. It is an

advanced version of an algorithm named SURFGEN which was developed by [Sheth et al., 2003, Sheth, 2005] to study large scale structure of the universe. The SURFGEN2 code has 3 main parts which are as follows:

1. The first component uses a 'Friends-of-Friends' (FoF) algorithm combined with periodic boundary conditions to identify all clusters (overdense or underdense segments) within the simulation box. When fragmented components of the same cluster are found at the box border (this occurs rarely), these parts are adjusted to reassemble the cluster.
2. The second section of the algorithm triangulates each cluster's surface and records the vertices of the triangles. [Sheth et al., 2003] and [Sheth, 2005] provide detailed descriptions of the triangulation approach, which employs the 'Marching Cube' algorithm [Chernyaev, 1995]. Instead, [Bag et al., 2018] employ the 'Marching Cube 33' [Lorensen and Cline, 1987] enhanced triangulation scheme, which avoids the problems that plagued the original Marching Cube algorithm. The Marching cube is a 3D version of the Marching Square algorithm which determines the implicit functions of geometrical shapes to identify boundaries between them.
3. The last section of the algorithm uses the stored triangle vertices to identify the MFs and Shapefinders for each cluster (ionized region). In the next part, we'll go over the algorithms for calculating the MFs.

#### 4.2.6 Determination of Shapefinders

After using the Marching Cube 33 algorithm to triangulate an isodensity surface, the MFs and Shapefinders can be determined using the following methods [Sheth and Tormen, 1999]:

- (i) The isodensity surface's surface area (A) may be calculated by adding all of the triangles that make up the surface,

$$A = \sum_{i=1}^{N_T} A_i, \quad (4.1)$$

Where  $N_T$  is the number of total triangles and  $A_i$  is the area of the  $i^{th}$  triangle

- (ii) The below equation gives the volume (V) encompassed by the isodensity surface.

$$V = \sum_{i=1}^{N_T} V_i, V_i = \frac{1}{3} A_i(\hat{n}_i \cdot P_i), \quad (4.2)$$

where  $V_i$  is the volume of the  $i^{th}$  tetrahedron, the base of which is the  $i^{th}$  triangle, and the apex of which is (an arbitrarily selected) origin. The centroid of the  $i^{th}$  triangle with normal in the  $\hat{n}_i$  direction is represented by  $P_i$ .

- (iii) Integrated Mean Curvature (IMC) is given by -

$$H = \sum_{i,j} \frac{1}{2} \epsilon l_{ij} \phi_{ij}. \quad (4.3)$$

Here, The length of the common edge and the angle between the normals of the neighbouring pair of triangles i,j are, respectively,  $l_{ij}$  and  $\phi_{ij}$ . The total is calculated over all neighbouring triangle pairs. When the triangular pair (i,j) is a part of a convex surface locally, it takes the value +1, and when the surface is concave, it takes the value -1.

- (iv) The Genus (G) and the Euler characteristic ( $\chi$ ) for the triangulated surface is given by,

$$\chi = N_T - N_E + N_V, G = 1 - \frac{\chi}{2}. \quad (4.4)$$

The number of triangles, vertices, and edges are represented as  $N_T, N_V, N_E$ , respectively. The Euler characteristic obtained using this approach is always precise since the above equation is essentially a representation of Euler's polyhedral formula. Each edge of a closed surface triangulated using the Marching Cube 33 method is always shared by two adjacent triangles. As a result, The total number number of edges, and total number of triangles, that make up a closed surface have the relationship  $N_E = \frac{3}{2} N_T$ . This connection can be used to evaluate if a triangulation technique is topologically accurate.

Because SURFGEN2 triangulates the surface of individual clusters, its accuracy is excellent, as demonstrated by [Sheth et al., 2003]Sheth et al.

(2003), and far superior to existing methods of estimating the MFs [Schmalzing and Buchert, 1997], such as using the Koenderink invariant [Koenderink, 1984] or Crofton’s formula [Crofton, 1868].

To summarize, these are the key steps of the analysis using SURFGEN2

1. Using the FoF algorithm, the code identifies all the isolated ionized spots in the simulated field according to the thresholding condition. It’s worth noting that SURFGEN2 only discovers areas that are compatible with periodic boundary conditions.
2. The code then uses the ‘Marching Cube 33’ triangulation technique [Lorensen and Cline, 1987] to simulate the surface of each ionized zone, which avoids the problems that the Marching Cube algorithm has [Chernyaev, 1995].
3. Finally, using the triangle vertices from the preceding stage, SURFGEN2 finds the Minkowski functionals and Shapefinders of each ionized volume independently.
4. The above steps are followed for all reionization scenarios for all the redshifts to find the topological and morphological features of the ionized regions.
5. The largest ionized region is tracked to calculate the development in LCS for each reionization environment.

#### **4.2.7 Variation in number of ionized regions ( $N_C$ ) with different parameters**

For various reionization situations, the number of ionized regions,  $N_C$  is displayed versus the neutral fraction  $\bar{x}_{\text{HI}}$  and the neutral fraction FF, respectively in 4.5 and 4.6. The neutral fraction drops from unity as reionization advances, but the filling factor rises.  $N_C$  rises with time as new ionized patches form at the commencement of reionization. On the other hand, when reionization progresses, ionized zones begin to overlap and the number of ionized regions decreases. As a result, in all reionization models,  $N_C$  has a maximum somewhere in the middle.

Furthermore, the way in which  $N_C$  achieves its maximum varies depending on the reionization scenario. The UIB model, for example, has the most ionized volumes, whereas  $N_C$  in the PL3 model has the fewest.  $N_C$  maxima occur at various  $\bar{x}_{\text{HI}}$  levels for different models, for example, early phases for PL3, Fiducial, and Clumping models, and late phases for the UIB dominated model. In contrast, the  $N_C$  maxima for various models are all found around similar  $FF \sim 10\%$  values. It is worth mentioning that for UIB model, the reproduced result is slightly different due to different threshold consideration.

## 4.2.8 Results obtained from Shapefinders

The topological and morphological evolution of the largest ionized region is calculated using the Shapefinders as mentioned in section 4.2.3 The results that were obtained are discussed below.

### 4.2.8.1 Genus

The variation in the Genus is observed with respect to neutral fraction for different reionization scenarios in 4.7 and 4.8 which are plotted till percolation occurs. The physical geometry of a percolating region cannot be described due to the periodic boundary constraint. As a result, the genus value of an area immediately following percolation is likewise poorly defined. However, when the ionized zone occupies the majority of the simulation volume, the genus value may be safely estimated in terms of per unit volume. It is to be noted that the x-axis is shifted by the critical neutral fraction value (The value of neutral fraction at which percolation takes place).

The greatest ionized patches in all of the situations grow more multiply linked as reionization occurs, and their genus rises. Despite the fact that ionized hydrogen percolates earlier in time in the clumping model, the genus of the greatest ionized patch in the clumping model is substantially larger by several orders than in the other situations. Due to strong non-uniform recombination in the clumping scenario, large pockets of neutral hydrogen persist, tunnelling through the ionized volumes, resulting in high genus values. Because the majority of these tunnels are negatively curved, the

overall integrated mean curvature is reduced. This breaks the natural order of the Shapefinders i.e  $T < B < L$  and yields high planarity value which might be inferred as a non-physical one. That is why Clumping is not included in the Shapefinder results.

#### 4.2.8.2 Planarity and Filamentarity

The variation in Planarity (P) and Filamentarity (F) of the largest ionized region for different reionization environment barring the clumping model is shown in 4.9 and 4.10. The filamentarity for all the models seem to increase with decreasing neutral fraction and reaches to 1 as percolation is achieved. On the contrary, the planarity is always less for all models. Hence, it can be concluded that the largest ionized regions become highly filamentary near the percolation which is consistent with the work seen in [Bag et al., 2018].

#### 4.2.8.3 T, B and L

As it is seen that the structure of the ionized region for all the models (excluding clumping) become filamentary at percolation, we consider it to be a filamentary object. In that case,  $T \times B$  is called the effective “cross section” for that filament.

In 4.11 and 4.12 It can be observed that length (L) has a sharp rise during percolation transition wherein the cross section ( $T \times B$ ) remain stable throughout. This indicates that the lengths are in general way higher in magnitude than the cross sections. Cross section has the highest value for PL3 model and lowest for the UIB model. It can be, thus, inferred that as the ionized regions for UIB model are smaller and uniform than PL3 or Fiducial model, the filamentarity of the inter-connected regions also appear smaller. As PL3 starts with a high value of LCS, the cross section also has a higher value.

### 4.3 Analysis on the mean subtracted field

The entire analysis as done in [Pathak et al., 2022] have to be done for the mean subtracted fields to mimic interferometric data. Zero-length baselines or visibilities with  $k = 0$  wave number are not present in radio interferometric

observations. As a result, the images produced by radio interferometric observations will lack an absolute flux calibration. This eventually results in an image where the mean has been subtracted..

### 4.3.1 Issues in analysing mean subtracted fields

Initially, when SURFGEN2 was put on run for the mean subtracted fields of the Fiducial model for different redshifts, they were consuming excessively high RAM resource  $\approx 190-200$  GB (In the non mean subtracted data, usually 7-10GB is consumed) which led in multiple segmentation faults. In these cases, the threshold values for each redshift was set to be the negative of the mean value for the corresponding non mean subtracted file as the  $\rho(x) = 0$  points will be mapped to the negative threshold values. It was observed that a particular part of the code was not able to handle the negative threshold values, so it was modified in a manner so that the mapping can be done correctly.

To analyse the issue, the variation in the threshold is checked for different parameters which are in Figure 4.13 and Figure 4.14.

As we go to lower neutral fraction, more number of ionized regions appear which implies that the threshold to identify neutral region is shifted lower which is consistent with our theoretical understanding of reionization. It is also consistent with the filling factor plot where as more and more regions get ionized, the threshold starts decreasing which is consistent with the plot of neutral fraction against the threshold. In the Figure 4.15 the variation of threshold is checked with Filling factor with changing  $\bar{x}_{\text{HI}}$  values. The vertical dotted line represents the mean threshold value which is 50 for our case. It is interesting to observe that for lower neutral fraction values, the threshold very quickly saturates as ionized regions are way too dominant for these cases.

Finally, we see the variation in threshold for the mean subtracted case for the Fiducial model with changing  $\bar{x}_{\text{HI}}$  values in Figure 4.16. In these cases, the threshold follows the same pattern as of in the case of non mean subtracted field but with negative values in threshold. It is important to note that, a constant value of threshold which is the mean of the individual thresholds of the mean subtracted fields are considered in this case. For a

better comparison, the value of the starting threshold has been shifted to the origin and similar results are found for this case as well.

### 4.3.2 Summary of this work

The below table ( 4.1 ) summarizes the results of the reproduced work of [Pathak et al., 2022] and the results are consistent with the Table 2 of [Pathak et al., 2022] which has the same set of parameters.

Reionization Scenarios	Critical $\bar{x}_{\text{HI}}$	$FF_C$	Planarity	Filamentarity
Fiducial	0.75	0.095	0.162	0.998
Clumping	0.81	0.063	0.837	0.922
UIB dominated	0.31	0.543	0.201	0.999
SXR dominated	0.50	0.318	0.126	0.984
UV+SXR+UIB	0.60	0.226	0.147	0.995
PL 3.0	0.75	0.095	0.872	0.997

Table 4.1: Summary of the reproduced work of [Pathak et al., 2022]. Critical values indicate the values where percolation transition has occurred.

The work done by [Pathak et al., 2022] was thoroughly verified in this work and it had yielded same results which confirms the analysis to be consistent with previously done work in literature.

The mean subtracted signal models were fed to SURFGEN2 to extract statistical parameters like LCS and topological parameters like the Shapefinders. But, the program was consuming excessive RAM and time. It was speculated that there might be a bug in the code or the problem had something to do with the machines that we were running our codes on. As it turned out, SURFGEN2 was initially unable to process negative threshold values as inputs to the mean subtracted hydrogen maps, hence, the mapping was changed. Alongside, a detailed analysis on the variation of threshold with changing filling factor was done for different redshifts or neutral fraction. This analysis was done only on the Fiducial model. After a second run, it was observed that SURFGEN2 was throwing garbage values, which, when analysed, pointed out another issue. After changing the optimization of the compiler from O3 to O2 it was put on run again, where some expected results were yielded. We then re-sample the brightness

temperature maps on a smaller grid within the SKA phase-I resolution limit to reduce the run-time of SURFGEN2 as more fixes on the code were being employed alongside.

### 4.3.3 Prospects

The points mentioned below are the prospective for this project:

1. A complete debugging of SURFGEN2 is required; to run the code in older kernel versions, optimization levels are reduced which has increased the runtime for the code. So, more bug fixes are required to reduce the time and space complexity of the algorithm.
2. An optimum threshold independent of ionization state of IGM has to be found. Our analysis on the threshold can be considered to be the first step to this.
3. The simulated brightness temperature maps of hydrogen can be resampled on a smaller grid to reduce the run-time of the codes. This degradation of resolution is expected to be the same as the resolution of the upcoming SKA phase-I observation data.
4. A testing is required to verify if varying threshold is better than having a constant threshold as we have considered in this work.
5. After a complete debugging of the SURFGEN2 algorithm, LCS is to be applied on the mean subtracted 21-cm hydrogen field.
6. Repeating the above analysis in the presence of SKA like noise in the signal is a necessary step to move towards a more realistic scenario. In this case the threshold will have random variation which is also required to be accounted for and analysed with different statistical methods. It will be challenging to distinguish between the ionized regions and the low density regions once noise is included in the signal space.
7. Along with noise different foreground models can also be included to have a better and practical analysis on the density fields.

8. Finally, the array synthesized beam effects and the instrumental effects on the signal have to be accounted for as well to have a complete realistic understanding of the evolving topology and morphology of the density field.

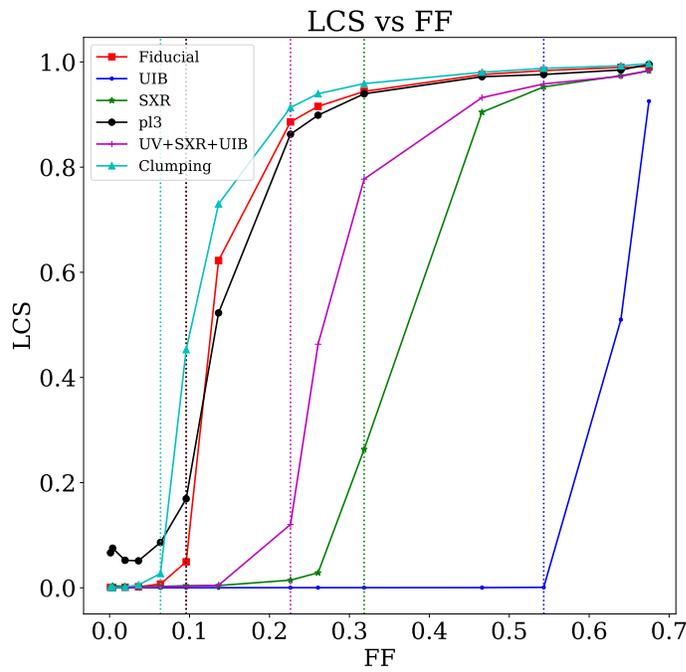
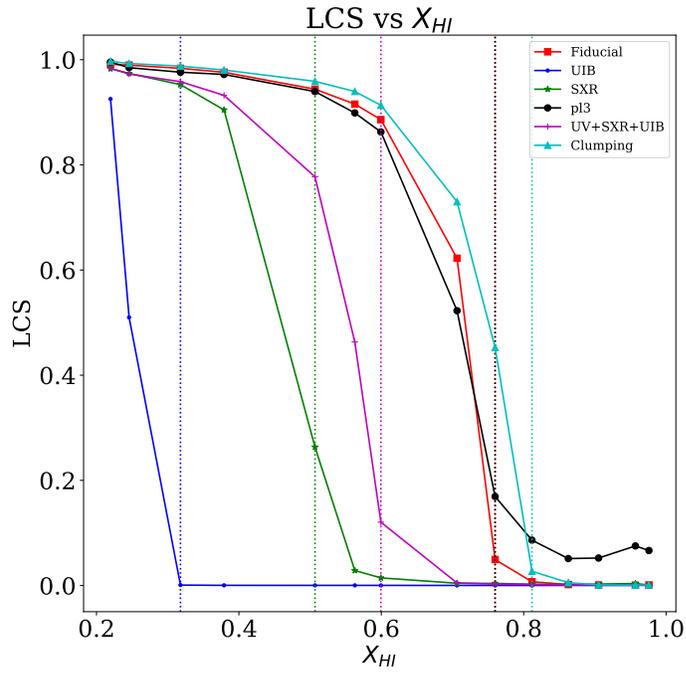


Figure 4.4: LCS vs FF and  $\bar{x}_{HI}$  as reproduced in this work.

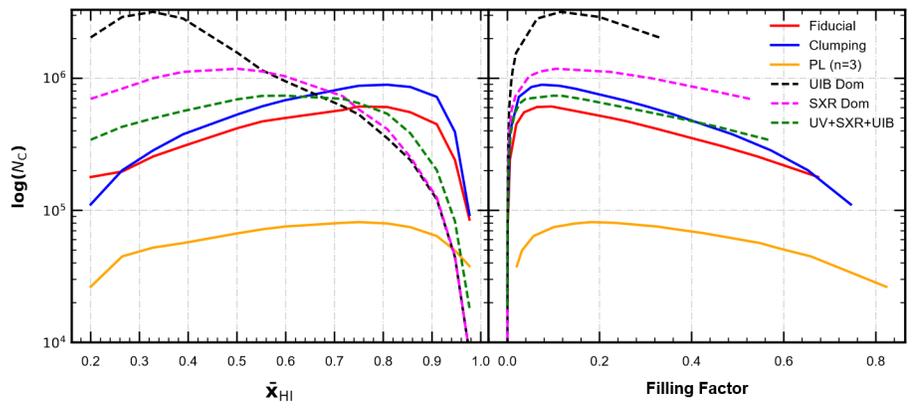


Figure 4.5: Variation of NC with FF and  $\bar{x}_{\text{HI}}$  as seen in [Pathak et al., 2022].

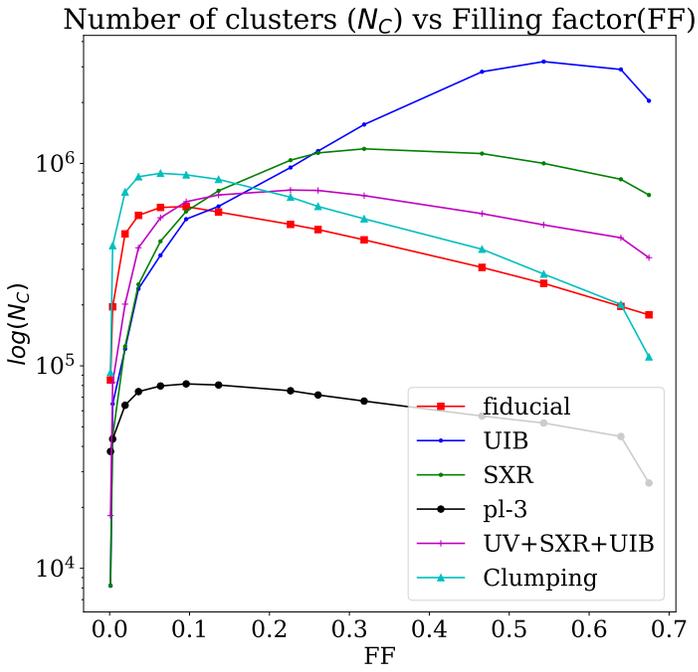
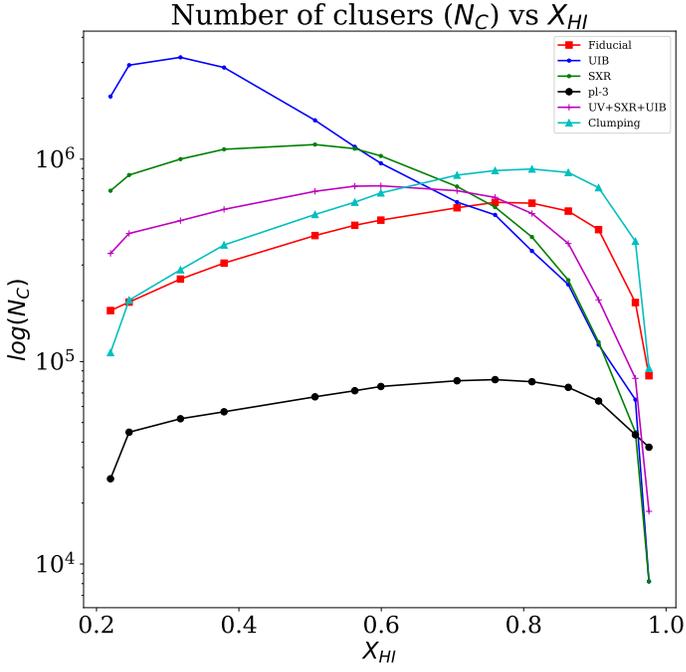


Figure 4.6:  $N_C$  vs FF and  $\bar{x}_{HI}$  as reproduced in this work.

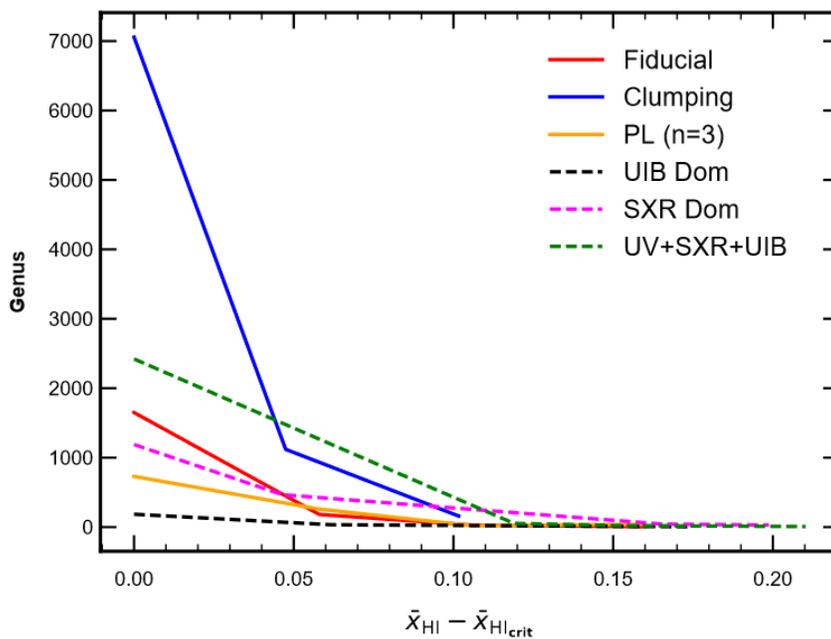


Figure 4.7: Variation of genus with  $\bar{x}_{\text{HI}}$  as seen in [Pathak et al., 2022].

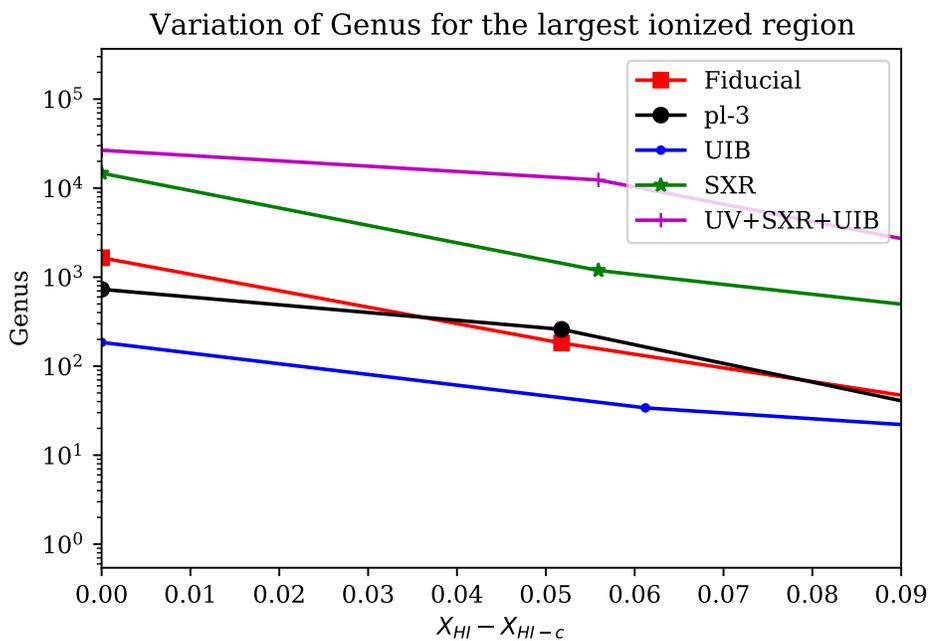


Figure 4.8: Variation of Genus as reproduced in this work.

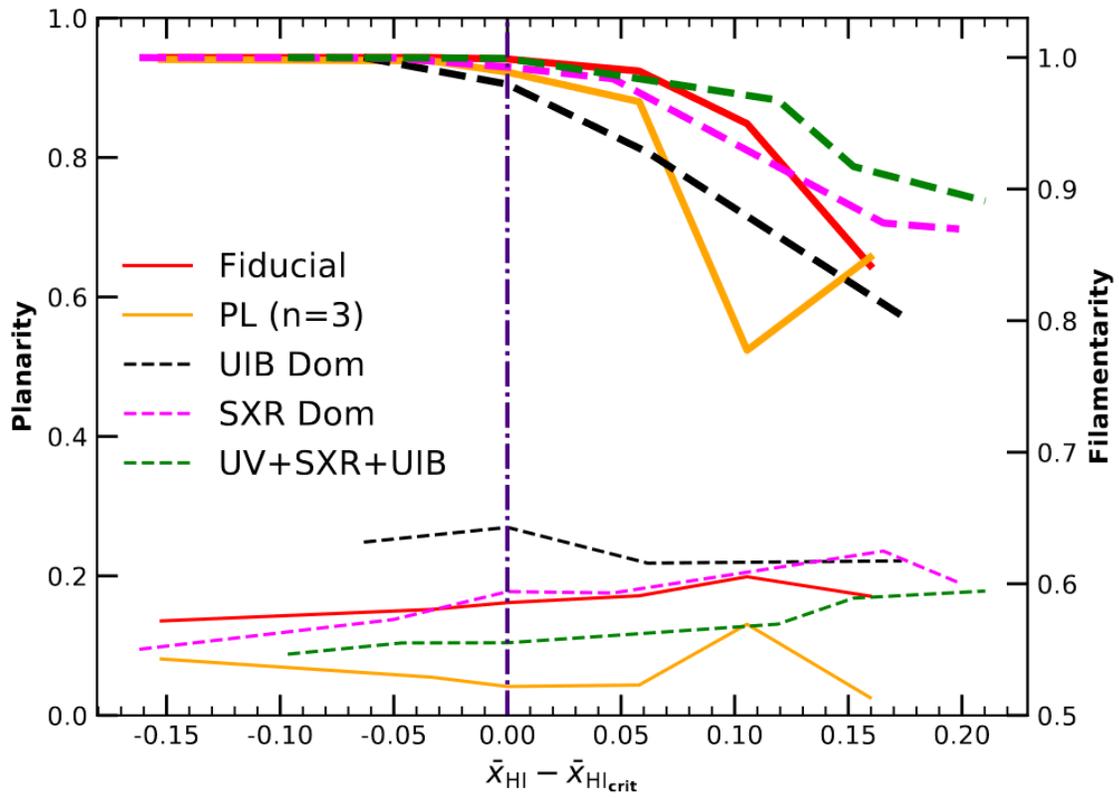


Figure 4.9: Variation of Planarity and Filamentarity with  $\bar{x}_{\text{HI}}$  as seen in [Pathak et al., 2022].

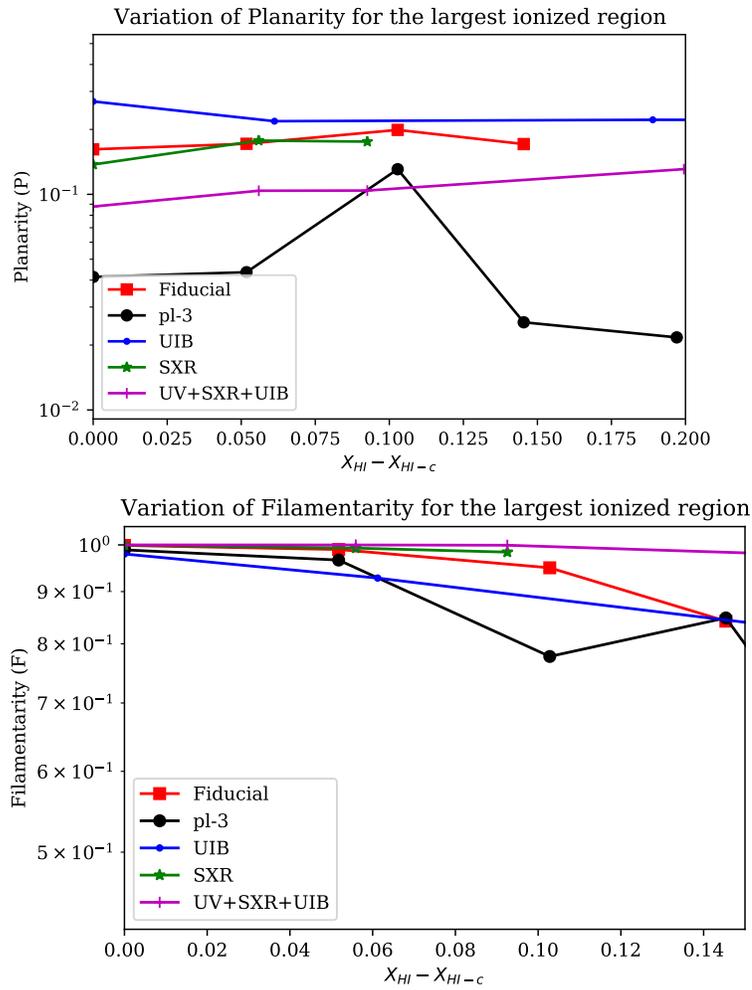


Figure 4.10: Planarity and Filamentarity vs  $\bar{x}_{HI}$  as reproduced in this work.

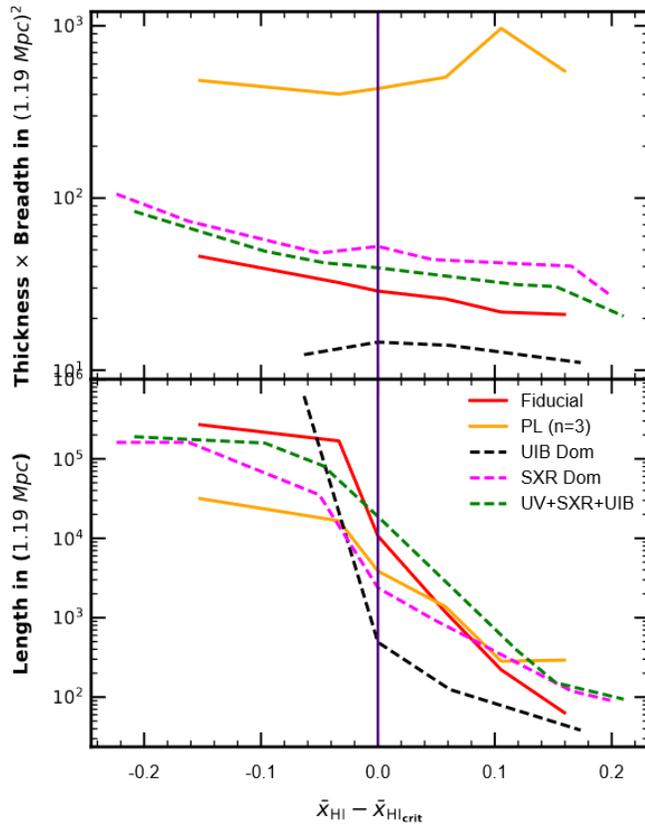


Figure 4.11: Variation of Length and  $T \times B$  with  $\bar{x}_{\text{HI}}$  as seen in [Pathak et al., 2022].

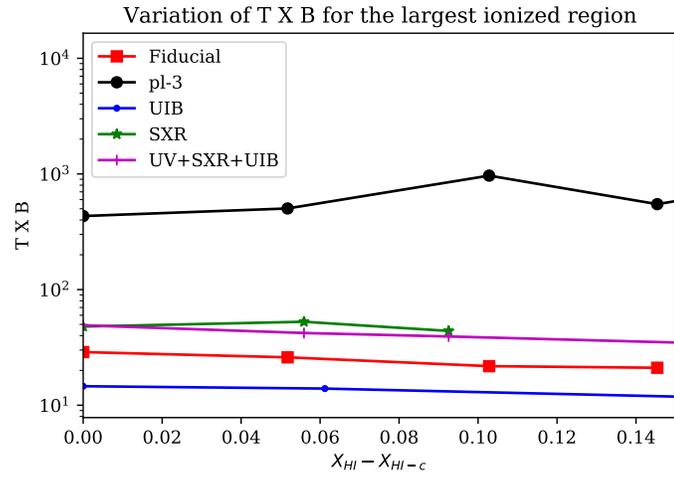
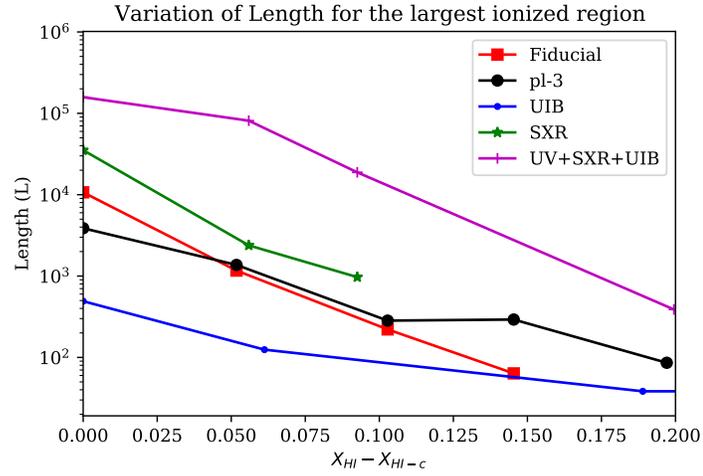


Figure 4.12: Variation of Length and  $T \times B$  with  $\bar{x}_{HI}$  as reproduced in this work.

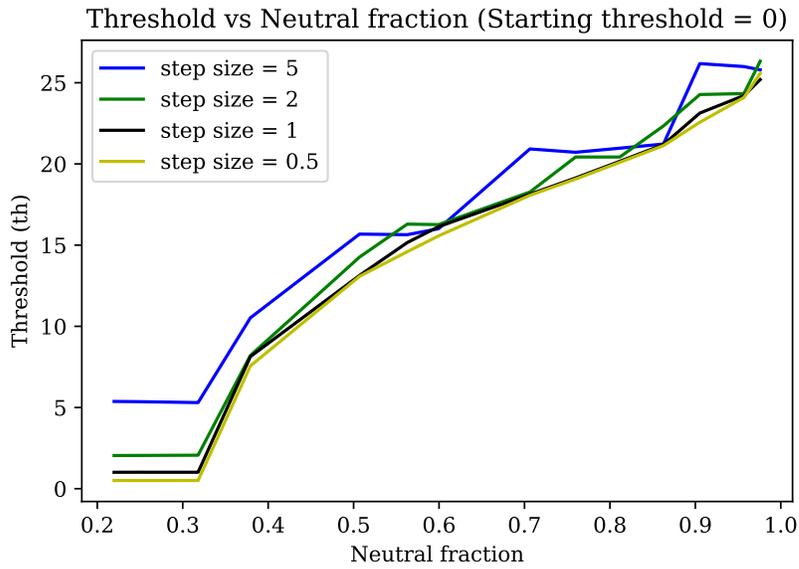


Figure 4.13: Threshold for non mean subtracted hydrogen map is plotted against changing neutral fraction ( $x_{HI}$ ) for different step sizes for Fiducial model.

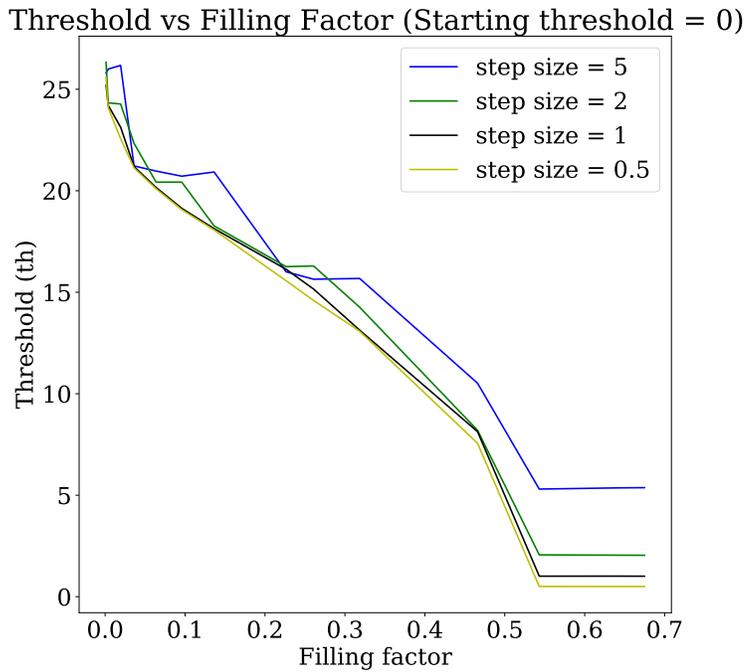


Figure 4.14: Threshold for non mean subtracted hydrogen map is plotted against changing filling factor for different step sizes for Fiducial model.

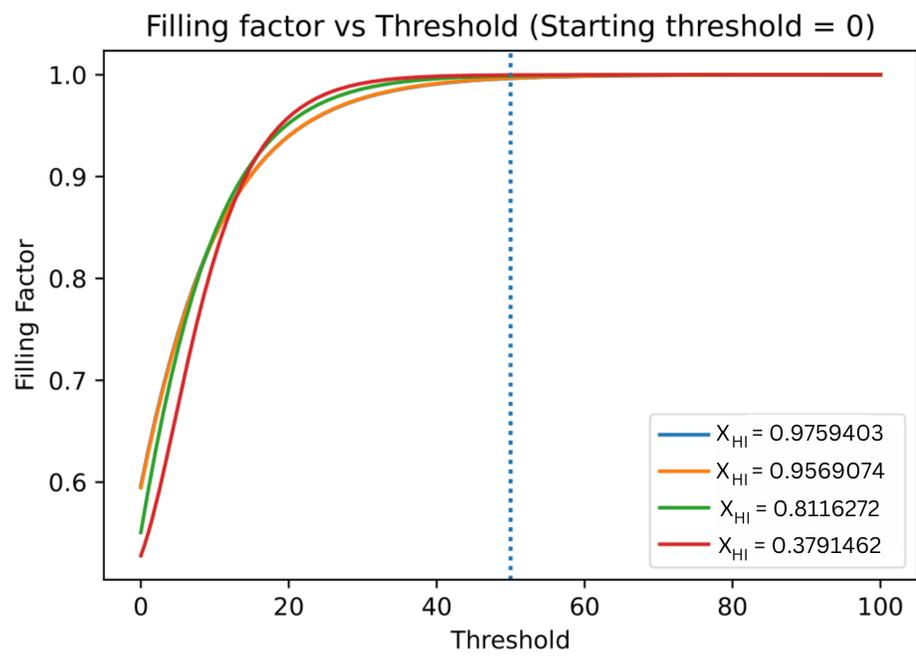


Figure 4.15: Threshold for non mean subtracted hydrogen map is plotted against changing filling factor for different step sizes for Fiducial model.

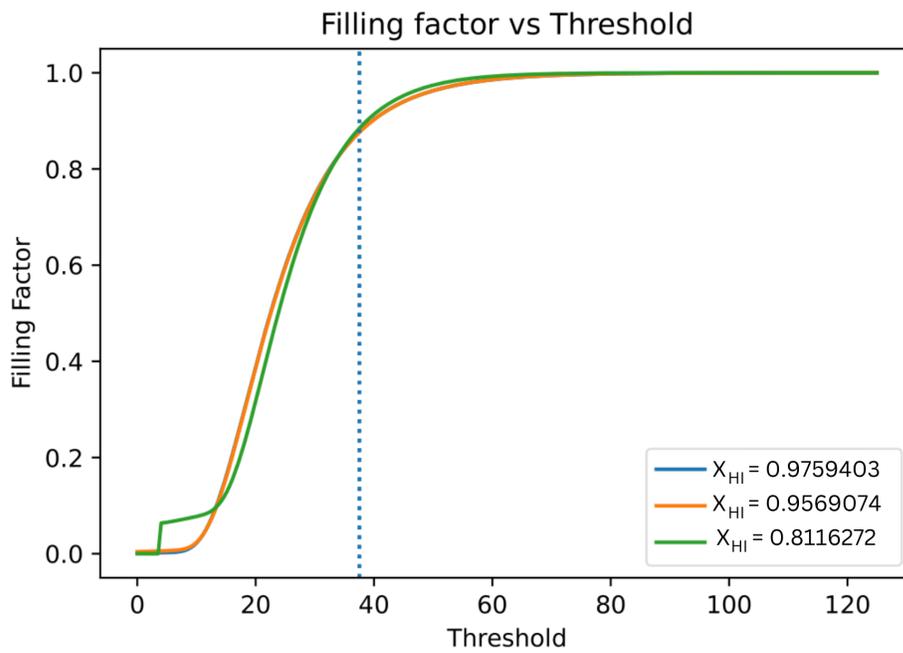


Figure 4.16: Threshold for mean subtracted hydrogen map is plotted against changing filling factor for different step sizes for Fiducial model.

# Chapter 5

## Analysis of LCS on sampled maps of lower resolution

The simulated maps that were initially used for this analysis had a resolution of  $1.19 \text{ cMpc}^3$  with  $600^3$  grid cells each side which is not at all realistic and far better than what SKA-1 Low is expected to capture. For this reason, the simulated maps of grid size  $600^3$  grid units were down-sampled to a coarser resolution of grid size  $300^3$  grid units, keeping the simulation volume same (i.e.  $714 \text{ cMpc}^3$ ). From the original 21-cm maps we have randomly pooled 8 neighboring cells and chose one random cell out of them and replaced it in place of the old eight cells. This way, our maps becomes twice ( $2.38 \text{ cMpc}$ ) as coarse as our original maps of  $600^3$  grid units which had a resolution of  $1.19 \text{ cMpc}$ . This type of sampling ensured that the new box that was generated follows the same reionization history. This was done to check the robustness of the geometrical estimations that are used for our particular analysis. This is done for three different models namely Fiducial, PL ( $n=3$ ) and SXR Dominated scenario. The analysis of LCS along with the Shapefinders was done and the results are shown in Figure 5.1 and Figure 5.2.

In Figure 5.1 we see that the original (solid/dashed curves) and low resolution (dotted curves) show similar behaviour. The left panel shows the variation of LCS with changing  $\bar{x}_{\text{HI}}$  and the right panel shows the corresponding changes in LCS with changing Filling Factor. For the three

different reionization scenarios, the results for the sampled maps follow the LCS curves as of the original simulated maps having  $600^3$  grid units. It is important to notice that the percolation transition remains unchanged for each of the reionization scenarios and the slight deviation that occurs in the low resolution maps occur due to the sparsely sampled discrete values of  $\bar{x}_{\text{HI}}$ . The difference would go away if we have enough number of samples in the percolation transition region. Alongside, it is to be noticed that PL ( $n=3$ ) is minimally affected by the sampling as it has the largest “Largest Ionized Region (LIR)” with the smallest value of Genus among all the three scenarios in consideration. It can also be concluded that the size of the LIR is unaffected by this kind of sampling.

In Figure 5.2 the cross-section ( $T \times B$ ) seems to be stable during percolation transition which is consistent with the previous findings of [Pathak et al., 2022].

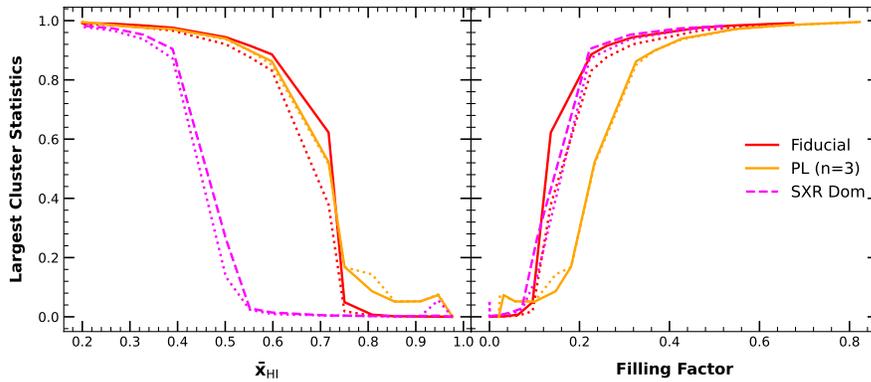


Figure 5.1: For three reionization source models, the LCS results from our original resolution maps (solid/dashed curves) are compared with those from the low-resolution 21-cm maps (dotted curves). The critical  $\bar{x}_{\text{HI}}$  at the start of percolation transitions in all the models is unchanged, and the LCS results are consistent for the different resolutions.

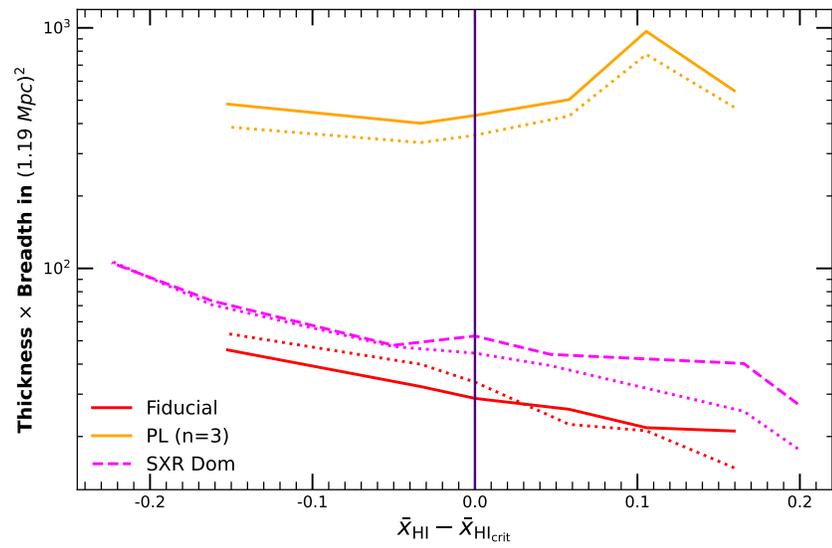


Figure 5.2: In three reionization situations, the growth of the largest ionized region’s “cross-section” has been analysed for low-resolution (dotted curves) and the original resolution fields (solid/dashed curves). Again, the Shapefinders results in the two resolutions are consistent.

## Chapter 6

# Testing LCS on realistic H<sub>I</sub> 21-cm maps

The analysis on threshold has shown that variable thresholds have to be chosen rather than choosing a constant threshold for the mean subtracted maps. However, in realistic scenarios, when telescope noise, foreground emissions, the effect of the array synthesized beam, effect of ionosphere and a lot of other factors are considered, the choice of threshold becomes even more complicated. In this particular analysis, a mean subtracted version of the lower resolution (2.38 cMpc) was chosen and on top of it different realistic effects like the effect of the array synthesized beam and noise were applied. Subsequently, a proper threshold was determined and thorough LCS analysis was done. All of these effects were added to the down-sampled, mean-subtracted maps to mimic the behaviour of the upcoming Square Kilometer Array (SKA) in its low frequency regime. A visual representation of these effects can be seen in Figure 6.1 which is the first slice from one of the Fiducial image cubes that has a neutral fraction  $\bar{x}_{\text{HI}} = 0.75$ . It can be seen that after addition of noise of RMS 3.10 mK, random variation in the pixels occur. A Gaussian smoothing filter that has a smoothing length of 7.14 cMpc was further added. This smoothing essentially works as a denoising process but a majority fraction of the pixels shifted to some certain intensity values that might be considered as “partially ionized regions”.

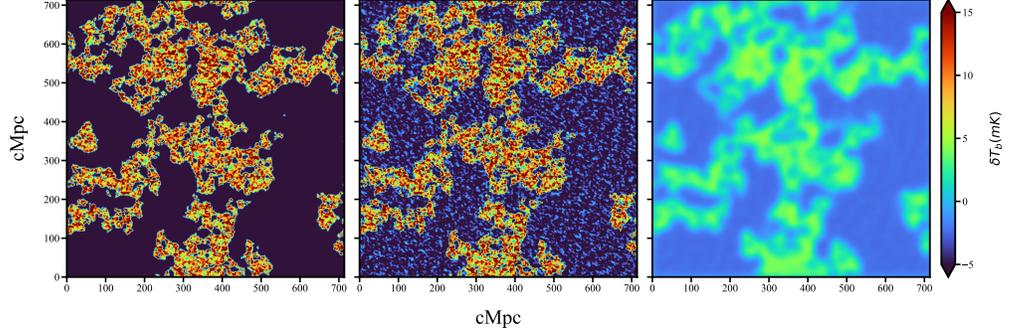


Figure 6.1: Slices of the brightness temperature maps for  $z = 7.221$  ( $\bar{x}_{\text{HI}} = 0.2$ ). *Left:* The original down-sampled map having a grid size of 300 cells along each side. *Center:* Map after adding a Gaussian random noise of rms 3.10mK. Random fluctuations in the pixel intensities can be seen. *Right:* Representation of the map after smoothing the noisy map with a Gaussian smoothing kernel of smoothing length 7.14 cMpc to mimic the SKA-I Low behaviour.

## 6.1 Addition of noise

A Gaussian random noise was generated and was added to the signal to see its effects on the LCS. The value of noise rms was chosen as per [Giri et al., 2018b]. In [Giri et al., 2018b], they assumed a 4hr observation at a declination of  $-30^\circ$ . The mean of this kind of noise is considered as zero and the noise sigma is given by the radiometer equation 6.1 [Ghara et al., 2016]

$$\sigma = \frac{\sqrt{2}K_B T_{\text{sys}}}{\epsilon A_D \sqrt{\Delta\nu t_{\text{int}}}} \quad (6.1)$$

$T_{\text{sys}}$  denotes the telescope system temperature and  $K_B$  denotes the Boltzmann constant. The antenna efficiency is quantified using  $\epsilon$  and it is dependent on the frequency of observation.  $A_D$  is the physical area of each dish.  $\Delta\nu$  and  $t_{\text{int}}$  denote the frequency resolution and the telescope integration time respectively.

The authors of [Giri et al., 2018b] have done the thorough analysis on noise and had chosen a noise rms value of 2.82 mK as per the capabilities of SKA1-low. For our case, as we are just dealing with the largest ionized bubble, we start with a higher value of noise rms which is 3.10 mK. We

gradually increase the noise and see its effects on LCS on Figure 6.3.

## 6.2 Determination of threshold

As mentioned earlier, the determination of a threshold that identifies the boundary between the two type of pixels i.e. ionized and neutral pixels in the H<sub>I</sub> map is crucial for running SURFGEN2. However, Addition of noise and smoothing the map shifts our threshold. When a random noise is added to the data, each pixel gets shifted by some random amount which is defined by the rms of the noise added. As a result, the boundary between the ionized and neutral pixels get randomly shifted for each map. Alongside, when these maps are smoothed with a Gaussian kernel of a particular length scale, an averaging effect on top of the noise takes place and the pixels are again shifted to some other value. This, as a result, shifts the threshold further.

The typical histogram of an ideal H<sub>I</sub> 21-cm image cube has a sharp bimodal feature intrinsic to it as shown in Figure 6.2. The two peaks in the histogram represent the ionized and neutral pixels in the given data. If the local minima between these two pixels can be calculated, then a boundary value between these two regions can be found. To do this, a Gradient descent method can be utilized and in this way, the threshold was determined for each of the maps. It is important to note that this entire analysis was done considering only the Fiducial model.

As the effect of noise and array synthesized beam was incorporated in the maps, the bimodal feature of these maps started to get smoothed out leading to a Gaussian-like feature. This effect was seen especially on the maps with higher neutral fraction ( $\bar{x}_{\text{HI}}$ ) values. Nevertheless, the bimodal feature could still be visible in the maps with lower neutral fraction values. This enabled us to utilize Gradient descent method and find our required thresholds. We verified the value found out from Gradient descent with the threshold values for our original maps which were already known ( i.e.  $\rho_{\text{HI}} \approx 0$  ).

This method of thresholding did not work as the noise rms and the smoothing length-scale was increased. It also failed to work when we run

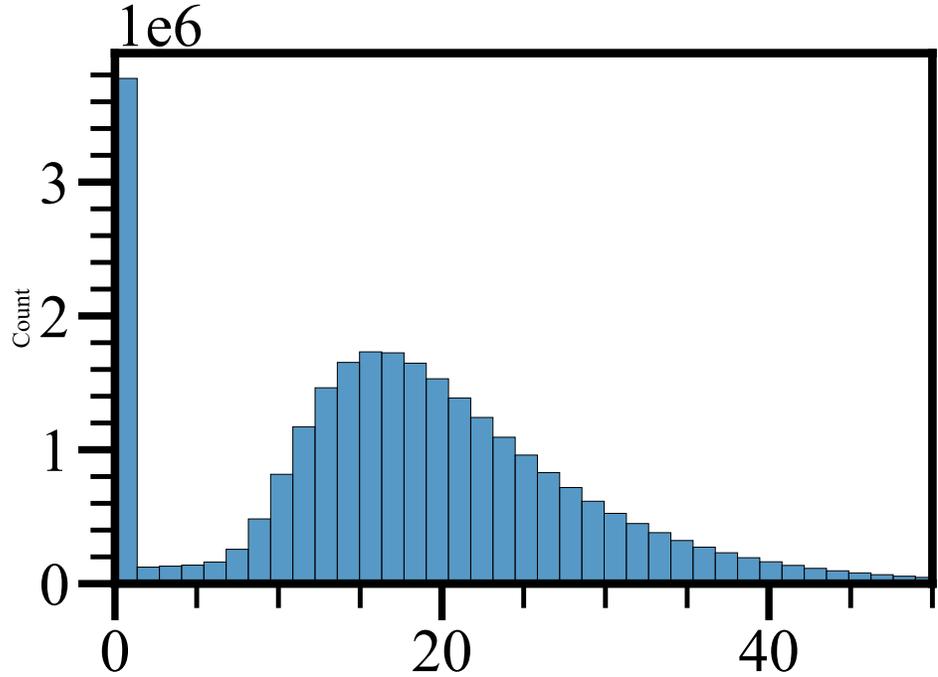


Figure 6.2: An image of a typical histogram of an ideal HI 21-cm image cube. It can be seen that there are sharp bimodal features. The peak on the left correspond to the ionized pixels ( $\rho_{HI} = 0$ ) and the peak on the right correspond to the neutral pixels.

our maps through the SKA1-low pipeline. To mitigate this, we used  $\sim 10\%$  of the maximum pixel value as our threshold. We also verified this with our initial results from the original maps.

### 6.3 Results after addition of noise and smoothing

After the addition of a Gaussian noise as per [Giri et al., 2018b], we see random fluctuations are introduced in the pixel values in the maps. It leads to a deviation in threshold which was taken care by the Gradient descent method that we discussed in Section 6.2.

We start from a noise rms of 3.10 mK and go upto 9.30 mK to test the robustness of LCS. As we increase the noise rms we see that due to the introduction of random fluctuations, the small ionized islands present in the maps of higher neutral fraction are lost. For this reason we lose the features

of LCS at higher neutral fraction as we keep on increasing the noise level as seen in Figure 6.3. However, for the lowest noise level 3.10 mK, we find that LCS could be completely recovered and we see no change in the percolation transition threshold and it remains the same. Also, as we increase the noise level, the bimodal feature is lost and alternate thresholding algorithms are required to solve this problem. To test the upper limit of LCS with Gradient descent type thresholding, it requires more number of noise samples between 3.10 mK to 4.50 mK.

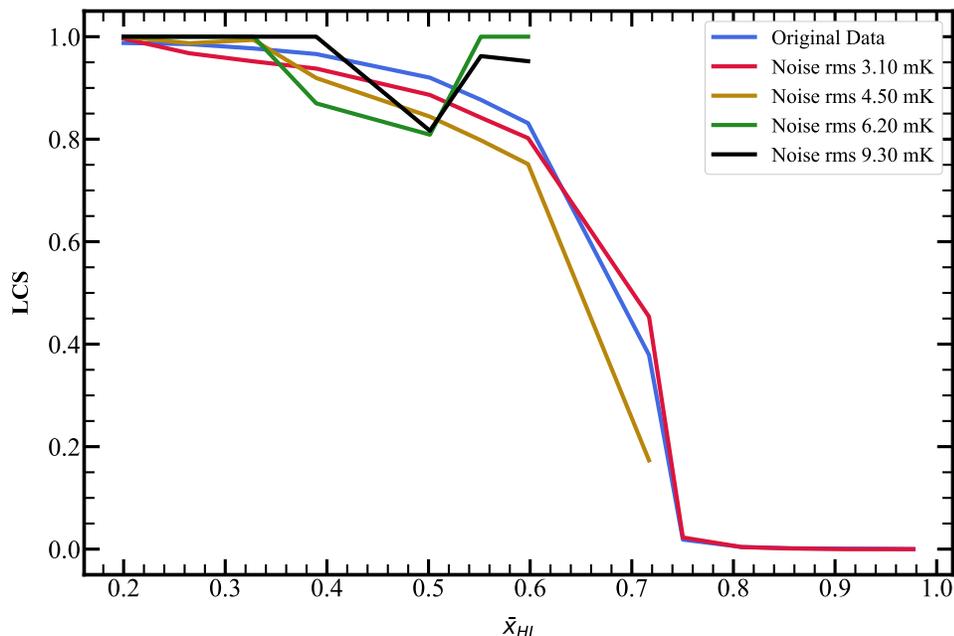


Figure 6.3: Variation of LCS with neutral fraction ( $\bar{x}_{\text{HI}}$ ) is shown. It is seen that as we go to higher noise levels, the sharp bimodal feature of the 21-cm maps are lost and LCS cannot be computed for higher  $\bar{x}_{\text{HI}}$  values.

To mimic the array synthesized beam effect of SKA1-low, we again refer to [Giri et al., 2018b] and find out that the resolution corresponding to the 2 km baseline of SKA1 would have a Full Width Half Maxima value of 3 grid units or a grid spacing of 7.14 Mpc in comoving units along each side of the simulation volume. We used a 3D Gaussian smoothing kernel of varying lengthscales to imitate this behaviour and plot LCS against neutral fraction for different smoothing lengthscales in Figure 6.4. It is observed that for lower smoothing lengthscales the LCS is not affected much. But, as we move to higher smoothing lengthscales, the small ionized regions in case of higher

neutral fraction gets smoothened out. This results in generation of a lot of partially ionized regions. As SURFGEN2 deals with a binarized 21-cm map, these partially ionized regions are treated as either completely neutral or completely ionized as per our threshold. Besides, as smoothing operation is done, the bimodal feature is also lost and the histograms tend to become Gaussian-like. These effects combined results in an erratic behaviour of the LCS curve for higher smoothing lengthscales. It necessitates a histogram independent threshold algorithm.

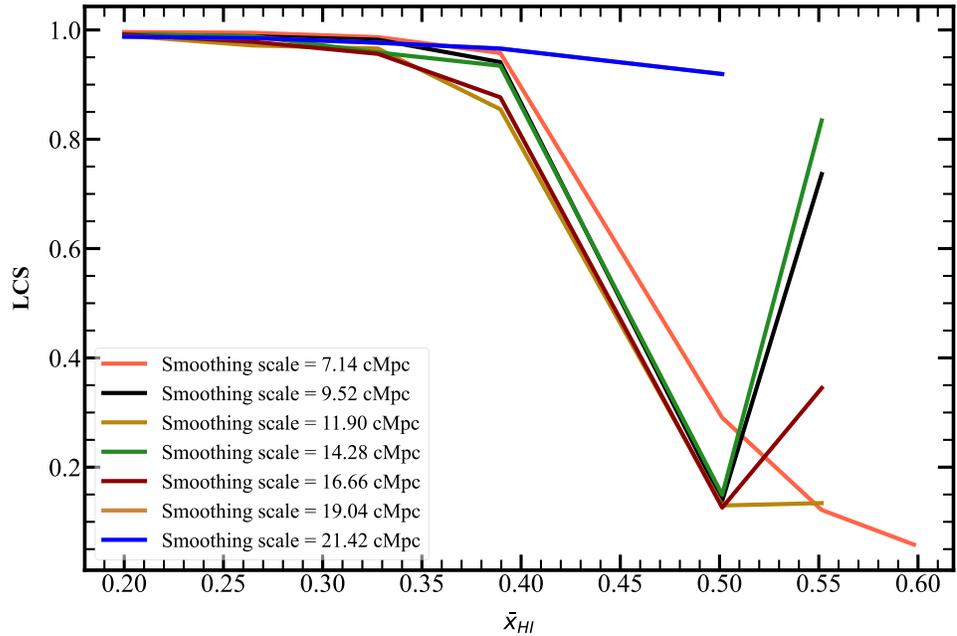


Figure 6.4: LCS vs neutral fraction ( $\bar{x}_{\text{HI}}$ ) for changing smoothing lengthscales. Due to the disappearance of small ionized regions at higher neutral fractions, the behaviour of LCS do not follow the expected trend.

Finally, when both noise and smoothing are taken into account simultaneously, all of the effects mentioned above come into play. Alongside, this affects the detection of number of clusters in SURFGEN2. We plot the LCS vs neutral fraction in Figure 6.5 to see the combined effect of noise and smoothing. We considered the minimum value of noise and the minimum lengthscale for our smoothing radius to check the evolution of the largest ionized region with LCS. We see that although Gradient descent could determine a threshold for maps at neutral fraction  $\bar{x}_{\text{HI}} \geq 0.72$ , the percolation transition point shifts heavily. It leads to an incorrect interpretation of the

reionization history.

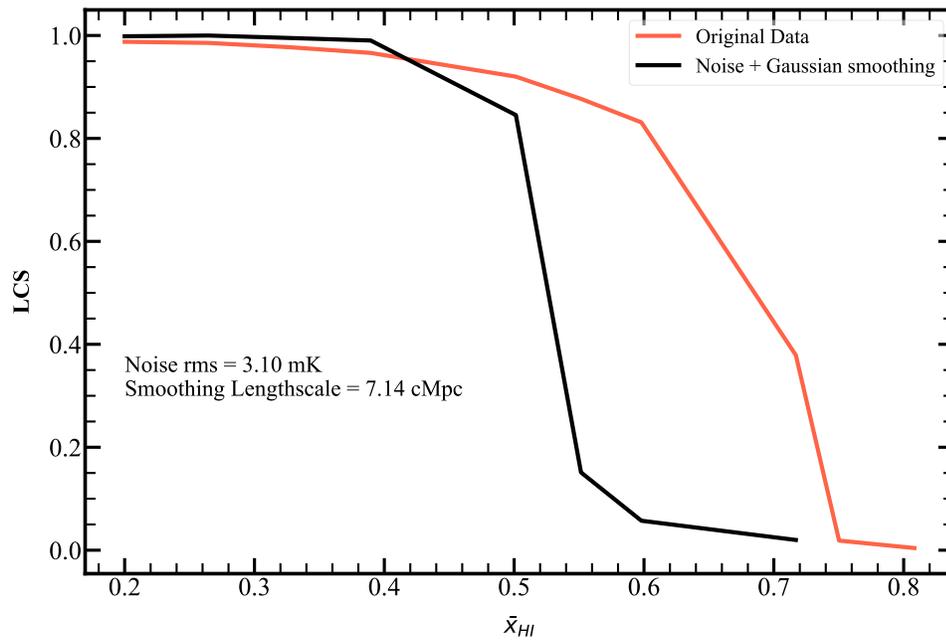


Figure 6.5: LCS vs  $\bar{x}_{HI}$  plotted for noise and Gaussian smoothing combined. It is seen that due to the incorrect threshold determination and error in the cluster-number detection, the percolation transition point is shifted to a later stage of reionization which is different from the value determined from simulation.

# Chapter 7

## Creating observational maps and determination of LCS

### 7.1 The End-to-End pipeline at hand

The pathway for developing synthetic observation and analysis is described in this section in detail. The building pieces which are needed to produce the synthetic observations are displayed in Figure 7.1. In Table 7.1, the observational parameters that are used are shown. Each block of the input parameter are also described afterwards. The simulations are done using **OSKAR**<sup>1</sup> software for SKA1-Low, and Common Astronomy Software Application (**CASA**<sup>2</sup>) is used to create 21-cm images from the visibility data. In our particular case, We have an observation for half an hour ( $\pm 0.5$  HA), with the phase centre at  $\alpha = 15\text{h}00\text{m}00\text{s}$  and  $\delta = -30^\circ$ . Cosmological parameters from the WMAP results [Komatsu et al., 2011] are used in this work:  $h = 0.7$ ,  $\Omega_m = 0.27$ ,  $\Omega_\Lambda = 0.73$ ,  $\Omega_b h^2 = 0.0226$  [Komatsu et al., 2009]. First, we divided the 21 cm cubes into different slices by considering one of the axes to be the frequency axis. Then, we converted these images from comoving coordinates to angular-frequency coordinates. These sliced images were taken as the input sky model for OSKAR to run the simulations. After the simulations are done, we stack the images back along the different frequency channels to form the final visibility cube.

---

<sup>1</sup><https://github.com/OxfordSKA/OSKAR/releases>

<sup>2</sup>[https://casaguides.nrao.edu/index.php?title=Main\\_Page](https://casaguides.nrao.edu/index.php?title=Main_Page)

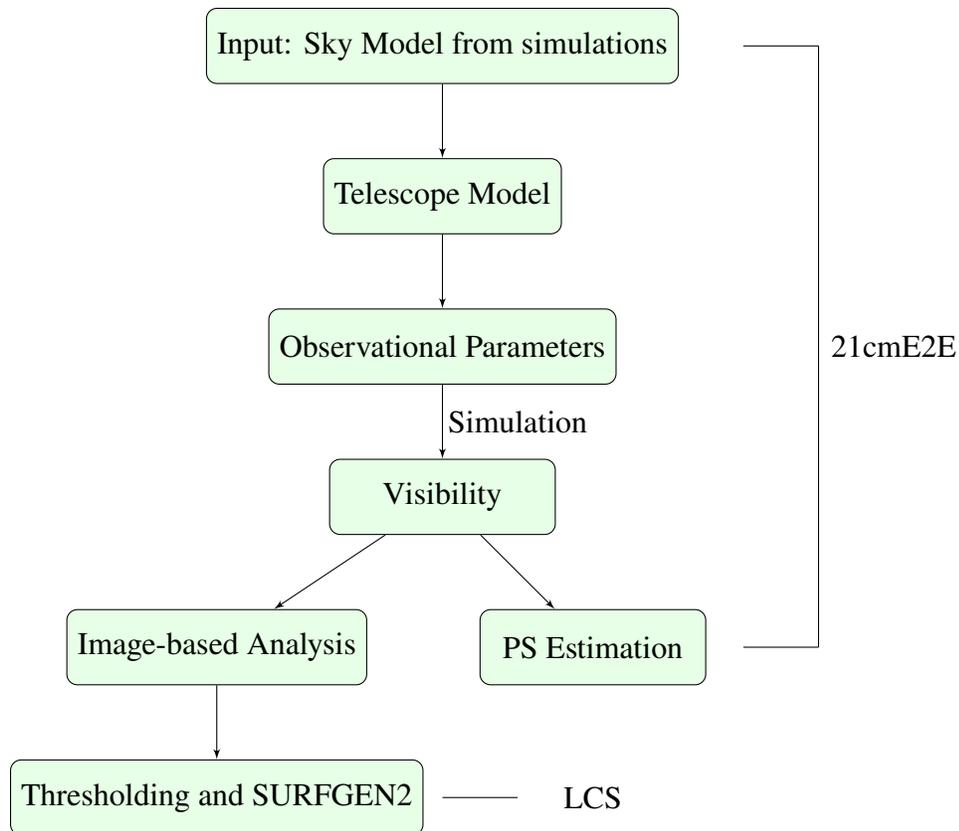


Figure 7.1: Schematic diagram of the 21cmE2E-pipeline used for analysing the H<sub>I</sub> 21-cm maps from the future SKA observations.

Table 7.1: Observational parameters used in the simulations

Parameter	Value
Field of view (FoV)	4°
Number of array elements	296
Maximum baseline	~ 2000 m
Polarization	Stokes I
Observation Time	30 minutes
No. of snapshots	15
Integration time per snapshot	2 minutes
Phase Center(J2000)	RA, DEC= 5 h, -30 °

### 7.1.1 Telescope model

The Square Kilometer Array (SKA1-Low, [Koopmans et al., 2015]), one of the most sensitive next-generation telescopes, is anticipated to statistically detect the signal and have just the right signal-to-noise ratio for tomography [Mellema et al., 2015]. A low-frequency array in Australia (working from 50 to 350 MHz) and a mid-frequency array in South Africa (operating from 350 to 14 GHz) will each be built as part of the SKA. The Murchison Radio-astronomy Observatory (MRO) location in Western Australia, where the MWA is situated, will house the low-frequency component of the SKA LOW. The first SKA1 LOW deployment will consist of 512 stations with a diameter of less than 40 metres, each with 256 antennas, scattered throughout a dense core, and spiral arms with a maximum baseline length of 65 kilometres. The compact core consists of 85% of all stations within a baseline of 2 km due to the target large-scale signal's sensitivity as shown in Figure 7.2. The corresponding  $uv$ -coverage for 30 minutes of observation is shown in Figure 7.3. The histogram for the  $uv$ -density distribution that shows the baseline density (number of baselines per unit area of the  $uv$ -plane) against  $U$  which is  $\sqrt{u^2 + v^2}$  is shown in Figure 7.4. Table 7.1 summarizes the telescope model that has been used.

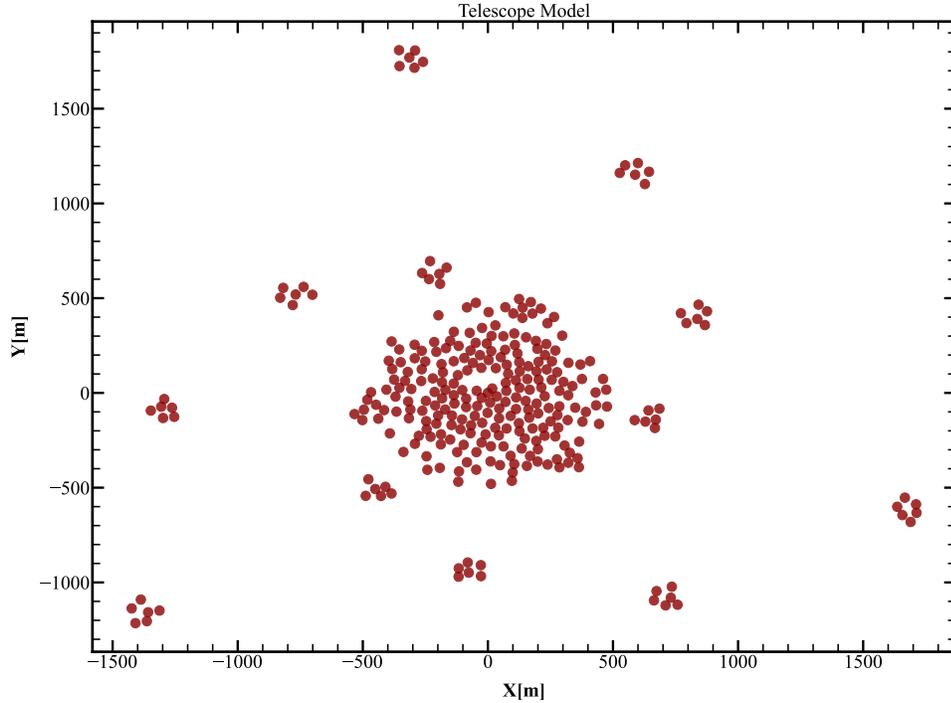


Figure 7.2: The telescope layout that has been used in this work. It comprises of stations around 2km from the central station.

### 7.1.2 OSKAR

A variety of applications for simulating astronomical radio interferometers are included in the OSKAR package. OSKAR was primarily created to produce simulated data from aperture array-using telescopes, such as those envisioned for the SKA. For an SKA-sized station's aperture array, OSKAR replicates digital beamforming. The simulator has been created as a flexible framework to research various computational approaches to the processing of the beamformer and their influence on the output beam quality [Mort et al., 2010].

For maximum flexibility, a straightforward computational structure, and quick switching between operational modes, OSKAR is built as a beamforming scheme with stringent computational needs in mind. The matrix-vector technique allows to keep this flexibility because it is assumed that just a fraction of the beams, frequencies, or polarizations are required for a specific observation.

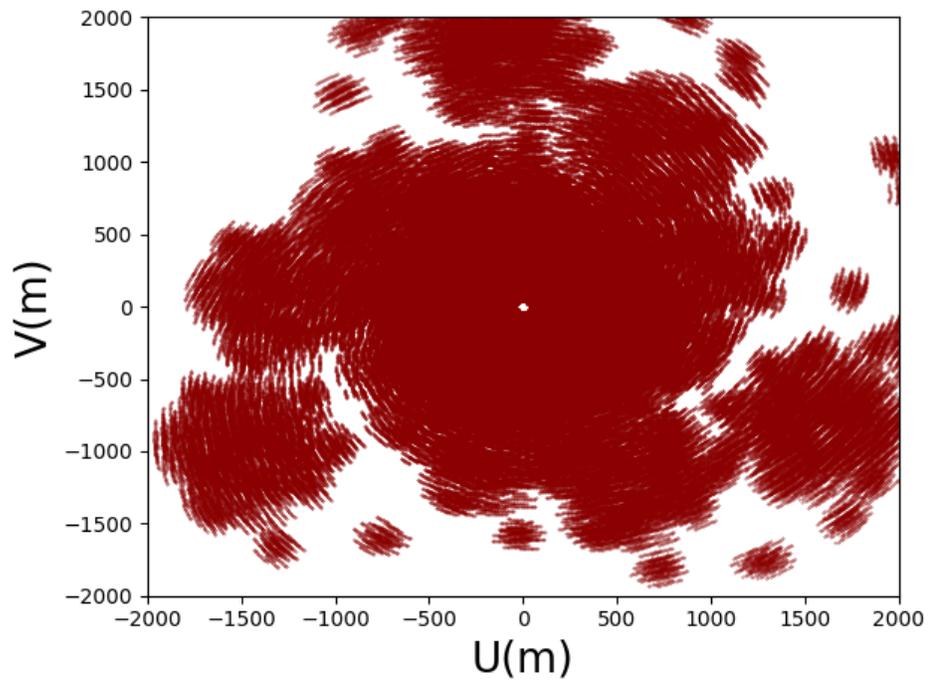


Figure 7.3:  $uv$ -coverage for 30 minutes observation.

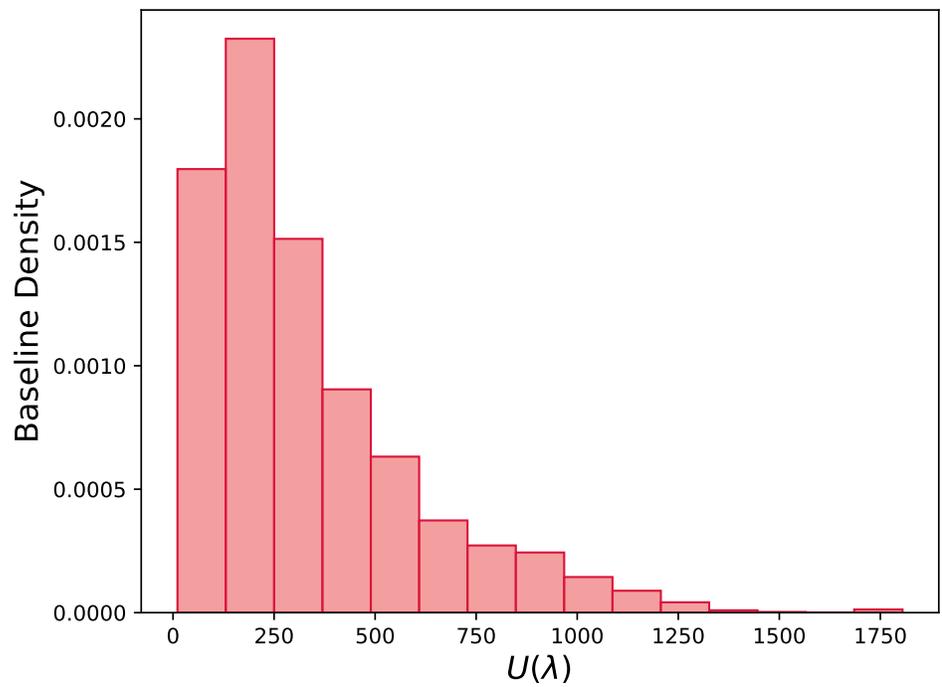


Figure 7.4:  $uv$ -density for the observation that has been taken into consideration.

### 7.1.2.1 OSKAR simulation

For an SKA-sized station's aperture array, OSKAR replicates digital beamforming. The simulator has been created as a flexible framework to research various computational approaches to the processing of the beamformer and their influence on the output beam quality. A flow chart of the working of OSKAR simulator is shown in Figure 7.5.

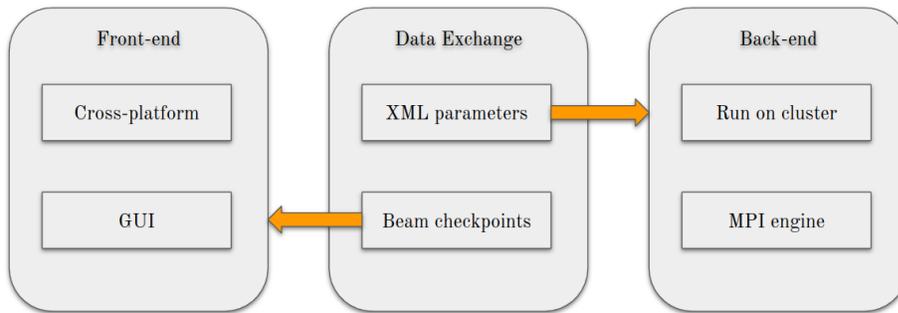


Figure 7.5: The OSKAR simulator's general architecture. The front end and back end are separate components that communicate with one another using a file interface. [Mort et al., 2010]

There are two primary parts to the simulator. The simulation settings, including antenna installation, the sky model, and beam positions, are set up using a front-end graphical interface. The front-end runs on the user's desktop or workstation and makes use of platform-neutral libraries. A supercomputer cluster-optimized, modular MPI-parallel computational back-end is used to run the actual beamforming simulation. Sky processing, antenna signal creation, beamforming, and complex weights production are all supported by modules. The modular design of OSKAR makes it simple to test out various beamforming algorithms by implementing additional weights generators. Data can be check-pointed at any level of the computational pipeline for use in other simulations if necessary. By intercepting the data stream, it is also possible to test out different hardware architectures.

The sky model is used by the antenna signal modules to generate a stream of complex-valued dual-polarization signals for each antenna. For each beam, the weights generators provide a set of complicated beamforming

weights that specify the direction of interest. It is possible to compute the weight coefficients away from the primary signal route since they vary considerably more slowly than the signal sample rate, which is how beam steering is accomplished. By merely performing multiply-accumulate operations between both the antenna signals and complex weights, this technique enables the computationally intensive beamforming modules to be fairly straightforward. We chose to employ the matrix-vector method in the OSKAR simulator for flexibility, but other beamforming techniques, such as the use of the FFT, may be simply added to OSKAR. OSKAR has two modes of operation: ‘beam-pattern’ mode and ‘end-to-end beamforming simulation’ mode.

For the specified array configuration, a single snapshot of the directional beam response is generated using the “beam-pattern” mode. The beamforming weights are maintained constant for the direction of interest while a test source with unit amplitude is swept across the model sky to assess the station response at each position because there is no time dependence in this mode. The effects of beam shaping, such as applying apodization or imposing null steering, as well as experimenting with various array designs and/or different antenna types, can all be studied in this mode.

The third mode of operation, called “end-to-end beamforming simulation”, creates antenna signals for each component of the aperture array using a point-source sky model. It is feasible to mimic beam tracking, numerous beams per level, and add noise and other time-variable effects in this mode because it models a full-size, genuine SKA station. This mode is ideal for testing various beamforming algorithms while analysing their performance and looking into the impacts of utilising a different processing hierarchy.

### 7.1.2.2 Sky model used

We have used 21 cm signal model as input generated from the Reion Yuga. We converted the sky brightness temperature (K) to specific intensity or brightness (Jy/beam) using the Rayleigh-Jean (RJ) formula. The RJ formula is given by:

$$S = \frac{2k_B T}{\lambda^2} \Omega \quad (7.1)$$

where the  $S$  is the specific intensity,  $k_B$  is the Boltzmann constant and  $\lambda$  is the wavelength of the corresponding central frequency of the observation. We have sliced the images from the third axis (along the line of sight or frequency axis) and converted to comoving Mpc to an angular conversion for radio observation. The formula for this conversion was taken from [Giri et al., 2018b] which is as follows:

$$\Delta\theta = \frac{\Delta x}{D_c(z)}, \quad (7.2)$$

where  $D_c(z)$  is the comoving distance to redshift  $z$ , and

$$\Delta\nu = \frac{\nu_0 H(z) \Delta x}{c(1+z)^2} \quad (7.3)$$

where  $\nu_0$  is the rest frequency of the 21-cm line,  $H(z)$  the Hubble parameter at redshift  $z$  and  $c$  the speed of light. For example, at a redshift  $z = 7.221$ , the 714 comoving Mpc to angular resolution corresponds to 0.015 deg and frequency resolution is 125 KHz. The slice, we used as an input and the perform the simulation on OSKAR, and we retrieve the visibility. As OSKAR uses an FFTW scheme for its operation it is better to use the input sky model to have a volume of the order of  $2^n$ . As we previously had maps of  $300^3$  grid units we needed to re-simulate new maps. For our particular case we newly simulate 21-cm maps of  $256^3$  grid units. We do this using  $1024^3$  particles of mass  $1.089 \times 10^8 M_\odot$  on a  $2048^3$  mesh grid. The final reionization maps were generated by coarsening the initial mesh grid with a factor of 8 resulting in a  $256^3$  grid size. The simulation volume for this realization is  $143.36^3$  Mpc in comoving length along each side of the cube. We again tweak the parameter  $N_{ion}$  (as mentioned in Section 3.4) in such a way that the reionization history remains consistent with the previous results as found in our sampled data cubes of  $300^3$  grid units. For this, we plot the LCS vs  $\bar{x}_{\text{HI}}$  between the maps of  $256^3$  grid units with the ones with  $300^3$  grid units in Figure 7.6. We see that for the same condition of threshold, our percolation transition point remains the same.

### 7.1.3 CASA

The Fourier relation between the primary beam patten  $A$ , sky intensity  $I$  and the visibility  $V$  observed with an interferometer which as given below:

$$A(l, m, \nu)I(l, m) = \int \int_{sky} V(u, v, \nu) \exp[2\pi i(ul + vm)] dudv \quad (7.4)$$

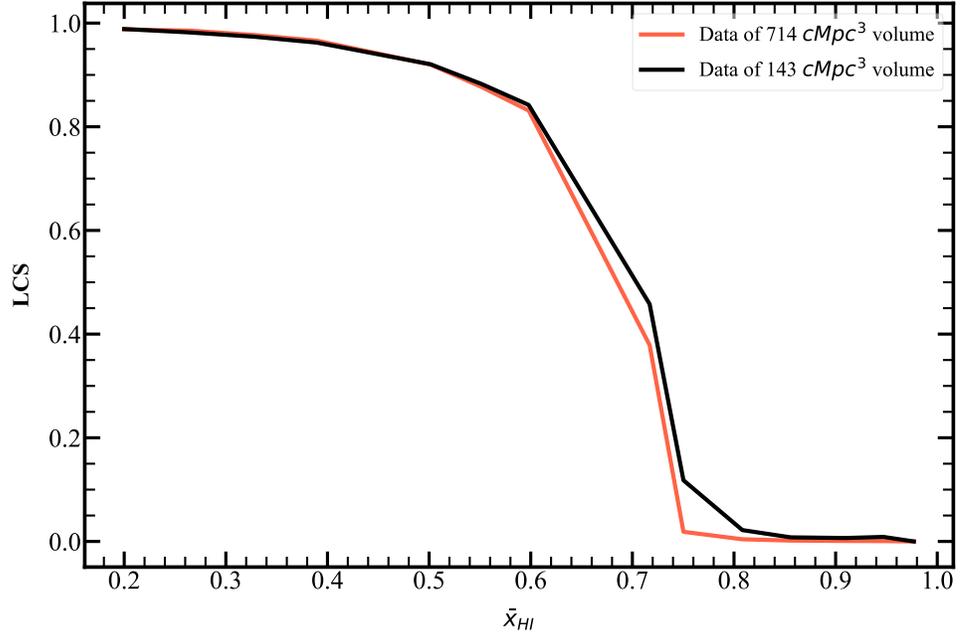


Figure 7.6: LCS vs  $\bar{x}_{HI}$  is plotted for maps of  $300^3$  grid volume and  $256^3$  grid volume. It can be seen that the percolation transition threshold remains the same even though the simulation volumes are different.

In our case primary beam correction is not considered ,i.e.  $A(l, m) = 1$ . Radiosynthesis imaging of astronomical sources was revolutionized by the invention of the CLEAN algorithm. This simple algorithm enabled synthesis imaging of complex objects even with relatively poor coverage of the Fourier plane, such as occurs with partial earth rotation synthesis or with arrays of a small number of antennas. In CLEAN, the sky image is assumed to be a collection of point sources. The algorithm is based on estimating the brightness and position of all point sources in the sky from the image in an iterative process. This is achieved in different ways in different variants of CLEAN. Here we used the Cotton-Schwab variant of CLEAN. In order to evaluate the Fourier transform of the measured visibilities, the observed visibilities are weighted and plotted on regularly spaced grid points. Two extreme weighting schemes are referred to as uniform and natural weights, respectively. With natural weighting, all measured values are weighted equally and added up. This type of weighting emphasizes the part of the visibility plane where there are more measurements. Note that when implementing a weighting scheme, the effective dirty beam is

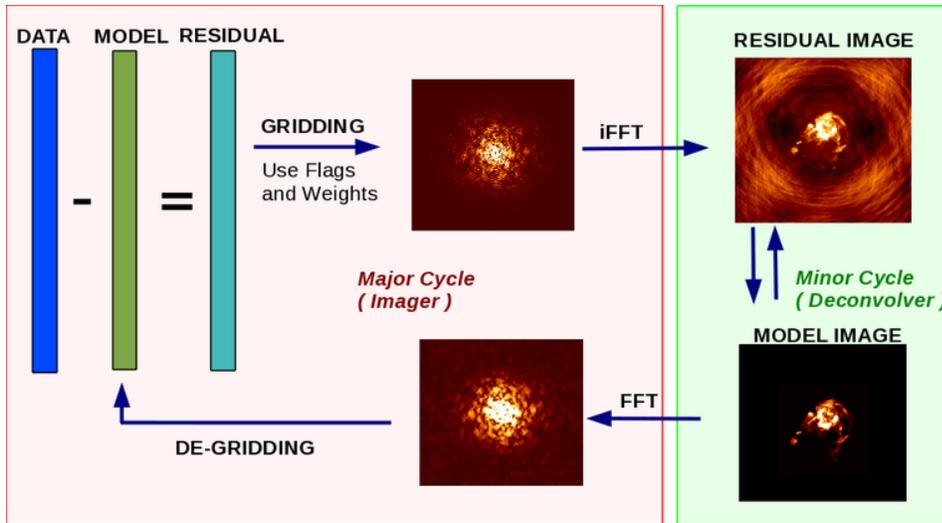


Figure 7.7: Iterative Image Reconstruction - Major and Minor Cycles. Image from [CASA tclean](#)

to be viewed as the Fourier transform of the sampling function multiplied by the weight kernel. As with most interferometers, baseline coverage is reduced at lower values of  $|U|$  better, and the PSF derived from the naturally weighted visibilities is wider but with fewer sidelobes. On the other hand, in another weighting scheme, called uniform weighting, the visibilities are weighted by the spatial density of measured visibilities before screening. This type of weighting results in a narrower PSF but has higher power in the sidelobes. The robust weighting scheme designed by [Briggs, 1995a] attempts to combine natural and uniform weighting. An iterative weighting scheme, called adaptive weighting, is discussed in [Yatawatta, 2014a] to support high-bandwidth, high-dynamic-range imaging by minimizing PSF variation across frequencies while maximizing sensitivity. The detailed process of gridding, various weights and their effects are discussed in [Thompson et al., 2017a].

We image the reweighted data using Multiscale Clean, which instead assumes that the sky can be represented by a set of Gaussian functions rather than delta functions. After using CASA CLEAN with Hogbom cleaning algorithm, we find artefact at the boundaries of the maps as shown in Figure 7.10. We suspect that this occurs due to the fact that our image size is not of the order of  $2^{3n}$ . As OSKAR employs an FFTW scheme to do its operation, the

	Robust/Uniform	Natural	Taper
resolution	higher	medium	lower
sidelobes	lower	higher	depends
point source sensitivity	lower	maximum	lower
extended source sensitivity	lower	medium	higher

Figure 7.8: The schematics representation of the CASA tclean weighting and tapering summary. Image from [CASA tclean](#)

optimum size for our maps should be of the order of  $2^{3n}$ . For this reason, we re-simulate maps of  $256^3$  grid units as mentioned in Section 7.1.2.2.

Along with the edge effect, we also encounter the effect of the PSF which is not properly deconvolved using hogbom cleaning. The artefacts rising due to this is shown in figure 7.9. We suspect that incorrect weighting in the extended sources lead to this kind of artefacts.

We go for a multiscale deconvolution algorithm scheme to mitigate these effects and we manually adjust the weights for the extended sources and the point sources. After doing so, we get a better-cleaned image as shown in 7.11 and can be verified with the original simulated image shown in above in 7.11. Here, even the small ionized regions are resolved.

Afterwards, we change the natural weighting scheme to Briggs weighting to recover the image cubes at high neutral fraction ( $\bar{x}_{\text{HI}}$ ) values. We have used the robust = 0.1 parameters to reconstruct the images from the actual image. In case of actual radio observations, we will not have any reference image to verify our cleaned images with. So, we have to use advanced algorithms to figure out the images from the real radio observations.

In this work, we try to recover the ionized region to run the largest cluster statistics. In figure 7.11, we have seen that the central part of the image gets distorted because of incorrect weighting. For this work, we will change the relative weights given to longer baselines. The default natural weight gives zero weight to portions of the u-v plane without samples.

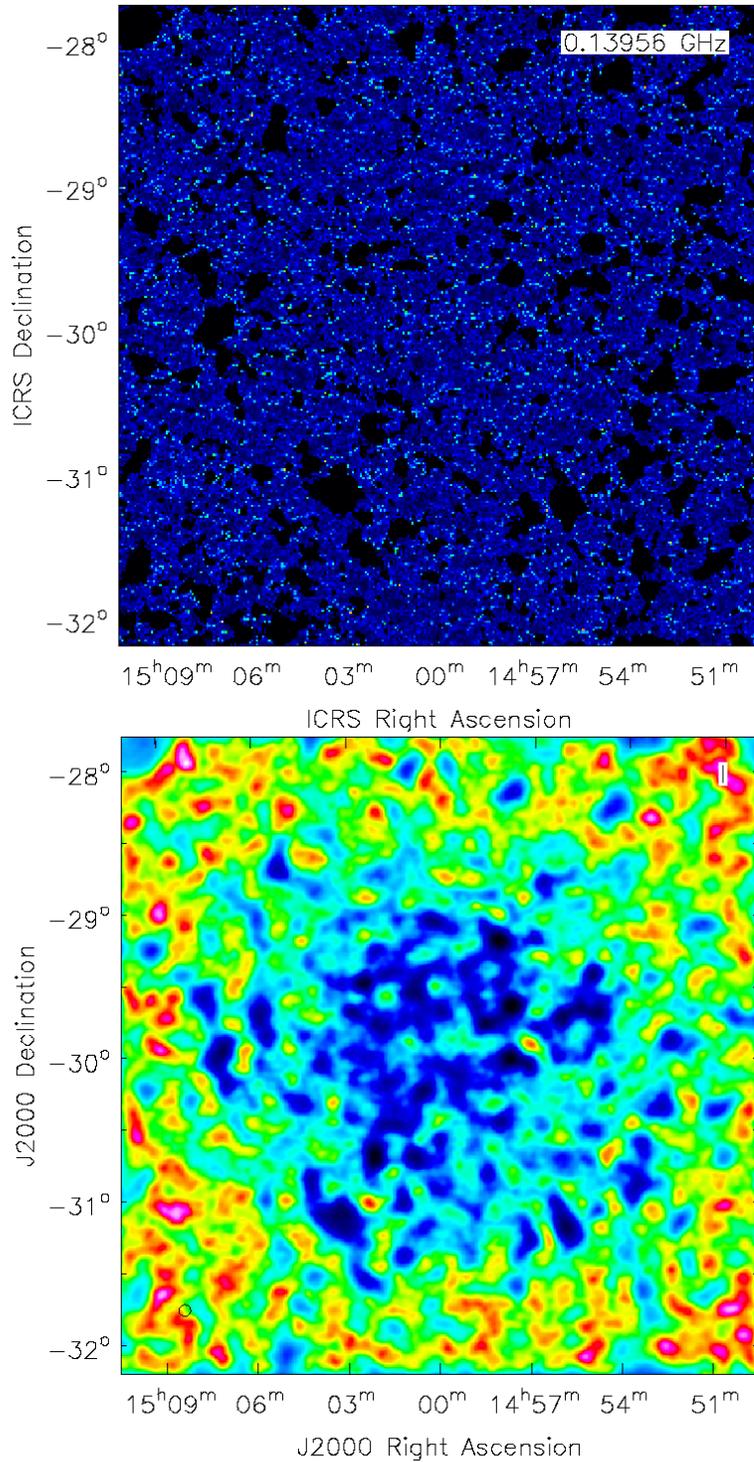


Figure 7.9: Slice of the original H<sub>I</sub> 21-cm signal map using ReionYuga at a neutral fraction value of  $\bar{x}_{\text{HI}} \approx 0.71$  from an image cube of size  $300^3$  grid units. The image below shows artefacts due to incorrect weighing in the extended source pixels in Hogbom CLEAN method with natural weighting.

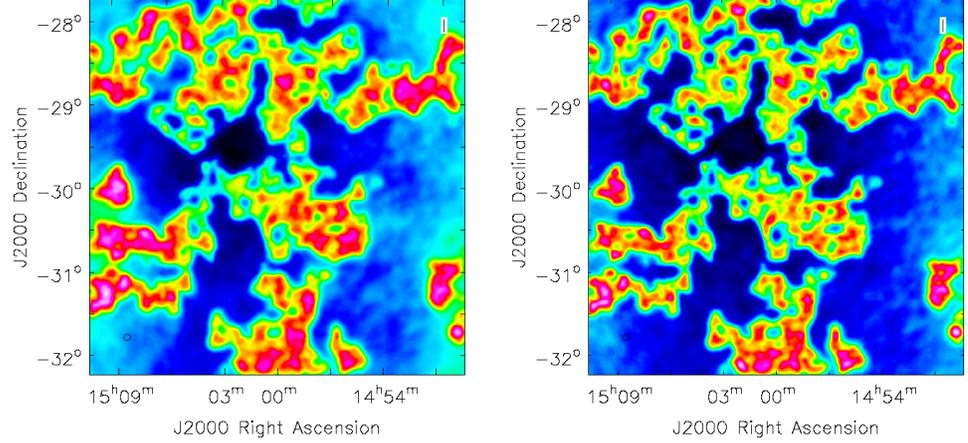


Figure 7.10: Comparison between the dirty image vs cleaned image slice of the simulated HI 21 -cm map at a neutral fraction  $\bar{x}_{\text{HI}} = 0.2$

The Briggs weight interpolates between samples, giving more weight to the interpolated regions for lower robustness values. We use Briggs weighting with robust=0.1 parameter to recover the ionized bubble for running the LCS. The Briggs weighting images are shown in figure 7.12.

#### 7.1.4 Results obtained

Initially, after performing Hogbom CLEAN algorithm on the image cubes of  $300^3$  grid units, we perform multiscale cleaning and plot the LCS with changing  $\bar{x}_{\text{HI}}$  in figure 7.13. We see that due to improper cleaning the ionized regions at higher neutral fractions where  $\bar{x}_{\text{HI}} \geq 0.6$  SURFGEN2 fails to perform with Gradient descent. Even when SURFGEN2 works for the neutral fraction values  $\bar{x}_{\text{HI}} \leq 0.6$ , we see a behaviour in the LCS plot that is erratic. We suspect that this may occur due to two reasons -

- Even after applying multiscale CLEAN to the images, as it is a manual process, some slices may contain artefacts due to incorrect pixel-weighting.
- As we have used the maps of  $300^3$  grid volume, they still contain the edge effect which SURFGEN2 might detect as neutral islands of hydrogen and calculates LCS accordingly.

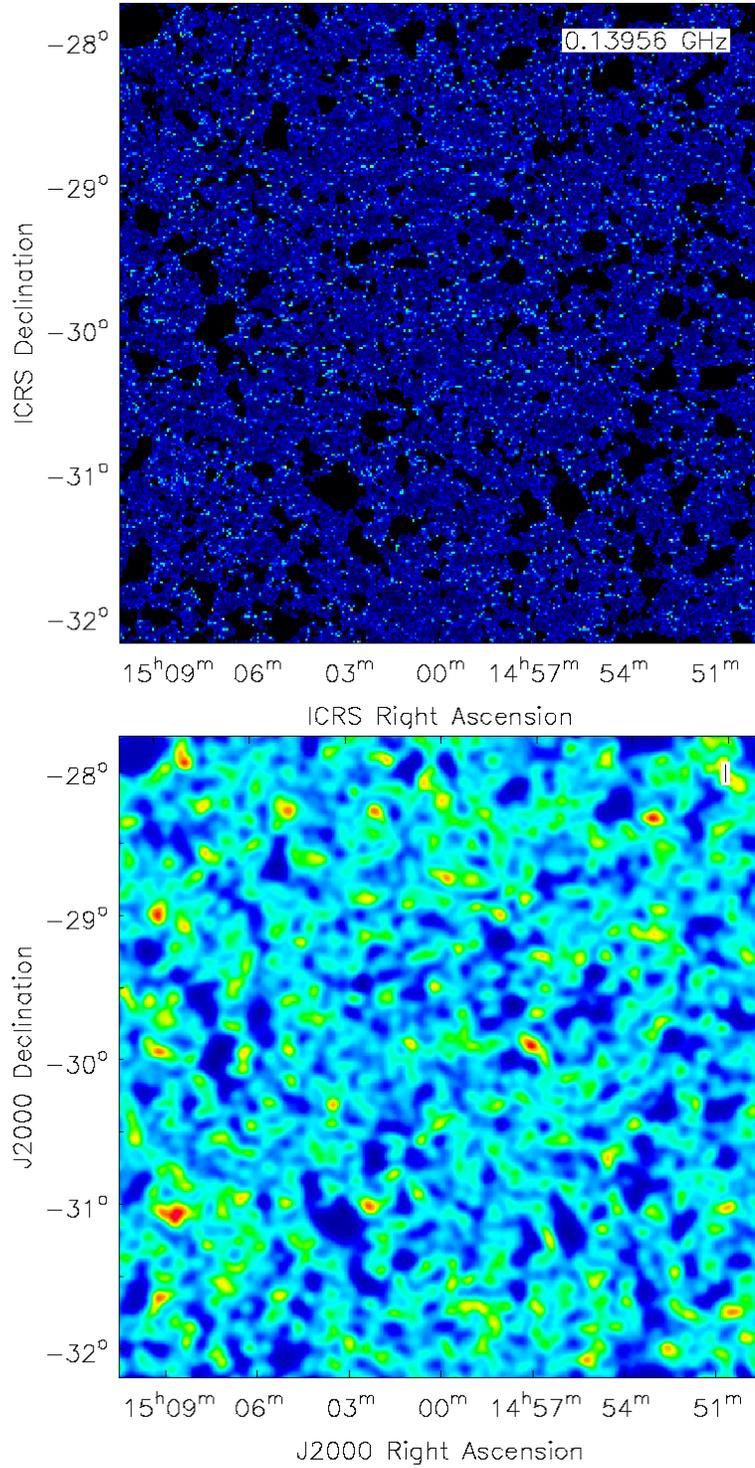


Figure 7.11: Comparison between the slice of the original H I 21-cm signal map using ReionYuga and the simulated image after Multiscale CLEANing at a neutral fraction value of  $\bar{x}_{\text{HI}} \approx 0.71$  from an image cube of size  $300^3$  grid units. It can be compared from the simulated image shown above that the features of the contrast difference could be recovered using Multiscale CLEAN.

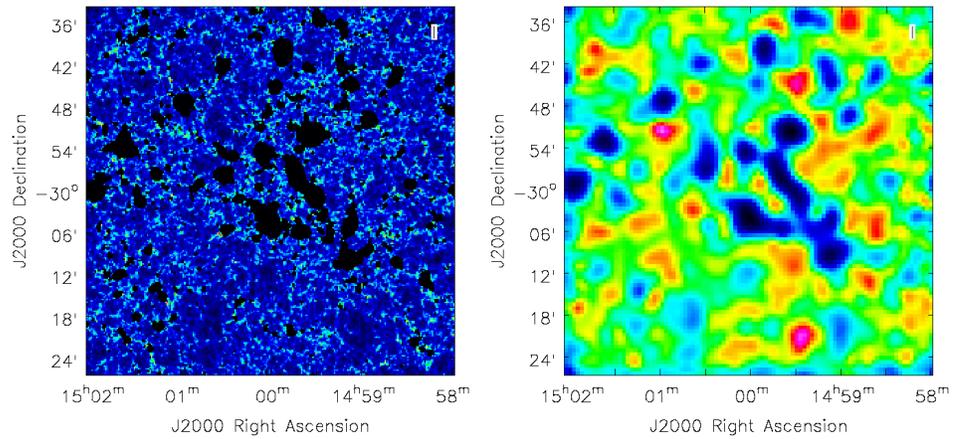


Figure 7.12: Comparison between the slice of the original H I 21-cm signal map and the simulated image after Multiscale CLEANing with the Briggs weighting scheme. It is important to note that these images have a grid volume of  $256^3$  grid units.

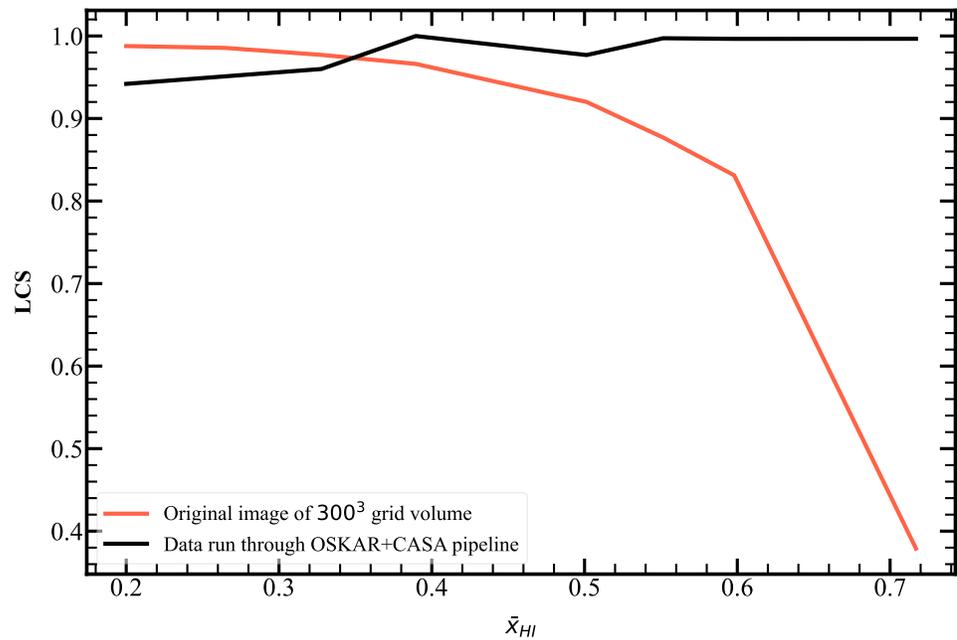


Figure 7.13: Effects of artefacts as seen in LCS vs  $\bar{x}_{HI}$ . Due to improper cleaning we see a trend in the LCS which is not expected and very different from our original case.

After creating new HI 21-cm cubes of gridsize  $256^3$  grid units we apply even more advanced CLEAN algorithm using Briggs weighing. Along with this, we change our threshold in the following manner - we choose  $\sim 10\%$  of the maximum pixel intensity value as our threshold for each map to perform SURFGEN2. The details of this thresholding is mentioned in section 6.2. After doing so, we calculate LCS and plot it against changing neutral fraction ( $\bar{x}_{\text{HI}}$ ) in Figure 7.14. The accuracy of the 21-cm images and LCS estimates significantly depends on how the dirty beam is removed from the observed data, as seen in Figure 7.14. However, due to the uncertain knowledge of the point spread function (PSF), recovering crucial features of the 21-cm maps that are necessary for identifying the largest ionized region's shape and volume is difficult. The conventional Hogbom CLEAN algorithm [Cornwell, 2008] is biased as it assumes that the sky emission is a set of delta functions. In contrast, the Multiscale CLEAN algorithm [Cornwell, 2008] can address this issue to some extent, but a non-trivial weighting scheme like Briggs weighting [Briggs, 1995b, Yatawatta, 2014b, Thompson et al., 2017b] is needed for the SURFGEN2 algorithm to be effectively applied to 21-cm images from the early stages of reionization. Despite using this advanced deconvolution algorithm, we find that LCS cannot be computed for the early stages of EoR where  $\bar{x}_{\text{HI}} \geq 0.7$  because of the absence of long baselines in the observations. This means that to be able to use LCS-like analyses, we need to have a better and denser sampling in the  $u - v$  plane coverage at higher baselines in the upcoming future SKA1-Low. To mitigate these issues we need to search for a better thresholding algorithm which we discuss in Section 8.2.

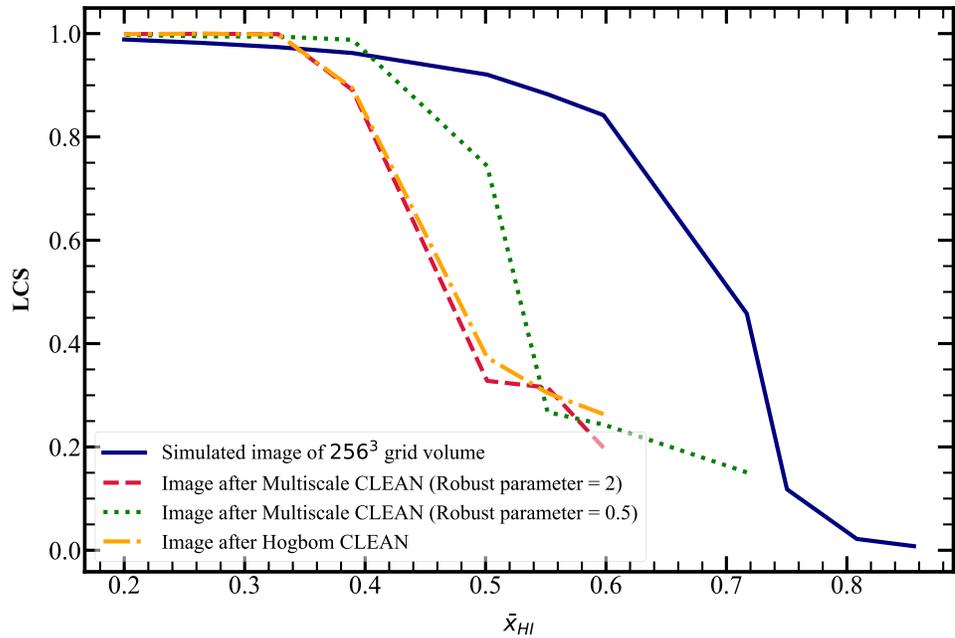


Figure 7.14: Estimated LCS for different neutral fractions using two Multiscale CLEAN techniques: natural weighted and Multiscale weighted with Briggs weighting (employing the robust parameter = 0.5). The LCS estimated using Briggs weighted CLEAN is superior to the natural weighted CLEAN in two ways. Firstly, we were able to calculate LCS values up to almost 0.7, which was not feasible with the natural weighted CLEAN. Secondly, the percolation transition was not shifted as significantly as in the case of natural weighted CLEAN.

# Chapter 8

## Summary and future work plan

### 8.1 Summary

- We first check the effect of instrumental noise by simulating a Gaussian Noise of different rms values. We check the LCS vs  $\bar{x}_{\text{HI}}$  plot and see that for a lower noise level of 3.10 mK we can recover the evolution of the largest ionized region correctly. But, as we move to higher noise level, we see that due to random fluctuations in the pixel values, threshold determination using Gradient descent does not work. This leads to us losing LCS features at higher neutral fraction for higher noise rms values.
- After that we mimic the effect of the array synthesized beam of the telescope using a 3D Gaussian filter of varying lengthscales of the smoothing kernel. For this also, we check the LCS against changing neutral fraction. We see that as smoothing comes into play the bimodal feature of the 21-cm maps are lost and Gradient descent fails to work. For higher smoothing lengthscales, the small ionized regions get smoothed out to partially ionized regions leading to a confusion for the SURFGEN2 code. Ultimately, this leads to an erratic behaviour in the LCS plots for higher smoothing lengthscales.
- The effect of noise and smoothing were combined and SURFGEN2 was run using a threshold determined by Gradient descent algorithm. Although some threshold could be determined, but due to the

combined effect of noise and smoothing, the LCS vs  $\bar{x}_{\text{HI}}$  plot gets heavily shifted. The percolation transition point shifts to a lower neutral fraction value as proper threshold could only be determined at lower neutral fractions. This leads to an incorrect inference of the reionization history. Also, it necessitates the requirement of a threshold algorithm independent of histogram.

- To test the robustness of LCS further we newly simulate 21-cm maps of gridsize  $256^3$  grid units and use it as an input sky-model for an SKA1-low pipeline using OSKAR. After the visibility dataset was generated from OSKAR, we perform an inverse fourier transform to get back our brightness temperature maps. We use advanced CASA CLEAN algorithm to deconvolve the array synthesized beam using existing beam models and generate our final HI 21-cm maps to run SURFGEN2 on them. To do this analysis we consider 10% of the maximum pixel intensity as our chosen threshold. After we run SURFGEN2 and find the variation of LCS against  $\bar{x}_{\text{HI}}$ . It is seen that after running our maps through this realistic pipeline, the percolation threshold is again shifted due to the effect of telescope. Which leads to the inference that no correct interpretation for the reionization history could be done. It would require denser sampling at higher baseline distances to have better resolution images to resolve these issues.

## 8.2 Future work-plan

- Firstly, a threshold finding algorithm which is histogram-independent has to be found. We explore the possibility of Superpixels and SegUnet as mentioned in [Giri et al., 2018b] for this case.
- We would also be using advanced machine learning algorithms to recover the true histograms from the corrupted histograms to get a better threshold for SURFGEN2.
- We would like to simulate the effect of noise from OSKAR to incorporate in the maps and check the robustness of LCS after we find a better thresholding algorithm.

- We would also like to incorporate the effect of Foregrounds and the effect of ionosphere and see at what extent LCS-like analyses could be valid.
- Finally, we would like to compare different telescope effects and check validity of LCS-like analyses on the observational maps

# Journal Publications

## Publications in peer-reviewed journals.

arXiv: [http://arxiv.org/a/dasgupta\\_s\\_1](http://arxiv.org/a/dasgupta_s_1)

Google Scholar:

<https://scholar.google.com/citations?user=em3YZfoAAAAJ&hl>

1. † **Dasgupta S.**, Pal S.K., Bag S., Dutta S., Majumdar S., Datta A., Pathak A., Kamran M., Mondal R., Sarkar P.; *Interpreting the  $H_I$  21-cm cosmology maps through Largest Cluster Statistics - I: Impact of the synthetic SKA1-Low observations*; accepted in the Journal of Cosmology and Astroparticle Physics, arXiv:2302.02727, DOI: [10.48550/arXiv.2302.02727](https://doi.org/10.48550/arXiv.2302.02727)  
**Contribution:** I am responsible for the majority of the work in this paper, creating the simulations, conducting the data analysis, creating most of the plots, drawing the majority of the conclusions, and writing most of the manuscript.
2. † Pathak A., Bag S., **Dasgupta S.**, Majumdar S., Mondal R., Kamran M., Sarkar P.; *Distinguishing reionization models using the largest cluster statistics of the 21-cm maps*; Journal of Cosmology and Astroparticle Physics, Edition: November, volume: 2022, Number: 11, Eid: 027, Page: 027. DOI: [10.1088/1475-7516/2022/11/027](https://doi.org/10.1088/1475-7516/2022/11/027),  
**Contribution:** I contributed to the simulation and analysis of low-resolution 21-cm maps.

**Note:** Publications 1 and 2 (marked with †) constitutes all of the major chapters of this thesis.

# Bibliography

- [Fan, 2003] (2003). Fan. *The Astronomical Journal*, 125:1649–1659.
- [Tin, 2013] (2013). Tingay. *Publications of the Astronomical Society of Australia*, 30:e007.
- [Shi, 2015] (2015). Shimabukuro. *Monthly Notices of the Royal Astronomical Society*, 451:467–474.
- [Wel, 2020] (2020). Fundamental physics with the square kilometre array.
- [Bag et al., 2019] Bag, S., Mondal, R., Sarkar, P., Bharadwaj, S., Choudhury, T. R., and Sahni, V. (2019). Studying the morphology of h  $\text{II}$  isodensity surfaces during reionization using shapefinders and percolation analysis. *Monthly Notices of the Royal Astronomical Society*, 485:2235–2251.
- [Bag et al., 2018] Bag, S., Mondal, R., Sarkar, P., Bharadwaj, S., and Sahni, V. (2018). The shape and size distribution of h  $\text{II}$  regions near the percolation transition. *Monthly Notices of the Royal Astronomical Society*, 477:1984–1992.
- [Becker and Bolton, 2013] Becker, G. D. and Bolton, J. S. (2013). New measurements of the ionizing ultraviolet background over  $z \sim 5$  and implications for hydrogen reionization. *Monthly Notices of the Royal Astronomical Society*, 436:1023–1039.
- [Bharadwaj and Ali, 2005] Bharadwaj, S. and Ali, S. S. (2005). On using visibility correlations to probe the h  $\text{I}$  distribution from the dark ages to the present epoch - i. formalism and the expected signal. *Monthly Notices of the Royal Astronomical Society*, 356:1519–1528.

- [Bharadwaj and Pandey, 2005] Bharadwaj, S. and Pandey, S. K. (2005). Probing non-gaussian features in the  $h$   $i$  distribution at the epoch of reionization. *Monthly Notices of the Royal Astronomical Society*, 358:968–976.
- [Bi and Davidsen, 1996] Bi, H. and Davidsen, A. F. (1996). Evolution of structure in the intergalactic medium and the nature of the Ly- $\alpha$  forest. *The Astrophysical Journal*, 479:523–542.
- [Bond et al., 1991] Bond, J. R., Cole, S., Efstathiou, G., and Kaiser, N. (1991). Excursion set mass functions for hierarchical gaussian fluctuations. *The Astrophysical Journal*, 379:440.
- [Bowman et al., 2013] Bowman, J. D., Cairns, I., Kaplan, D. L., Murphy, T., Oberoi, D., Staveley-Smith, L., Arcus, W., Barnes, D. G., Bernardi, G., Briggs, F. H., et al. (2013). Science with the Murchison widefield array. *Publications of the Astronomical Society of Australia*, 30:e031.
- [Bradt, 2008] Bradt, H. (2008). *Astrophysics Processes*.
- [Briggs, 1995a] Briggs, D. (1995a). High fidelity interferometric imaging: Robust weighting and nlns deconvolution. In *American Astronomical Society Meeting Abstracts*, volume 187, pages 112–02.
- [Briggs, 1995b] Briggs, D. (1995b). High fidelity interferometric imaging: Robust weighting and nlns deconvolution. In *American Astronomical Society Meeting Abstracts*, volume 187, pages 112–02.
- [Buchert, 1994] Buchert, T. (1994). Robust morphological measures for large-scale structure.
- [Chardin et al., 2012] Chardin, J., Aubert, D., and Ocvirk, P. (2012). A study of simulated reionization histories with merger trees of HII regions. *Astronomy & Astrophysics*, 548:A9.
- [Chen, 2004] Chen, B. (2004). A simplified elementary proof of Hadwiger’s volume theorem. *Geometriae Dedicata*, 105:107–120.
- [Chernyaev, 1995] Chernyaev, E. (1995). Marching cubes 33: Construction of topologically correct isosurfaces. Technical report.

- [Choudhuri et al., 2020] Choudhuri, S., Ghosh, A., Roy, N., Bharadwaj, S., Intema, H. T., and Ali, S. S. (2020). All sky angular power spectrum: I. estimating brightness temperature fluctuations using tgss 150 mhz survey. *Monthly Notices of the Royal Astronomical Society*, 494:1936–1945.
- [Choudhury et al., 2016] Choudhury, T. R., Datta, K., Majumdar, S., Ghara, R., Paranjape, A., Mondal, R., Bharadwaj, S., and Samui, S. (2016). Modelling the 21-cm signal from the epoch of reionization and cosmic dawn.
- [Choudhury and Ferrara, 2006] Choudhury, T. R. and Ferrara, A. (2006). Updating reionization scenarios after recent data. *Monthly Notices of the Royal Astronomical Society: Letters*, 371:L55–L59.
- [Choudhury et al., 2009] Choudhury, T. R., Haehnelt, M. G., and Regan, J. (2009). Inside-out or outside-in: The topology of reionization in the photon-starved regime suggested by  $\text{Ly}\alpha$  forest data. *Monthly Notices of the Royal Astronomical Society*, 394:960–977.
- [Choudhury et al., 2000] Choudhury, T. R., Padmanabhan, T., and Srianand, R. (2000). Semi analytic approach to understanding the distribution of neutral hydrogen in the universe. *Monthly Notices of the Royal Astronomical Society*, 322:561–575.
- [Cooray et al., 2008] Cooray, A., Li, C., and Melchiorri, A. (2008). Trispectrum of 21-cm background anisotropies as a probe of primordial non-gaussianity. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 77:103506.
- [Cornwell, 2008] Cornwell, T. J. (2008). Multiscale clean deconvolution of radio synthesis images. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):793–801.
- [Crofton, 1868] Crofton, M. W. (1868). Vii. on the theory of local probability, applied to straight lines drawn at random in a plane; the methods used being also extended to the proof of certain new theorems in the integral calculus. *Philosophical Transactions of the Royal Society of London*, (158):181–199.

- [Datta et al., 2010] Datta, A., Bowman, J. D., and Carilli, C. L. (2010). Bright source subtraction requirements for redshifted 21cm measurements. *Astrophysical Journal*, 724:526–538.
- [Datta et al., 2016] Datta, K. K., Ghara, R., Majumdar, S., Choudhury, T. R., Bharadwaj, S., Roy, H., and Datta, A. (2016). Probing individual sources during reionization and cosmic dawn using square kilometre array hi 21-cm observations.
- [Davis et al., 1985] Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292:371.
- [Di Matteo et al., 2002] Di Matteo, T., Perna, R., Abel, T., and Rees, M. J. (2002). Radio foregrounds for the 21 centimeter tomography of the neutral intergalactic medium at high redshifts. *The Astrophysical Journal*, 564(2):576.
- [Dominus, 2010] Dominus, M. J. (2010). Short introduction to topology. *Topology*, 14:1–11.
- [Edelsbrunner and Harer, 2010] Edelsbrunner, H. and Harer, J. J. (2010). *Computational topology : an introduction*. American Mathematical Society.
- [Elbers and van de Weygaert, 2019] Elbers, W. and van de Weygaert, R. (2019). Persistent topology of the reionisation bubble network. i: Formalism & phenomenology. *Monthly Notices of the Royal Astronomical Society*, 486:1523–1538.
- [Fragos et al., 2013] Fragos, T., Lehmer, B. D., Naoz, S., Zezas, A., and Basu-Zych, A. (2013). Energy feedback from x-ray binaries in the early universe. *The Astrophysical Journal Letters*, 776(2):L31.
- [Friedrich et al., 2011] Friedrich, M. M., Mellema, G., Alvarez, M. A., Shapiro, P. R., and Iliev, I. T. (2011). Topology and sizes of h ii regions during cosmic reionization. *Monthly Notices of the Royal Astronomical Society*, 413:1353–1372.

- [Furlanetto et al., 2006] Furlanetto, S., Oh, S. P., and Briggs, F. (2006). Cosmology at low frequencies: The 21 cm transition and the high-redshift universe. *Physics Reports*, 433:181–301.
- [Furlanetto and Oh, 2016] Furlanetto, S. R. and Oh, S. P. (2016). Reionization through the lens of percolation theory. *Monthly Notices of the Royal Astronomical Society*, 457:1813–1827.
- [Furlanetto et al., 2004] Furlanetto, S. R., Zaldarriaga, M., and Hernquist, L. (2004). The growth of h ii regions during reionization. *The Astrophysical Journal*, 613:1–15.
- [Gardner et al., 2006] Gardner, J. P., Mather, J. C., Clampin, M., Doyon, R., Greenhouse, M. A., Hammel, H. B., Hutchings, J. B., Jakobsen, P., Lilly, S. J., Long, K. S., et al. (2006). The james webb space telescope. *Space Science Reviews*, 123:485–606.
- [Ghara et al., 2015] Ghara, R., Choudhury, T. R., and Datta, K. K. (2015). 21 cm signal from cosmic dawn: imprints of spin temperature fluctuations and peculiar velocities. *Monthly Notices of the Royal Astronomical Society*, 447(2):1806–1825.
- [Ghara et al., 2016] Ghara, R., Choudhury, T. R., Datta, K. K., and Choudhuri, S. (2016). Imaging the redshifted 21 cm pattern around the first sources during the cosmic dawn using the SKA. *Monthly Notices of the Royal Astronomical Society*, 464(2):2234–2248.
- [Giri and Mellema, 2021] Giri, S. K. and Mellema, G. (2021). Measuring the topology of reionization with betti numbers. *Monthly Notices of the Royal Astronomical Society*, 505:1863–1877.
- [Giri et al., 2018a] Giri, S. K., Mellema, G., Dixon, K. L., and Iliev, I. T. (2018a). Bubble size statistics during reionization from 21-cm tomography. *Monthly Notices of the Royal Astronomical Society*, 473:2949–2964.
- [Giri et al., 2018b] Giri, S. K., Mellema, G., and Ghara, R. (2018b). Optimal identification of h ii regions during reionization in 21-cm

observations. *Monthly Notices of the Royal Astronomical Society*, 479(4):5596–5611.

[Gnedin and Ostriker, 1997] Gnedin, N. Y. and Ostriker, J. P. (1997). Reionization of the universe and the early production of metals. *The Astrophysical Journal*, 486.

[Gorce et al., 2021] Gorce, A., Hutter, A., and Pritchard, J. R. (2021). Using the sample variance of 21 cm maps as a tracer of the ionisation topology. *Astronomy and Astrophysics*, 653:A58.

[Gunn and Peterson, 1965] Gunn, J. E. and Peterson, B. A. (1965). On the density of neutral hydrogen in intergalactic space. *The Astrophysical Journal*, 142.

[Guth, 1997] Guth, A. H. (1997). *The inflationary universe : the quest for a new theory of cosmic origins*. Addison-Wesley Pub.

[Hutter et al., 2020] Hutter, A., Watkinson, C. A., Seiler, J., Dayal, P., Sinha, M., and Croton, D. J. (2020). The 21 cm bispectrum during reionization: a tracer of the ionization topology. *Monthly Notices of the Royal Astronomical Society*, 492(1):653–667.

[Iheanetu et al., 2019] Iheanetu, K., Girard, J. N., Smirnov, O., Asad, K. M., Villiers, M. D., Thorat, K., Makhathini, S., and Perley, R. A. (2019). Primary beam effects of radio astronomy antennas - i. modelling the karl g. jansky very large array (vla) l-band beam using holography. *Monthly Notices of the Royal Astronomical Society*, 485:4107–4121.

[Iliev et al., 2014] Iliev, I. T., Mellema, G., Ahn, K., Shapiro, P. R., Mao, Y., and Pen, U.-L. (2014). Simulating cosmic reionization: how large a volume is large enough? *Monthly Notices of the Royal Astronomical Society*, 439(1):725–743.

[Iliev et al., 2006] Iliev, I. T., Mellema, G., Pen, U. L., Merz, H., Shapiro, P. R., and Alvarez, M. A. (2006). Simulating cosmic reionization at large scales - i. the geometry of reionization. *Monthly Notices of the Royal Astronomical Society*, 369:1625–1638.

- [Institute, 2012] Institute, J. B. (2012). Analysis of betti numbers and persistence diagrams of 2-dimensional gaussian random fields kapteyn astronomical institute .,
- [Jelić et al., 2014] Jelić, V., De Bruyn, A., Mevius, M., Abdalla, F., Asad, K., Bernardi, G., Brentjens, M., Bus, S., Chapman, E., Ciardi, B., et al. (2014). Initial lofar observations of epoch of reionization windows-ii. diffuse polarized emission in the elais-n1 field. *Astronomy & astrophysics*, 568:A101.
- [Jelić et al., 2008] Jelić, V., Zaroubi, S., Labropoulos, P., Thomas, R. M., Bernardi, G., Brentjens, M. A., De Bruyn, A., Ciardi, B., Harker, G., Koopmans, L. V., et al. (2008). Foreground simulations for the lofar–epoch of reionization experiment. *Monthly Notices of the Royal Astronomical Society*, 389(3):1319–1335.
- [Kaaret et al., 2011] Kaaret, P., Schmitt, J., and Gorski, M. (2011). X-rays from blue compact dwarf galaxies. *The Astrophysical Journal*, 741(1):10.
- [Kakiichi et al., 2017] Kakiichi, K., Majumdar, S., Mellema, G., Ciardi, B., Dixon, K. L., Iliev, I. T., Jelić, V., Koopmans, L. V., Zaroubi, S., and Busch, P. (2017). Recovering the hii region size statistics from 21-cm tomography. *Monthly Notices of the Royal Astronomical Society*, 471:1936–1954.
- [Kalirai, 2018] Kalirai, J. (2018). Scientific discovery with the james webb space telescope. *Contemporary Physics*, 59.
- [Kamran et al., 2021] Kamran, M., Majumdar, S., Ghara, R., Mellema, G., Bharadwaj, S., Pritchard, J. R., Mondal, R., and Iliev, I. T. (2021). Probing igm physics during cosmic dawn using the redshifted 21-cm bispectrum.
- [Kapahtia et al., 2019] Kapahtia, A., Chingangbam, P., and Appleby, S. (2019). Morphology of 21cm brightness temperature during the epoch of reionization using contour minkowski tensor. *Journal of Cosmology and Astroparticle Physics*, 2019:053.
- [Kim et al., 2016] Kim, S., Choi, S., Oh, E., Byun, J., Kim, H., Lee, B., Lee, S., and Hong, Y. (2016). Revisit to three-dimensional percolation theory:

Accurate analysis for highly stretchable conductive composite materials. *Scientific reports*, 6(1):1–10.

[Klypin and Shandarin, 1993] Klypin, A. and Shandarin, S. F. (1993). Percolation technique for galaxy clustering. *The Astrophysical Journal*, 413:48.

[Koenderink, 1984] Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception*, 13(3):321–330.

[Komatsu et al., 2009] Komatsu, E., Dunkley, J., Nolta, M., Bennett, C., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Limon, M., Page, L., et al. (2009). Five-year wilkinson microwave anisotropy probe\* observations: cosmological interpretation. *The Astrophysical Journal Supplement Series*, 180(2):330.

[Komatsu et al., 2011] Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2011). Seven-year wilkinson microwave anisotropy probe (wmap\*) observations: Cosmological interpretation. *Astrophysical Journal, Supplement Series*, 192:18.

[Koopmans et al., 2015] Koopmans, L., Pritchard, J., Mellema, G., Aguirre, J., Ahn, K., Barkana, R., van Bemmell, I., Bernardi, G., Bonaldi, A., Briggs, F., de Bruyn, A. G., Chang, T. C., Chapman, E., Chen, X., Ciardi, B., Dayal, P., Ferrara, A., Fialkov, A., Fiore, F., Ichiki, K., Illiev, I. T., Inoue, S., Jelic, V., Jones, M., Lazio, J., Maio, U., Majumdar, S., Mack, K. J., Mesinger, A., Morales, M. F., Parsons, A., Pen, U. L., Santos, M., Schneider, R., Semelin, B., de Souza, R. S., Subrahmanyam, R., Takeuchi, T., Vedantham, H., Wagg, J., Webster, R., Wyithe, S., Datta, K. K., and Trott, C. (2015). The Cosmic Dawn and Epoch of Reionisation with SKA. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 1.

[Kuhlen and Faucher-Giguère, 2012] Kuhlen, M. and Faucher-Giguère, C.-A. (2012). Concordance models of reionization: implications for faint

galaxies and escape fraction evolution. *Monthly Notices of the Royal Astronomical Society*, 423(1):862–876.

[Kuronya, 2010] Kuronya, A. (2010). Introduction to topology - google books.

[Lewis, 2011] Lewis, A. (2011). The real shape of non-gaussianities. *Journal of Cosmology and Astroparticle Physics*, 2011(10):026.

[Liu, 2008] Liu, L. (2008). The hydrogen 21-cm line and its applications to radio astrophysics.

[Loeb, 2010] Loeb, A. (2010). *How Did the First Stars and Galaxies Form?*

[Loeb and Furlanetto, 2013] Loeb, A. and Furlanetto, S. R. (2013). *The first galaxies in the universe*.

[Lorensen and Cline, 1987] Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3d surface construction algorithm. *ACM siggraph computer graphics*, 21(4):163–169.

[Majumdar et al., 2016] Majumdar, S., Jensen, H., Mellema, G., Chapman, E., Abdalla, F. B., Lee, K.-Y., Iliev, I. T., Dixon, K. L., Datta, K. K., Ciardi, B., et al. (2016). Effects of the sources of reionization on 21-cm redshift-space distortions. *Monthly Notices of the Royal Astronomical Society*, 456(2):2080–2094.

[Majumdar et al., 2020] Majumdar, S., Kamran, M., Pritchard, J. R., Mondal, R., Mazumdar, A., Bharadwaj, S., and Mellema, G. (2020). Redshifted 21-cm bispectrum—i. impact of the redshift space distortions on the signal from the epoch of reionization. *Monthly Notices of the Royal Astronomical Society*, 499(4):5090–5106.

[Majumdar et al., 2014] Majumdar, S., Mellema, G., Datta, K. K., Jensen, H., Choudhury, T. R., Bharadwaj, S., and Friedrich, M. M. (2014). On the use of seminumerical simulations in predicting the 21-cm signal from the epoch of reionization. *Monthly Notices of the Royal Astronomical Society*, 443(4):2843–2861.

- [Majumdar et al., 2018] Majumdar, S., Pritchard, J. R., Mondal, R., Watkinson, C. A., Bharadwaj, S., and Mellema, G. (2018). Quantifying the non-gaussianity in the eor 21-cm signal through bispectrum. *Monthly Notices of the Royal Astronomical Society*, 476(3):4007–4024.
- [McQuinn, 2012] McQuinn, M. (2012). Constraints on x-ray emissions from the reionization era. *Monthly Notices of the Royal Astronomical Society*, 426(2):1349–1360.
- [McQuinn et al., 2006] McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., and Furlanetto, S. R. (2006). Cosmological parameter estimation using 21 cm radiation from the epoch of reionization. *The Astrophysical Journal*, 653(2):815.
- [Mecke et al., 1993] Mecke, K. R., Buchert, T., and Wagner, H. (1993). Robust morphological measures for large-scale structure in the universe.
- [Mellema et al., 2006] Mellema, G., Iliev, I. T., Pen, U.-L., and Shapiro, P. R. (2006). Simulating cosmic reionization at large scales—ii. the 21-cm emission features and statistical signals. *Monthly notices of the royal astronomical society*, 372(2):679–692.
- [Mellema et al., 2015] Mellema, G., Koopmans, L., Shukla, H., Datta, K. K., Mesinger, A., and Majumdar, S. (2015). HI tomographic imaging of the Cosmic Dawn and Epoch of Reionization with SKA. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 10.
- [Mertens et al., 2020] Mertens, F. G., Mevius, M., Koopmans, L. V. E., Offringa, A. R., Mellema, G., Zaroubi, S., Brentjens, M. A., Gan, H., Gehlot, B. K., Pandey, V. N., Sardarabadi, A. M., Vedantham, H. K., Yatawatta, S., Asad, K. M. B., Ciardi, B., Chapman, E., Gazagnes, S., Ghara, R., Ghosh, A., Giri, S. K., Iliev, I. T., Jelić, V., Kooistra, R., Mondal, R., Schaye, J., and Silva, M. B. (2020). Improved upper limits on the 21-cm signal power spectrum of neutral hydrogen at  $z \approx 9.1$  from LOFAR. *Monthly Notices of the Royal Astronomical Society*, 493(2):1662–1685.

- [Merz et al., 2005] Merz, H., Pen, U.-L., and Trac, H. (2005). Towards optimal parallel pm n-body codes: Pmfast. *New Astronomy*, 10(5):393–407.
- [Mesinger and Furlanetto, 2007] Mesinger, A. and Furlanetto, S. (2007). Efficient simulations of early structure formation and reionization. *The Astrophysical Journal*, 669(2):663.
- [Mesinger et al., 2011] Mesinger, A., Furlanetto, S., and Cen, R. (2011). 21cmfast: a fast, seminumerical simulation of the high-redshift 21-cm signal. *Monthly Notices of the Royal Astronomical Society*, 411(2):955–972.
- [Milnor, 1969] Milnor, J. (1969). *Morse Theory. (AM-51), Volume 51*. Princeton University Press.
- [Mitra et al., 2015] Mitra, S., Choudhury, T. R., and Ferrara, A. (2015). Cosmic reionization after planck. *Monthly Notices of the Royal Astronomical Society: Letters*, 454(1):L76–L80.
- [Mondal et al., 2016] Mondal, R., Bharadwaj, S., and Majumdar, S. (2016). Statistics of the epoch of reionization 21-cm signal–i. power spectrum error-covariance. *Monthly Notices of the Royal Astronomical Society*, 456(2):1936–1947.
- [Morales, 2005] Morales, M. F. (2005). Power spectrum sensitivity and the design of epoch of reionization observatories. *The Astrophysical Journal*, 619(2):678.
- [Mort et al., 2010] Mort, B. J., Dulwich, F., Salvini, S., Adami, K. Z., and Jones, M. E. (2010). Oskar: Simulating digital beamforming for the ska aperture array. In *2010 IEEE International Symposium on Phased Array Systems and Technology*, pages 690–694. IEEE.
- [Murphy, 2016] Murphy, M. (2016). Definition > lyman-alpha forest.
- [Pathak et al., 2022] Pathak, A., Bag, S., Majumdar, S., Mondal, R., Kamran, M., and Sarkar, P. (2022). Distinguishing reionization models using the largest cluster statistics of the 21-cm maps.

- [Poole et al., 2016] Poole, G. B., Angel, P. W., Mutch, S. J., Power, C., Duffy, A. R., Geil, P. M., Mesinger, A., and Wyithe, S. B. (2016). Dark-ages reionization and galaxy formation simulation—i. the dynamical lives of high-redshift galaxies. *Monthly Notices of the Royal Astronomical Society*, 459(3):3025–3039.
- [Pranav et al., 2016] Pranav, P., Edelsbrunner, H., van de Weygaert, R., Vegter, G., Kerber, M., Jones, B. J. T., and Wintraecken, M. (2016). The topology of the cosmic web in terms of persistent betti numbers. *Monthly Notices of the Royal Astronomical Society*, 465:4281–4310.
- [Press and Schechter, 1974] Press, W. H. and Schechter, P. (1974). Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *The Astrophysical Journal*, 187:425.
- [Pritchard and Loeb, 2012] Pritchard, J. R. and Loeb, A. (2012). 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901.
- [Qin et al., 2020] Qin, Y., Poulin, V., Mesinger, A., Greig, B., Murray, S., and Park, J. (2020). Reionization inference from the cmb optical depth and e-mode polarization power spectra. *Monthly Notices of the Royal Astronomical Society*, 499.
- [Rauch, 1998] Rauch, M. (1998). The lyman alpha forest in the spectra of qsos. *arXiv.org*, astro-ph.
- [Ricotti et al., 2002] Ricotti, M., Gnedin, N. Y., and Shull, J. M. (2002). The fate of the first galaxies. i. self-consistent cosmological simulations with radiative transfer. *The Astrophysical Journal*, 575(1):33.
- [Rote and Vegter, 2006] Rote, G. and Vegter, G. (2006). Computational topology: An introduction.
- [Ryden, 2003] Ryden, B. (2003). *Introduction to cosmology*.
- [Sagan, 1997] Sagan, C. (1997). Pale blue dot.

- [Sahni et al., 1998] Sahni, V., Sathyaprakash, B. S., and Shandarin, S. F. (1998). Shapefinders: A new shape diagnostic for large-scale structure. *The Astrophysical Journal*, 495:L5–L8.
- [Sanders, 2013] Sanders, G. H. (2013). The thirty meter telescope (tmt): An international observatory. *Journal of Astrophysics and Astronomy*, 34:81–86.
- [Schmalzing and Buchert, 1997] Schmalzing, J. and Buchert, T. (1997). Beyond genus statistics: a unifying approach to the morphology of cosmic structure. *The Astrophysical Journal*, 482(1):L1.
- [Shaw et al., 2019] Shaw, A. K., Bharadwaj, S., and Mondal, R. (2019). The impact of non-gaussianity on the error covariance for observations of the epoch of reionization 21-cm power spectrum. *Monthly Notices of the Royal Astronomical Society*, 487(4):4951–4964.
- [Sheth, 2005] Sheth, J. V. (2005). Exploring the geometry, topology and morphology of large scale structure using minkowski functionals. *Current Science*, pages 1101–1116.
- [Sheth et al., 2003] Sheth, J. V., Sahni, V., Shandarin, S. F., and Sathyaprakash, B. S. (2003). Measuring the geometry and topology of large-scale structure using surfgen: methodology and preliminary results. *Monthly Notices of the Royal Astronomical Society*, 343(1):22–46.
- [Sheth and Tormen, 1999] Sheth, R. K. and Tormen, G. (1999). Large scale bias and the peak background split. *Monthly Notices of the Royal Astronomical Society*, 308:119–126.
- [SNELL, 2021] SNELL, R. L. K. (2021). *FUNDAMENTALS OF RADIO ASTRONOMY : astrophysics*. CRC PRESS.
- [Soille et al., 1999] Soille, P. et al. (1999). *Morphological image analysis: principles and applications*, volume 2. Springer.
- [Sokolowski et al., 2015] Sokolowski, M., Wayth, R. B., Tremblay, S. E., Tingay, S. J., Waterson, M. F., Tickner, J., Emrich, D., Schlagenhauser, F., Kenney, D., and Padhi, S. K. (2015). The impact of the ionosphere on

- ground-based detection of the global epoch of reionization signal. *The Astrophysical Journal*, 813:18.
- [Songaila and Cowie, 2010] Songaila, A. and Cowie, L. L. (2010). The evolution of Lyman limit absorption systems to redshift six. *The Astrophysical Journal*, 721(2):1448.
- [Swarup, 1991] Swarup, G. (1991). Giant metrewave radio telescope (gmrt). *International Astronomical Union Colloquium*, 19:376–380.
- [Thomas and Zaroubi, 2008] Thomas, R. M. and Zaroubi, S. (2008). Time-evolution of ionization and heating around first stars and miniquasars. *Monthly Notices of the Royal Astronomical Society*, 384(3):1080–1096.
- [Thomas et al., 2009] Thomas, R. M., Zaroubi, S., Ciardi, B., Pawlik, A. H., Labropoulos, P., Jelić, V., Bernardi, G., Brentjens, M. A., De Bruyn, A., Harker, G. J., et al. (2009). Fast large-scale reionization simulations. *Monthly Notices of the Royal Astronomical Society*, 393(1):32–48.
- [Thompson et al., 2017a] Thompson, A. R., Moran, J. M., and Swenson, G. W. (2017a). *Interferometry and synthesis in radio astronomy*. Springer Nature.
- [Thompson et al., 2017b] Thompson, A. R., Moran, J. M., and Swenson, G. W. (2017b). *Van Cittert–Zernike Theorem, Spatial Coherence, and Scattering*, pages 767–786. Springer International Publishing, Cham.
- [Tiwari et al., 2022] Tiwari, H., Shaw, A. K., Majumdar, S., Kamran, M., and Choudhury, M. (2022). Improving constraints on the reionization parameters using 21-cm bispectrum. *Journal of Cosmology and Astroparticle Physics*, 2022(04):045.
- [van Haarlem et al., 2013a] van Haarlem, M. P., Wise, M. W., Gunst, A., Heald, G., McKean, J. P., Hessels, J. W., de Bruyn, A. G., Nijboer, R., Swinbank, J., Fallows, R., et al. (2013a). Lofar: The low-frequency array. *Astronomy & astrophysics*, 556:A2.
- [van Haarlem et al., 2013b] van Haarlem, M. P., Wise, M. W., Gunst, A., Heald, G., McKean, J. P., Hessels, J. W., de Bruyn, A. G., Nijboer, R.,

- Swinbank, J., Fallows, R., et al. (2013b). Lofar: The low-frequency array. *Astronomy & astrophysics*, 556:A2.
- [Watkinson et al., 2021] Watkinson, C. A., Trott, C. M., and Hothi, I. (2021). The bispectrum and 21-cm foregrounds during the epoch of reionization. *Monthly Notices of the Royal Astronomical Society*, 501.
- [Yamauchi, 2022] Yamauchi, D. (2022). Signature of primordial non-gaussianity on 21-cm power spectrum from dark ages.
- [Yatawatta, 2014a] Yatawatta, S. (2014a). Adaptive weighting in radio interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 444(1):790–796.
- [Yatawatta, 2014b] Yatawatta, S. (2014b). Adaptive weighting in radio interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 444(1):790–796.
- [Yatawatta et al., 2013] Yatawatta, S., De Bruyn, A., Brentjens, M. A., Labropoulos, P., Pandey, V., Kazemi, S., Zaroubi, S., Koopmans, L., Offringa, A., Jelić, V., et al. (2013). Initial deep lofar observations of epoch of reionization windows-i. the north celestial pole. *Astronomy & Astrophysics*, 550:A136.
- [Yau, 2008] Yau, S.-T. (2008). A survey of calabi-yau manifolds. *Surveys in Differential Geometry*, 13:277–318.
- [Yess et al., 1997] Yess, C., Shandarin, S. F., and Fisher, K. B. (1997). Percolation analysis of a wiener reconstruction of the iras 1.2 jy redshift catalog. *The Astrophysical Journal*, 474(2):553.
- [Zahn et al., 2007] Zahn, O., Lidz, A., McQuinn, M., Dutta, S., Hernquist, L., Zaldarriaga, M., and Furlanetto, S. R. (2007). Simulations and analytic calculations of bubble growth during hydrogen reionization. *The Astrophysical Journal*, 654:12–26.