

School of Physics and Astronomy  
Queen Mary University of London

# **Hydrogen Epoch of Reionization Array Experiment: Effect of Radio Frequency Interference Excision on 21-cm Power Spectrum**

Ankur Dev (160244199)

23<sup>rd</sup> April 2020

Supervisor: Dr. Phil Bull

SPA7016U Physics Research Project  
45 Credit Units

Submitted in partial fulfilment of the requirements for the degree of  
Master in Science (Programme: MSci Astrophysics) from Queen Mary University  
of London

# **Declaration**

I hereby certify that this project report, which is approximately 11500 words in length, has been written by me at the School of Physics and Astronomy, Queen Mary University of London, that all material in this dissertation which is not my own work has been properly acknowledged, and that it has not been submitted in any previous application for a degree.

Ankur Dev (160244199)

# Acknowledgements

I would like to thank my supervisor Dr. Phil Bull for his technical guidance and support throughout the project. The QMUL library resources have been helpful in completing this project. I want to thank Dr. Samir Choudhuri for his assistance in setting up Python environments and helping me understand the basics of radio astronomy. I also thank the entire HERA research group at School of Physics and Astronomy for the weekly meetings which helped me to develop an overall knowledge in the subject. I acknowledge the constructive feedback from my peers which has been of immense help. In particular, I thank my friend Beata Drabik for helping me to proofread the text.

# Abstract

The Epoch of Reionization (EoR) is an era in the history of the Universe when neutral Hydrogen was ionized by radiation from the first stars and galaxies. Experiments such as the Hydrogen Epoch of Reionization Array (HERA) are dedicated to measure redshifted 21-cm emission of neutral Hydrogen from this epoch. One of the biggest challenges of the experiment is mitigating foreground radiation from our Galaxy which is  $\sim 5 - 6$  orders of magnitude brighter than the EoR signal. Primary objective of HERA is to measure the statistical power spectrum of the EoR signal rather than direct imaging. However, 21-cm power spectrum can be affected by fluctuations in visibilities due to sporadic removal of radio frequency interference (RFI) flagged data samples. RFI is any source of radio-frequency transmission other than the sky signal itself. In this project, we investigate the effect of excising RFI on the 21-cm power spectrum within the context of HERA experiment. We analyse techniques to mitigate the contaminating effects of interference which may affect EoR signal detection. In particular, we use a delay filter based on linear least-squares solver to fit a smooth model to flagged datasets. We present results of statistical power spectrum analysis for various simulated test cases which are relevant to the HERA experiment. We find that the RFI mitigation method used in this project works well for cases where fraction of RFI flagged frequencies is less than 15% and that excess power increases as number of flagged channels increase.

# Contents

<b>List of Figures</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Brief history of the Universe . . . . .	1
1.2 Open Questions . . . . .	4
<b>2 Epoch of Reionization: Theory and Observation</b>	<b>7</b>
2.1 Using 21-cm line to probe the EoR . . . . .	7
2.1.1 21-cm Observation Experiments . . . . .	9
2.2 Challenges in measuring Epoch of Reionization signal . . . . .	9
2.2.1 Foreground . . . . .	9
2.2.2 Radio Frequency Interference . . . . .	11
2.3 Basics of Interferometry . . . . .	12
2.4 Hydrogen Epoch of Reionization Array (HERA) . . . . .	14
<b>3 RFI Mitigation: Simulation Approach</b>	<b>17</b>
3.1 Problem Statement . . . . .	17
3.2 Methodology . . . . .	18
3.3 Simulation Set-up . . . . .	20
<b>4 Results and Conclusion</b>	<b>21</b>
4.1 Simulation Plots . . . . .	21
4.2 Conclusion . . . . .	35
<b>Bibliography</b>	<b>37</b>

# List of Figures

1.1	A schematic outline of the cosmic history . . . . .	3
1.2	Cosmic archaeology of the observable volume of the Universe . . . . .	4
2.1	The 21-centimeter cosmic hydrogen signal . . . . .	8
2.2	HERA: EoR Window and “Wedge” . . . . .	10
2.3	RFI flags detected from a LOFAR observation . . . . .	11
2.4	Two element interferometer . . . . .	12
2.5	Geometric representation of <i>uv</i> -plane . . . . .	13
2.6	HERA Radio Telescope . . . . .	14
2.7	Single HERA Element . . . . .	15
4.1	Simulated noise for 1 LST bin as a function of frequency . . . . .	22
4.2	Simulated EoR signal for 1 LST bin as a function of frequency . . . . .	22
4.3	Waterfall plot of simulated EoR visibilities . . . . .	23
4.4	Ideal sky data including foregrounds, noise and EOR signal without RFI flags . . . . .	23
4.5	Delay power spectrum calculated for un-flagged data . . . . .	24
4.6	Waterfall plot of visibilities after adding RFI flags . . . . .	24
4.7	Comparison of visibilities of flagged and un-flagged data for 1 LST bin . . . . .	25
4.8	Best-fit model (for 1 LST bin) obtained from the in-painting process .	26
4.9	Comparison of visibilities of in-painted data with flagged and un-flagged data . . . . .	26
4.10	Waterfall plot of visibilities of in-painted data . . . . .	27
4.11	Difference of the best-fit model and the un-flagged data for one LST bin . . . . .	28
4.12	Average over time for difference of the best-fit model and the un-flagged data . . . . .	28
4.13	Standard deviation of noise residual obtained from difference of the best-fit model and the un-flagged data . . . . .	29
4.14	Waterfall plot of difference of smooth model and un-flagged data . . . . .	29

4.15	Difference of the in-painted data and un-flagged data for one LST bin	30
4.16	Waterfall plot of difference of in-painted data and un-flagged data . . . . .	31
4.17	Simulation without noise: Difference of the in-painted data and un-flagged data for one LST bin . . . . .	32
4.18	Simulation without noise: Delay power spectrum of in-painted and un-flagged data . . . . .	32
4.19	Simulation with noise: Delay power spectrum of in-painted and un-flagged data . . . . .	33
4.20	Simulation with RFI flagged fraction at 15%: Delay power spectrum of in-painted and un-flagged data . . . . .	34
4.21	Simulation with RFI flagged fraction at 30%: Delay power spectrum of in-painted and un-flagged data . . . . .	34



# 1 Introduction

Cosmology is the science of understanding the evolution and origin of the Universe and aims to understand the large scale structure of the Universe. Physical cosmology is based on two principles: the Universe is homogeneous and isotropic, which means that the Universe appears homogeneous at large scales ( $\sim 100$  Mpc) and all directions are same (Liddle 2003). About a century ago, not much was known about the distribution of galaxies and matter. In fact, knowledge of the local Universe was very limited until early 1900s. Over the past hundred years, astronomers have peered into the sky using telescopes in almost the entire electromagnetic spectra from gamma rays to radio waves. Electromagnetic waves travel at a finite speed which means that by observing distant objects in the sky, astronomers can understand different epochs of the Universe. In 1929, Edwin Hubble (Hubble 1929) studied the distances and velocities of extragalactic sources and showed that galaxies were receding away from us with a velocity proportional to their distances. This came to be known as the *Hubble's Law*. Almost all galaxies are moving away from us and this is an important foundation of observational cosmology. The emitted wavelength from astronomical sources gets stretched due to expansion of the Universe and this forms the basis of *cosmological redshift* denoted hereafter by  $z$ .

In the next sections, a brief overview of the different epochs in the evolution of the Universe (section 1.1) is presented followed by a review of the unanswered questions related to the earliest epochs of the Universe in section 1.2. In chapter 2 we describe the theory as well as the practical aspects involved in observing the Epoch of Reionization, while chapter 3 delves into the simulation approach and methodology of this project. Finally, the results are shown in chapter 4 followed by a conclusion.

## 1.1 Brief history of the Universe

The expanding Universe tells us that in the distant past all matter and radiation had a common origin. The well accepted model of the evolution of the early Universe

is known as the *Big Bang model* which states that the Universe originated from a highly dense and hot state. A review of the model is presented in [Uzan \(2016\)](#). In the earliest phases after the Big Bang, the Universe expanded exponentially caused by the *cosmic inflation*. At approximately  $10^{-6}$  seconds after the Big Bang, baryons such as protons and neutrons formed ([Mathews et al. 2017](#); [Tytler et al. 2000](#)). The Universe was radiation dominated at this point. The cosmic soup consisted of protons, neutrons, electrons and light nuclei and the Universe was completely ionized. This epoch was opaque to electromagnetic radiation because the photons were scattered by free electrons and had short mean free path - a process known as *Thomson scattering*.

The Universe cooled down as it expanded and around 400,000 years after the Big Bang, the Universe cooled down enough to allow slowly moving protons and electrons to combine and form neutral atomic hydrogen (H<sub>I</sub>) ([Loeb 2013](#)). This is referred to as the *epoch of recombination*<sup>1</sup> and occurred at a redshift of around  $z \approx 1100$ . This allowed the photons to decouple from matter and travel freely through space. The radiation from this epoch has been redshifted to microwave wavelengths due to cosmic expansion and we can observe it today as the *cosmic microwave background* (CMB). CMB is the relic radiation that was discovered in 1964 accidentally by Arno Penzias and Robert Wilson at Bell Telephone Laboratories and forms a key piece of observational evidence for cosmological models and the Big Bang model ([Penzias and Wilson 1965](#)). CMB has an almost perfect black body spectrum as predicted by the Big Bang model. Recent observations from the Wilkinson Microwave Anisotropy Probe (WMAP) and Planck satellite show that the temperature fluctuation in the CMB is extremely small, in the order of  $\sim 10^{-5}$  and nearly same in all directions. The Universe was homogeneous and smooth in its early phases and was in thermal equilibrium.

During the epoch of recombination, the Universe is matter dominated and the small density perturbations in the early Universe eventually get amplified due to gravity. As the Universe continues to evolve and expand, the overdense regions collapsed to form the first galaxies and clusters which led to the large scale structures that we observe in the present day Universe. The distribution of galaxies on the large scale have the same statistical properties as found in the CMB ([Loeb 2013](#)). This is an evidence of the theory that the lumpy distribution of galaxies on smaller scales observed today evolved out of early Universe density perturbations.

Shortly after the epoch of recombination, the intergalactic medium (hereafter,

---

<sup>1</sup>Recombination is a misnomer as electrons and protons combined for the first time in cosmic history.

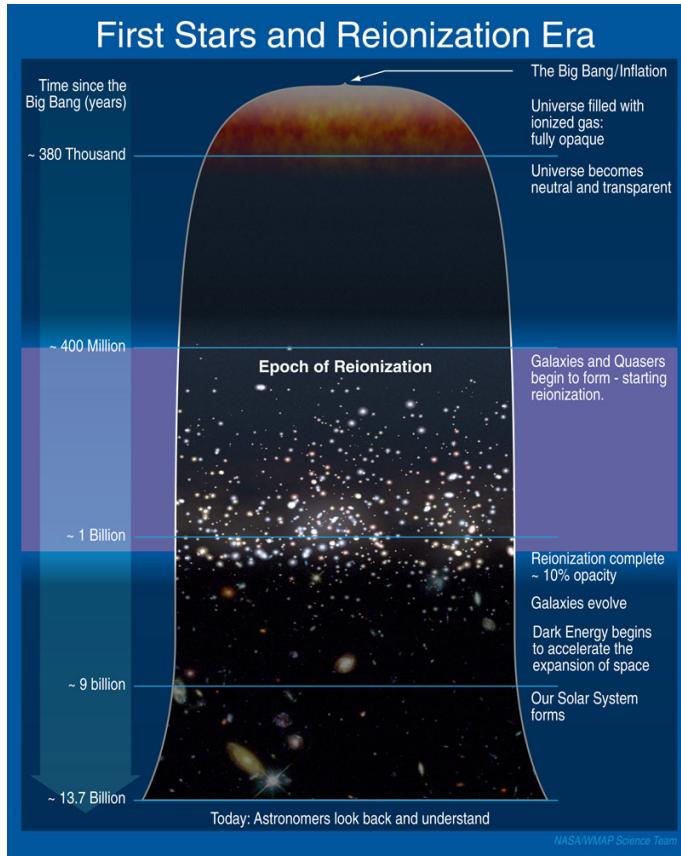


Figure 1.1: A schematic outline of the cosmic history is shown here. The different epochs of evolution of the Universe are shown along with the time elapsed since Big Bang. Around 400 million years after the Big Bang, the Cosmic Dawn begins marked by the Epoch of Reionization. (Image Credit: NASA/WMAP Science Team, [Astrophysics Science Division, NASA/GSFC \(2018\)](#))

IGM) was filled with neutral hydrogen ( $\text{H}\text{I}$ ). Now the Universe became neutral and transparent to electromagnetic radiation. This era is referred to as the *dark ages* because there were no light sources in the Universe yet other than the photons from the CMB. The dark ages lasted for about 400 million years until the first galaxies appeared ([Zaroubi 2013](#)).

Fortunately,  $\text{H}\text{I}$  has a characteristic spectral line emission with a wavelength of 21 cm and this radiation of neutral Hydrogen is referred to as the *21-cm line* in astronomy. Astronomers can use the 21-cm line of  $\text{H}\text{I}$  to study the early universe and it is used as a cosmological probe to understand these epochs. The earliest galaxies, low metallicity stars and quasars started emitting high energy UV photons which changed the entire IGM landscape eventually. The IGM that was completely neutral during the dark ages slowly started getting ionized. Pockets of ionized regions filled with ionized hydrogen ( $\text{H}\text{II}$ ) started appearing around the first luminous sources.

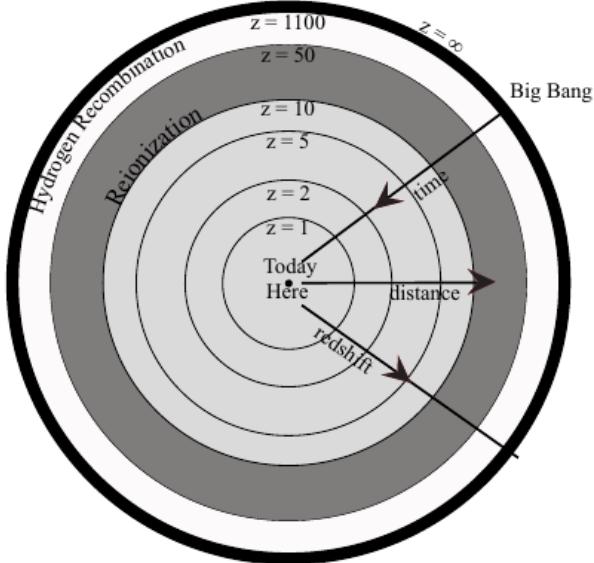


Figure 1.2: Cosmic archaeology of the observable volume of the Universe is shown at respective redshifts. The grey area marks the Epoch of Reionization when the first luminous sources started to reionize the neutral Hydrogen in the IGM. Adapted from: [Loeb \(2013\)](#)

This period when the IGM became reionized from being completely neutral is known as the *Epoch of Reionization* (EoR) ([Pritchard and Loeb 2012](#); [Zaroubi 2013](#)).

EoR is one of the most important periods in the evolution of the Universe because the first luminous sources appeared and marked the end of the dark ages. We also called this period as the beginning of the *Cosmic Dawn* because the earliest light sources started appearing now. [Figure 1.1](#) and [Figure 1.2](#) show the different epochs of the Universe as described above. Prior to EoR, large scale structure formation was dominated by dark matter alone. However, after this era of transition baryonic matter played an crucial role as well. Although a trove of cosmological and astrophysical information remains buried in this epoch, it has not yet been imaged directly because of the requirements of high sensitivity and resolution. Understanding EoR is key to answering the fundamental questions regarding large scale structure formation as well as evolution of galaxies.

## 1.2 Open Questions

In the past decade, numerical simulations and high-redshift experiments such as [Ali et al. \(2015\)](#) and [Bowman et al. \(2018\)](#) have started to constrain reionization.

Although galaxy formation in the low redshift universe has been studied extensively ([Benson 2010](#); [Conselice 2012](#)), the high-redshift regime is practically untouched. Several questions remain unanswered regarding EoR and properties of IGM during this transition period. In particular, it is not clear what sources drove the reionization. It is theoretically understood that UV radiation from stars inside small protogalaxies reionized the surrounding neutral IGM, producing bubbles of HII regions which grew until the entire Universe was filled with ionized hydrogen ([Furlanetto et al. 2006](#)). Observational evidence for this process comes from the discovery of Gunn-Peterson trough in a quasar at  $z = 6.28$  ([Becker et al. 2001](#)). The presence of HI in the IGM creates the Gunn-Peterson trough, which is a feature in the spectra of quasars and was predicted in [Gunn and Peterson \(1965\)](#). The presence of HI trough in the  $z = 6.28$  quasar and the missing trough in spectra of quasars at  $z < 6$  shows that the Hydrogen was ionised around  $z = 6$ . However, a unified model that takes into account realistic small scale structure and reionization is yet to be developed.

In the hierarchical model of galaxy formation, it is believed that small halos and galaxies formed first which merged to form bigger galaxies and clusters. This process would have been too slow and possibly feedback mechanisms played an important role in early structure formation during the era of cosmic dawn. It is also unclear what role did supermassive black holes play during EoR and when did they first appear. The first stars (Population III) were probably massive and hot stars with negligible metal content ([Ferrara 2018](#)). When did the first heavy metals appear and what processes drove the chemical enrichment of galaxies and IGM? Was the IGM lumpy at these epochs ([Furlanetto et al. 2006](#))? These questions still need to be answered by future surveys and experiments.

The earliest generations of cosmic structures are the seeds of the underlying distribution of baryonic and dark matter that we see today and they probably had profound effects on the evolution of the Universe at later stages. Therefore, efforts to study the global reionization history are urgently needed. Redshifts  $z > 6$  must be studied using the 21-cm signal as well as synergies with other observational probes like Lyman- $\alpha$  forest <sup>2</sup> ([Ouchi et al. 2009](#); [Zaroubi 2013](#)), Gamma-ray bursts ([Bromm 2006](#)), high-redshift metal abundance ([Rudie et al. 2012](#)) and intensity mapping of CII emission ([Yue et al. 2015](#)) are important. James Webb Space Telescope (JWST) to be launched in 2021 will provide researchers with spectra from galaxies formed during EoR ([Wang et al. 2019](#)).

---

<sup>2</sup>An absorption feature found in the spectra of quasi-stellar objects



## 2 Epoch of Reionization: Theory and Observation

Epoch of Reionization (hereafter, EoR) is one of the unexplored periods in the history of evolution of the Universe. As stated in section 1.1, during this epoch, the first light sources appeared and reionized the IGM. There are several processes that affect the reionization history and studying the electromagnetic signal from EoR can inform us about properties of the IGM from this frontier. During EoR, ionized bubbles of HII first started to appear around the densest regions and the voids became ionized later towards the end of this period of transition. By studying the morphology of the ionization state of the IGM, we can learn about how the dark matter is clustered as well as where the ionising sources are and when do they first appear in the evolution of the Universe.

### 2.1 Using 21-cm line to probe the EoR

In the past decade, it has become clear that the 21-cm line of HI is an important cosmological probe and can be used to map the density fluctuations. The 21-cm line has some great advantages over other methods: it can be directly used to probe the IGM that contains a large fraction of baryonic matter and as a redshifted spectral line it can be used to extract the three-dimensional history of the evolution. Observing the sky in a bandwidth gives the average over cosmic volume.

The 21-cm radiation is produced due to the difference in energy in the two hyperfine levels of Hydrogen. The excited state has parallel spins whereas the lower energy level has anti-parallel spins. The transition from the higher energy state to the lower is called *spin-flip transition* and has an extremely low probability of occurring. Despite it being a forbidden line, 21-cm line exists because of the huge amount of Hydrogen in the Universe.

The fundamental quantity that is used to measure 21-cm signal is *brightness temperature*,  $T_b(\nu)$  for observing at *frequency*  $\nu$ . The rest frequency ( $\nu_0$ ) for hyperfine

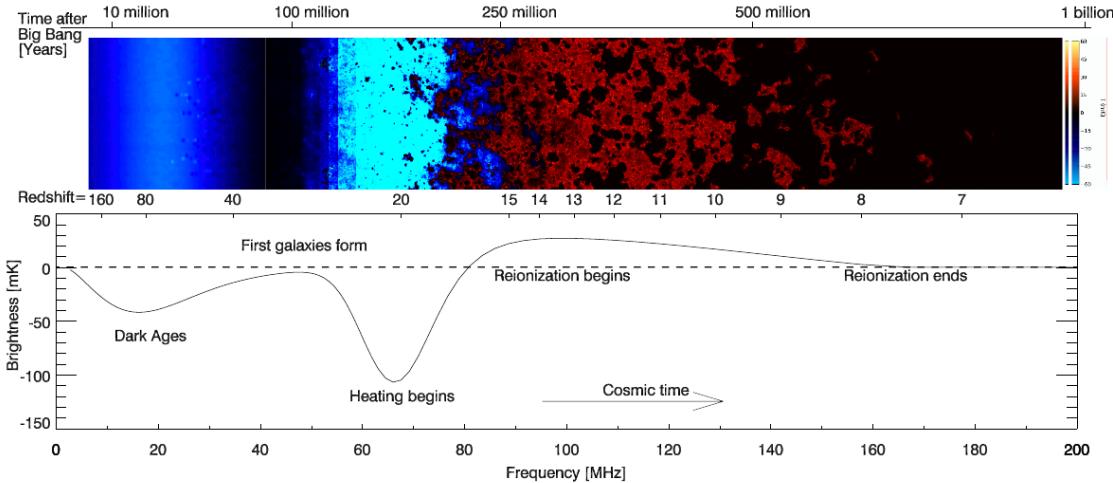


Figure 2.1: The upper panel shows the evolution of intergalactic Hydrogen as a function of redshift. The red colour represents the 21-cm radiation relative to the CMB and the black regions show ionization of Hydrogen. Around redshift  $z \approx 15$ , pockets of ionized Hydrogen start to appear around the first galaxies and stars. Gradually these regions grow and the IGM is almost completely ionized at  $z \approx 7$ . The expected global 21-cm signal, shown here as a sky-averaged spectrum (solid black line), from the dark ages till end of reionization is presented in the lower panel. (Adapted from: [Pritchard and Loeb \(2012\)](#))

splitting of the ground state of the Hydrogen atom is known accurately and measured to be  $1\,420\,405\,751.7667 \pm 0.0009$  Hz ([Dupays et al. 2003](#)). Brightness temperature is measured relative to a background source, usually CMB. The observable quantity is given by  $\delta T_b$  which is the brightness temperature of 21-cm line measured against the CMB ([Mesinger et al. 2011](#)):

$$\delta T_b = \frac{T_s - T_\gamma}{1 + z} (1 - e^{-\tau_{\nu_0}}) \quad (2.1)$$

where  $T_s$  is the gas spin temperature,  $T_\gamma$  is the CMB temperature, rest frequency is given by  $\nu_0$  and  $\tau_{\nu_0}$  is the optical depth at  $\nu_0$ . Throughout the evolution of the Universe, the 21-cm line is affected by a number of different astrophysical phenomena at various epochs. [Figure 2.1](#) shows the 21-cm signal at different epochs as well as shows a simulation of how the HI gets ionised during EoR. The HI gas temperature is close to the CMB temperature during the EoR as a result of Compton scattering off electrons. During the dark ages, the gas temperature is quite different from the CMB temperature and it depends on the large scale evolution of fluctuations. During the cosmic dawn, the Lyman- $\alpha$  photons and x-rays from the first luminous sources change the mean gas temperature substantially. Finally during the later

stages of EoR, HII regions grow due to ionizing radiation from these sources that leads to a drop in the mean signal reaching at  $z \approx 6$ . This makes the 21-cm line the ultimate cosmological probe to study a range of redshifts. These astrophysical effects have been studied using semi-analytical models widely. The 21-cm signal as a function of redshift for 193 different models is presented in [Cohen et al. \(2017\)](#).

### 2.1.1 21-cm Observation Experiments

Since the past decade or so, a number of international scientific groups are attempting to detect the 21-cm signal and trace the evolution of  $\delta T_b$ . There has been significant progress in the past few years due to advancements in technology and computing facilities. Some recent experiments targetting 21-cm detection are low frequency arrays like LOFAR ([Patil et al. 2017](#)), MWA ([Dillon et al. 2015](#)), PAPER ([Ali et al. 2015](#)), HERA ([DeBoer et al. 2017](#)) as well as those involving dish arrays like the Giant Metrewave Radio Telescope ([Paciga et al. 2013](#)). In principle, there are two categories of 21-cm experiments: one for measuring the global sky averaged signal and the other for measuring the brightness temperature fluctuations. Experiments such as MWA, PAPER, LOFAR and HERA are examples of the latter category and attempt to measure the fluctuation in  $\delta T_b$  caused by the Lyman- $\alpha$  and x-ray heating fluctuations.

## 2.2 Challenges in measuring Epoch of Reionization signal

The signal for EoR is very weak and several astrophysical as well as systematic problems need to be addressed in order to make an observation. A significant challenge at low radio frequencies is to mitigate the bright foreground emissions which are  $\sim 5$  orders of magnitude brighter than the EoR signal. Other important issues concerning 21-cm observation are radio frequency interference (RFI), ionospheric distortion, noise and instrument calibration.

### 2.2.1 Foreground

The foreground is composed of mainly three components, namely the galactic synchrotron emission, extragalactic sources and galactic free-free emission. The foreground signal is predominantly spectrally smooth because the electrons emit synchrotron or free-free emission over a broad range of frequencies ([Furlanetto et al.](#)

2006). Foreground cleaning techniques involve removing bright point sources such as the supernova remnants and extragalactic radio sources, and foreground subtraction. In foreground subtraction method, a smooth spectral model is fit and subtracted because foregrounds have smooth spectral signature. Typically, smooth polynomials of arbitrary order are fit and the best-fit model is subtracted. However, this method suffers from over-fitting and may destroy the signal or residuals may remain stronger than the EoR even after subtraction (Weltman et al. 2020).

An alternative to this method, known as the *foreground avoidance* technique has been applied to the PAPER and HERA experiments. It involves leaving out the parameter space that is mostly occupied by the strong foregrounds. We can choose to represent the data in Fourier space as a power spectrum. The spatial correlations across the sky is given by the wavenumber vector,  $\mathbf{k}$  (DeBoer et al. 2017).  $\mathbf{k}$  has two components:  $k_{\parallel}$ , related to bandwidth and  $k_{\perp}$  which relates to antenna baselines.  $k_{\parallel}$  and  $k_{\perp}$  correspond to line of sight and angular scales respectively. The foregrounds occupy low- $k_{\parallel}$  modes and EoR signal is relatively brighter in the low- $k_{\perp}$  modes. Hence, the “EoR window” which is almost free from foreground contamination, has a sharp demarcation and drives instrument design in EoR experiments. Figure 2.2 uses simulations from Mesinger et al. (2011) to show that 21-cm signal dominates in the EoR window and smooth-spectrum foregrounds that create a “wedge” in the power spectrum do not leak into the EoR window. We shall revisit the concept of power spectrum in the context of HERA experiment in section 3.1.

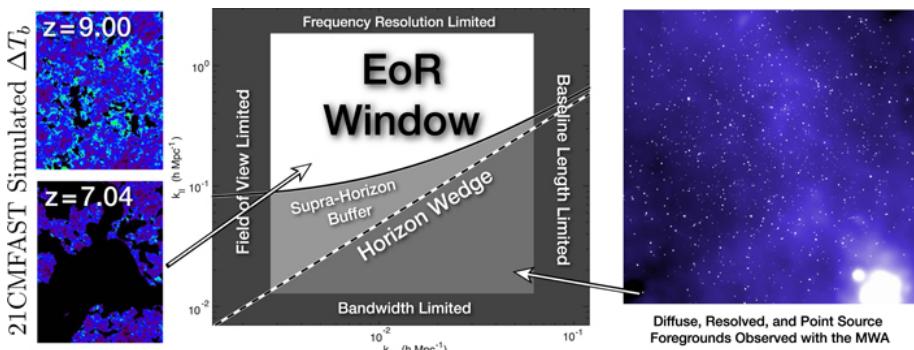


Figure 2.2: Simulations from 21CMFAST show that “EoR window” is not affected by smooth foregrounds. Diffuse and point source foregrounds create a “wedge” beyond which they do not contaminate the EoR signal. (Adapted from: DeBoer et al. (2017))

## 2.2.2 Radio Frequency Interference

In the present day, the world is entirely dependent on communication and this is a major issue for radio astronomy. Almost all communication devices make use of frequencies that are in the radio end of the spectrum. Radio Frequency Interference (RFI) is any signal of terrestrial origin present in radio frequencies that interfere with the radio signal from the sky. These may be frequencies allocated for FM and TV broadcasts, or communication frequencies used by mobile phones, aeroplanes and satellites. In addition to these, RFI may also be generated from spurious emission from electronic devices. In general, radio telescopes are built in remote locations that are sparsely populated. However, due to the high sensitivity requirements of these telescopes, they still suffer from RFI contamination as they detect high-power broadcasts. There are several RFI mitigation techniques that have been developed which involve automated detection and removal of RFI flagged data or characterising the interfering signals and subtracting them. [Figure 2.3](#) shows an example of RFI occupancy for an observation from the LOFAR EoR project. RFI mitigation is particularly important for low-frequency arrays and 21-cm experiments and there has been some recent progress in these fronts ([Offringa et al. 2015](#)). However, RFI still remains a major challenge for EoR detection and a new method has been presented in [chapter 3](#).

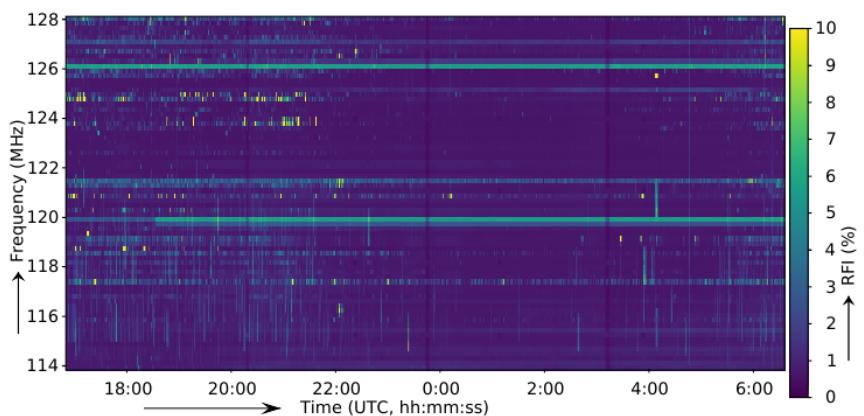


Figure 2.3: Shown here is an example of RFI occupancy over time and frequency obtained for a LOFAR observation of the North Celestial Pole (NCP) which is one of the target fields of LOFAR EoR Project (Adapted from: [Offringa et al. \(2019\)](#))

## 2.3 Basics of Interferometry

An interferometer telescope is an array of radio antennas with the most basic element being a pair of antennas separated by a *baseline* distance  $\mathbf{B}$ . The signal arrives at one antenna time-shifted by a delay  $\tau$  relative to the other antenna because the two antennas are separated by a distance and speed of propagation of electromagnetic signals is constant. A basic schematic diagram showing the components of a two-element interferometer is shown in Figure 2.4. The induced voltages by received electromagnetic signal on antennas  $T_1$  and  $T_2$  are  $V_1$  and  $V_2$  respectively and is given by:

$$V_1 \propto E e^{i\omega t} \quad (2.2)$$

and,  $V_2 \propto E e^{i\omega(t-\tau)}$

where the electromagnetic wave has amplitude  $E$  and the direction of propagation of wave causes an delay of  $\tau$  due to the separation  $\mathbf{B}$ .

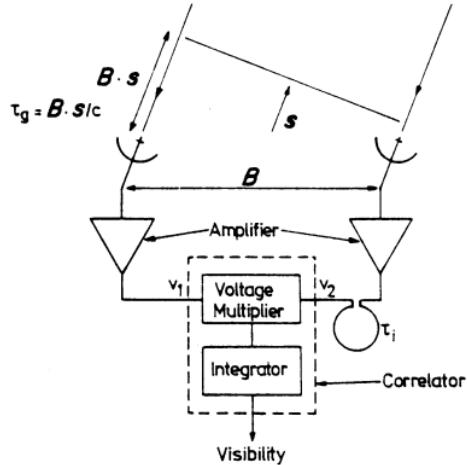


Figure 2.4: A schematic diagram of a two element interferometer is shown here. The two antennas, separated by baseline  $\mathbf{B}$ , produce voltages  $V_1$  and  $V_2$ .  $\tau_g$  and  $\tau_i$  are the geometric and instrument delays respectively. (Adapted from: [Wilson \(2011\)](#))

Let  $A(\mathbf{s})$  be the primary beam response function and  $\mathbf{s}$  is the direction vector to the source. If the radio sky brightness distribution is given by  $I_\nu(\mathbf{s})$ , the power received per bandwidth  $d\nu$  from the source element  $d\Omega$  is  $A(\mathbf{s})I_\nu(\mathbf{s})d\Omega d\nu$  ([Rohlfs and Wilson 2000](#)). The total interferometric response over source S is

$$R(\mathbf{B}) = \iint_{\Omega} A(\mathbf{s})I_\nu(\mathbf{s}) \exp \left[ i 2\pi\nu \left( \frac{1}{c} \mathbf{B} \cdot \mathbf{s} - \tau \right) \right] d\Omega d\nu \quad (2.3)$$

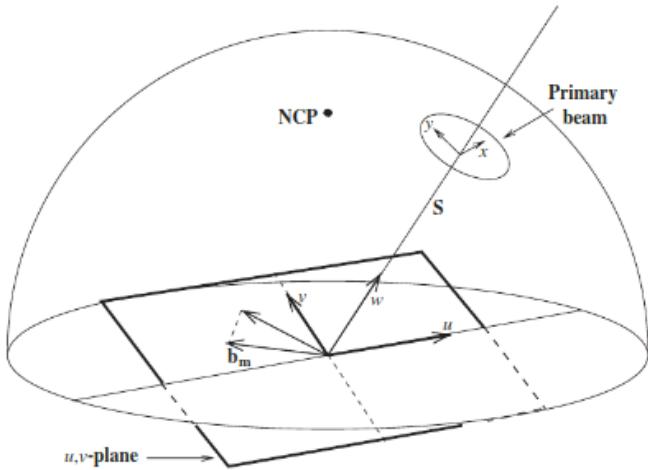


Figure 2.5: The celestial sphere and the projection onto the  $uv$ -plane is shown here. The primary beam is directed towards the source and NCP is the North Celestial Pole. Component  $w$  is directed towards the source. (Image from: [Burke and Graham-Smith \(1996\)](#))

where  $\tau$  is the delay and  $\Omega$  is the solid angle for the surface integral. We can introduce a right-handed rectilinear coordinate system  $(u, v, w)$  expressed in units of wavelength, and the baseline vector  $\mathbf{B}$  is related to this new coordinate system as  $\mathbf{B}/\lambda \equiv (u, v, w)$ . The component along  $w$  is parallel to direction vector to the source, and  $u, v$  lie in the perpendicular plane projected along East and North directions respectively. This plane which is perpendicular to the source direction is referred to as *uv-plane* in radio astronomy ([Burke and Graham-Smith 1996](#)). Figure 2.5 shows the *uv*-plane relative to the celestial sphere. [Equation 2.3](#) can then be expressed in terms of the ***visibility function***  $V$  of the intensity distribution as:

$$V_{ij}(u, v) = \iint A(l, m) I_\nu(l, m) \exp(i 2\pi(u l + v m)) d^2\Omega \quad (2.4)$$

where  $(l, m, n)$  are the direction cosines of  $\mathbf{s}$  with respect to *uv*-plane and component  $n$  has no contribution because  $w$  is along  $\mathbf{s}$ . In practice, angles  $l$  and  $m$  are small and angular size of source is smaller than the antenna beam size. Using small angle approximation, the coordinates  $(l, m)$  can be replaced with  $(x, y)$ , which simplifies [Equation 2.4](#) giving:

$$V(u, v) \approx \iint I(x, y) \exp(i 2\pi(ux + vy)) dx dy \quad (2.5)$$

or,  $V(u, v) \stackrel{\text{F.T.}}{=} I(x, y)$

This shows that visibility is in fact the Fourier transform of the observed sky. The

interferometer observes the Fourier transform of the source brightness distribution at a given frequency, which is measured in the *uv*-plane. A single observation gives only one Fourier mode. In order to obtain the entire Fourier transform of the source brightness distribution, many different baselines are required and building an array of antennas gives the required *uv* coverage.

## 2.4 Hydrogen Epoch of Reionization Array (HERA)



Figure 2.6: Shown here is an image of the HERA photographed in March 2020. The array is located in Karoo desert in South Africa. (Image Credit: Dara Storer, HERA Collaboration)

The Hydrogen Epoch of Reionization Array (HERA)<sup>1</sup> is a radio telescope located in Karoo desert in South Africa with the primary science objective of observing the 21-cm signal from EoR. Science from HERA data would allow us to characterise the 21-cm signal from redshift  $z \sim 30$  to  $z \sim 6$  and study the IGM heating and reionization history (DeBoer et al. 2017). HERA is an array of 350 14-metre diameter non-tracking parabolic dishes arranged in a hexagonal grid. The array is designed to

---

<sup>1</sup><http://reionization.org/>

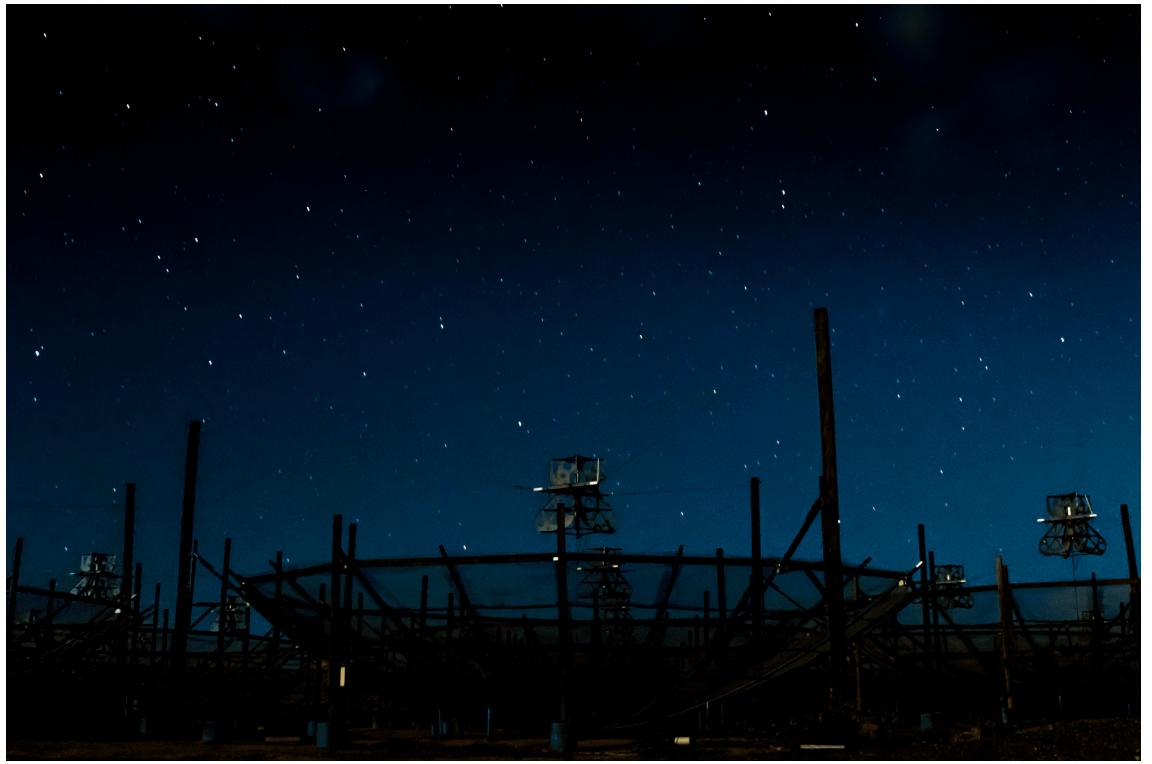


Figure 2.7: A close-up view of a parabolic HERA dish with the Southern night sky in the background. When complete, the array shall consist of 350 such elements. The dishes are constructed from PVC pipes and wire mesh. The support pipes are fixed to the center in a concrete cylinder and at the rim with wooden pillar supports. The feed is hung from three equally spaced poles which are arranged in a hexagonal lattice ([HERA Collaboration 2016](#)) (Image Credit: Dara Storer, HERA Collaboration)

optimise signal to noise ratio and provide high sensitivity for detection EoR signal. HERA is a successor to MWA and PAPER and is a precursor instrument for Square Kilometre Array (SKA). The HERA dish element has a collecting area of  $150 \text{ m}^2$  which is 40 times more collecting area than the PAPER element. HERA 350-element total collecting area will be  $53878 \text{ m}^2$ , which is greater than that of MWA, PAPER and LOFAR while providing robust sensitivity comparable to SKA. [Figure 2.6](#) and [Figure 2.7](#) show the HERA and a dish element, photographed recently by the HERA team in March 2020. Initial objective of the HERA experiment is to measure the statistical power spectrum of the 21-cm signal rather than direct imaging. The compact configuration of HERA is driven by the fact that short baselines are more sensitive to EoR detection and is designed to detect 21-cm power spectrum while implementing foreground avoidance techniques. HERA radio data shall be used in conjunction with other reionization probes such as JWST, WFIRST, CMB maps, line intensity mapping of CO, CII and Ly- $\alpha$  as well as semi-analytical models to

constrain the EoR (DeBoer et al. 2017; Lidz et al. 2009; Vrbanec et al. 2016).

# 3 RFI Mitigation: Simulation Approach

As discussed in section 2.2, there are several challenges in radio astronomy and handling RFI in the data reduction process is a critical step. Due to increased sensitivity requirements of next-generation telescopes and increasing number of RFI sources, RFI mitigation techniques are important for EoR detection. Radio telescopes such as HERA, LOFAR, MWA and several others use automated RFI flaggers (Offringa et al. 2010b). Toolkits like Astronomical Image Processing System<sup>1</sup>, MIRIAD<sup>2</sup> or their Python counterparts like Astronomical Interferometry in PYthon<sup>3</sup> have been developed to process radio data and flag RFI. One of the preliminary steps in the pipeline for analysing data from interferometers is RFI detection, which is known as ‘*data flagging*’. Usually RFIs are flagged, classified and then processed for removal. Some RFI mitigation techniques involved are curve fitting, smoothing, thresholding and line detection (Offringa et al. 2010a, 2015).

## 3.1 Problem Statement

The success of 21-cm EoR experiments depends on how well the foregrounds can be separated from the EoR signal. EoR projects attempt to detect a weak signal which means that we must be careful to ensure that a detection is not affected by RFI leakage. In case, unwanted features that are introduced by either instrument instability or RFI are not properly modelled, it may become impossible to extract the EoR signal.

The HERA experiment shall measure the EoR spatial power spectrum and uses a delay-spectrum approach. This approach is based on the “delay transform” which is defined as the Fourier transform of the visibility spectrum measured by a single

---

<sup>1</sup>AIPS: <http://www.aips.nrao.edu/dec19.shtml>

<sup>2</sup>MIRIAD: a data reduction package for Australia Telescope Compact Array, <https://www.atnf.csiro.au/computing/software/miriad/>

<sup>3</sup>AIPY: <https://pypi.org/project/aipy/>

baseline along the frequency axis (Parsons et al. 2012):

$$\tilde{V}_b(\tau) = \int A(l, m, n) I(l, m, n) \exp(-2\pi i n \tau) dl dm dn \quad (3.1)$$

where  $A(l, m, n)$  describe the response of an interferometric pair of antennas,  $I(l, m, n)$  is the specific intensity,  $\tilde{V}_b(\tau)$  is the *delay-transformed* visibility and coordinates  $(l, m, n)$  are as defined in section 2.3. The power spectrum  $P(\mathbf{k})$  is then obtained from the product of two delay-transformed visibilities and is given by:

$$P(\mathbf{k}) \approx \frac{X^2 Y}{4K_B^2} \left[ \frac{\tilde{V}_b^2(\tau)}{\Omega_b B/\lambda^4} \right] \quad (3.2)$$

where  $\tilde{V}_b$  is given in Equation 3.1,  $K_B$  is Boltzmann's constant,  $X$  and  $Y$  are parameters relating wavenumber to physical cosmic volume,  $\lambda$  is the observation wavelength,  $B$  is the effective bandwidth, and  $\Omega_b$  is the integrated beam response (DeBoer et al. 2017).

This project investigates the effect of RFI on 21-cm power spectrum. The contaminating effects caused by interference creates problems for 21-cm power spectrum measurement. The power spectrum is affected by fluctuations in visibilities due to sporadic removal of RFI flagged data samples. Sharp cuts in the data introduce higher modes in the Fourier space. The effect of RFI on power spectrum needs to be studied because spurious higher modes can contaminate the signal from EoR. This is known as ‘ringing’ in the power spectrum. In addition to this, gaps in the interferometric data also affect the power spectrum. The approach this project takes is to fit a smooth model to the RFI flagged dataset, which shall be referred to as *in-painting process*, and then analyse the power spectra of the fitted data to that of the un-flagged dataset.

## 3.2 Methodology

We take a simulation based approach to analyse the effects of RFI on the power spectrum. We use `hera_sim`<sup>4</sup> which is a Python library developed by the HERA team. We use it to simulate visibilities and include diffuse foreground, point sources, thermal noise and a noise-like EoR signal. The core idea is to produce an ideal sky visibility data, hereafter called the *un-flagged data*, so that the power spectrum of the un-flagged data can then be compared to that after introducing the RFIs. The

---

<sup>4</sup>[https://github.com/HERA-Team/hera\\_sim](https://github.com/HERA-Team/hera_sim)

power spectrum is calculated using another `Python` package, `hera_pspec`<sup>5</sup>, which provides the data structures and tools needed to perform a delay spectrum analysis on radio data. The delay power spectrum is calculated using the *optimal quadratic estimator* as described in [Ali et al. \(2015\)](#).

At the next step of the simulation, randomly generated RFIs are added to the data. These RFIs are similar to what we expect from FM channels, DTV frequency bands or satellite communication signals. In case the RFI populated data is flagged and removed, the power spectrum is significantly affected. The gaps and cuts in data causes higher  $\mathbf{k}$  modes in the power spectrum. In order to mitigate the effects, we fit a smooth model to the flagged data using a linear least-squares solver. The model is a Fourier series and the number of modes can be chosen as required. The best fit model is calculated and the data with flagged frequencies is filled in with the smooth solution. This process is called *in-painting* and the modified data shall be referred to as *in-painted data*.

The process of fitting a sample of data  $y_i$  containing  $N$  number of data points requires converging on a minimum value in the parameter space. In linear least-squares fitting method ( $\chi^2$  fit), the model  $P_i = ax_i + b$  is linear, where the uncertainty on each point is  $\sigma(y_i)$ .

$$\chi^2 = \sum_{i=1}^N \left( \frac{y_i - ax_i - b}{\sigma(y_i)} \right)^2 \quad (3.3)$$

$\chi^2$  is then minimised with respect to both  $a$  and  $b$  simultaneously. In linear algebra, this is equivalent to solving the matrix equation  $\mathbf{AX} = \mathbf{B}$  which gives us the coefficients for the best-fit curve. In the *in-painting process*, the best-fit coefficients are calculated using a *log-likelihood* function assuming Gaussian distribution of data. The coefficients are used to build a smooth model to which an uncorrelated noise realization can be added in the flagged regions. The algorithm allows us to specify the noise standard deviation and the maximum order of Fourier modes to fit. The algorithm returns the best-fit model which composed of a sum of Fourier modes, the coefficients of Fourier modes and the in-painted data. Then we compare this in-painted data to the data from the ideal case simulation without the RFI by analysing the difference in both data. We also compare their respective power spectra to understand how the in-painted power spectrum depends on the fraction of frequency channels that have been flagged.

---

<sup>5</sup>[https://github.com/HERA-Team/hera\\_pspec](https://github.com/HERA-Team/hera_pspec)

### 3.3 Simulation Set-up

In order to simulate the steps as discussed in section 3.2, we simulate visibilities using `hera_sim` for 50 LST (Local Sidereal Time) bins and 128 frequency bins. We use one baseline and 16 hours of total observation time to keep the simulation computationally feasible. We use the `Simulator` class from `hera_sim` for the above defined configuration to instantiate a `UVData` object. `UVData` is a class built in `pyuvdata`<sup>6</sup> which supports interferometric visibilities and associated metadata.

To create the un-flagged ideal sky model, we add diffuse foreground and 500 point sources to the simulation. The EoR signal is expected to be noise-like because regions of HI are randomly distributed over a wide range of redshifts. This allows us to simulate the EoR signal as a noise-like model. Thermal noise model is also added to the simulation and EoR signal is assumed to be several orders of magnitude smaller than the noise. The `Simulator` class calculates visibilities for each of models and adds it to the underlying data and returns a `UVData` object. Interferometric visibility is typically measured in units of Jansky ( $10^{-23}$  ergs sec $^{-1}$  cm $^{-2}$  Hz $^{-1}$ ). It is useful to map visibilities into a brightness temperature scale from a flux density scale. A conversion from Jansky (Jy) scale to milli-Kelvin (mK) scale is given by:

$$V_{\text{mK}}(\lambda) = V_{\text{Jy}}(\lambda) 10^{-23} \frac{\lambda^2}{2 K_B \Omega_p(\lambda)} 10^3 \quad (3.4)$$

where  $\Omega_p$  is the integral of the beam power response over a solid angle  $\Omega$ ,  $K_B$  is the Boltzmann's constant and  $\lambda$  is the wavelength. We shall use mK units for calculating the power spectra (mK $^2$  h $^{-3}$  Mpc $^3$ ). Python package `hera_pspec` is used to calculate a delay spectrum which is stored as a `UVPSpec` object. We refer to this as the *un-flagged power spectrum*. We use the Blackman-Harris tapering window function and apply identity weighting matrix to the input data.

RFI flagged frequencies are randomly generated and visibilities in bins with RFI flags are set to zero. We add the same RFI flags for all LST bins. The fraction of RFI flagged frequencies is set to 7% for the first round of simulations and thereafter increased to 15% and 30%. We set maximum number of Fourier modes to fit to 50 and keep  $\sigma = 1$ , which is the noise standard deviation. The in-painting process is applied to the RFI-flagged data by using a smooth model to fit the flagged regions. An unconstrained noise realization is added to the in-painted bins. The *in-painted power spectrum* is then calculated and compared to the previously obtained un-flagged power spectrum.

---

<sup>6</sup><https://github.com/RadioAstronomySoftwareGroup/pyuvdata>

# 4 Results and Conclusion

We present the simulation results in this chapter. In these set of simulations we use 128 frequency channels and 50 LST bins. We look into 1D plots of single LST bins as a function of frequency to analyse the data at each step of the simulation before comparing the power spectra. A waterfall plot, which is a colour-map plot of visibilities with Time and Frequency axes, is created for each case to visualise the data. The waterfall plot shows the amplitude and phase values of the complex-valued dataset.

## 4.1 Simulation Plots

[Figure 4.1](#) and [Figure 4.2](#) show the noise and EoR signal respectively for a single time slice as a function of frequency. The noise is  $\sim 3$  orders of magnitude higher than the EoR signal. [Figure 4.3](#) shows the waterfall plot for the noise-like EoR signal. The waterfall plot of the complete ideal sky visibility data which includes the diffuse foreground, 500 foreground points sources, noise and EoR signal is shown in [Figure 4.4](#). There is no RFI in this data. The calculated delay power spectrum averaged over time for the *un-flagged* dataset is shown in [Figure 4.5](#). The smooth part of the power spectrum at lower delays corresponds to the foregrounds while the noise and EoR signal correspond to the higher delays.

Next, RFI flags are randomly added and data from the flagged regions are set to zero. For this set of simulation, RFI flagged percentage was set to 7%. [Figure 4.6](#) shows the visibility of the flagged data set. Frequency bins with no data correspond to the RFI flags. In order to analyse the flagged data set we plot visibility for the same time slice for both *un-flagged* and the *flagged* datasets in [Figure 4.7](#). The dashed-green line shows the modified data after adding RFI flags and the underlying dotted-blue line shows the original visibility. This is similar to actual observational data that gets contaminated by RFI flags.

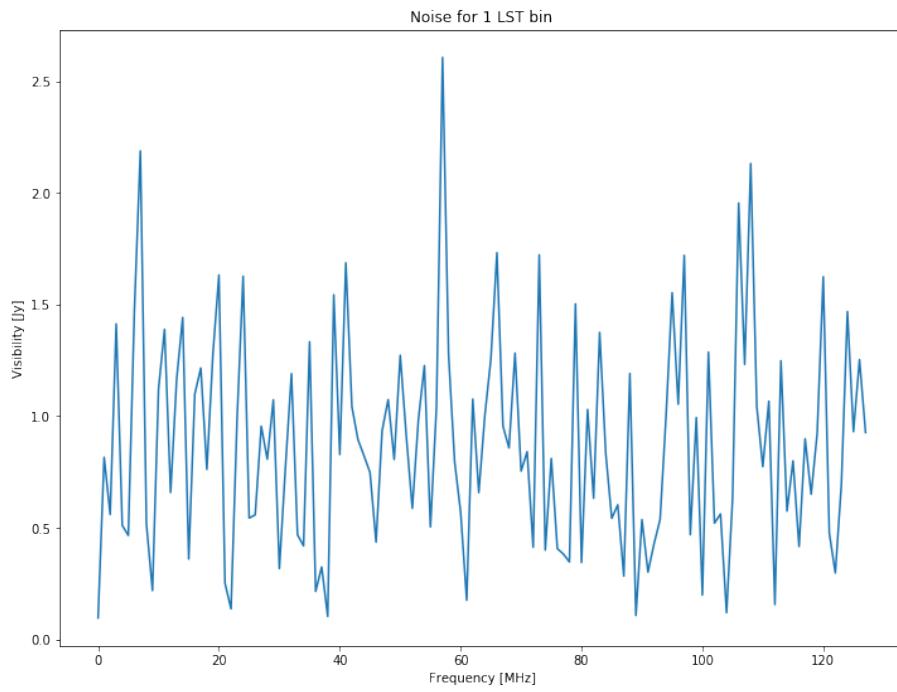


Figure 4.1: Simulated noise visibility [Jy] for 1 LST bin as a function of frequency [MHz] is shown. The mean of the noise is around  $10^3$  times the EoR signal.

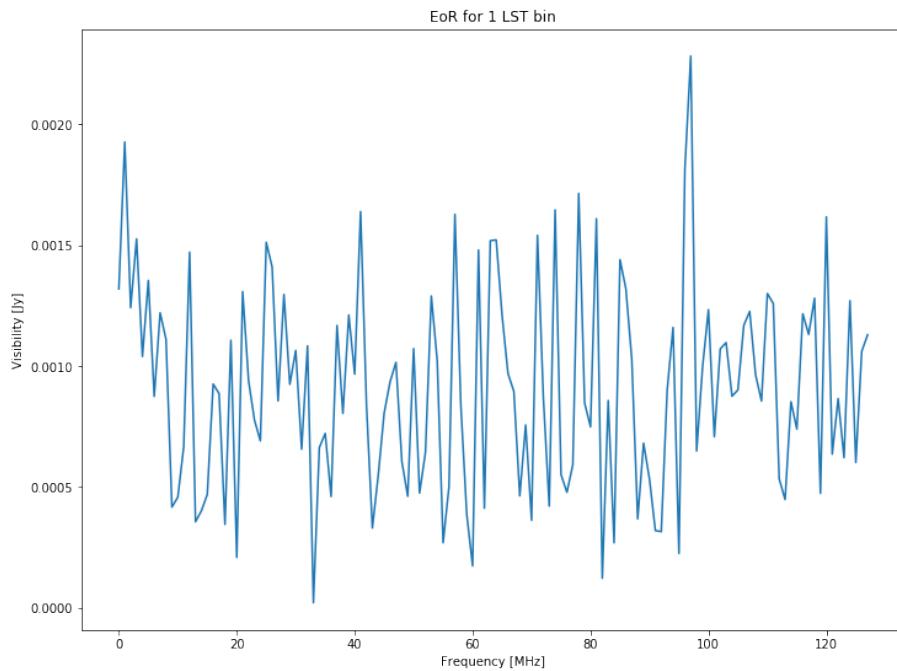


Figure 4.2: Plotted here is the simulated noise-like EoR signal [Jy] for 1 LST bin as a function of frequency [MHz]

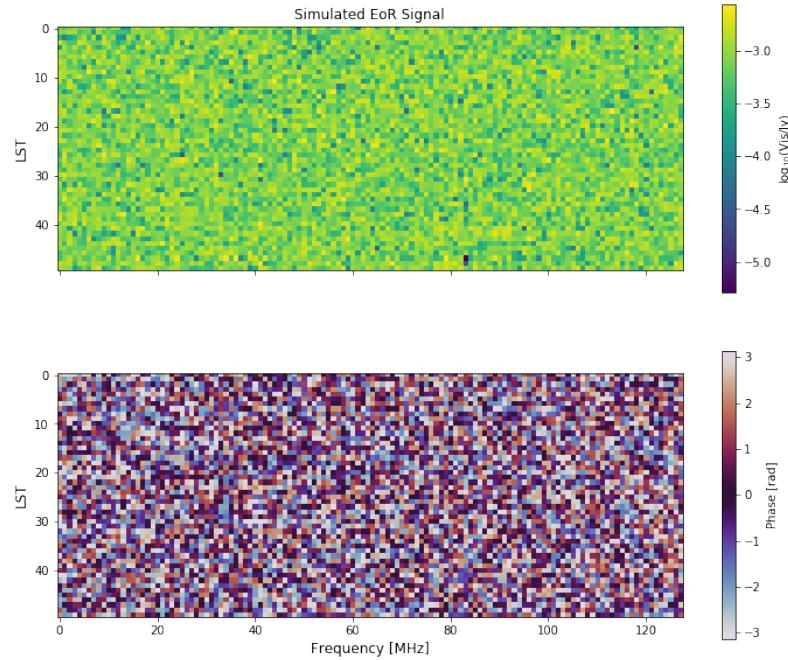


Figure 4.3: Waterfall plot showing the amplitude and phase of the simulated noise-like EoR signal

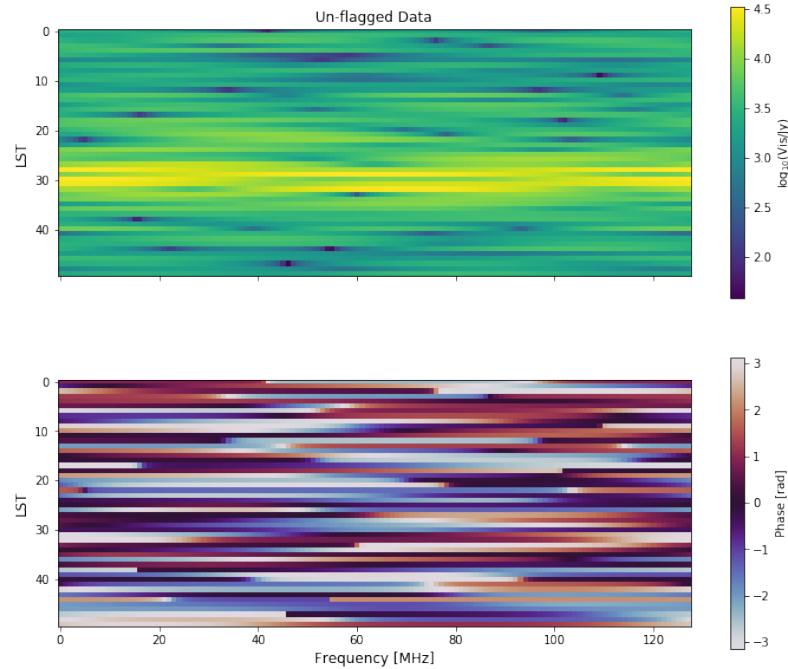


Figure 4.4: Waterfall plot of the un-flagged ‘ideal’ sky visibilities after including diffuse foreground, 500 point sources, thermal noise and EoR signal. The data is obtained from simulating 50 LST bins and 128 frequency channels at 1 baseline. The phase patterns in each LST bin is from the several point sources moving in and out of the antenna beam as the sky rotates.

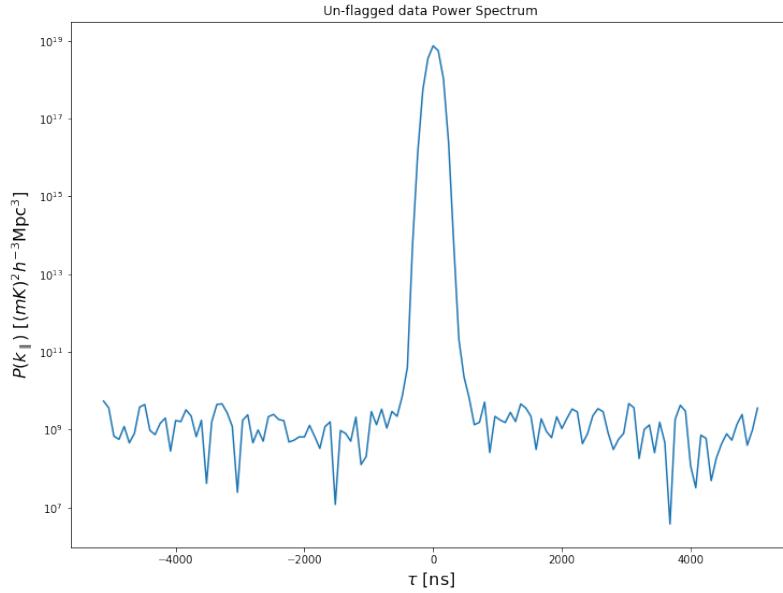


Figure 4.5: Delay power spectrum [ $\text{mK}^2 \text{ h}^{-3} \text{ Mpc}^3$ ] along the frequency axis for 1 baseline, averaged over time for the un-flagged data is shown.

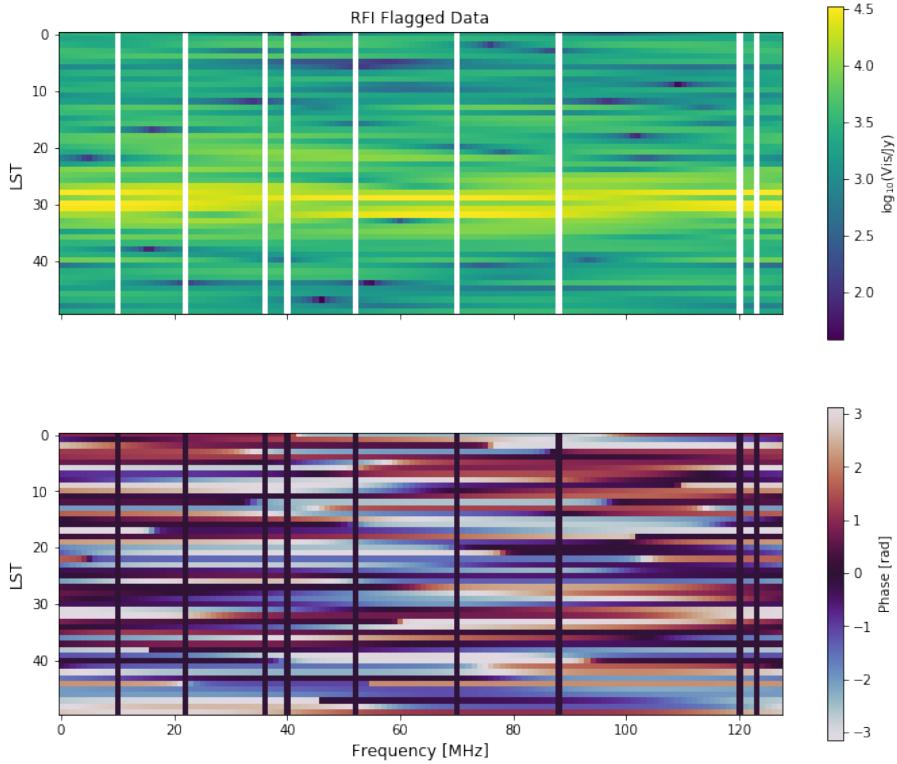


Figure 4.6: RFI flags are randomly added to the ideal sky data. The data from the flagged frequency bins are removed. Waterfall plot of flagged data is shown here. Same RFI flags are applied for all LST bins. Fraction of RFI flagged channels here is 7%

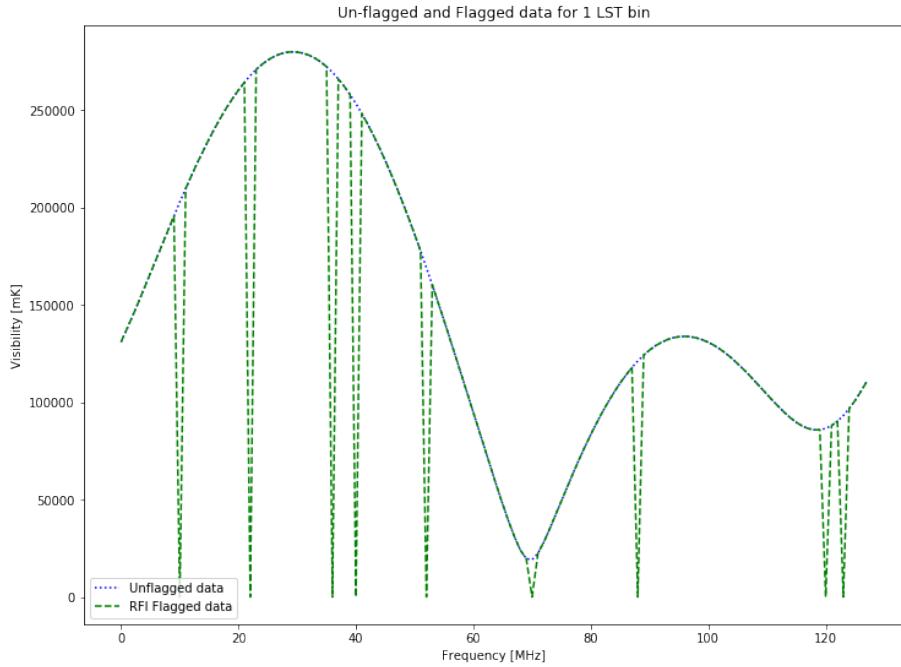


Figure 4.7: A comparison of the visibilities of flagged and un-flagged data for 1 LST bin is shown. The flagged curve has sharp cuts at regions of RFI flags.

Our aim is to eliminate the sharp features in the RFI flagged data and we want to fit a smooth model and recover the ideal sky visibility data without the flags. At the next step, we take this *flagged* dataset, and apply the *in-painting* process to it. [Figure 4.8](#) shows the best-fit model for a single LST bin that is fit to the flagged data. As discussed in [section 3.2](#), the smooth model comes from a linear least-squares solver. The data with flagged regions is filled in with the smooth solution and therefore is called ‘*in-painted*’ data. [Figure 4.9](#) shows the un-flagged (dotted-blue line), flagged (dashed-green line) as well as the in-painted data (dotted-orange line) for the same LST bin as a function of frequency. This plot clearly shows that the in-painting process is able to recover the smooth features of the original un-flagged ideal data from the RFI flagged data. The in-painted data is similar to the un-flagged data and the smooth model bridges the gaps in the data caused by RFI flags. We also add an uncorrelated noise realisation to the in-painted areas.

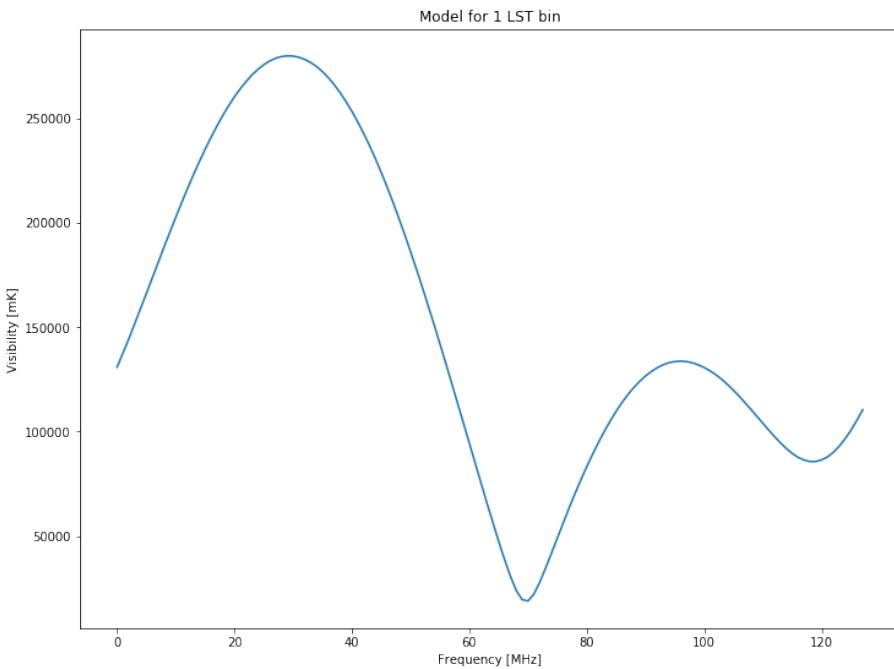


Figure 4.8: Visibility [mK] of the smooth best-fit model (for 1 LST bin) obtained by solving for the coefficients of Fourier modes is shown.

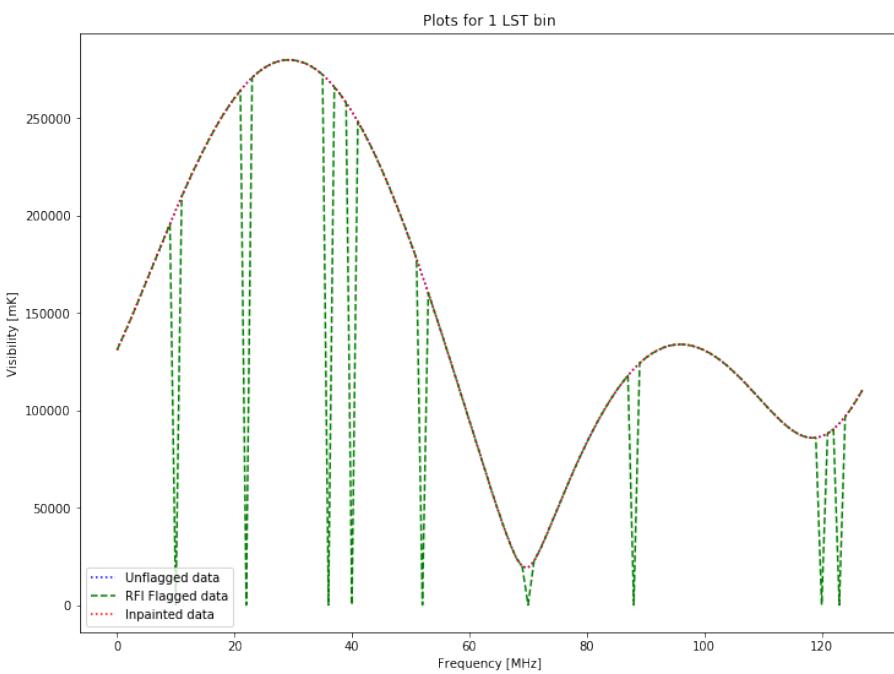
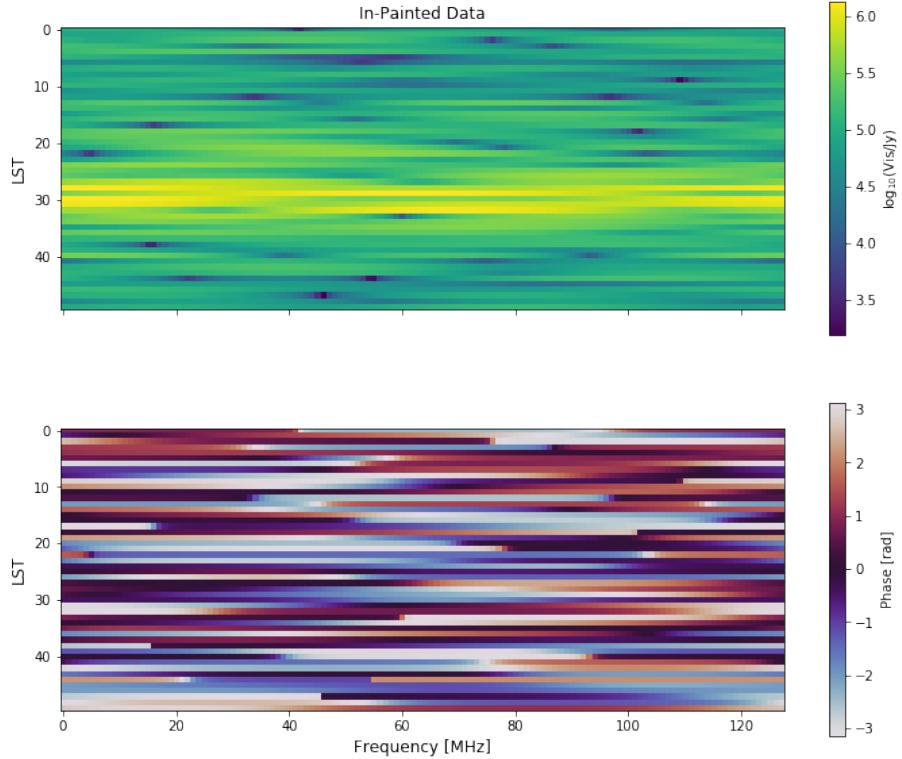


Figure 4.9: Shown here are the visibilities [mK] for the same LST bin from un-flagged, flagged and in-painted data. The in-painted data has ‘filled-in’ the gaps in the RFI flagged regions.

The complete two-dimensional in-painted dataset is shown as a waterfall plot in [Figure 4.10](#). When compared to [Figure 4.6](#), we see that the initial RFI flagged blanked bins have now been replaced with realistic visibilities and the plot looks close to what we started with in the ideal case scenario in [Figure 4.4](#). At this point, it is interesting to analyse and compare the un-flagged dataset and the best-fit model to gain an insight into how well the in-painting process works. [Figure 4.11](#) shows the difference of the best-fit model and the un-flagged data for one LST bin. The initial flagged frequencies are marked in the plot in light red for comparison. It is important to note that the smooth model does not include the noise. So the difference of model and un-flagged data gives us the noise ‘residual’. As shown in [Figure 4.12](#), we then average the noise residual over all LST bins. The average shows that the residual is in the order of  $10^2$  in the bins that were initially flagged and this is of the same order as the noise in the ideal sky simulation. The standard deviation of the residual is plotted in [Figure 4.13](#). The error on the average is within an order of magnitude of the noise level. [Figure 4.14](#) gives the waterfall plot of difference of model and un-flagged data. This shows that the residual is in fact the noise and no extra features are introduced by the in-painting process.



[Figure 4.10](#): Waterfall plot of visibilities of in-painted data. The visibilities are similar to the ideal sky data in [Figure 4.4](#).

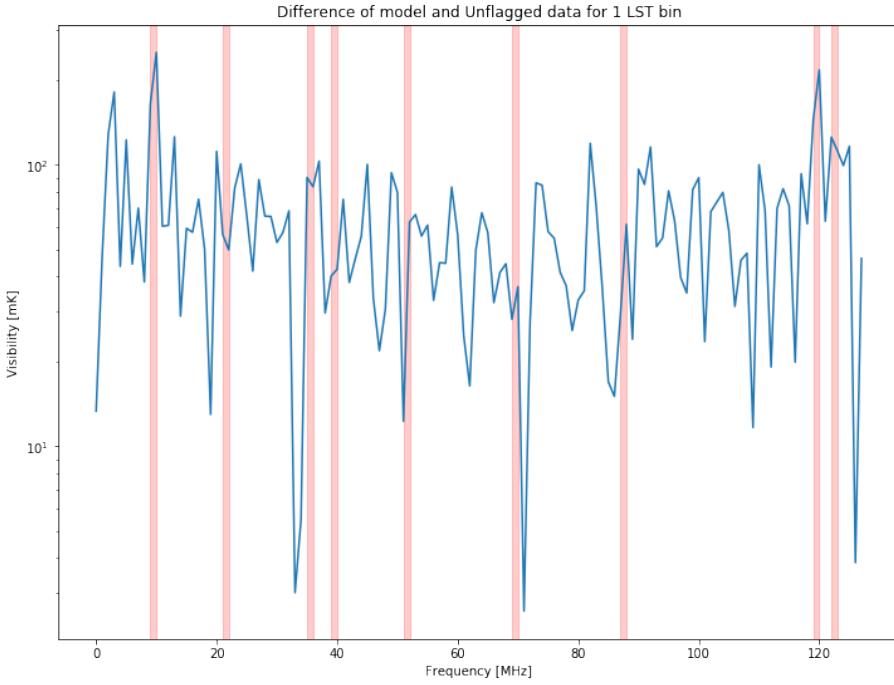


Figure 4.11: Difference of the best-fit model and the un-flagged data for one LST bin shows a noisy residual. This is because the smooth model does not include noise. Flagged regions are marked in light red.

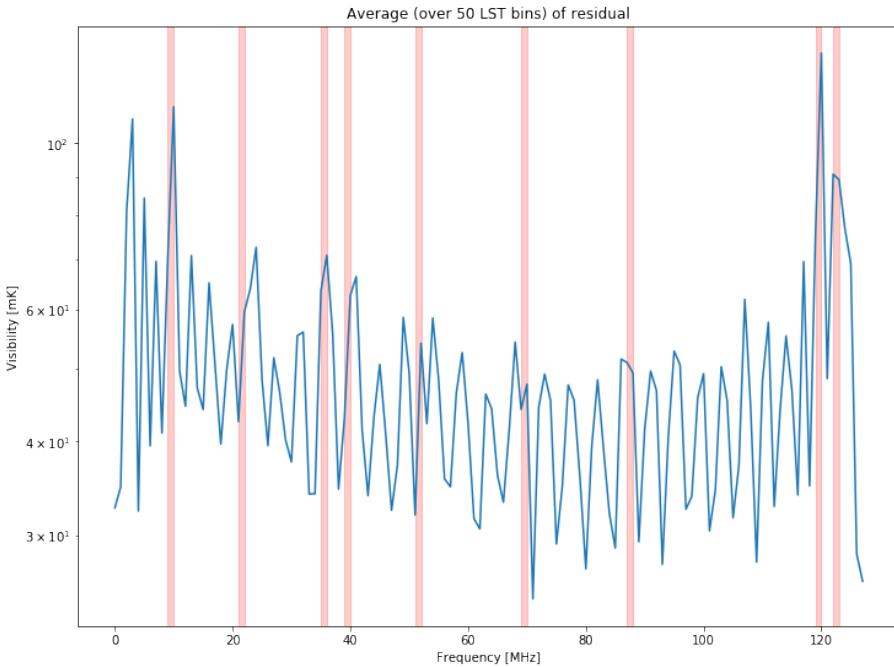


Figure 4.12: The noise residual averages down over time and is at the same level for both flagged and un-flagged regions. Previously flagged frequencies are highlighted for visual clarity.

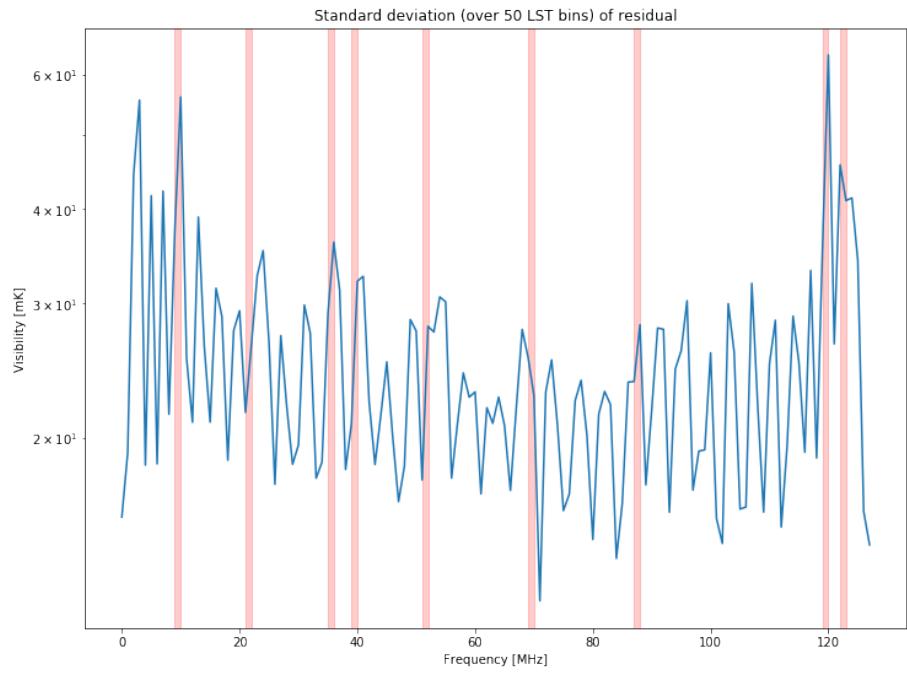


Figure 4.13: Standard deviation of noise residual obtained from difference of the best-fit model and the un-flagged data shows the error on the noise average.

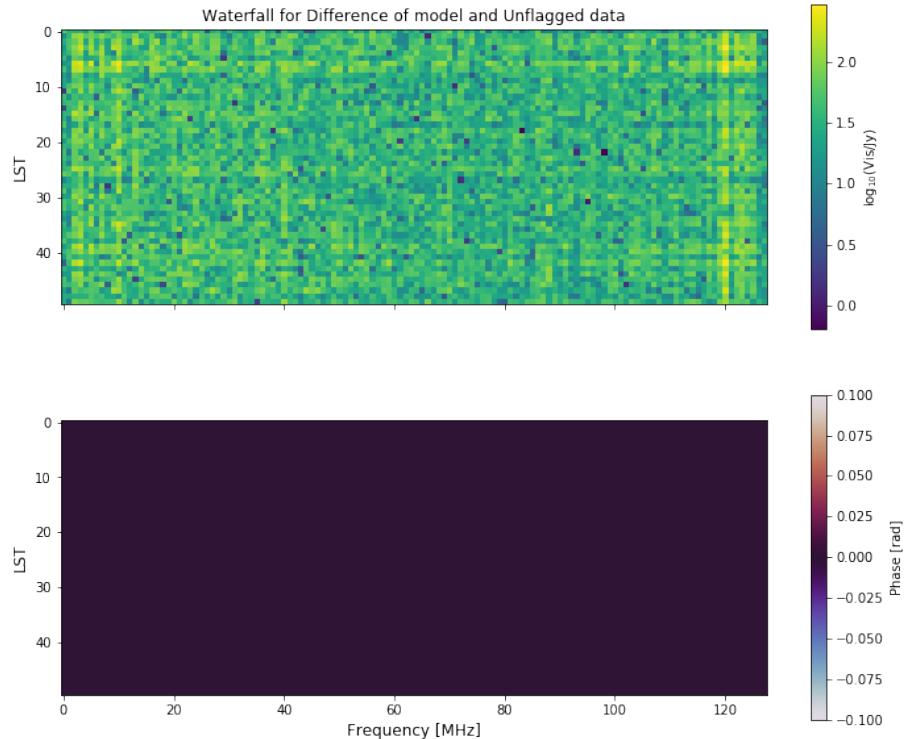
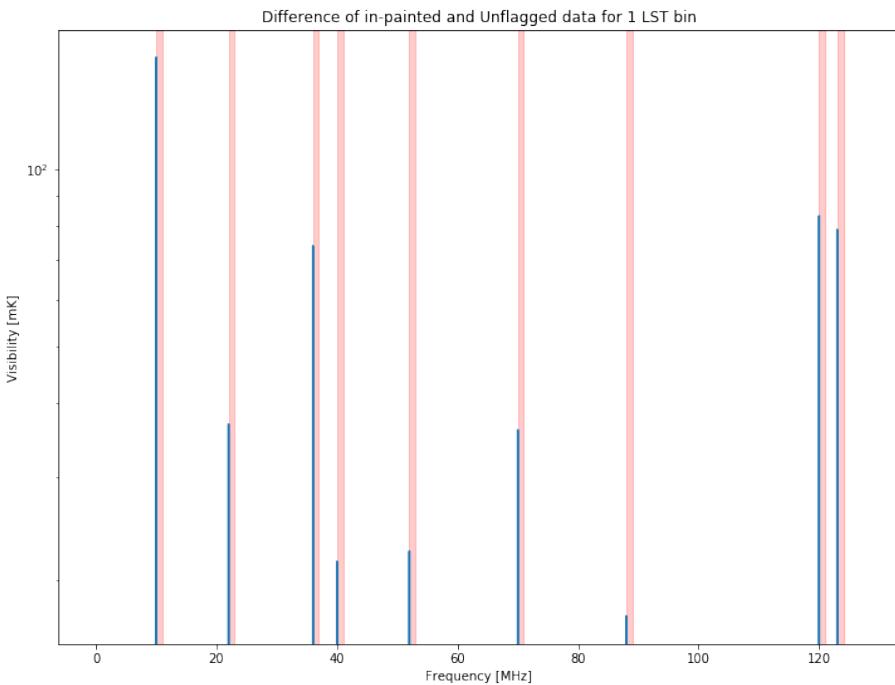


Figure 4.14: Waterfall plot of difference of smooth model and un-flagged data. This shows the noisy residual in the data, which is at the same level as the thermal noise introduced earlier in the simulation.

Next we study the difference of in-painted data and ideal sky data. [Figure 4.15](#) and [Figure 4.16](#) show the difference of the in-painted and un-flagged data for a single LST bin and the entire array respectively. We see that at flagged regions the difference is of the same order as that of the noise and the difference is zero elsewhere. This shows that the in-painting process adds a smooth model along with an unconstrained noise realization in the flagged regions and does not destroy the non-flagged regions of the dataset. We emphasise that this was the original goal and we aimed to recover the smooth dataset from RFI contaminated data.



[Figure 4.15](#): Difference of the in-painted data and un-flagged data for one LST bin. The noise peaks are at the same locations as the RFI flagged frequencies.

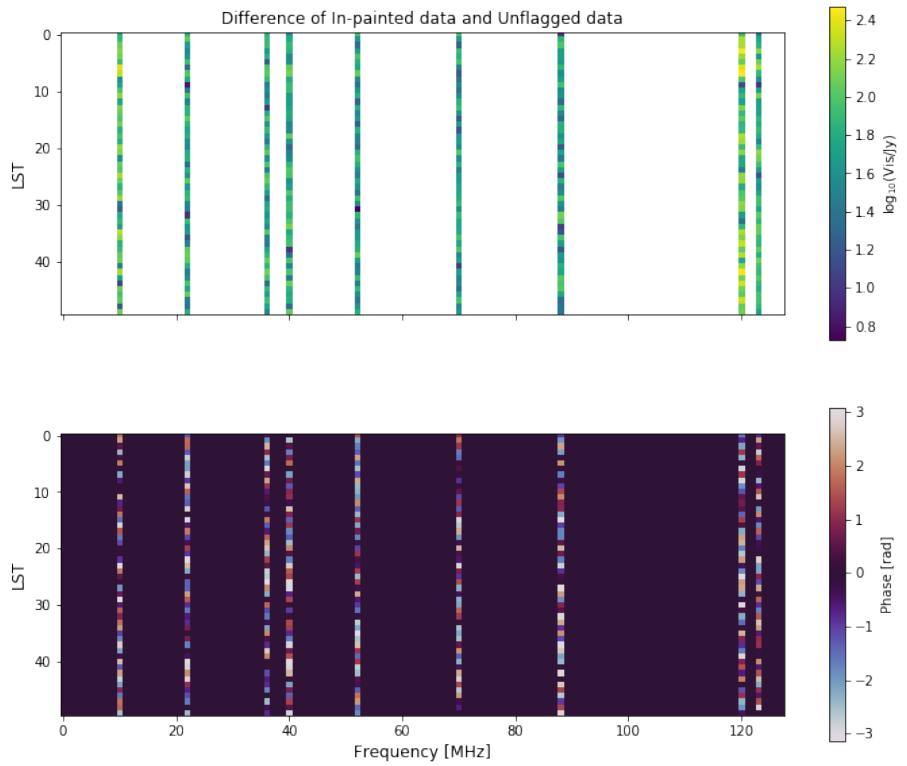


Figure 4.16: Waterfall plot of difference of in-painted data and un-flagged data. This shows that the in-painted data is same as the un-flagged data at all regions except at the RFI flags. At regions of RFI flags, the difference comes from the noise in the data.

In order to validate the fact that no other features except the noise is added to the in-painted data, the simulation is run without including the thermal noise. In this scenario, we add the diffuse foreground, 500 foreground points sources and EoR signal while noise is left out of the simulation. The results for this set of simulation is expected to have residuals free from noise. [Figure 4.17](#), which is the difference of in-painted data and the un-flagged data from the new simulation, shows that in fact there is no noise in the in-painted data and is similar to the data from the ideal sky. The difference in visibilities is negligible for all flagged frequencies and the mean is 0.03 mK with a standard deviation of 0.14.

We then plot the delay power spectra for both the scenarios. From [Figure 4.18](#), we see that for the simulation without noise, the power spectra for in-painted and un-flagged data are extremely similar. The power falls off quickly in the higher delays because there is no noise in the simulation. [Figure 4.19](#) shows the delay power spectra for the simulation including the noise. The difference in the higher delays comes from the noise realization in the in-painted data. For both these simulations, RFI flagged frequency occupancy was kept at 7%.

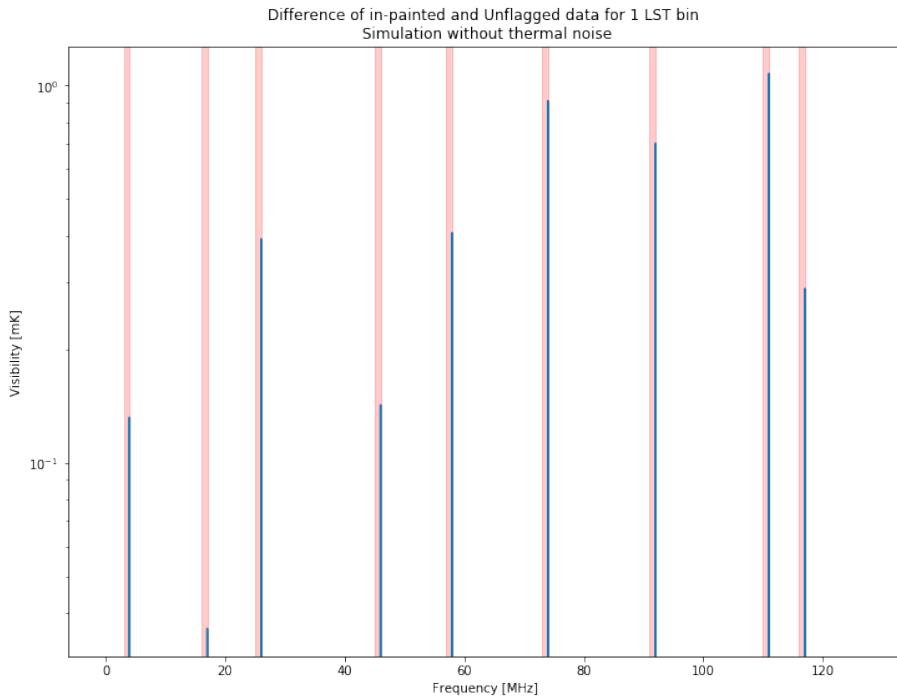


Figure 4.17: Simulation without noise and with 7% RFI flagged fraction: Difference of the in-painted data and un-flagged data for one LST bin. The difference is extremely small because there is no noise.

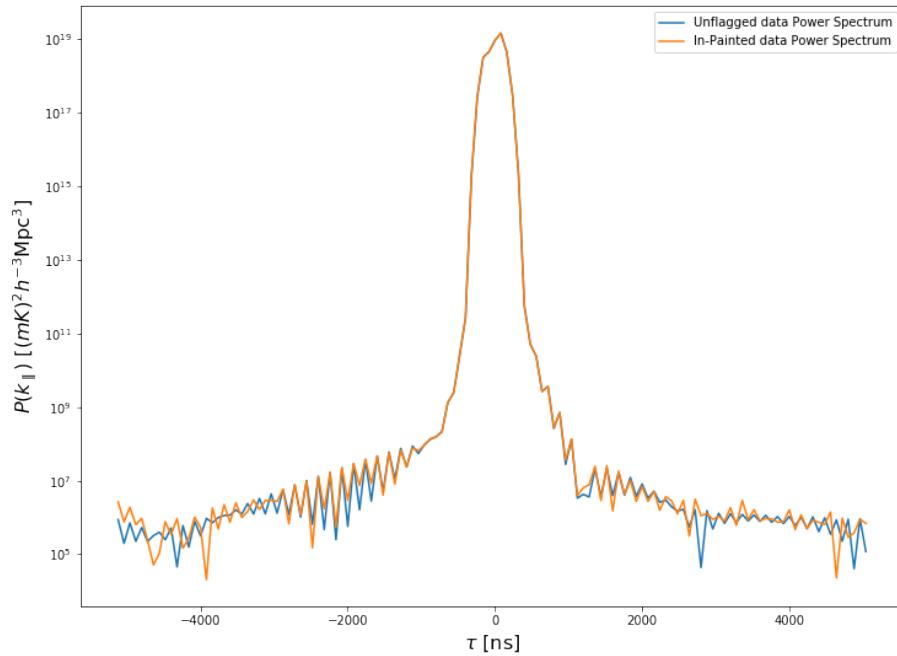


Figure 4.18: Simulation without noise: Delay power spectrum [ $\text{mK}^2 \text{ h}^{-3} \text{ Mpc}^3$ ] averaged over 50 times for in-painted and un-flagged data.

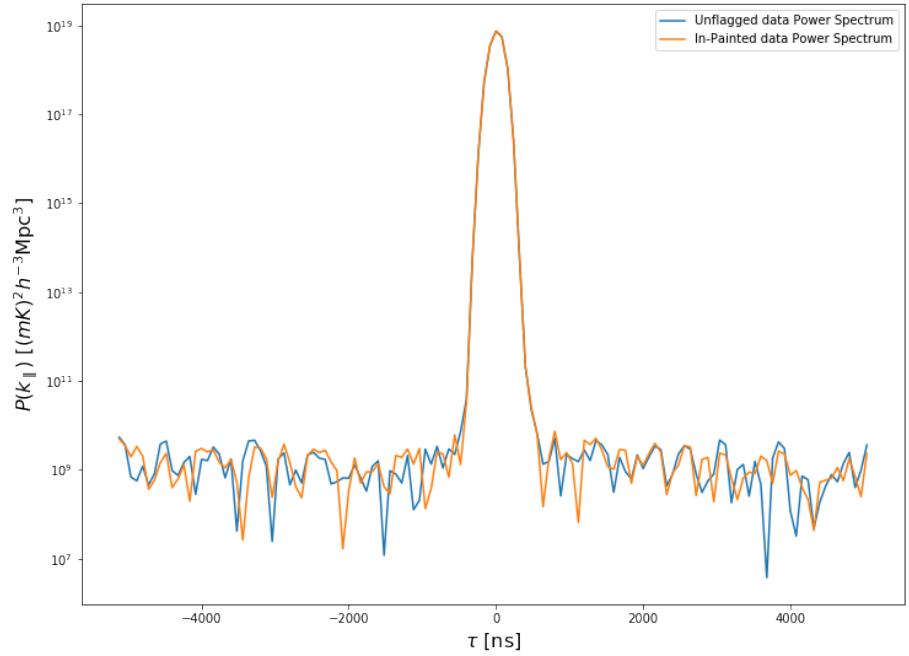


Figure 4.19: Simulation with noise and 7% RFI flagged fraction: Delay power spectrum of in-painted and un-flagged data using maximum number of Fourier modes = 50

We then increase the fraction of flagged frequencies in the data to 15% and 30%, while keeping rest of the parameters same and including thermal noise, to investigate the effect of increased flagging on the in-painted power spectra. As seen in Figure 4.20 in the case of 15% RFI flagged fraction, the power spectrum is at the same order as that of 7%. However, when the fraction on RFI flagged frequencies increases to 30% (Figure 4.21), there is a significant increase in power as compared to previous cases as well the un-flagged data. This happens due to the increase in the gaps in the flagged data, which leads to excess power after fitting the model. In both these simulations, maximum number of Fourier modes was kept at 50 and dataset was created for 128 frequency channels and 50 LST bins using 1 baseline.

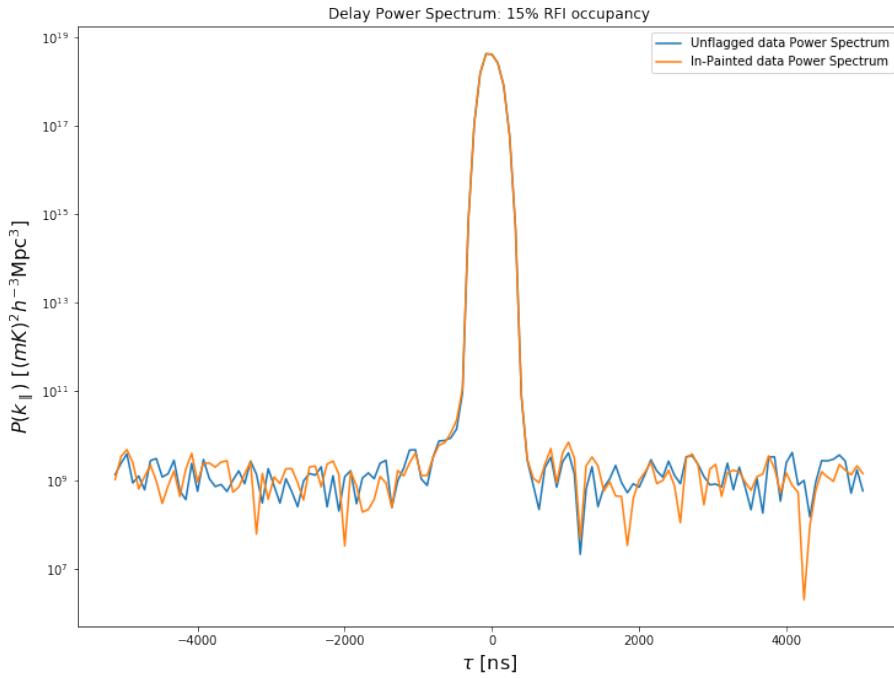


Figure 4.20: Simulation with noise and 15% RFI flagged fraction: Delay power spectrum averaged over 50 times for in-painted and un-flagged data using maximum number of Fourier modes = 50

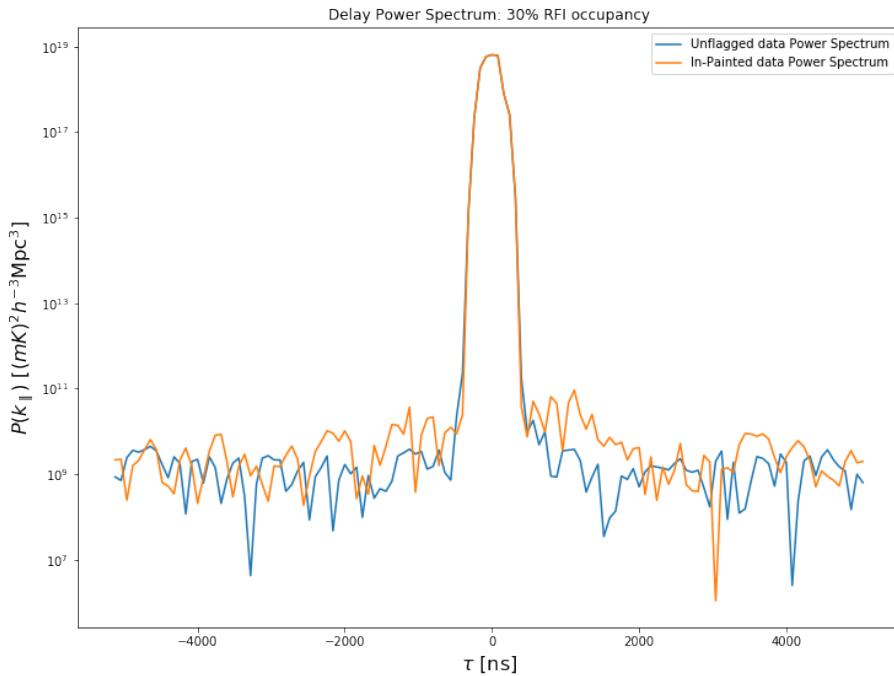


Figure 4.21: Simulation with noise and 30% RFI flagged fraction: Delay power spectrum averaged over 50 times for in-painted and un-flagged data. There is an increase in the in-painted power spectra in this case.

## 4.2 Conclusion

We started with the aim to understand the effect of RFI flagged frequencies on the delay power spectrum and to mitigate the effects by fitting a smooth model using the in-painting process. We showed that the smooth model ‘fills-in’ the RFI flagged bins and that the in-painted data is similar to the un-flagged data. The residual noise levels, from the difference of the smooth model and un-flagged data, average down over time. This shows that differences in the two datasets are within the noise levels. When the fraction of RFI flagged frequencies is 7% and 15%, we find that the power in the in-painted delay power spectra is similar to that of the un-flagged data. However, for the case of 30% RFI flagged frequencies, the in-painted delay power spectrum has significant excess power as compared to the un-flagged power spectrum. This is due to the fact that noise levels in the in-painted data increase as the number of flagged channels increase. This is in agreement with [Offringa et al. \(2019\)](#) where it was found that flagging causes excess power to scale linearly with the number of flagged samples.

We applied the in-painting process to a simulated set of visibilities with 500 point sources, a diffuse foreground, noise-like EoR signal and thermal noise with randomly added RFI flags. Having understood that the in-painted data is very similar to the ideal sky data when fraction of RFIs is less than 15%, we can apply the model to real observational data, provided the detected flagged frequencies is not higher than 15%. Although in the simulations, assumptions and parameters specific to the HERA were adopted, the in-painting process itself is based on first principles of curve fitting and therefore, can be applied to most EoR experiments including LOFAR and MWA. We also assumed the same RFI flagged frequencies for all times. However, the model is fit for each LST bin independently (while applying some time saving pre-computations) and therefore is applicable even for cases when flags are scattered in time and frequency. In general, fitting a smooth model as shown in this project is a better approach for RFI mitigation than replacing missing samples by the mean of the averaging bin. The in-painting process is also effective to be implemented in a radio data reduction pipeline because it is computationally cheap and straightforward.



# Bibliography

- Ali, Z. S., Parsons, A. R., Zheng, H., Pober, J. C., Liu, A., Aguirre, J. E., Bradley, R. F., Bernardi, G., Carilli, C. L., Cheng, C., DeBoer, D. R., Dexter, M. R., Grobbelaar, J., Horrell, J., Jacobs, D. C., Klima, P., MacMahon, D. H. E., Maree, M., Moore, D. F., Razavi, N., Stefan, I. I., Walbrugh, W. P., and Walker, A. PAPER-64 Constraints on Reionization: The 21 cm Power Spectrum at  $z = 8.4$ . *ApJ*, 809(1):61, Aug. 2015. doi: 10.1088/0004-637X/809/1/61.
- Astrophysics Science Division, NASA/GSFC. The End of the Dark Ages, 2018. URL <https://imagine.gsfc.nasa.gov/observatories/satellite/webb/darkages.html>.
- Becker, R. H., Fan, X., White, R. L., Strauss, M. A., Narayanan, V. K., Lupton, R. H., Gunn, J. E., Annis, J., Bahcall, N. A., Brinkmann, J., Connolly, A. J., Csabai, I., Czarapata, P. C., Doi, M., Heckman, T. M., Hennessy, G. S., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., McKay, T. A., Munn, J. A., Nash, T., Nichol, R., Pier, J. R., Richards, G. T., Schneider, D. P., Stoughton, C., Szalay, A. S., Thakar, A. R., and York, D. G. Evidence for reionization at [ITAL][CLC]z[/CLC][/ITAL]  $\sim 6$ : Detection of a gunn-peterson trough in a [ITAL][CLC]z[/CLC][/ITAL] = 6.28 quasar. *The Astronomical Journal*, 122(6):2850–2857, dec 2001. doi: 10.1086/324231. URL <https://doi.org/10.1086%2F324231>.
- Benson, A. J. Galaxy formation theory. *Phys. Rep.*, 495(2-3):33–86, Oct. 2010. doi: 10.1016/j.physrep.2010.06.001.
- Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., and Mahesh, N. An absorption profile centred at 78 megahertz in the sky-averaged spectrum. *Nature*, 555(7694):67–70, Mar. 2018. doi: 10.1038/nature25792.
- Bromm, V. High-redshift gamma-ray bursts from population III progenitors. *Astrophys. J.*, 642:382, 2006. doi: 10.1086/500799.
- Burke, B. and Graham-Smith, F. *An Introduction to Radio Astronomy*. Cambridge University Press, Cambridge, 1996. doi: 10.1017/9781316987506.
- Cohen, A., Fialkov, A., Barkana, R., and Lotem, M. Charting the parameter space of the global 21-cm signal. *MNRAS*, 472(2):1915–1931, Dec. 2017. doi: 10.1093/mnras/stx2065.
- Conselice, C. J. Galaxy Formation: Where Do We Stand? *arXiv e-prints*, art. arXiv:1212.5641, Dec. 2012.

DeBoer, D. R., Parsons, A. R., Aguirre, J. E., Alexander, P., Ali, Z. S., Beardsley, A. P., Bernardi, G., Bowman, J. D., Bradley, R. F., Carilli, C. L., Cheng, C., de Lera Aedo, E., Dillon, J. S., Ewall-Wice, A., Fadana, G., Fagnoni, N., Fritz, R., Furlanetto, S. R., Glendenning, B., Greig, B., Grobbelaar, J., Hazelton, B. J., Hewitt, J. N., Hickish, J., Jacobs, D. C., Julius, A., Kariseb, M., Kohn, S. A., Lekalake, T., Liu, A., Loots, A., MacMahon, D., Malan, L., Malgas, C., Maree, M., Martinot, Z., Mathison, N., Matsetela, E., Mesinger, A., Morales, M. F., Neben, A. R., Patra, N., Pieterse, S., Poher, J. C., Razavi-Ghods, N., Ringuette, J., Robnett, J., Rosie, K., Sell, R., Smith, C., Syce, A., Tegmark, M., Thyagarajan, N., Williams, P. K. G., and Zheng, H. Hydrogen Epoch of Reionization Array (HERA). *PASP*, 129(974):045001, Apr. 2017. doi: 10.1088/1538-3873/129/974/045001.

Dillon, J. S., Neben, A. R., Hewitt, J. N., Tegmark, M., Barry, N., Beardsley, A. P., Bowman, J. D., Briggs, F., Carroll, P., de Oliveira-Costa, A., Ewall-Wice, A., Feng, L., Greenhill, L. J., Hazelton, B. J., Hernquist, L., Hurley-Walker, N., Jacobs, D. C., Kim, H. S., Kittiwisit, P., Lenc, E., Line, J., Loeb, A., McKinley, B., Mitchell, D. A., Morales, M. F., Offringa, A. R., Paul, S., Pindor, B., Poher, J. C., Procopio, P., Riding, J., Sethi, S., Shankar, N. U., Subrahmanyan, R., Sullivan, I., Thyagarajan, N., Tingay, S. J., Trott, C., Wayth, R. B., Webster, R. L., Wyithe, S., Bernardi, G., Cappallo, R. J., Deshpande, A. A., Johnston-Hollitt, M., Kaplan, D. L., Lonsdale, C. J., McWhirter, S. R., Morgan, E., Oberoi, D., Ord, S. M., Prabu, T., Srivani, K. S., Williams, A., and Williams, C. L. Empirical covariance modeling for 21 cm power spectrum estimation: A method demonstration and new limits from early Murchison Widefield Array 128-tile data. *Phys. Rev. D*, 91(12):123011, June 2015. doi: 10.1103/PhysRevD.91.123011.

Dupays, A., Beswick, A., Lepetit, B., Rizzo, C., and Bakalov, D. Proton zemach radius from measurements of the hyperfine splitting of hydrogen and muonic hydrogen. *Phys. Rev. A*, 68:052503, Nov 2003. doi: 10.1103/PhysRevA.68.052503. URL <https://link.aps.org/doi/10.1103/PhysRevA.68.052503>.

Ferrara, A. First stars, reionization and Gamma-Ray Bursts. *Mem. Soc. Astron. Italiana*, 89:168, Jan. 2018.

Furlanetto, S. R., Oh, S. P., and Briggs, F. H. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. *Phys. Rep.*, 433(4-6):181–301, Oct. 2006. doi: 10.1016/j.physrep.2006.08.002.

Gunn, J. E. and Peterson, B. A. On the Density of Neutral Hydrogen in Intergalactic Space. *ApJ*, 142:1633–1636, Nov. 1965. doi: 10.1086/148444.

HERA Collaboration. HERA Technical Design, 2016. URL <http://reionization.org/science/technical-design/>.

Hubble, E. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929. ISSN

0027-8424. doi: 10.1073/pnas.15.3.168. URL <https://www.pnas.org/content/15/3/168>.

Liddle, A. *An Introduction to Modern Cosmology*. Wiley, 2003. ISBN 9780470848340.

Lidz, A., Zahn, O., Furlanetto, S. R., McQuinn, M., Hernquist, L., and Zaldarriaga, M. Probing Reionization with the 21 cm Galaxy Cross-Power Spectrum. *ApJ*, 690(1):252–266, Jan. 2009. doi: 10.1088/0004-637X/690/1/252.

Loeb, A. *The Basic Theoretical Framework*, volume 396 of *Astrophysics and Space Science Library*, page 3. 2013. doi: 10.1007/978-3-642-32362-1\_1.

Mathews, G. J., Kusakabe, M., and Kajino, T. Introduction to big bang nucleosynthesis and modern cosmology. *International Journal of Modern Physics E*, 26(8):1741001, Jan. 2017. doi: 10.1142/S0218301317410014.

Mesinger, A., Furlanetto, S., and Cen, R. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. *MNRAS*, 411(2):955–972, Feb. 2011. doi: 10.1111/j.1365-2966.2010.17731.x.

Offringa, A. R., de Bruyn, A. G., Biehl, M., Zaroubi, S., Bernardi, G., and Pandey, V. N. Post-correlation radio frequency interference classification methods. *MNRAS*, 405(1):155–167, June 2010a. doi: 10.1111/j.1365-2966.2010.16471.x.

Offringa, A. R., de Bruyn, A. G., Zaroubi, S., and Biehl, M. A LOFAR RFI detection pipeline and its first results. *arXiv e-prints*, art. arXiv:1007.2089, July 2010b.

Offringa, A. R., Wayth, R. B., Hurley-Walker, N., Kaplan, D. L., Barry, N., Beardsley, A. P., Bell, M. E., Bernardi, G., Bowman, J. D., Briggs, F., Callingham, J. R., Cappallo, R. J., Carroll, P., Deshpande, A. A., Dillon, J. S., Dwarakanath, K. S., Ewall-Wice, A., Feng, L., For, B. Q., Gaensler, B. M., Greenhill, L. J., Hancock, P., Hazelton, B. J., Hewitt, J. N., Hindson, L., Jacobs, D. C., Johnston-Hollitt, M., Kapińska, A. D., Kim, H. S., Kittiwisit, P., Lenc, E., Line, J., Loeb, A., Lonsdale, C. J., McKinley, B., McWhirter, S. R., Mitchell, D. A., Morales, M. F., Morgan, E., Morgan, J., Neben, A. R., Oberoi, D., Ord, S. M., Paul, S., Pindor, B., Pober, J. C., Prabu, T., Procopio, P., Riding, J., Udaya Shankar, N., Sethi, S., Srivani, K. S., Staveley-Smith, L., Subrahmanyan, R., Sullivan, I. S., Tegmark, M., Thyagarajan, N., Tingay, S. J., Trott, C. M., Webster, R. L., Williams, A., Williams, C. L., Wu, C., Wyithe, J. S., and Zheng, Q. The Low-Frequency Environment of the Murchison Widefield Array: Radio-Frequency Interference Analysis and Mitigation. *PASA*, 32:e008, Mar. 2015. doi: 10.1017/pasa.2015.7.

Offringa, A. R., Mertens, F., and Koopmans, L. V. E. The impact of interference excision on 21-cm epoch of reionization power spectrum analyses. *MNRAS*, 484 (2):2866–2875, Apr. 2019. doi: 10.1093/mnras/stz175.

Ouchi, M., Ono, Y., Egami, E., Saito, T., Oguri, M., McCarthy, P. J., Farrah, D., Kashikawa, N., Momcheva, I., Shimasaku, K., Nakanishi, K., Furusawa, H.,

- Akiyama, M., Dunlop, J. S., Mortier, A. M. J., Okamura, S., Hayashi, M., Cirasuolo, M., Dressler, A., Iye, M., Jarvis, M. J., Kodama, T., Martin, C. L., McLure, R. J., Ohta, K., Yamada, T., and Yoshida, M. Discovery of a Giant Ly $\alpha$  Emitter Near the Reionization Epoch. *ApJ*, 696(2):1164–1175, May 2009. doi: 10.1088/0004-637X/696/2/1164.
- Paciga, G., Albert, J. G., Bandura, K., Chang, T.-C., Gupta, Y., Hirata, C., Odegova, J., Pen, U.-L., Peterson, J. B., Roy, J., Shaw, J. R., Sigurdson, K., and Voytek, T. A simulation-calibrated limit on the H I power spectrum from the GMRT Epoch of Reionization experiment. *MNRAS*, 433(1):639–647, July 2013. doi: 10.1093/mnras/stt753.
- Parsons, A. R., Pober, J. C., Aguirre, J. E., Carilli, C. L., Jacobs, D. C., and Moore, D. F. A Per-baseline, Delay-spectrum Technique for Accessing the 21 cm Cosmic Reionization Signature. *ApJ*, 756(2):165, Sept. 2012. doi: 10.1088/0004-637X/756/2/165.
- Patil, A. H., Yatawatta, S., Koopmans, L. V. E., de Bruyn, A. G., Brentjens, M. A., Zaroubi, S., Asad, K. M. B., Hatef, M., Jelić, V., Mevius, M., Offringa, A. R., Pandey, V. N., Vedantham, H., Abdalla, F. B., Brouw, W. N., Chapman, E., Ciardi, B., Gehlot, B. K., Ghosh, A., Harker, G., Iliev, I. T., Kakiichi, K., Majumdar, S., Mellema, G., Silva, M. B., Schaye, J., Vrbanec, D., and Wijnholds, S. J. Upper Limits on the 21 cm Epoch of Reionization Power Spectrum from One Night with LOFAR. *ApJ*, 838(1):65, Mar. 2017. doi: 10.3847/1538-4357/aa63e7.
- Penzias, A. A. and Wilson, R. W. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421, July 1965. doi: 10.1086/148307.
- Pritchard, J. R. and Loeb, A. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, Aug. 2012. doi: 10.1088/0034-4885/75/8/086901.
- Rohlfs, K. and Wilson, T. *Tools of Radio Astronomy*. 01 2000. doi: 10.1007/978-3-662-05394-2.
- Rudie, G. C., Steidel, C. C., Trainor, R. F., Rakic, O., Bogosavljević, M., Pettini, M., Reddy, N., Shapley, A. E., Erb, D. K., and Law, D. R. The Gaseous Environment of High-z Galaxies: Precision Measurements of Neutral Hydrogen in the Circumgalactic Medium of z ~2-3 Galaxies in the Keck Baryonic Structure Survey. *ApJ*, 750(1):67, May 2012. doi: 10.1088/0004-637X/750/1/67.
- Tytler, D., O’Meara, J. M., Suzuki, N., and Lubin, D. Review of Big Bang nucleosynthesis and primordial abundances. *Phys. Scripta T*, 85:12, 2000. doi: 10.1238/Physica.Topical.085a00012.
- Uzan, J.-P. The big-bang theory: construction, evolution and status. *arXiv e-prints*, art. arXiv:1606.06112, June 2016.

- Vrbanec, D., Ciardi, B., Jelić, V., Jensen, H., Zaroubi, S., Fernandez, E. R., Ghosh, A., Iliev, I. T., Kakiichi, K., Koopmans, L. V. E., and Mellema, G. Predictions for the 21 cm-galaxy cross-power spectrum observable with LOFAR and Subaru. *MNRAS*, 457(1):666–675, Mar. 2016. doi: 10.1093/mnras/stv2993.
- Wang, L., Mould, J., Baade, D., Baron, E., Bromm, V., Cooke, J., Fan, X., Foley, R., Fruchter, A., Gal-Yam, A., Heger, A., Höflich, P., Howell, D. A., Kashlinsky, A., Koekemoer, A., Mather, J., Mazzali, P., Pacucci, F., Patat, F., Pian, E., Perlmutter, S., Rest, A., Rubin, D., Sand , D., Stubbs, C., Suntzeff, N., Wang, X. F., Whalen, D., Wheeler, J. C., and Yue, B. JWST: Probing the Epoch of Reionization with a Wide Field Time-Domain Survey. *BAAS*, 51(3):399, May 2019.
- Weltman, A., Bull, P., Camera, S., Kelley, K., Padmanabhan, H., Pritchard, J., Raccanelli, A., Riemer-Sørensen, S., Shao, L., Andrianomena, S., Athanassoula, E., Bacon, D., Barkana, R., Bertone, G., Boehm, C., Bonvin, C., Bosma, A., Brüggen, M., Burigana, C., Calore, F., Cembranos, J. A. R., Clarkson, C., Connors, R. M. T., Cruz-Dombriz, Á. d. l., Dunsby, P. K. S., Fonseca, J., Fornengo, N., Gaggero, D., Harrison, I., Larena, J., Ma, Y. Z., Maartens, R., Méndez-Isla, M., Mohanty, S. D., Murray, S., Parkinson, D., Pourtsidou, A., Quinn, P. J., Regis, M., Saha, P., Sahlén, M., Sakellariadou, M., Silk, J., Trombetti, T., Vazza, F., Venumadhav, T., Vidotto, F., Villaescusa-Navarro, F., Wang, Y., Weniger, C., Wolz, L., Zhang, F., and Gaensler, B. M. Fundamental physics with the Square Kilometre Array. *PASA*, 37:e002, Jan. 2020. doi: 10.1017/pasa.2019.42.
- Wilson, T. L. Techniques of Radio Astronomy. *arXiv e-prints*, art. arXiv:1111.1183, Nov. 2011.
- Yue, B., Ferrara, A., Pallottini, A., Gallerani, S., and Vallini, L. Intensity mapping of [C II] emission from early galaxies. *MNRAS*, 450(4):3829–3839, July 2015. doi: 10.1093/mnras/stv933.
- Zaroubi, S. *The Epoch of Reionization*, volume 396 of *Astrophysics and Space Science Library*, page 45. 2013. doi: 10.1007/978-3-642-32362-1\_2.