Centre for Modeling and Simulation
Savitribai Phule Pune University

Master of Technology (M.Tech.)
Programme in Modeling and Simulation

Project Report

A Simulation-based Modeling and study of the effects of Broadband RFI on Astronomical Data

Jafar Ali Habshee
CMS1721

Academic Year 2017-19

Centre for Modeling and Simulation
Savitribai Phule Pune University

Certificate

This is certify that this report, titled

A Simulation-based Modeling and study of the effects of Broadband RFI on Astronomical Data,

authored by

Jafar Ali Habshee (CMS1721),

describes the project work carried out by the author under our supervision during the period from January 2019 to June 2019. This work represents the project component of the Master of Technology (M.Tech.) Programme in Modeling and Simulation at the Center for Modeling and Simulation, Savitribai Phule Pune University.

Divya Oberoi, Reader-F
National Centre for Radio Astrophysics, (NCRA-TIFR)
Pune 411007 India

Kaushal D. Buch, Engineer-E
Giant Metrewave Radio Telescope, National Centre for Radio Astrophysics (NCRA-TIFR)
Pune 411007 India

Bhalchandra Pujari, Faculty
Centre for Modeling and Simulation
Savitribai Phule Pune University
Pune 411007 India

Mihir Arjunwadkar, Director
Centre for Modeling and Simulation
Savitribai Phule Pune University
Pune 411007 India

Centre for Modeling and Simulation
Savitribai Phule Pune University

Author's Declaration

This document, titled

A Simulation-based Modeling and study of the effects of Broadband RFI on Astronomical Data,

authored by me, is an authentic report of the project work carried out by me as part of the Master of Technology (M.Tech.) Programme in Modeling and Simulation at the Center for Modeling and Simulation, Savitribai Phule Pune University. In writing this report, I have taken reasonable and adequate care to ensure that material borrowed from sources such as books, research papers, internet, etc., is acknowledged as per accepted academic norms and practices in this regard. I have read and understood the University's policy on plagiarism (http://unipune.ac.in/administration_files/pdf/Plagiarism_Policy_University_14-5-12.pdf).

Jafar Ali Habshee
CMS1721

Dedicated to my Brother, Mother and Father.

Abstract

The Giant Metrewave Radio Telescope (GMRT), located at Khodad, Pune is one of state of the art facilities used by scientists all over the world for doing Radio Astronomy. In addition to the astronomical signals the highly sensitive parabolic reflectors pick up unwanted signals from the radio frequency interference (RFI) sources present in the telescopes operating area. These signals can severely corrupt observation data, leading to inaccurate or misleading results. Signals from moving sources like farm machinery, vehicular traffic, cell phone and signals from constant sources like power-line interference and radiation from electric fences are examples of RFI sources.

High tension (HT) powerlines in the vicinity of the GMRT antennas produce RFI harmful to radio astronomy observations. It was noted that measurements made by Swarup and Raybole in April 2008 showed that the power-line RFI seems to be much higher than that recorded in 1998 [1]. It is known that severe radio noise (RFI) is generally caused by corona discharges on the HT lines of voltage greater than 65 kV. For lines of lower voltage, say of 11 kV and 33 kV, gap discharges at the insulators located at the electrical poles, poor grounding of the support arms and loose contacts in the joints gives rise to severe RFI. Similarly, RFI is also observed near the transformers of the irrigation pumps located near the antennas of the GMRT.

The aim of this study is to do a robust statistical modeling the RFI pulses through simulation and quantitatively estimate its effect on the measured cross correlation functions. It is observed that the profile of the RFI is a function of numerous physical parameters, hence modeling them from a single distribution will not serve the purpose. The pulses / signal are modelled using various statistical measures the majority of which are user defined, and enable us to mimic the characteristics of the RFI as observed at GMRT. In parallel a simulation of 2-element radio interferometer is also designed which would take these modeled/artificial signal with RFI as inputs and see its behaviour on the correlated outputs which are obtained. The understanding thus obtained would be used to scrutinize our work on real-time RFI filtering by optimizing RFI filtering techniques and also provide an upper-bound on the efficacy of the existing techniques.

Acknowledgements

Many thanks!

I would like to express my gratitude to my eminent guides at the National Centre for Radio Astrophysics (NCRA) and Giant Metrewave Radio Telescope (GMRT), Dr. Divya Oberoi and Mr. Kaushal Buch, who gave me the opportunity to work with them and learn with them. Despite my limited knowledge and experience of the field, Dr. Divya Oberoi agreed to take me for this project. The door to Dr. Divya Oberoi's office was always open whenever I ran into a trouble spot or had a question about my approach and implementation. Through his busy schedule he has always found time and presented his work around approach in such a crisp manner that was enough to keep me engaged and wrap my head around his methods. It was through his methods that I was able to catch the subtleness of this work and understand a large part of Radio Astronomy techniques. Is is clear to me that this opportunity has cleared many academic and professional roadblocks that a recent graduate would encounter.

Anyone who has been around Mr. Kaushal Buch knows that how convivial his nature is. Me and Mr. Kaushal majorly had our conversation over phone but it was during my brief stint at GMRT, I got the chance to know him and learn a major portion of this project. By and large his conceived ideas, technical discussions over tea and implementation skills have steered the course of this project through headwinds. He provided me with all the necessary physical and non physical amenities that would become an integral part of this work. At GMRT he hosted me at his dinner tables with our never ending discussions on music, language and definitely this work as well.

I am grateful to my guide at the Centre for Modeling and Simulation (CMS), Dr. Bhalchandra Pujari who chose to guide me with this course. He never failed to amaze me with his skills and frolic comments. During our project meetings, he terrifically hit the sweetest spots with his valuable comments and suggestions. Dr. Pujari was a constant encouragement for me and played a pivotal role for this study to rise to existence. Thanks to him that he gave me a chance to be a part of interesting side projects that I pursued during this course which cemented my understanding of code logic and mathematics. I can swear by the Old Gods and the New that his lectures on ***Transforms*** and ***Complex Analysis*** have proven to be the First Men and Protectors of this project.

When I go back to the days when I started at CMS, the subjects taught by Dr. Sukratu Barve, Dr. BhalCandra Gore, Dr. V.K. Jayaraman, Dr. Rajeev, Dr. Aamod Sane, Dr. Dinesh Nehete, Dr. Snehal Shekatkar, Dr. Ankita Katre, Dr. T.V. Ramanathan, Dr. Akanksha Kashikar and Dr. Mihir Arjunwadkar have entrenched my learning, insights and abstract understanding of the realistic problems. I find myself privileged to have known and work with them. Because of them I have learned how a given problem can be branched in many smaller parts and gunned down easily! Though going through CMS was a tough time, but when I compare when I started

then and what I am now, I see only gain and improvement. This growth is not just academic, but I have grown as a mentor, a speaker and a much more confident person.

I would like to thank the staff of the CMS, Prachi Gaikwad, Satish Dada, Mausi, Gajanan Rakhunde, Mrunalini Dharmadhikari, Vaibhav Rajkumar for providing me with all that I could ask for. CMS has provided a 24 hour open premises with all accessible facilities which has been a great enabler in making my research skills reach a level to handle a project independently.

I would also like to thank my mates at NCRA, Arpit Behra, Rajiv Rathour, Avishek Basu, Minhazur Rehman and Krishna Shinde, with whom I had many eye opening discussions about life and physics! GMRT is a quiet and serene place in a small village but I found people sharing my niche and hosting me warmly. I am hugely indebted to Bela Dixit, Kshitija Pawar and Gotya for not letting me bang my head just on work. Though for a short span I had also spent my time with Dr. Narendra Patra (RRI), Dr. Kshitija Kelkar (RRI) and Mr. Shogato Chaterjee at GMRT. I am grateful to them as well.

I cannot less thank the *luminiferous ethers* of my life Adhreja Dey, Divye Singh, Ganesh Shinde, Anushree More, Kanchan Kalunge, Ashwini Mali, Abhishek Daundkar, Srinath Birajdar, Ankita Patel, Yogita Manka, and Rishi Bhosle for their constant and untiring efforts throughout this project to pull me out of my chair and push me back again into it!

I would as well take a moment to thank Led Zeppelin, Lucky Ali and late Louis Armstrong who had made all the necessary moments that I needed.

Contents

Chapter 1

Introduction

Astronomy is one of the oldest sciences in existence, though the radio sky was not discovered before 1931, when Karl Jansky for the first time detected the radio signals from our Galaxy. This event marked a new beginning for astrophysics research. A lot of new discoveries followed, many of which had a large impact on the field of astronomy and our knowledge of the Universe, such as the detection of the hydrogen spectral line at a wavelength of 21 cm, discovery of pulsars, detection of Fast Radio Bursts and the latest imaging of M87 Black Hole. The field of radio astronomy has been flourishing for decades, and scientists continue to stretch the limits on the possibilities of observing in the radio domain. Modern telescopes like the Giant Metrewave Radio Telescope (GMRT), Very Large Array (VLA), Low-Frequency Array (LOFAR) are incredibly sensitive machines which observe the sky in great depth and detail to generate high quality radio images. With technical advancements, the observing bandwidth of these telescopes have also increased proportionally. These radio telescopes are passive devices, which only receive radio signals and do not transmit any radiation themselves. Due to the very long distances between astronomical objects and the telescopes, the radiated signals from these objects arriving at the receivers are extremely weak. This necessitates the use of very sensitive receivers to detect them above the noise floor level. Due to increased the bandwidth the radio spectrum has become more and more crowded and seriously disturbs radio observations. This would either impede observations altogether or lead the astronomers into making erroneous data interpretations.

1.1 Radio Signals and Interference

1.1.1 Cosmic Sources

The radio frequency emissions from cosmic sources arise due to underlying random processes happening within sources due to interactions with their near-environment. As this signal is generated independently from various parts of a large object, at different distances from our location, and containing many frequencies, this signal is completely random in character. Therefore, applying the Central Limit Theorem, cosmic source signals are assumed to be centered, temporally i.i.d. with a Gaussian distribution. All the processes involve the movement of electrons, in particular, changes in their velocity and losing energy, which is radiated away as a radio wave. The record of the electric field $E(t)$, received at a point on earth from a source of radio waves can be called a signal, so long as we do not take this to imply intelligence at the transmitting end [2].

1.1.2 Radio Frequency Interference (RFI)

Interference is a very broad term that depends on the application. Basically, an interference is any kind of signal that corrupt the signal of interest. RFI for radio astronomy can be of various kinds. Impulsive thunderstorm emissions, emissions from household appliances, aircraft distance measuring equipment, cellular phones and overhead powerline emissions are typical examples of RFI of sources. Most of the persistent interference signals encountered in radio astronomy are due to constant/non-mobile man-made sources like home automation, power-line faults etc. The main powerline radiation are due to corona and gap-type discharges from a nearby electric fence, sparking noise, and local interference from the telescopes instrumentation. The signals from these instruments are much stronger than the feeble astronomical signals and affect the observations adversely.

Corona noise/discharge

The corona discharges occur mostly due to the presence of irregularities on the conductor surface. These could be some form of a defect, caused by damage or poor design, which protrudes above the conductor surface and thus increasing the electric field intensity around the conductor. Water droplets, insects, snow, and dust/dirt formed on the conductor surfaces form bulges and irregular structures which can result in corona discharge. This conductor corona noise is generated due to partial electrical discharge caused by ionization of air on the surfaces of electrical conductors of high voltage powerlines. This ionization generates an avalanche of electrons, which carry electric charges into the surrounding air. The presence of corona is characterized by a violet/bluish-coloured light, due to free photons released during the ionization process. An audible noise in the form of a hissing sound is also produced during the corona discharge process. This kind of radio interference can be seen in the time domain as repetitive pulses having a fast rise time and short duration [3].

Spark Gap Noise

The principle behind the generation of a spark gap electric discharge is that when the potential difference between the conductors exceeds the breakdown voltage of the gas within the gap. First the gas becomes electrically conducting, the electric current heats up the gas, causing the formation of a very high impulsive signal. As this insulation breaks down an avalanche of electrons cross the ionized medium, this results in the voltage potential across the gap between the two surfaces being temporarily diminished and the spark is stopped. The breakdown is extinguished for a short time before the potential across the gap starts to increase again to the breakdown point, thus generating another spark. The electrodes involved in sparking could be a guy wire, a loose nut, a corroded wire on the line, or a loose clamp or connection on a power line device. The gaps are most commonly found on loose, un-bonded and/or corroded hardware on distribution wooden structured power lines. Micro-sparking results from a very small spark, as small as 0.02 mm, while sparking results from electrical breakdown between two conducting parts that are separated by a gap that is as small as a fraction of a mm or up to 20 mm long [3].

1.2 Motivation of project

Understanding the RFI and its properties has become one of the key requirement for radio astronomers. A successful observation would require analysis of sky signals in their most pristine form. Through this project we would try to understand the various aspects of RFI and how it corrupts the astronomical signals. We would try to quantify some statistical properties and

try to regenerate statistically similar signals which would mimic the actual RFI signals to a high degree of accuracy. The reproduced signals are expected to have similar properties as its parent. Every observed RFI signal has some distinct shape and different underlying dominant frequencies. Thus these signals, when represented on the time-frequency plot show bizarre patterns. Understanding these patterns could let the scientists know the underlying behaviour of a particular RFI source. To facilitate such an endeavour, this project could potentially serve as a facility which would provide the ability to generate realistic RFI affected signals as well as algorithms to characterise and identify the RFI.

Chapter 2

Radio Astronomy: Fundamentals

Radio astronomy is the study of celestial objects that emit radio waves. With radio astronomy, we study astronomical phenomena that are often invisible or hidden in the optical and other portions of the electromagnetic spectrum. Many objects in the universe, including stars, galaxies, and nebulae, as well as a wide variety of peculiar, fascinating, and often mysterious objects, emit radio waves through naturally occurring processes. Radio waves are generated by several phenomenon, depending on the physical conditions in the radio-emitting object as outlined in Section 1.1.1.

The frequency band within which radio waves can be transmitted through the earth's atmosphere without significant attenuation extends from about 10 MHz to 100 GHz and enables radiation in this range from celestial radio sources to be picked up by radio telescopes on the Earth.

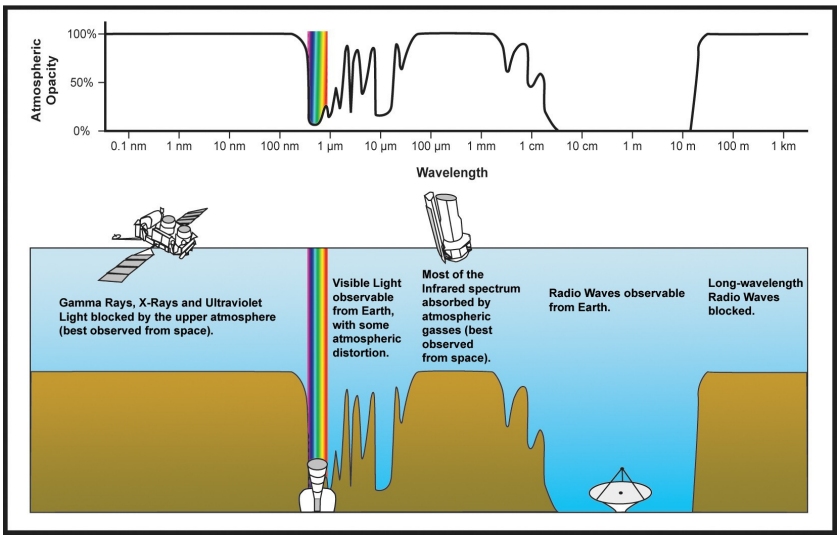


Figure 2.1: Radio Window source: coolcosmos

2.1 Equations and Fundamentals

Consider electromagnetic radiation from the sky falling on a flat horizontal area A at the surface of the Earth, as shown in the elevation view of Figure 2a and perspective view of Figure. 2b. The infinitesimal power dW from a solid angle $d\Omega$ of the sky incident on a surface area dA is

given by [4]

$$dW = B\ cos\theta\ d\Omega\ dA\ d\nu \tag{2.1}$$

where,

dW = infinitesimal power, $watts$
B = Brightness of sky at position of $d\Omega$, $watts\ m^{-2}\ Hz^{-1}\ rad^{-2}$
$d\Omega$ = infinitesimal solid angle of sky ($= sin\theta\ d\theta\ d\phi$), rad^2
θ = angle between $d\Omega$ and zenith, rad
dA = infinitesimal area of surface, m^2
$d\nu$ = infinitesimal element of bandwidth, Hz

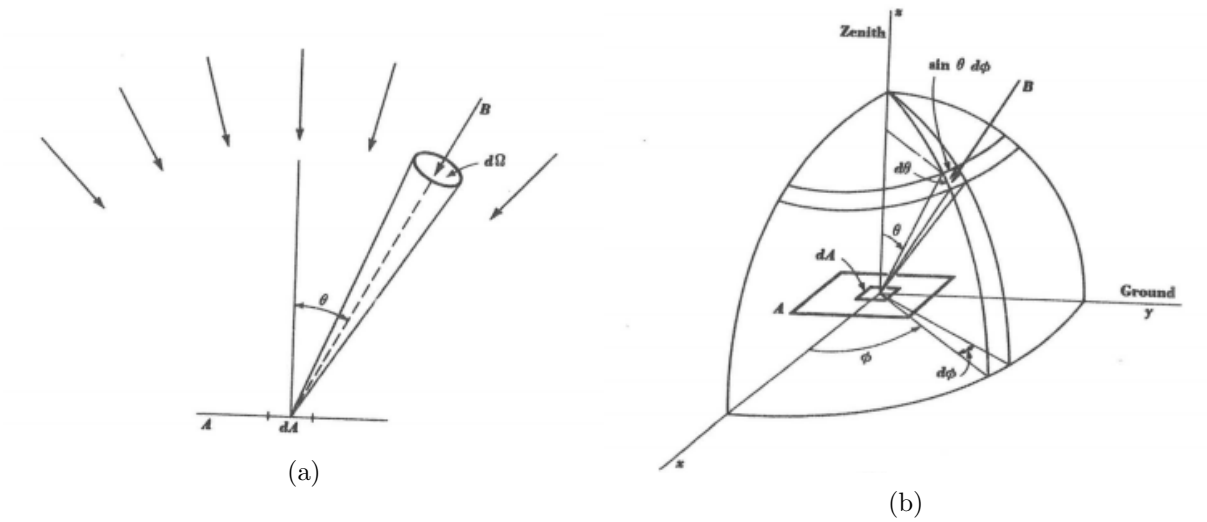


Figure 2.2: Brightness distribution source: Radio Astronomy by Kraus

Integrating (1.1), one can obtain the power W received over a bandwidth $\Delta\nu$ from a solid angle $d\Omega$ of the sky. Thus,

$$W = A \int_{\nu}^{\nu+\Delta\nu} \int_{\Omega} B\ cos\theta\ \ d\Omega\, d\nu \tag{2.2}$$

In the above equation the power received is considered independent of the area.

$$B' = \int_{\nu}^{\nu+\Delta v} B\ d\nu \tag{2.3}$$

where,

B′= Total Brightness $watts\ m^{-2}\ rad^{-2}$
B = Brightness per unit bandwidth, $watts\ m^{-2}\ Hz^{-1}\ rad^{-2}$
$d\nu$ = infinitesimal bandwidth, Hz
ν = frequency, Hz
$\Delta\nu$ = finite bandwidth, Hz

In many situations the power per unit bandwidth is more pertinent than the power contained in an arbitrary bandwidth $\Delta\nu$. This power per unit bandwidth is often called the spectral power, since its variation with frequency constitutes the power spectrum. Thus, introducing the concept of spectral power, (1.1) becomes,

$$dW' = AB\ cos\theta\ d\Omega \tag{2.4}$$

$$w = AB \iint_{\Omega} cos\theta \, d\Omega \tag{2.5}$$

Since the sky brightness varies with direction, it may be expressed explicitly by the functions of angles in the sky as $B(\theta, \phi)$. Thus, from (2.5) the spectral power, or power per unit bandwidth, becomes,

$$w = A \iint_{\Omega} B(\theta, \phi) \; cos\theta \, d\Omega \tag{2.6}$$

$$w = \frac{1}{2} \; A \iint_{\Omega} B(\theta, \phi) \; P_n(\theta, \phi) \; cos\theta \, d\Omega \tag{2.7}$$

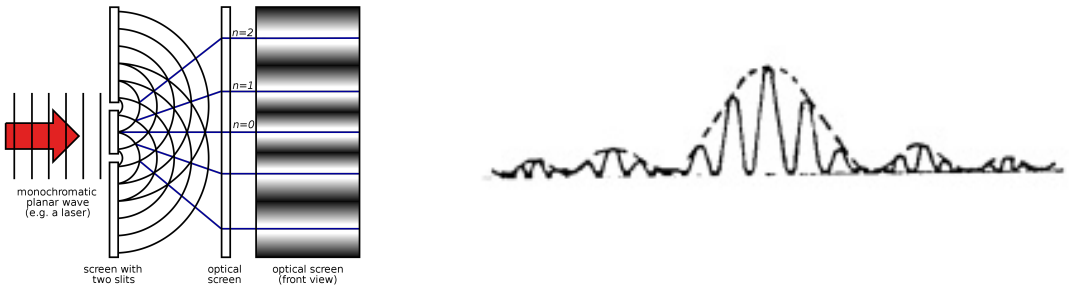
where,

$P_n(\theta, \phi)$ = normalized power pattern of antenna, dimensionless

If the incident radiation is of an incoherent, unpolarized nature then only one half of the incident power will be received, since any antenna is responsive to only one polarization component. A factor $\frac{1}{2}$ introduced in (2.7) is for this context.

2.2 Interferometry Principle

Radio interferometry is based on the principle of interference of light. When waves (generally Electromagnetic (EM) Waves) are superimposed to produce the fringe pattern which can be used to extract meaningful information. This works because when two waves with the same frequency combine, the resulting intensity pattern is determined by the phase difference between the two waves. Waves that are in phase will undergo constructive interference while waves that are out of phase will undergo destructive interference. Waves which are not completely in phase nor completely out of phase will have an intermediate intensity pattern, which can be used to determine their relative phase difference. A general idea can be taken from the Young's double slit experiment. When the wavelets from the two monochromatic sources combine, it produces light and dark pattern on the screen.



(a) Young's Experiment

(b) Fourier transform of a $sinc^2$ function

Figure 2.3: Young's experiment analogy with the Fourier transform. source: shmoop.com

The intensities of the light and dark bands keeps on decreasing as one moves away from centre of the screen as shown in Figure 2.3. This is analogous to taking a Fourier transform of two square wave functions and plotting their intensities. The $(\frac{sin\ x}{x})^2$ function also displays a similar mathematical behaviour in the Fourier space. In a similar fashion a two element radio telescope takes up the EM waves combines them in a processing unit and produces a pattern.

If a single antenna (radio telescope) can do the job then why do we need an array of such telescopes? The answer here is angular resolution. The angular resolution is defined as the ability of a telescope to separate and distinguish details of far away objects which it sees. For example take three images in Figure 2.4:

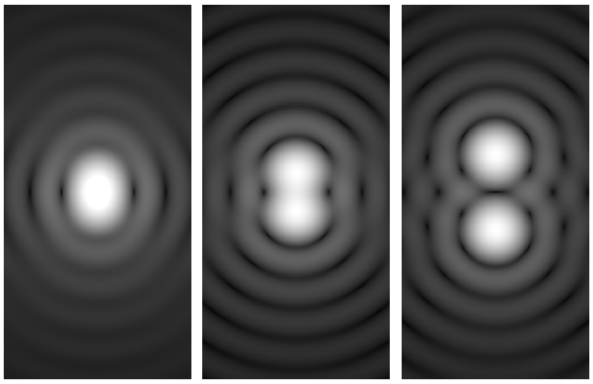


Figure 2.4: An illustration of resolution of a source object source: Wikipedia

On moving form left two right, a clear and distinct transition can be seen on how the sources are separated (resolved). From a far away distance the two sources would appear as a single source. However, as one moves near a clear distinction between the objects is noticed. The angular resolution of telescope of any lens is given by,

$$\theta_{res} \simeq 1.22\frac{\lambda}{D} \quad radians \tag{2.8}$$

where,

λ is wavelength of the EM wave.
D is diameter of lens.

As the diameter D of the telescope increases, an instrument can resolve an object to much finer detail. For an interferometer the effective size of the 'lens' correspond to the separation between the farthest dishes. Therefore an array of such telescopes is used to increase the effective aperture diameter.

A single radio antenna is referred to as an *element* in an interferometer array. The vector distance between two such elements is called a *baseline*. A two element interferometer consists of two radio dishes, the signals from these two dishes are multiplied, accumulated a stored in form of complex numbers termed as complex visibilities. In principle multiplication and accumulation of signals means finding the correlation between the signals. The simplest way to measure the correlation between the two signals is to simply let them interfere.

$$V_{12} = \langle E_1^*(t)E_2(t)\rangle \tag{2.9}$$

where,

V_{12} is complex visibility
E_1^* and E_2 are signal from antenna 1 and antenna 2.

Weiner Khinchin Theorem

From the principles of Fourier transform we know that the auto-correlation function is the Fourier transform of the power spectrum. This is known as the Wiener-Khinchin Theorem [2]. For any signal $x(t)$, its power spectrum $S_{xx}(f)$ can be defined as,

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft}\, dt \tag{2.10}$$

$$S_{xx}(f) = |X(f)|^2 \tag{2.11}$$

$$r(\tau) = \int_{-\infty}^{\infty} S_{xx}(f)\; e^{i2\pi f\tau}\; df \tag{2.12}$$

where,
$X(f)$ is the Fourier transform of $x(t)$.
$r(\tau)$ is the auto correlation function.

If one needs to find the cross correlation of the signal, then taking the Fourier transform of the cross-correlated spectral power density would give the desired result. The cross correlated power spectral density of two signals in the transform domain would be $X_1(f) \cdot X_2^*(f)$.

Van Cittert - Zernike Theorem

The Van Cittert-Zernike theorem is a generalization of the Wiener-Khinchin Theorem. Suppose one cosmic point source emits an electromagnetic field received on two antennas located on two spatial positions **u** and **v** as $E(u,t)$ and $E(v,t)$. The mutual coherence between two locations defined by the spatial vectors **u** and **v** is defined as [5] [6],

$$\Gamma(u,v,\tau) = \langle E(u,t,\tau).E^*(v,t,\tau)\rangle \tag{2.13}$$

In other words this mutual coherence function is also called the cross-covariance function and the cross correlation function can be obtained by,

$$\gamma(u,v,\tau) = \frac{\Gamma(u,v,\tau)}{r_{uu}r_{vv}} \tag{2.14}$$

where,
$\gamma(u,v,\tau)$ is the cross correlation function at lag τ.
$\Gamma(u,v,\tau)$ is the mutual coherence or cross co-variance function.
r_{uu} is the auto-correlation of signal $E(u,t)$.
r_{vv} is the auto-correlation of signal $E(v,t)$.

The coherence function $\Gamma(u,v,0)$ is linked to an intensity spatial distribution, e.g. the skymap, following the Van Cittert-Zernike Theorem:

$$\Gamma(u,v,0) = \iint I(l,m) e^{-i2\pi(ul+vm)}\, dl\, dm \tag{2.15}$$

where,

l and m defining spatial coordinates in the field of view of the instrument.
u and v are the position vectors of radio dishes on earth.

2.3 GMRT Antenna Positions and its uv coverage

The interferometery technique takes the advantage of earth's rotation to make a radio map of the sky. The position vectors u and v are said to be in a $u-v$ plane. There is third vector as well ,w which is directed towards the source.

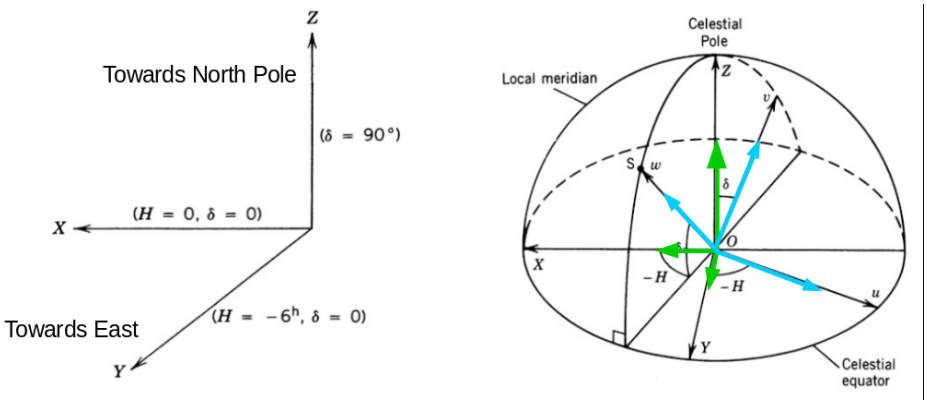


Figure 2.5: uvw reference frame source: Low Frequency Radio Astronomy

Relationships between the $X\ Y\ Z$ and $u\ v\ w$ coordinate systems. The $u\ v\ w$ system is defined for observation in the direction of the point source S, which has hour angle H and declination δ. In the above Figure 2.5, S is in the eastern half of the hemisphere and H is therefore negative. The direction cosines in the transformation matrix follow from the relationships in this equation [7].

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} \sin H & \cos H & 0 \\ -\sin\delta\cos H & -\sin\delta\sin H & \cos\delta \\ \cos H\cos\delta & -\cos\delta\sin H & \sin\delta \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.16}$$

Here H and δ are usually the hour angle and declination of the phase reference position.

2.3.1 uv Coverage of GMRT

The following Figures 2.6 to 2.9 show the u-v coverage of the GMRT antennas for few observing durations and declinations. The figures are plotted only for the positive plane. The correspondng conjugate points are not plotted.

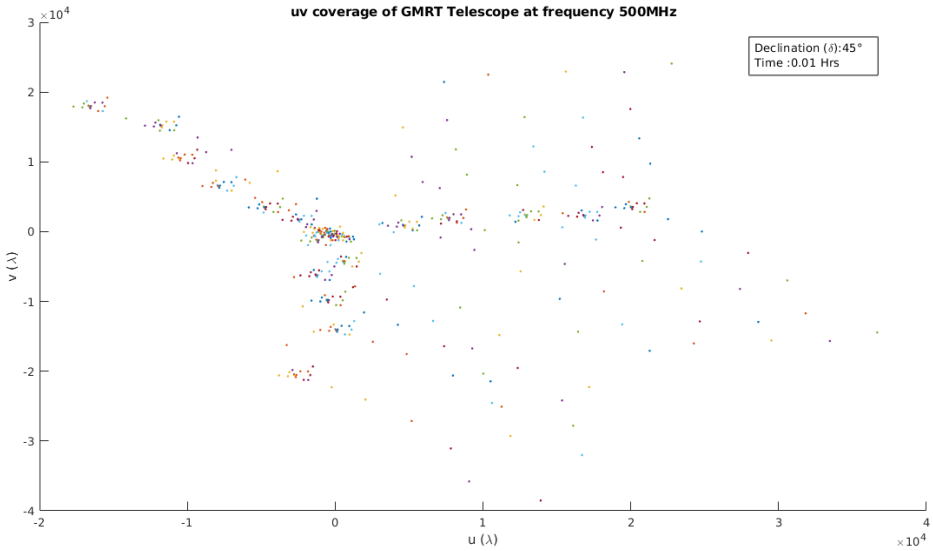


Figure 2.6: uv snapshot at 45° declination and 0th time frame

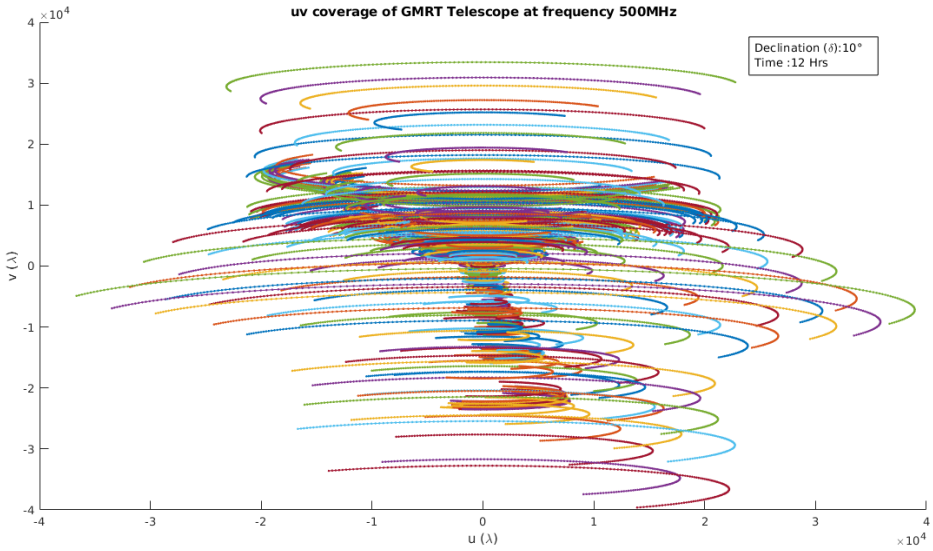


Figure 2.7: uv snapshot at 10° declination and 12 hr observation

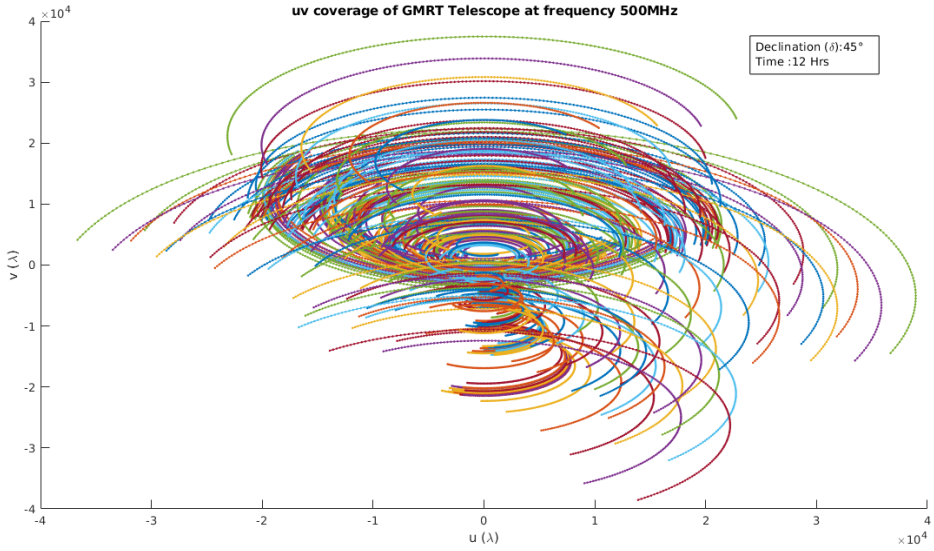


Figure 2.8: uv snapshot at 45° declination and 12 hr observation

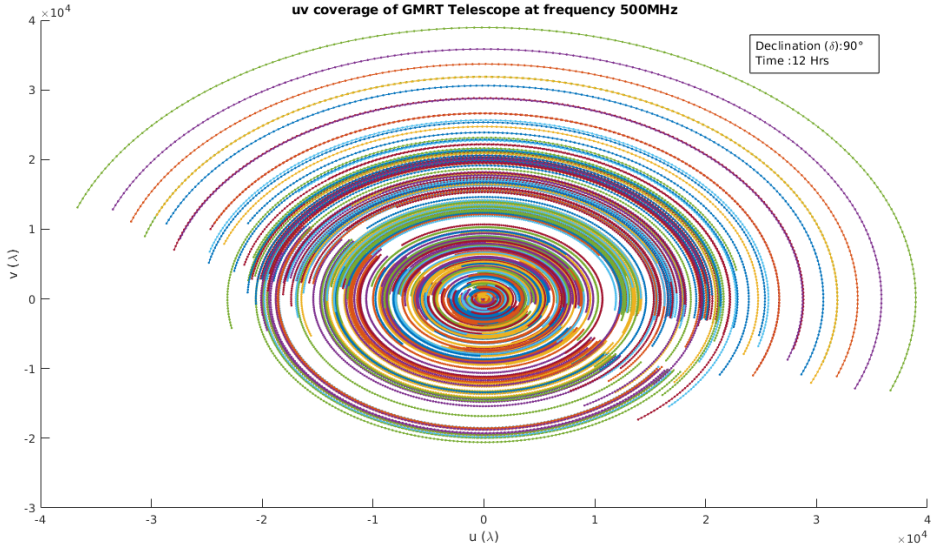


Figure 2.9: uv snapshot at 90° declination and 12 hr observation

Chapter 3

Interferometer Components

For a point source at infinity the sky signals received on the widely separated dishes have a plane wave front incident on them. The signals received on the dishes will have a time delay τ which is caused due to the relative positions of the dishes with respect to the wavefront of the distant source. Figure 3.1 shows the basic schematic of a two element interferometer. When the signals reach the correlation stage, they have to be corrected for this delay and the mutual coherence function is computed at 0 lag. However, before reaching the correlator the signals pass through a chain of instrumentation starting from the pre-amplifier, low noise amplifiers, pass band filters, local oscillator, sample digitizer and filter banks. A detailed schematic of the Legacy GMRT receiver is depicted in Figure 3.2

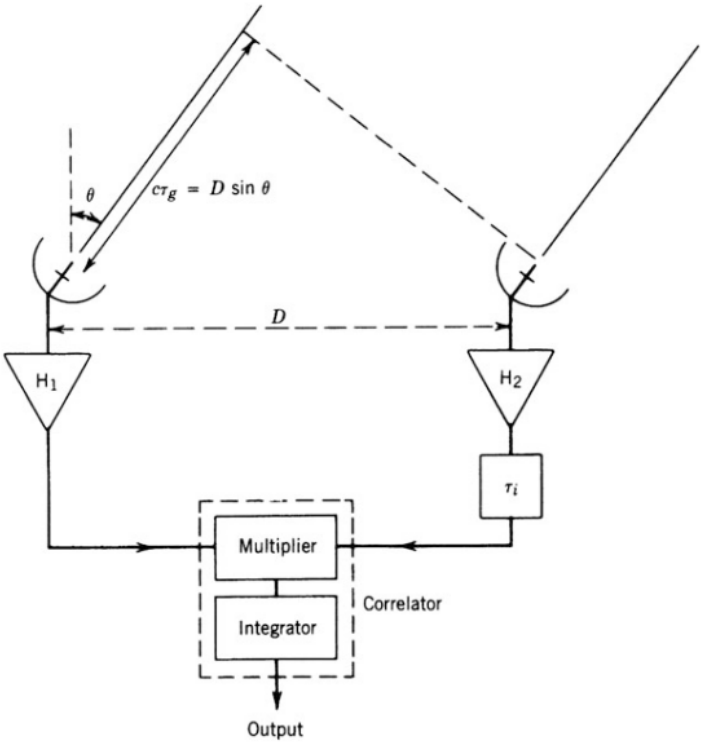


Figure 3.1: Elementary interferometer schematic diagram source: ShoutWiki

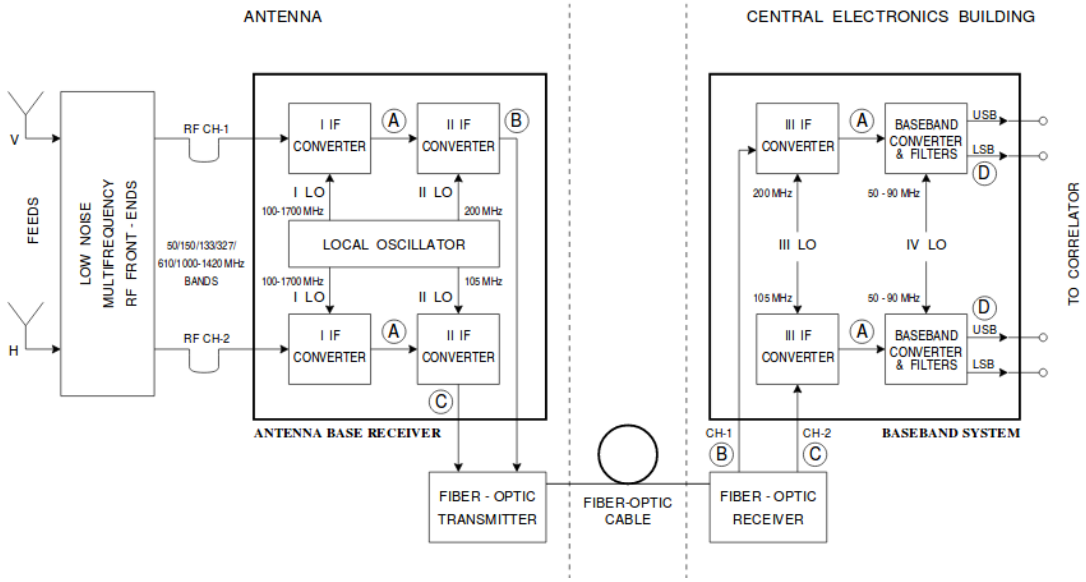


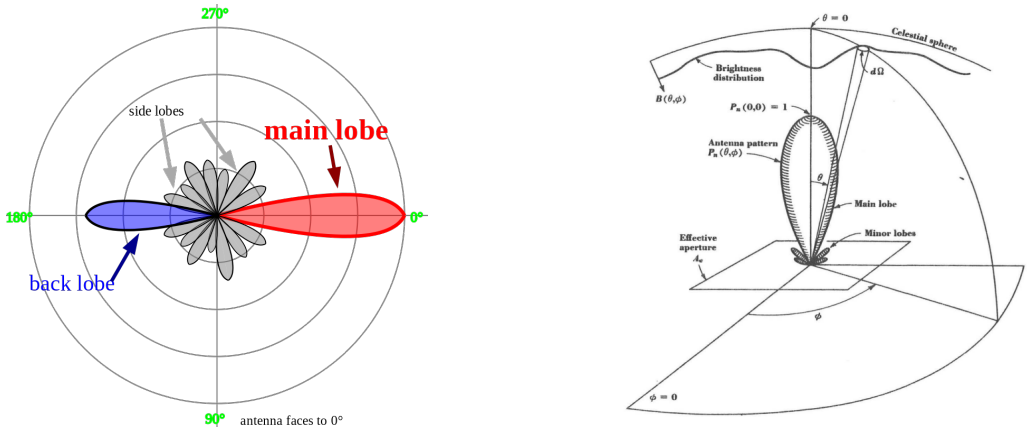
Figure 3.2: Legacy GMRT Receiver System source: Low Frequency Radio Astronomy

3.1 Antenna

The antenna is the interface for collecting the radio waves and then passing it to the receiver system for data processing. This antenna is analogous to a mirror as used in an optical telescope. A radio telescope uses large dishes to increase the effective aperture area. The radio telescopes are called passive antenna as they do not transmit any signals. The frequency range over which the GMRT dishes operate is 50-1500MHz.

3.1.1 Radiation Pattern and Antenna Directivity

The radiation pattern of an antenna is a plot of the relative field strength of the radio waves emitted by the antenna at different angles in the far-field [8]. It is typically represented by polar plots of the horizontal and vertical cross sections. The pattern of a highly directional antenna is represented in the Figure 3.3.



(a) Radiation Pattern

(b) 3D Radiation Pattern

Figure 3.3: Radiation Pattern of Antenna source: AntennaTheory.com

An antenna has a directivity of unity if it's radiation pattern is same in all the directions. Directivity of an antenna is a measure how much power density it radiates is concentrated in a single direction. It is measured as how much power the antenna radiates in a single direction as compared to an isotropic antenna.

3.1.2 Antenna Gain

Antenna gain is a measure of how well an antenna converts a radio wave arriving from a specific direction. A high-gain antenna will radiate/receive most of its power in a particular direction, while a low-gain antenna will radiate/receive over a wider angle. The antenna gain, or power gain of an antenna is defined as the ratio of the intensity radiated by the antenna in the direction of its maximum output, at an arbitrary distance, divided by the intensity radiated at the same distance by a omni-directional antenna [8].

3.1.3 Aperture Efficiency

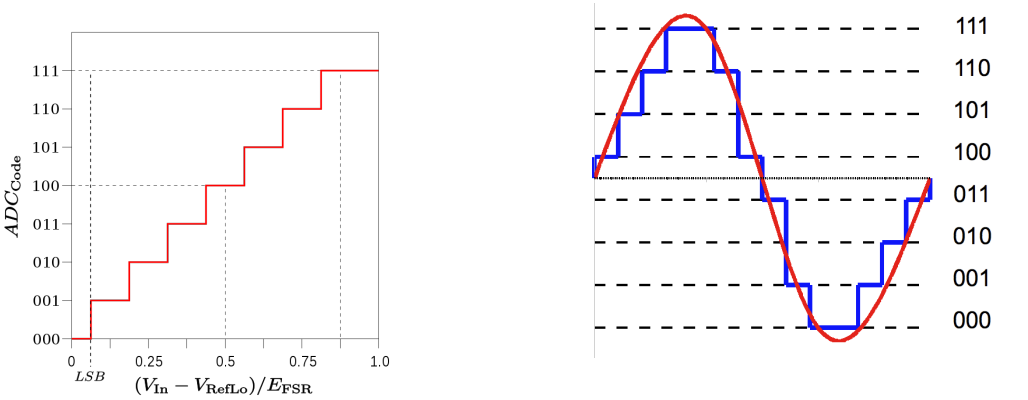
A parameter for calculating the receiving power of an antenna is the effective area or effective aperture. Aperture efficiency measures how much the antenna comes close to convert all the radio wave entering its physical aperture.

3.2 Filters

A filter is a device that removes some unwanted frequencies or features or amplitudes from a signal. Filtering results in partial suppression of the signal allowing only the signals of interest to paas through the receiver chain. Some common filters used in signal processing are low pass, high pass and bandpass filters. The low pass filters allow frequencies below some specific threshold/cut-off frequency. The high pass filter allow signals which are above a specific cut-off frequency. A band-pass filter passes frequencies within a certain range and rejects frequencies outside that range. A bandpass can also be formed using a combination of a low pass and a high pass filter [9].

3.3 Analog to Digital Converter

A radio antenna receives continuous, time varying sky signals, these are converted to discrete voltage amplitude form using an Analog to Digital Converter (ADC). The ADC converts the analog signals to digital by quatization. Quantization defines a mapping from the uncountable continuous domain to a discrete countable domain [10]. To decrease the quantization error the resolution of the quantization is kept as higher as possible however, it is application dependent as well. The resolution of the converter indicates the number of discrete values it can take over the range of analog values. Though any number of quantization levels is possible, common word-lengths are 8-bit (2^8 levels), 16-bit (2^{16} levels) and 24-bit (2^{24} levels).

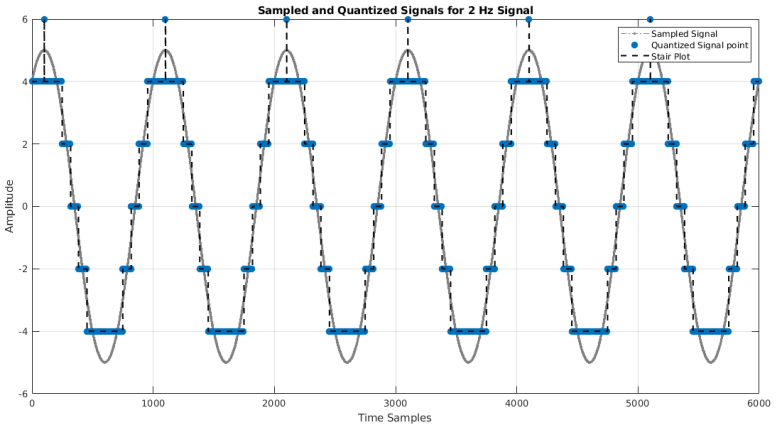


(a) 3 Bit Quantization levels

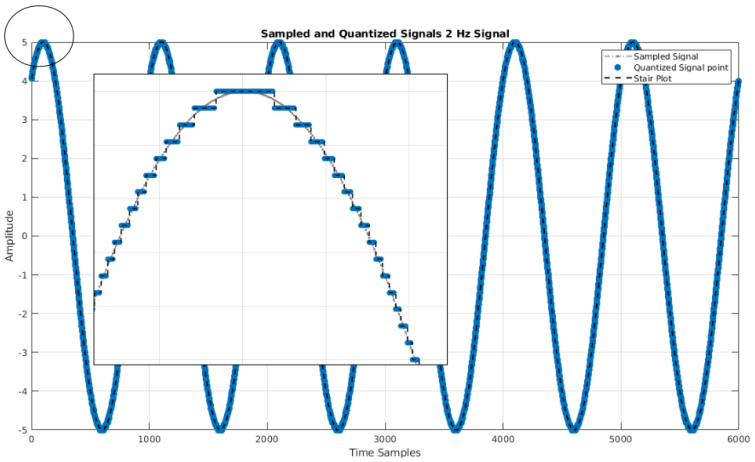
(b) Quantization of an example wave

Figure 3.4: 3 bit Quantization representation. source: Wikipedia

The Figures 3.4b and 3.5 shows the quanization level distribution uniformly over the sin wave. However the quantization can spread unevenly as well. For the purpose of this work we have considered uniform quantization only.



(a) 3 Bit Quantization



(b) 8 bit Quantization

Figure 3.5: Quantization representation of sin wave.

The artificial signals generated are quantized as per Figure 3.6. The RFI signal generated for

below graph has a very wide normal distribution with σ as 50 units. Choosing the number bits is a very crucial element as it is decisive of the amplitude range of the signal (Section 4.4).

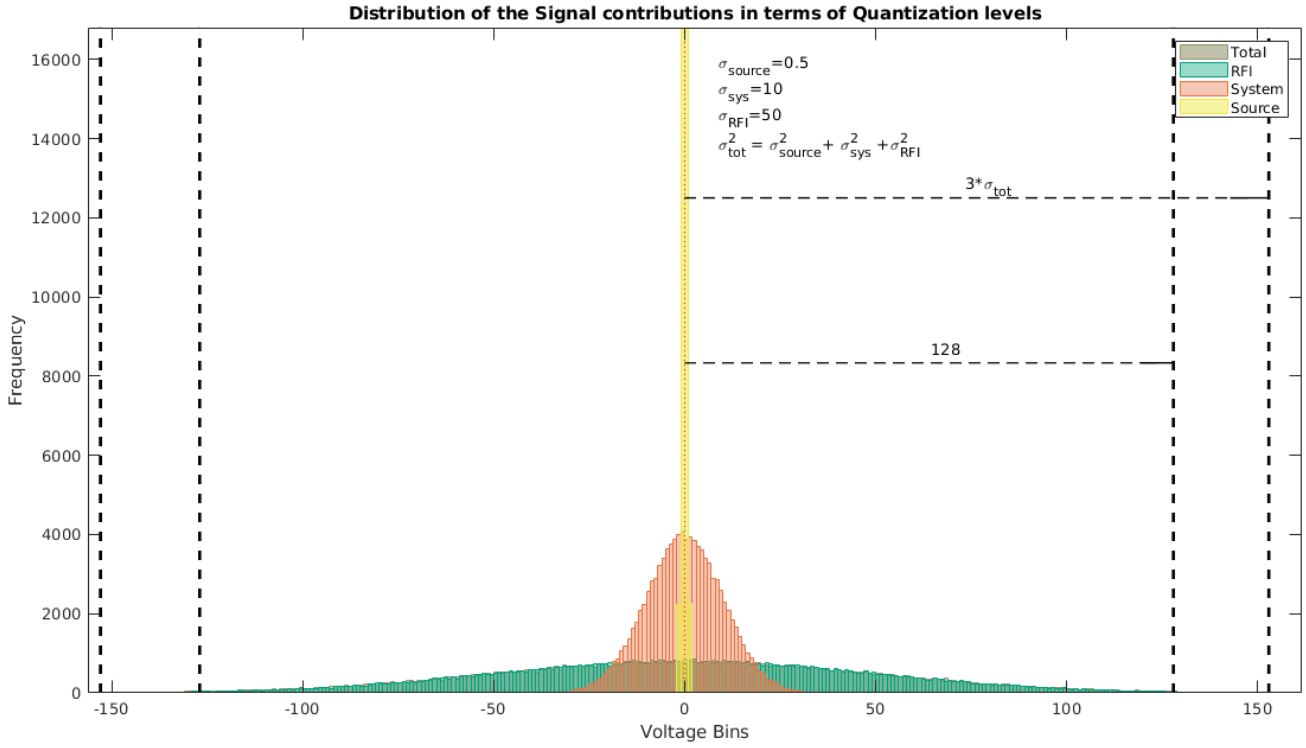


Figure 3.6: Quantization of Signal Components

3.4 A primer to Fourier Transform

Fourier decomposition of a function would result in all the basis functions that construct that particular function. It is powerful mathematical tool that breakdowns a function in to sin and cosine. If the input curve is in time domain then the Fourier transform results in frequencies constituting that time signal [11].

$$X(f) = \int_{-\infty}^{\infty} f(t) e^{-i2\pi ft}\, dt \tag{3.1}$$

$$f(t) = \int_{-\infty}^{\infty} X(f) e^{i2\pi ft}\, df \tag{3.2}$$

3.4.1 The Fast Fourier Transform

The Fast Fourier Transform (FFT) operates by decomposing an N point time domain signal into N time domain signals each composed of a single point as shown in Figure 3.7. The second step is to calculate the N frequency spectra corresponding to these N time domain signals. Lastly, the N spectra are synthesized into a single frequency spectrum [11].

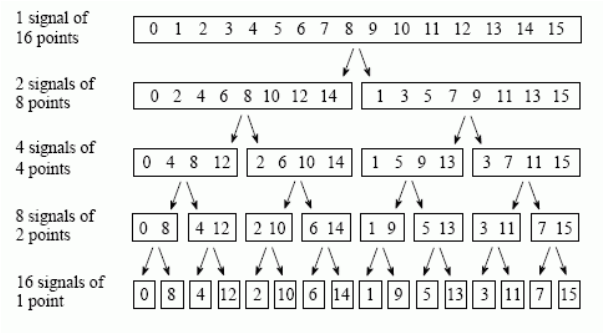


Figure 3.7: One 16 point time signal divided into 16 time signals source: www.dspguide.com

The FFT technique is applied on a signal in single shot. This means for a complete time domain signal, an FFT spectrum would give the user its components however, it will not give information about when these components are occurring or simply put it the FFT spectrum lacks the information of time itself. To overcome this a technique called Short Time Fourier Transform (**STFT**) is used. STFT determines the frequency and phase content of localised sections of the signal as it changes over time. See Figure 3.8

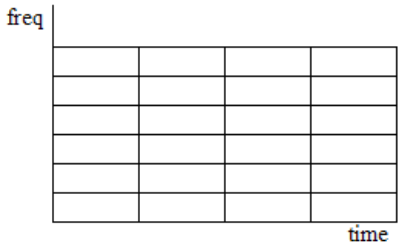


Figure 3.8: A STFT plot source: Wikipedia

3.4.2 Cross correlation

The cross correlation is very much related to convolution technique of the Fourier transforms. A convolution describes the output of signal in the transform domain when they are multiplied in the time domain. Cross correlation is similar process where one signal slides over the another signal without time reversal of the signal. The cross correlation outputs gives a similarity measure of two signals. This is covered in Chapter 2 Section 2.2 on use of Weiner Khinchin Theorem to find auto correlation and cross correlation of signals in time and Fourier space.

3.5 FX Correlator

Any book explaining about Radio Astronomy calls the correlator as the *Heart of an Interferometer*. The function it performs is computing the correlation of two incoming data streams at the interferometer. The FX type correlator is one of the architectures through which the cross correlation functions of the signals are obtained. The ***F*** stands for Fourier Transform and ***X*** for multiplication or correlation. A schematic of an FX correlator is shown in Figure 3.9.

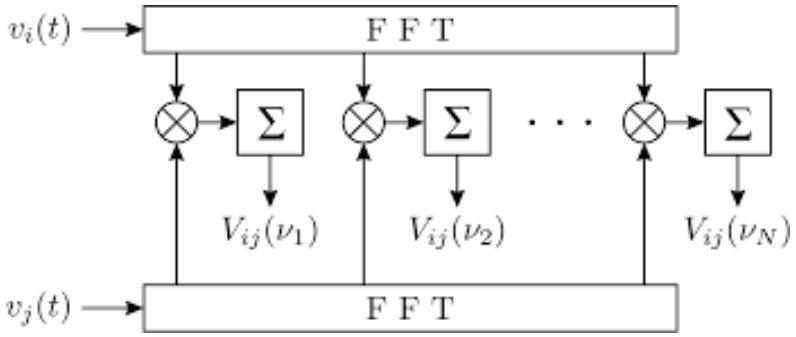


Figure 3.9: Scheme Diagram of FX Correlator source: http://www.atnf.csiro.au

The digitized time series of the signal from the ADC is first delay corrected and sent to the filter bank (FFT) which decomposes signal into its frequency components or channels. The output of each spectrum channel of the two data streams is multiplied channel by channel give a ***multiplied*** (**M**) output. At this stage the auto and the cross correlation functions are obtained. This multiplied output is time averaged based on the chosen averaging time and ***accumulated*** (**AC**). This is known as one **MAC** cycle. This is calculated for each of the baseline. [12]. This architecture of one MAC cycle can be very well visualised in the Figure 3.10.

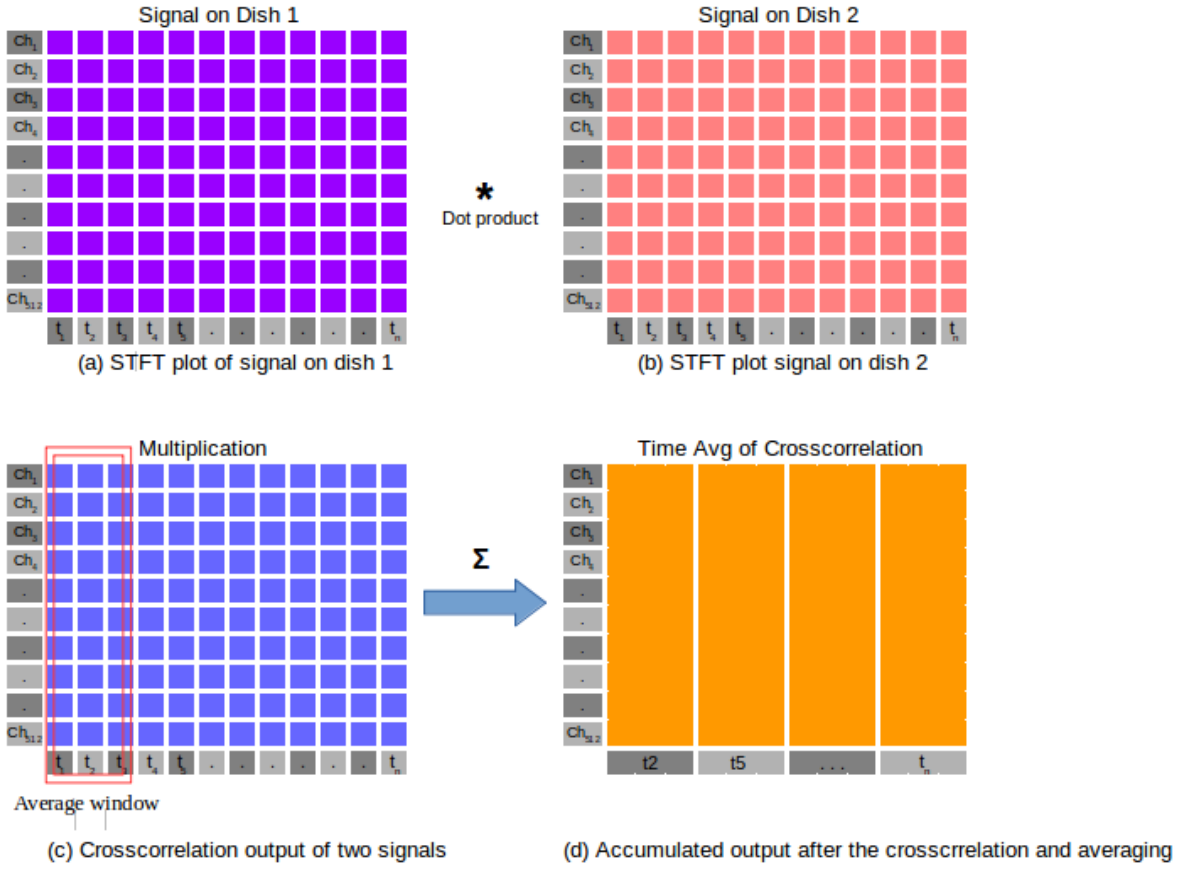


Figure 3.10: Schematic of one MAC cycle

Chapter 4

Modeling and Simulation

To investigate and quantify the effects of powerline RFI, various statistical tests were performed to learn about its time domain characteristics. Prior to devising a simulation methodology we checked for RFI properties of the actual recorded data from the GMRT. To do the analysis we used available data streams of 50ms each at the operating bands of 150 MHz, 325 MHz and 610 MHz. Although situated in a relatively quiet radio environment the radiometers receive these objecttionable signals which corrupt the observations. The nature of these signals can be both narrowband and broadband. A narrowband signal will appear as spikes in the frequency spectrum. Whereas a wideband signal will appear as impulses in the time domain and cover a large portion of the frequency spectrum. The Figures 4.1 to 4.6 show some specimens how the RFI occurs in the signal. These signals snapshots were captured at the GMRT antennae using a LeCroy Oscilloscope with a sampling time of 5ns (sampling frequency : 200 MHz).

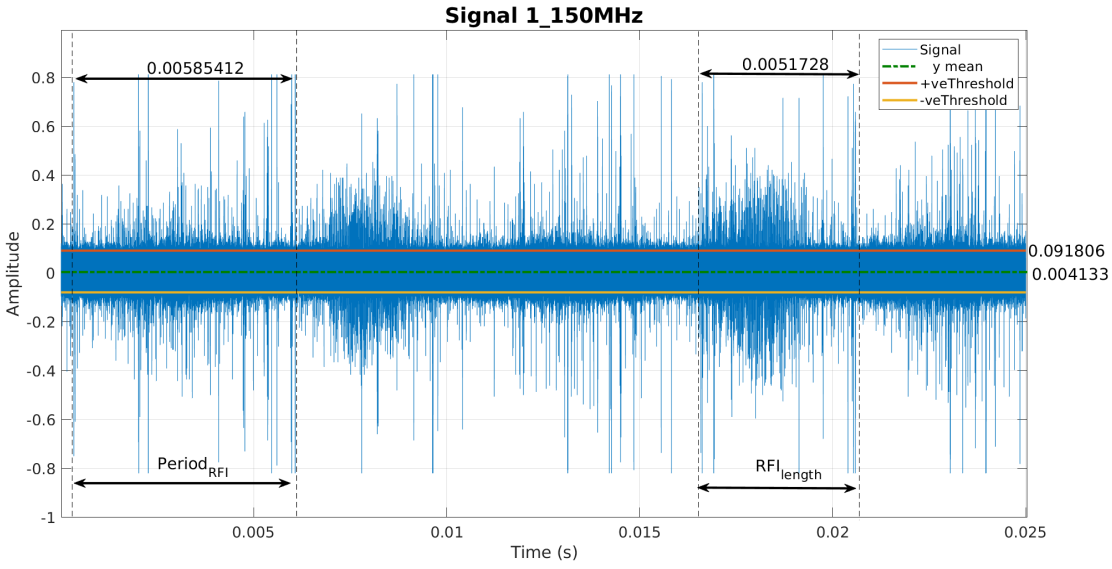


Figure 4.1: Signal at 150 MHz

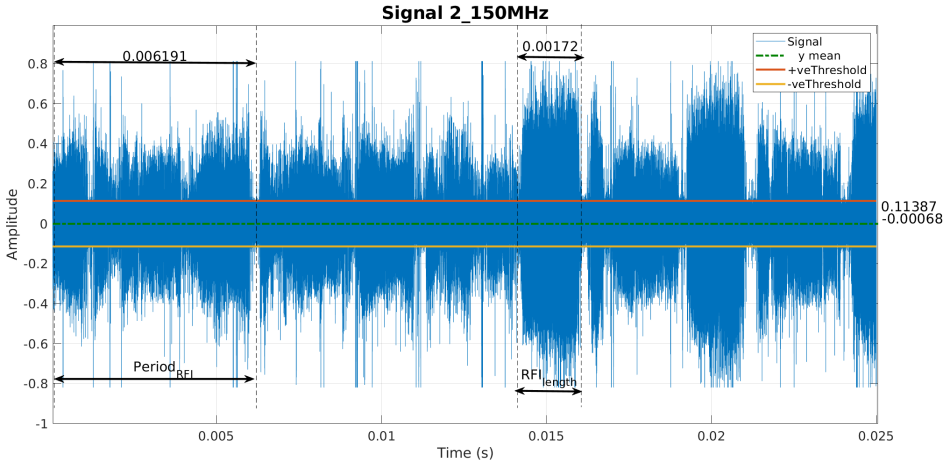


Figure 4.2: Signal at 150 MHz

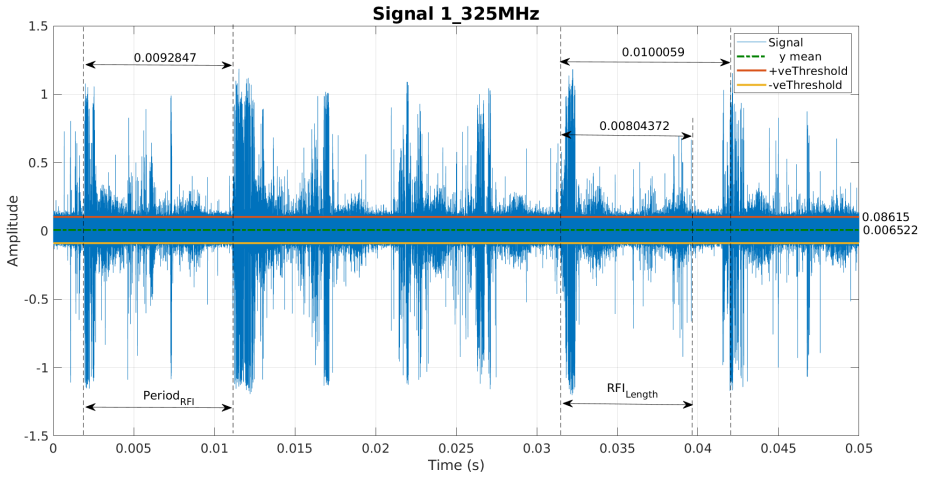


Figure 4.3: Signal at 325 MHz

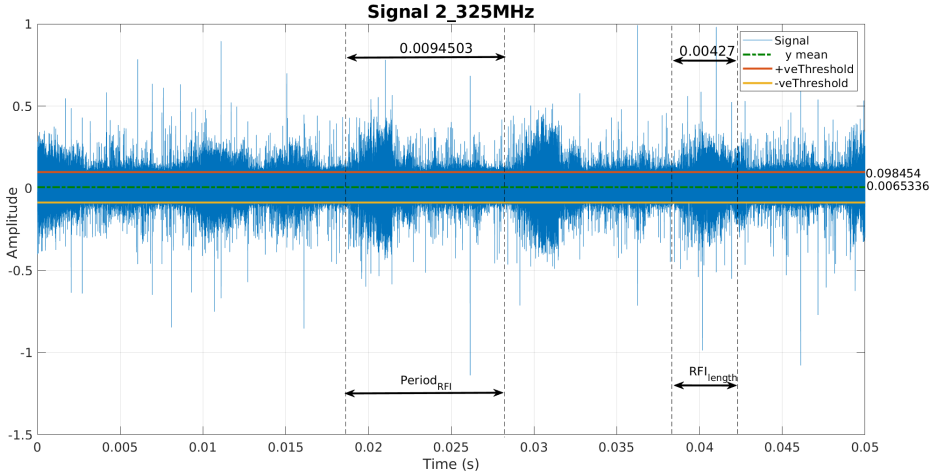


Figure 4.4: Signal at 325 MHz

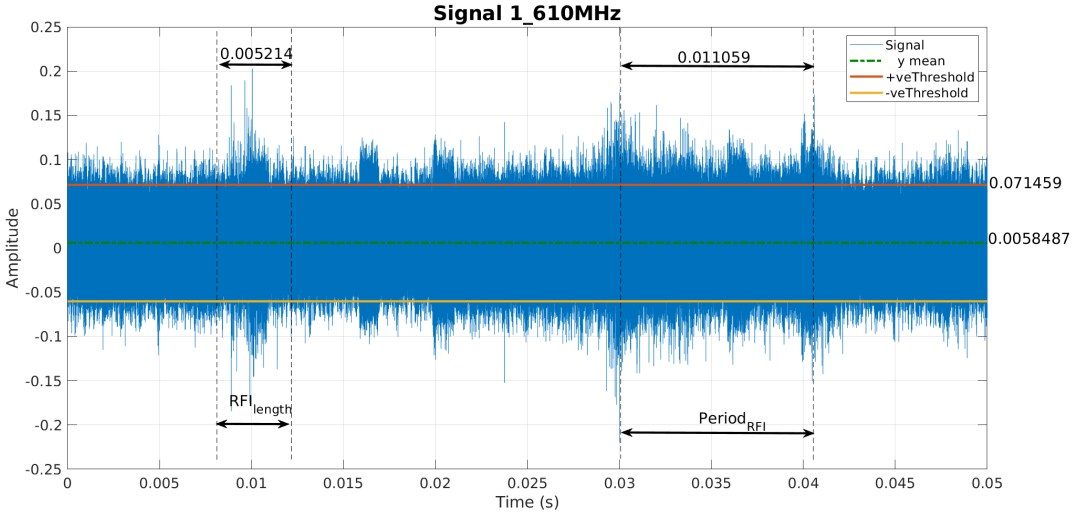


Figure 4.5: Signal at 610 MHz

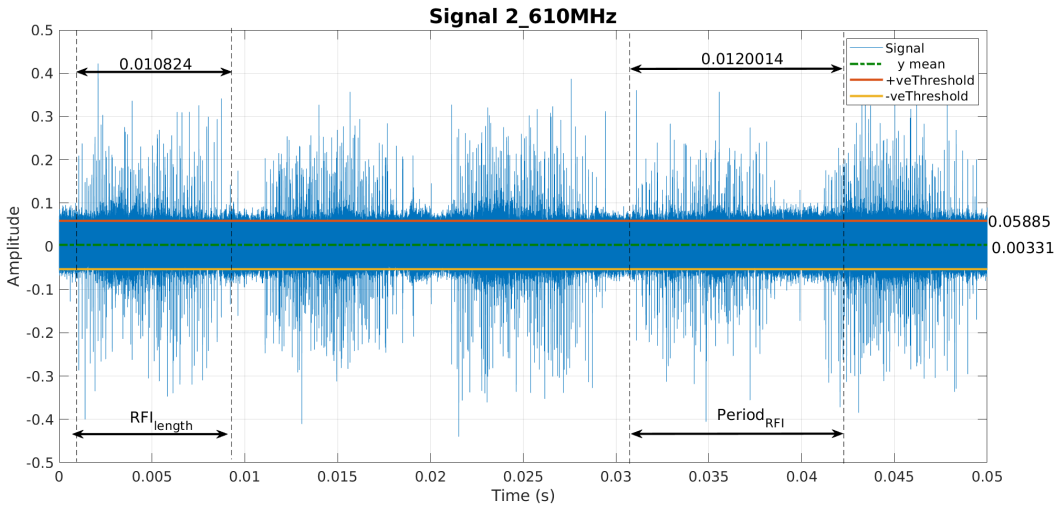


Figure 4.6: Signal at 610 MHz

Going forward in this chapter we would discuss how the different forms of RFI can be simulated. The various forms of simulated/artificial RFI profiles will be passed through the interferometer simulation tool (discussed in next chapter) to examine its effect on the cross-correlations.

4.1 Understanding the RFI Business

As seen from the above figures, the observed RFI at GMRT site occurs in small bunches and these bunches seems to follow a pattern. However, this pattern does not last long and persists just for few milliseconds. After this, there is an abrupt shift in the RFI characteristics. This is possibly due to rapid changes in the voltage breakdown levels at the sparking location, contribution from the spark plugs of vehicles or discharge from the switches and home appliances etc. Therefore an antenna receives a superposition of objectionable signals from a a variety of sources contributing to form these rapidly changing RFI profiles. This makes the predictive modeling of the RFI impossible and application of any time series forecasting model is likely to

be unsuccessful. However, doing statistical analysis over these bunches and signals did reveal some interesting insights which paved a way for modeling these signals.

4.1.1 Inter arrival Pulse Times

Currently the GMRT Wideband Backend (GWB) system uses Median Absolute Deviation (MAD) based thresholding for real time RFI excision. The pulses whose amplitudes are above and below the the upper and lower thresholds, τ_U and τ_L respectively, are considered to be plausible RFI candidates. For a input data vector X, the upper and lower thresholds are calculated using the sample median $M(X)$ as [13]:

$$\tau_U = M(X) + n \times (1.4826 \times MAD) \tag{4.1}$$

$$\tau_L = M(X) - n \times (1.4826 \times MAD) \tag{4.2}$$

The value of n is the under the control of user and depends how much robust RFI clipping (s)he requires. However, since the process of filtering is done real time applying MAD over large data streams is limited due to hardware constraints. Therefore a more robust estimate for filtering Median of MAD (MoM) is used [14]. The inter arrival time of the consecutive pulses is calculated using its respective time stamp and its density variation is plotted in Figures 4.7, 4.8 and 4.9 for 150 MHz, 325 MHz and 610 MHz respectively.

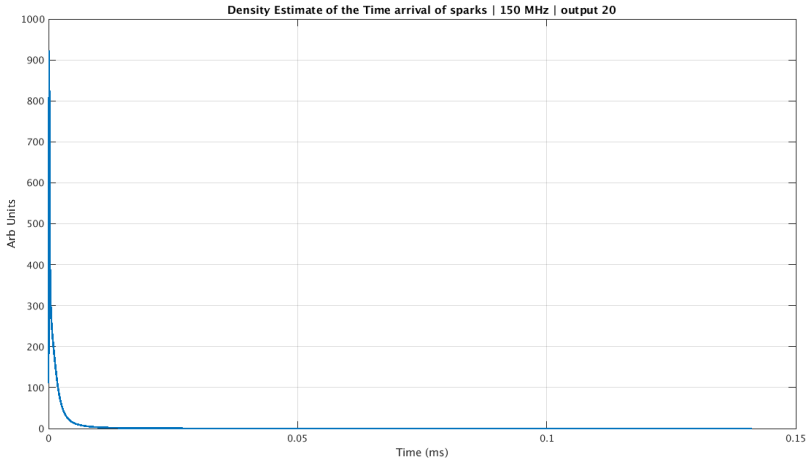


Figure 4.7: Density of inter arrival time of pulses at 150 MHz band

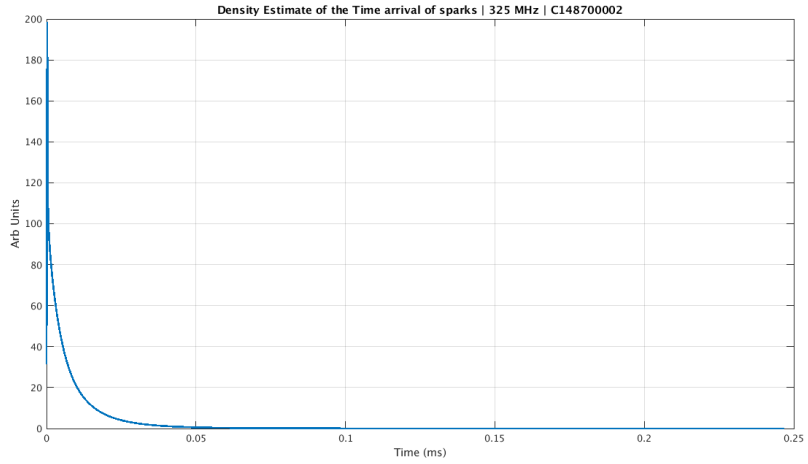


Figure 4.8: Density of inter arrival time of pulses at 325 MHz band

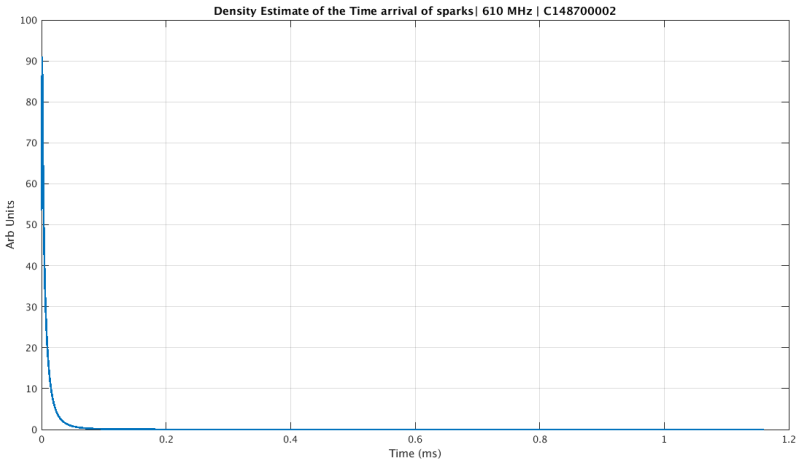


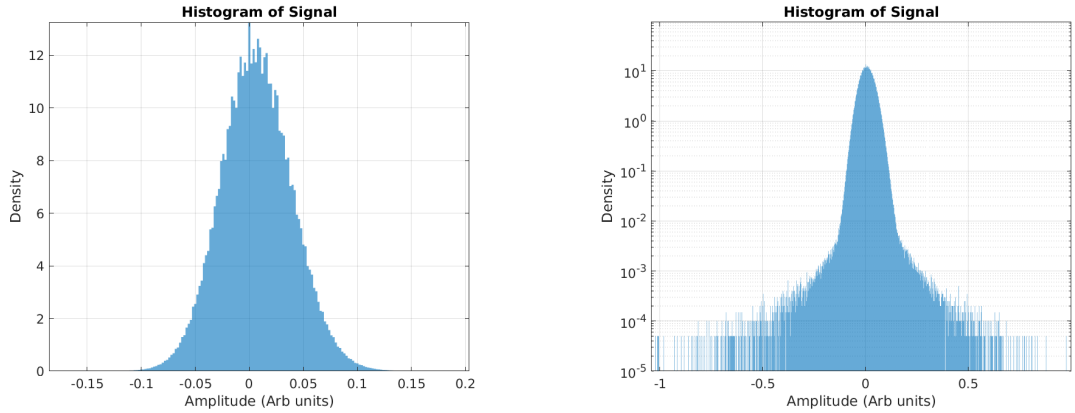
Figure 4.9: Density of inter arrival time of pulses at 610 MHz Band

A visual inspection of the time of arrival densities for each of the frequency band clearly depicts an exponential decay trend. The variation of time of arrival in the 150 MHz band show the pulses occurring at very short inter arrival time as compared to 325 MHz and 610 MHz band. This is consistent with the fact that the signals suffer with more RFI at lower frequencies at the GMRT site.

4.1.2 Gaussianity Checks

The basic test for testing if data chosen from a normal distribution is to check if >99 percentage of data points lying within the standard deviation of $\pm 3\sigma$. For all the signals the 99% of the data fell within the slab of $\pm 3\sigma$. Further, the Gaussianity of the signals was checked using the χ^2 Test under the assumption that the signals are drawn from a normal distribution with a mean and standard deviation derived from the data. The null hypothesis that the signal is derived from a Gaussian failed when testing using the χ^2 Test. This is due to the so evident heavy tailed nature of the signal. This said heavy tailed-ness is not visible on a linear scale but log scale. Figure 4.10 represents the same. For a standard normal distributed data the expected value of kurtosis is 3 [15]. The data observed by us had values greater than 3. Figures 4.11 and

4.12 represents two datasets recorded on 325 MHz and 610 MHz band with their kurtosis and MAD (Section 4.1.1) filtered values respectively.



(a) Histogram of signal on a linear scale

(b) Histogram of signal on a log scale

Figure 4.10: Linear and log scale histograms of signal

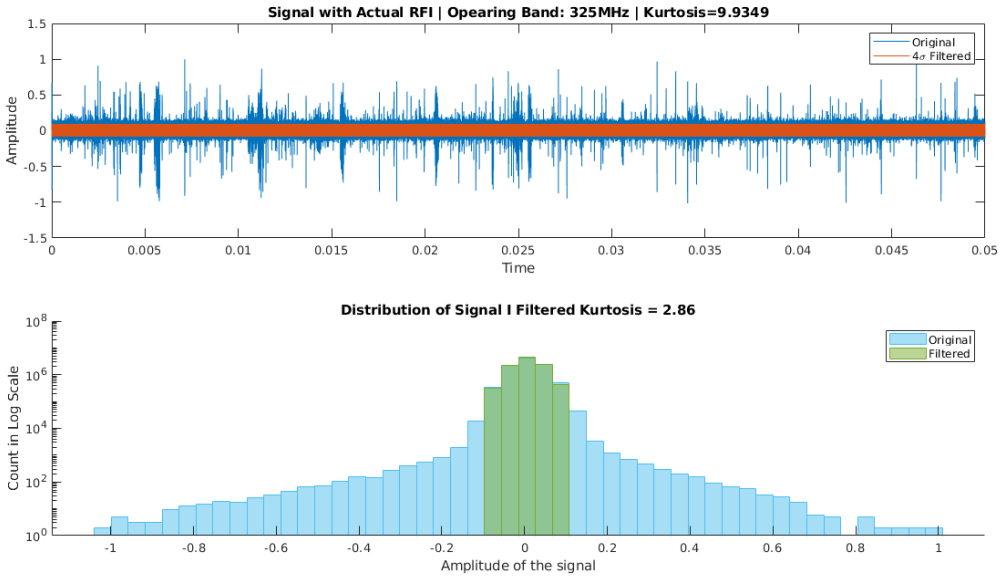


Figure 4.11: MAD filtered signal with kurtosis at 325 MHz band

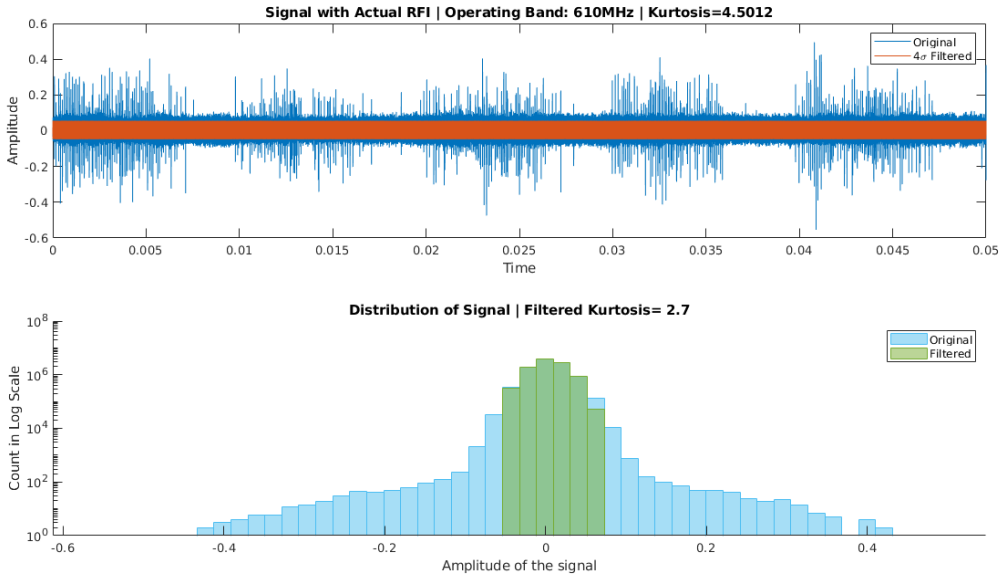
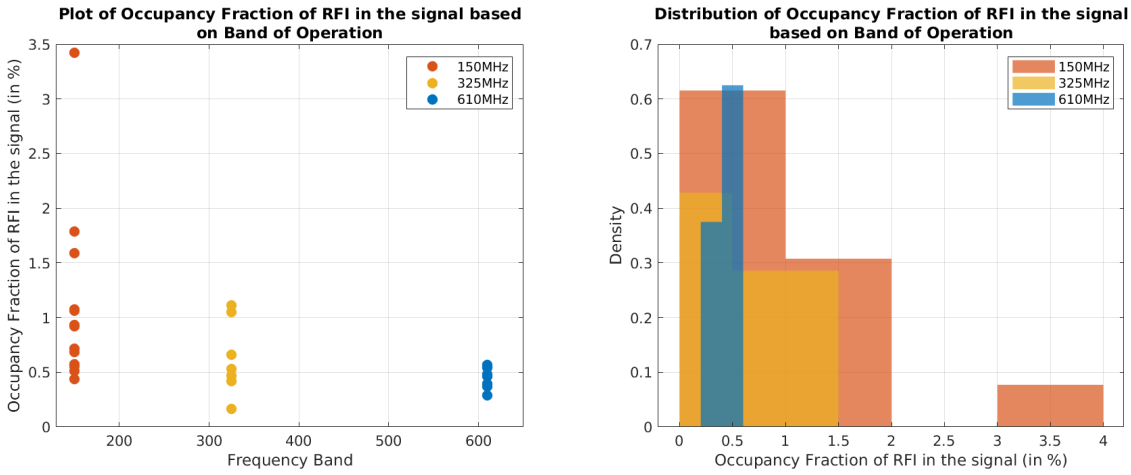


Figure 4.12: MAD filtered signal with kurtosis at 610 MHz band

4.1.3 RFI Occupancy Fraction in Signal

The RFI occupancy fraction describes the percentage occupancy of RFI in the signal based on the operating frequency band. Figure 4.13 represents the variation of occupancy fraction with frequency band. The occupancy for lower frequency shows a higher spread of RFI occupancy in the signal. It was observed that at lower frequencies there is a very high variability from one data set to another. This variability can be contributed to the many sources active at this frequency band. However, as one moves from the lower to higher band there is a significant drop in the RFI fraction. This could due to the RFI received at these frequencies at the telescope is of low amplitude.

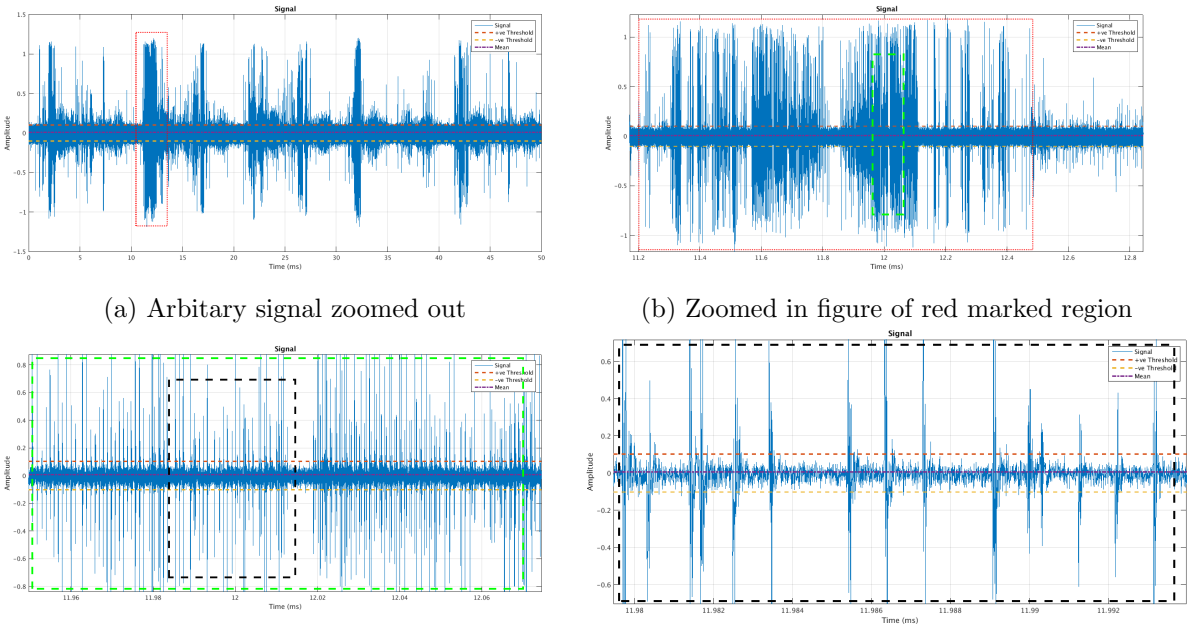


(a) (b)

Figure 4.13: RFI Occupancy based on band of operation.

4.1.4 Understanding individual Impulsive Sparks

In the time domain the RFI do not occur as continuous phenomenon but as bunches of individual sparks as described in Section 4.1.1. We tried to investigate if there was an underlying shape for these sparks. This was one of the fundamental expectation on the basis of which we wanted to model these spark bunches.



(a) Arbitary signal zoomed out

(b) Zoomed in figure of red marked region

(c) Zoomed in figure of green marked region

(d) Zoomed in figure of black marked region

Figure 4.14: Zooming on individual sparking events

To investigate of any behaviour that the sparks are following, we just visualised them by stacking them over each other and look for a pattern. An example of how these sparks are staked is shown in the Figure 4.15. A full stacking of 1000 sparks is shown in Figure 4.16 where a decay after the spike is clearly visible.

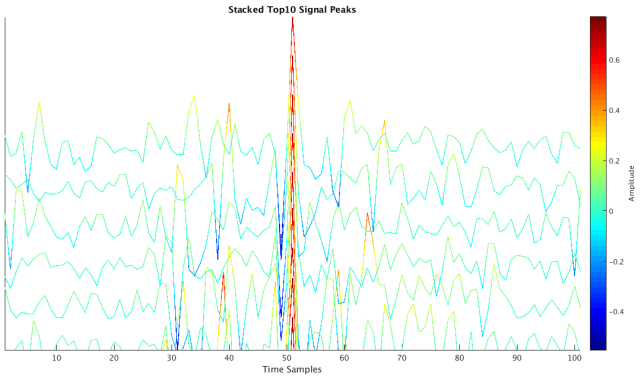


Figure 4.15: An example of stacking of sparks of signal

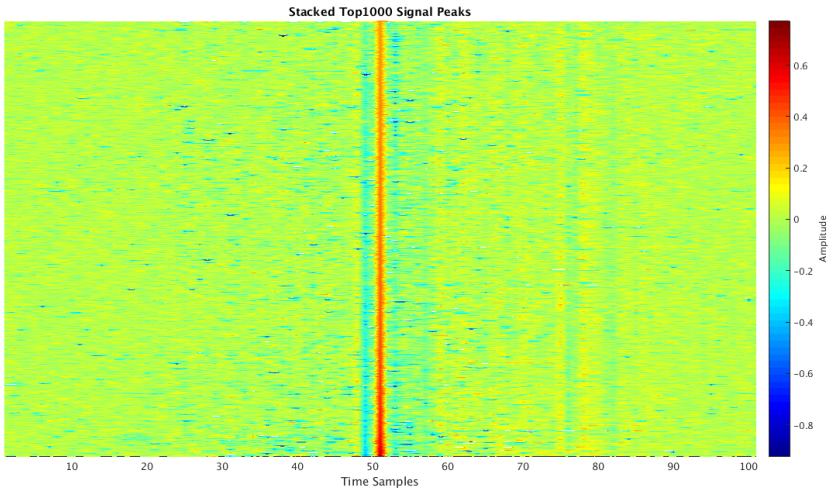


Figure 4.16: Stacked peaks of the sparks of a signal

To further understand the underlying behaviour of the decay of the spark, we took some samples from a data set and saw if the decay pattern follow some defined measure. Figures 4.17 to 4.20 depicts some sparking events to the envelope of which a curve was fit. To find the curve equation only data points above the positive threshold were considered. It was observed that the decay pattern of these sparking events followed a power function rather than an exponential or a double exponential function. Power function seems to best describe this behavior because its roll off reaches towards 0 into the noise floor level and does not rise again. An expected pattern in a sparking event will also be that the decay of the spark assimilates in the noise region of the signal. However, due to high variability it was not be possible to quantify a defined set of interval over the rate parameter (exponent). And therefore generalising that each spark follows a predefined curve will outdo the sanity of this stochastic phenomenon.

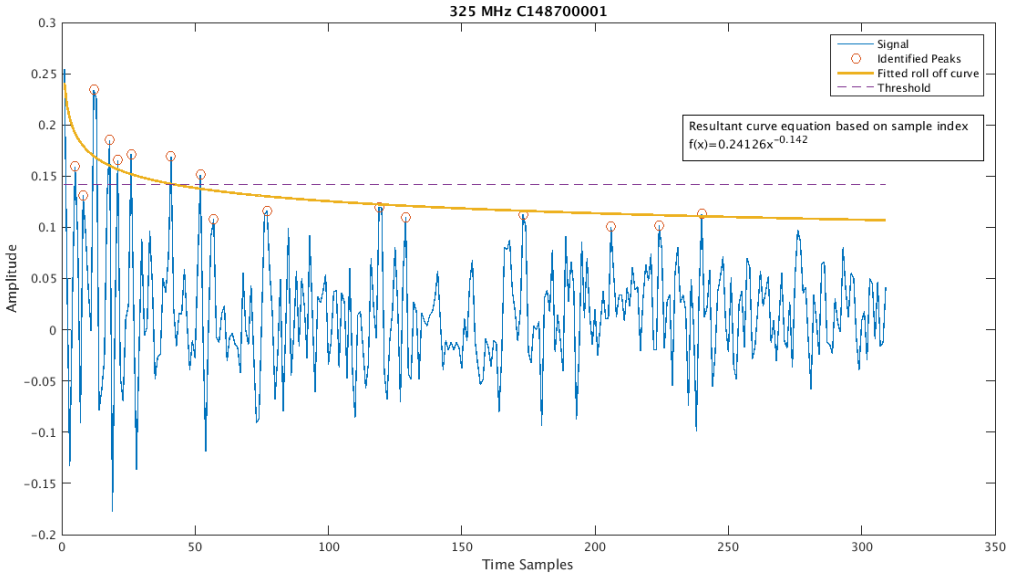


Figure 4.17: Power curve envelope over a spark event in a signal at 325 MHz

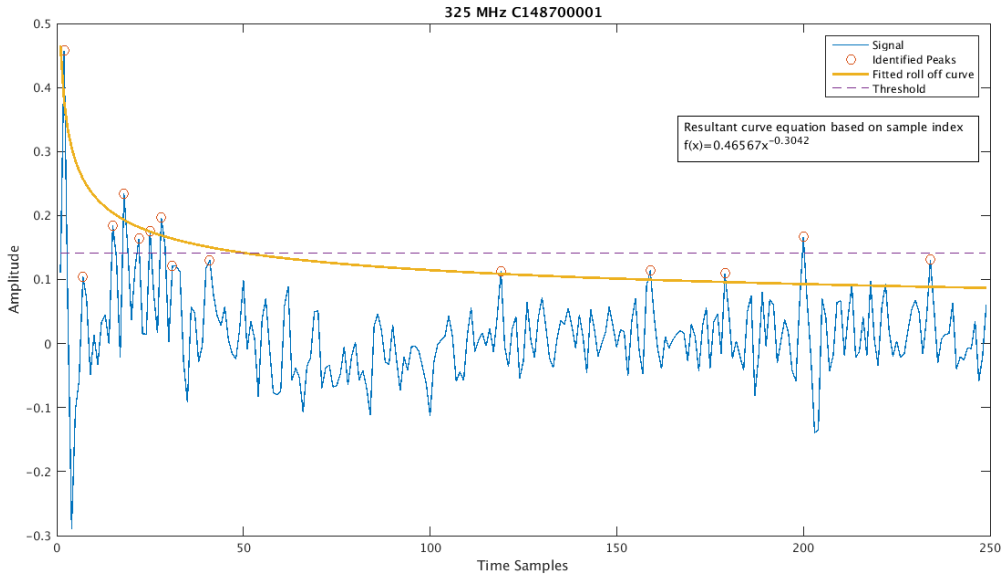


Figure 4.18: Power curve envelope over a spark event in a signal at 325 MHz

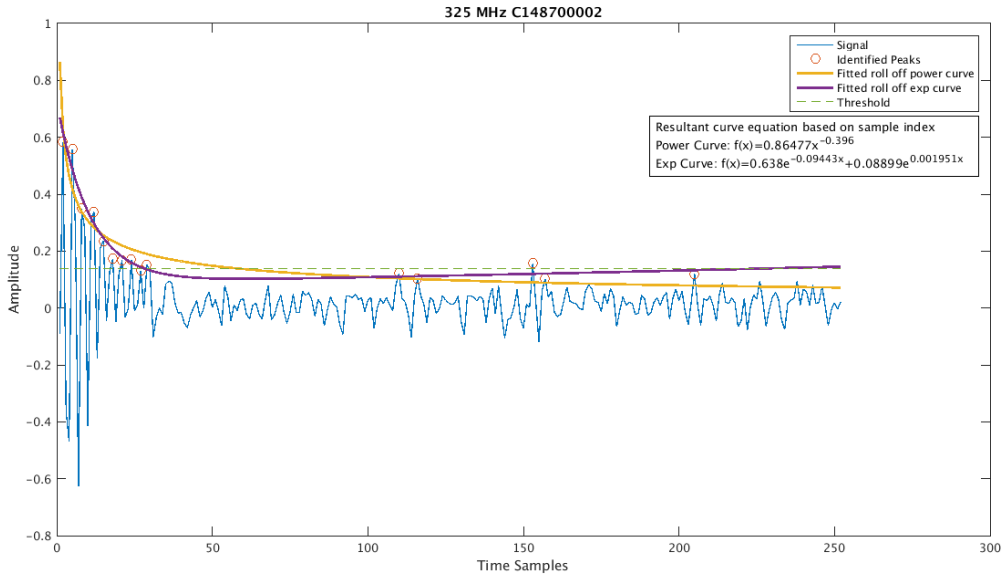


Figure 4.19: Power curve and double exponential envelope over a spark event in a signal at 325 MHz

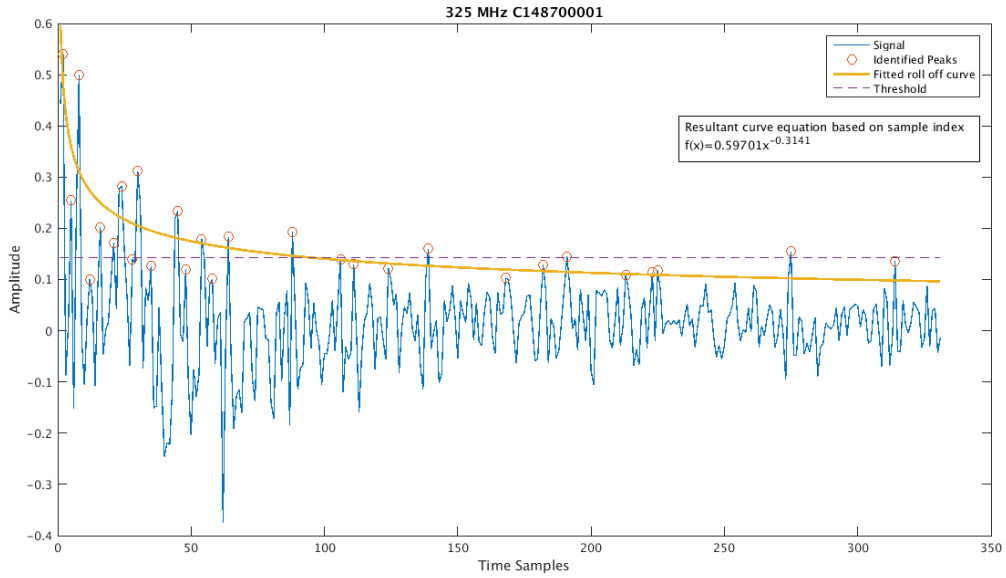


Figure 4.20: Power curve envelope over a spark event in a signal at 325 MHz

4.2 Modeling Start: Key Assumptions

The developed simulation tool will give user a handle on various parameters to create an RFI profile of desired shape and characteristics. The basic assumptions that on which the tool was developed were:

- Observing a single unresolved radio source.
- Signal from the radio source is assumed to be entirely Gaussian with a pre-defined handle on its standard deviation.
- Antennas are affected by a single RFI source. (subjected to change. Inputs for multiple RFI sources can also be previded).
- For convenience the radio source is assumed to be at zenith i.e. the geometric delay, τ_{geom} is 0.
- The delay in the signal is only due to the position of the RFI source.
- The system was tested for a 2 element interferometer, however the architecture allows as many antennas user wants to simulate subjected to computational expense.

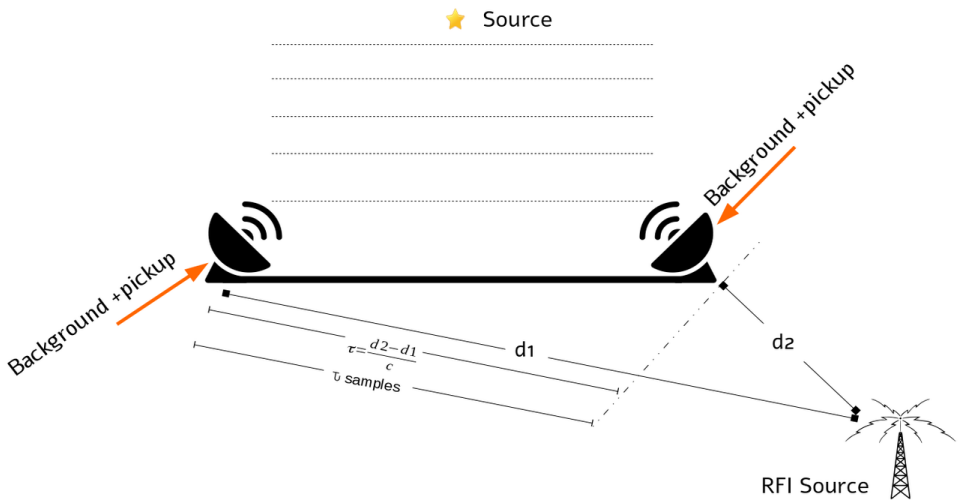


Figure 4.21: Model assumption schematic

In the figure 4.21 the signal from the *sky source* is assumed to be a white Gaussian noise with a predefined value of the standard deviation of the signal. The *background and noise* pickup are taken as independent Gaussian signals with specified standard deviations and same goes for the RFI signal as well. These standard deviations of different components of the signal are chosen in such a way that the ratio of their standard deviations is proportional to their corresponding Black Body Temperature. The locations of the antennae are chosen based on the spatial coordinates of the GMRT antennae. The position of the RFI source is relative to the centre of the Earth $\langle 0, 0, 0 \rangle$.

4.3 Signal Generation

The signal picked up by the antenna can be decomposed into a) astronomical signal, b) receiver temperature, and c) signal from the RFI source. A generalised time domain model of the signal with the above contributions can be given as [16]:

$$x(t) = x_{sky}(t) + x_{rec}(t) + x_{RFI}(t) \tag{4.3}$$

The sky signal and the receiver noise are zero mean random signals with user specified standard deviation. The RFI signal is generated as detailed below.

To capture the behavior of the RFI the following types of pulses were simulated using the following methods and their respective cross correlation plots were examined. This section describes how these specific pulses were generated and the parameters used to generate these pulses.

4.3.1 Pulse Type: Gaussian

Given the information of time and sampling frequency the user can generate as long data stream as required. The signal can be categorised into two factions- the time during which the RFI duty cycle is ON, and the time for which the RFI duty cycle is OFF. The duration for which the RFI duty cycle is ON will be termed as RFI_{length} hereafter and time after which the RFI bunch repeats again is called $Period_{RFI}$. This can be observed in Figures 4.1 to 4.6. Figure 4.22 depicts an artificial signal with RFI. The parameters of the same are described in the figure.

The amplitudes of the RFI pulses are drawn from a Gaussian distribution, $X \sim N(\mu, \sigma)$ and the generated Gaussian series of specific RFI length are added at the start of every RFI period.

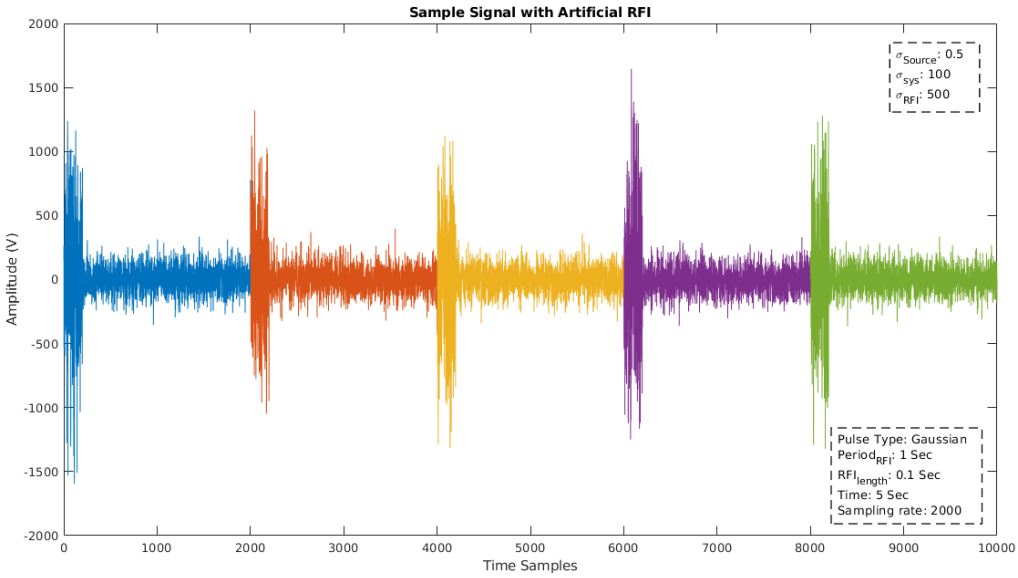


Figure 4.22: Artificial Gaussian RFI pulse. Each color in the signal represents one $Period_{RFI}$.

4.3.2 Pulse Type: Uniform

Using the same idea as described in the Section 4.3.1, pulses with amplitude range $\pm 3\sigma_{RFI}$ can be drawn from a uniform distribution, $X \sim U(-3\sigma_{RFI}, 3\sigma_{RFI})$

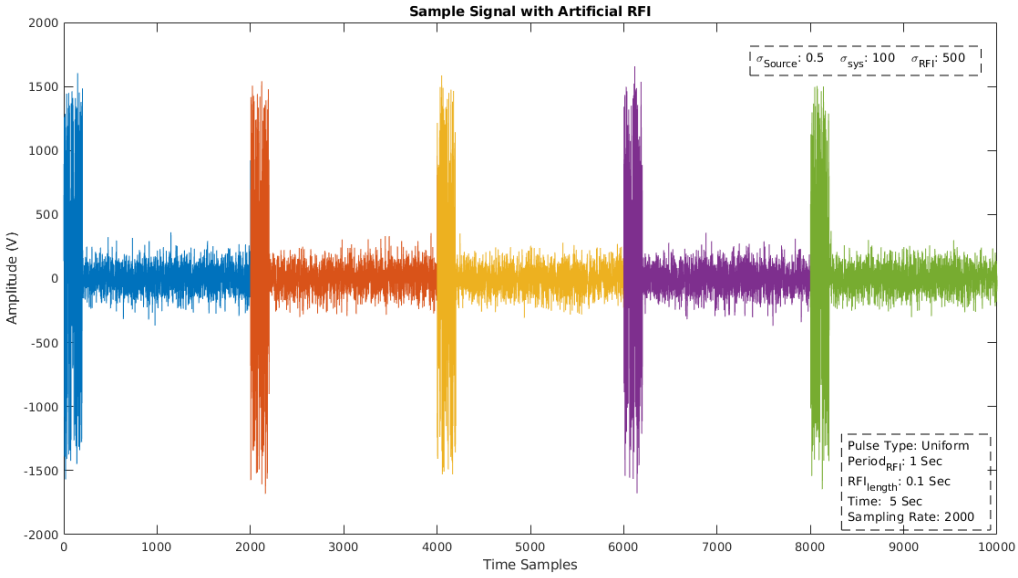


Figure 4.23: Artificial Uniform RFI pulses. Each color in the signal represents one $Period_{RFI}$. 4.22

4.3.3 Pulse Type: Exp-Poisson

The exponential-Poisson pulse uses two distributions to generate the RFI pulse profiles. The Poisson distribution determines the number of pulses to be generated in an RFI period and the exponential distribution models the inter-arrival time of the pulses. The rate parameters of both the exponential and Poisson distribution are user tunable.

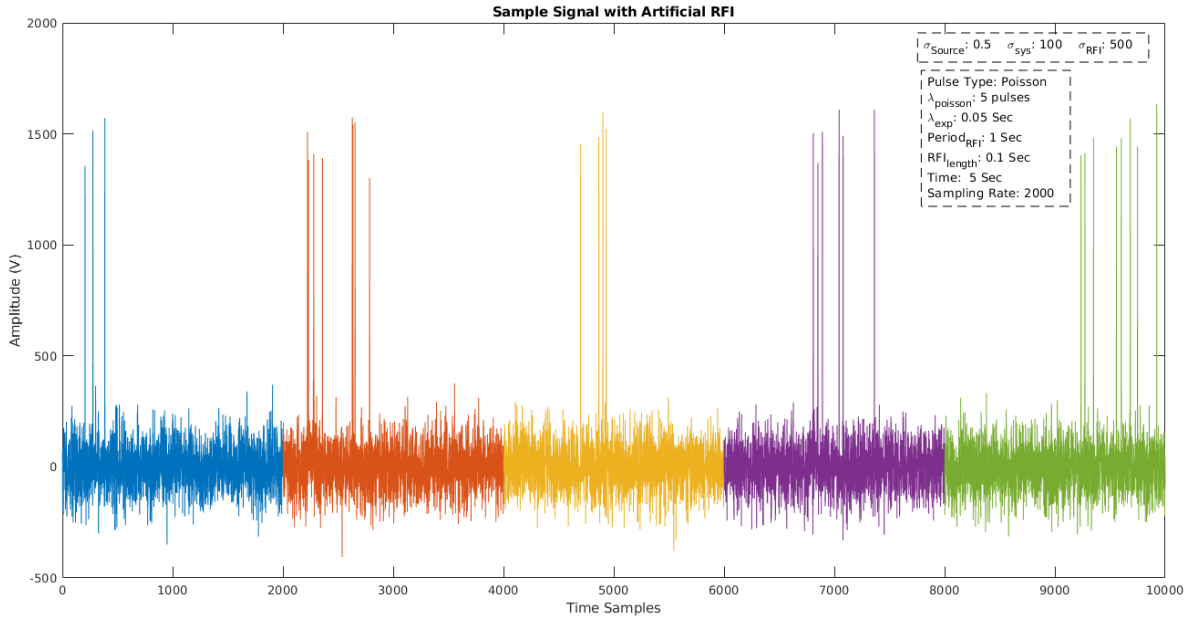


Figure 4.24: Artificial RFI pulse with exp-Poisson distribution. Each color in the signal represents one $Period_{RFI}$.

4.3.4 Pulse Type: User Defined 1

Another type of pulse generation based on the observed characteristics of the real data was also developed. Given the information of inter-arrival pulse times, average number of pulses in the RFI period and the shape of the spark events as described in Section 4.1.4, a variety of pulses can be generated with shapes that follow a power curve profile. The methodology to implement this is as follows:

1. Find the average number of spark pulses for a particular RFI period using Poisson distribution.
2. Find the inter-arrival pulse time of each of the pulses.
3. Fix the number of samples for which the sparks of RFI signal lasts.
4. Fix the range of exponents for power curve.
5. Draw the amplitudes of the spark signal from the choice of distribution with a defined σ_{RFI}.
6. Multiply each of the drawn spark profile with its individual power curve estimate.

Figure 4.25 represents the RFI pulses generated using this method. The signal contains no contribution from the source and the receiver, hence it is a pure RFI signal.

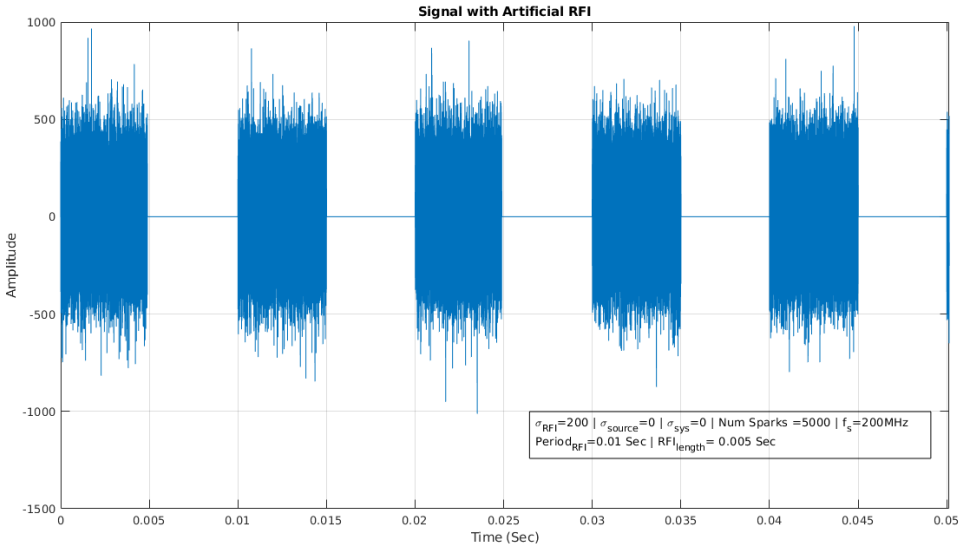


Figure 4.25: Signal with artificial RFI using user defined 1 method

4.3.5 Pulse Type: User Defined 2

The pulses describe in the above sections are based on the premise that user is given the information of parameters of the spark shape and the inter-arrival times. This section provides the recipes for generation of spark pulses based on the characteristics of the observed data. The general parameters to be taken into consideration for modeling an RFI region are the inter-arrival time of sparks, shape of the sparks and number of sparks in each RFI period.

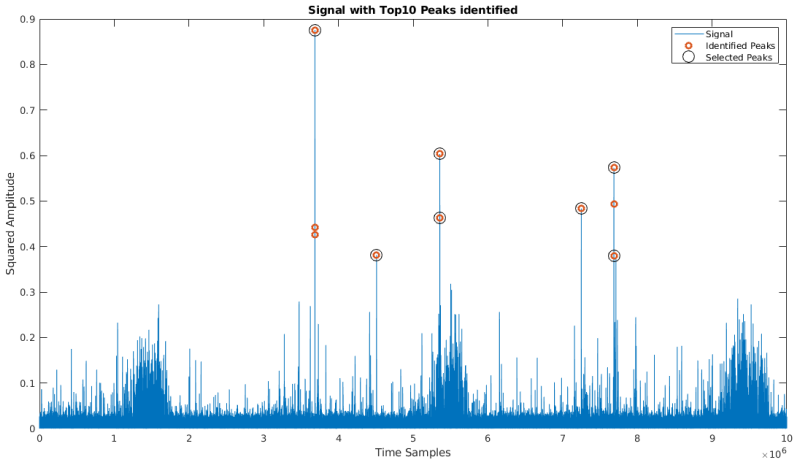
Deriving densities of the inter-arrival time

In the Section 4.3.3 the inter-arrival time was assumed to follow a exponential distribution. This requires fixing of a rate parameter of the exponential distribution. If the rate parameter is not directly available to the user then the inter-arrival time can be inferred from inter-arrival density using the following approach. The inter-arrival density is formed by taking a region of RFI and finding the inter-arrival time of sparks in that region. The density estimate of these inter arrival times can be found using kernel density estimation. Then, the time of arrival of each pulse can be drawn from this distribution using importance sampling [17].

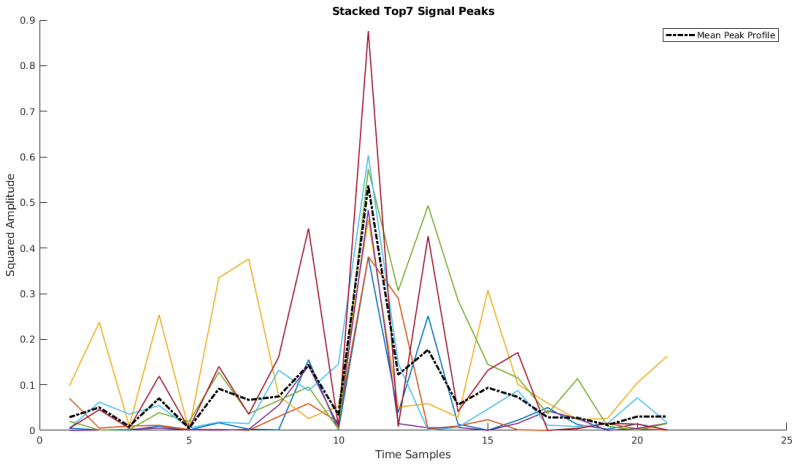
Spark Density estimation

The estimation of spark profile resembles selecting a curve from a distribution. What if there existed a distribution from which these curves can be chosen? A continuous curve can be called a function. This notion of making a distribution of functions comes under the domain of Functional Data Analysis (FDA) [18]. To pick a sparking profile from a distribution, a distribution of spark profiles was made. To make such a density let us assume that each time point of a spark is derived from a particular distribution and hence the corresponding distribution for each time point can be made. However, this is a very crude application of Functional Data Analysis. Defining a distribution for a functional data is not straight forward. A meaningful density estimation would require doing a functional Principal Component Analysis for dimension constraining of the functions. The Figures 4.26 and 4.27 show a toy example how the density of these sparks can be formed and how the spark profiles can be simulated. Figure 4.26a shows the selection of example peaks with 21 neighbourhood points. Figure 4.26b reveals

a zoomed in picture of the selected peaks of the sparking events, stacked at a common point where all the peaks are aligned to one sample. Each curve in Figure 4.27a is a Cumulative Density Function (CDF) of one single time point. Since there are 21 time points, there are 21 CDF curves. At last Figure 4.27b shows the sparks simulated from the CDFs using importance sampling. Figure 4.28 show the simulated RFI signal from this method.

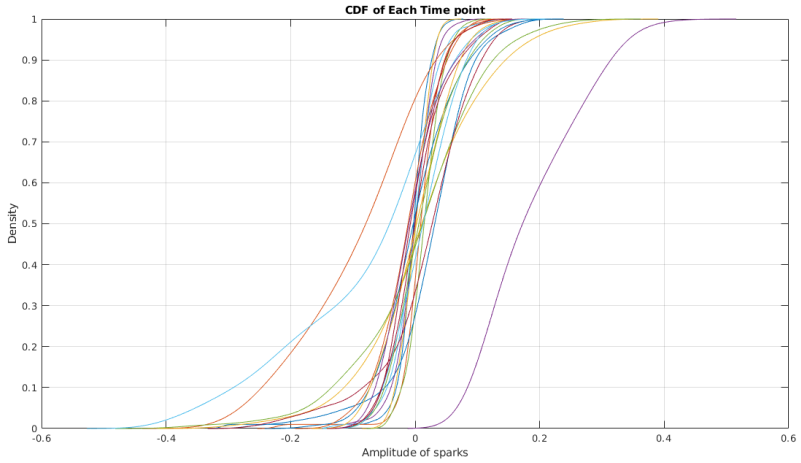


(a) Selected peaks for spark profile generation

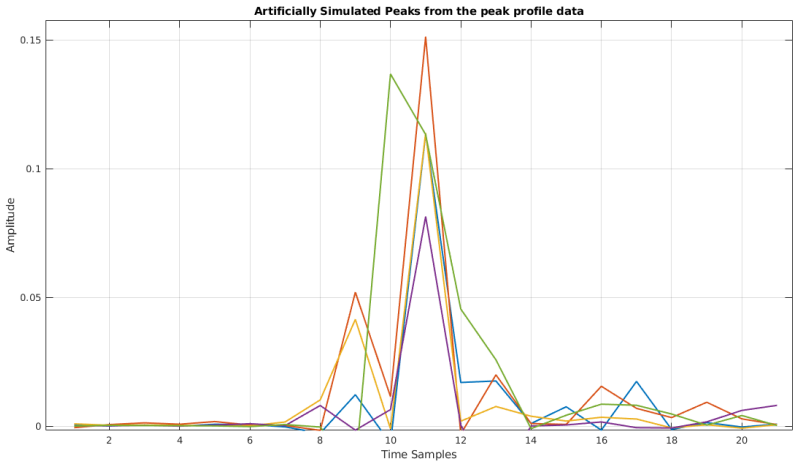


(b) Peak profiles of the selected peaks

Figure 4.26: Peak stacking for density estimation



(a) CDF of the time points of the sparks



(b) Simulated peak from the peak profile data

Figure 4.27: CDF of spark profiles and simulated sparks

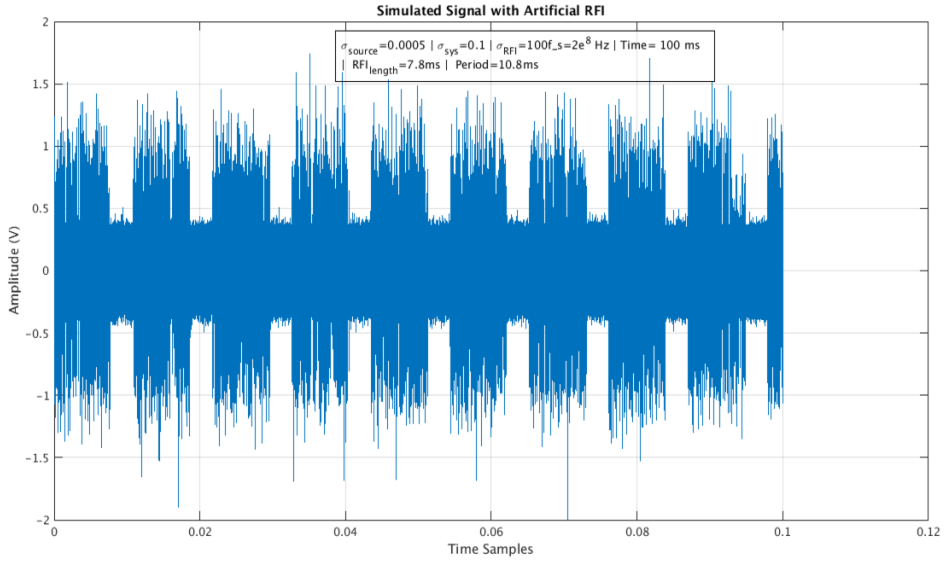


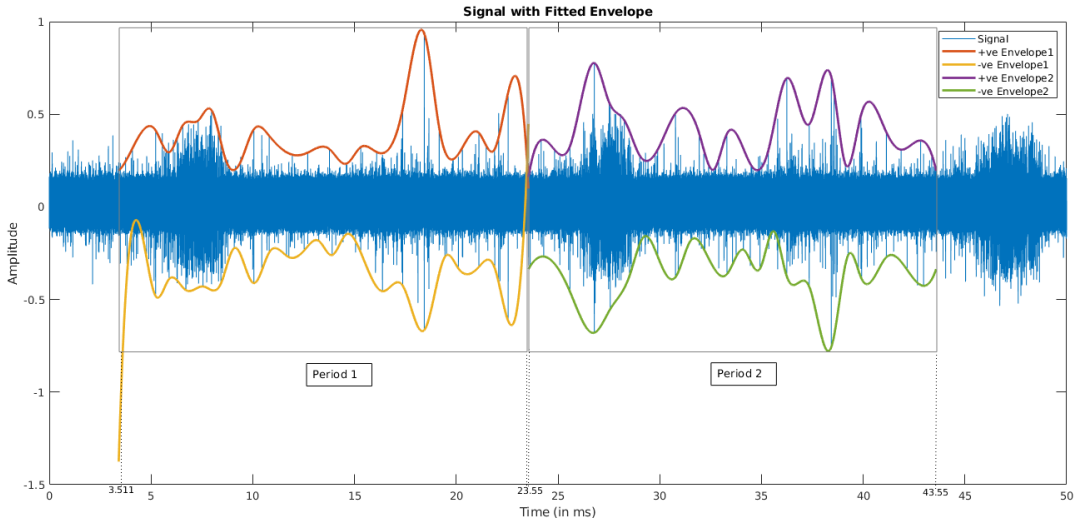
Figure 4.28: Artificial signal generated using the above method

Signal Envelopes

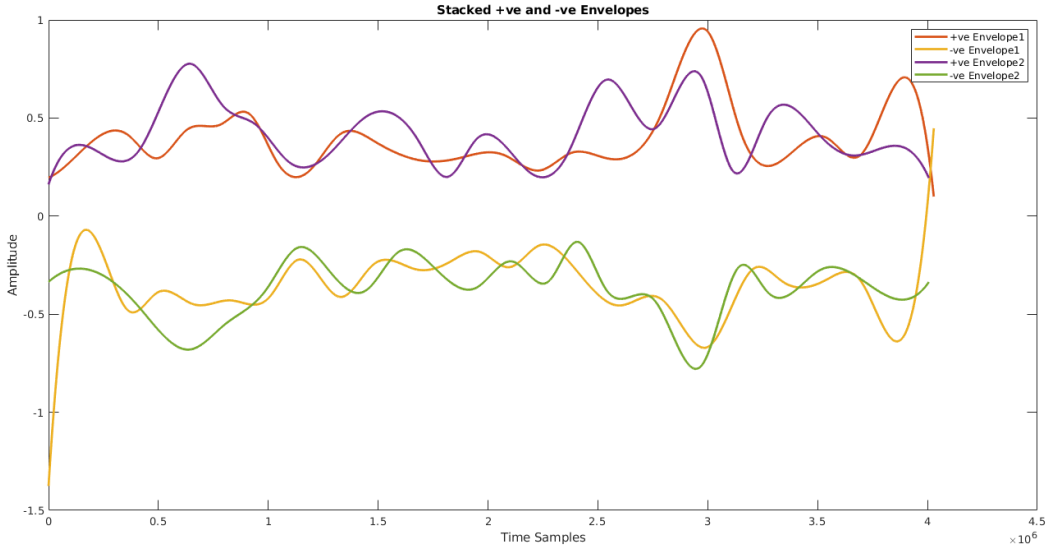
In the Sections 4.3.4 and 4.3.5 the spark profiles are generated disregarding the underlying shape of the signals. The spark generation with underlying shape can be modeled as a conditional probability distribution.

$$P(spark|shape) = \frac{P(spark \cap shape)}{P(shape)} \tag{4.4}$$

Using the FDA technique as described in the Section 4.3.5 a random signal underlying profile can be obtained. The final shape of the signal is obtained by multiplying the envelope function to the generated signal. Figure 4.29 show signal envelopes over a real signal observed at GMRT and Figure 4.30 signal formed used various different envelopes.

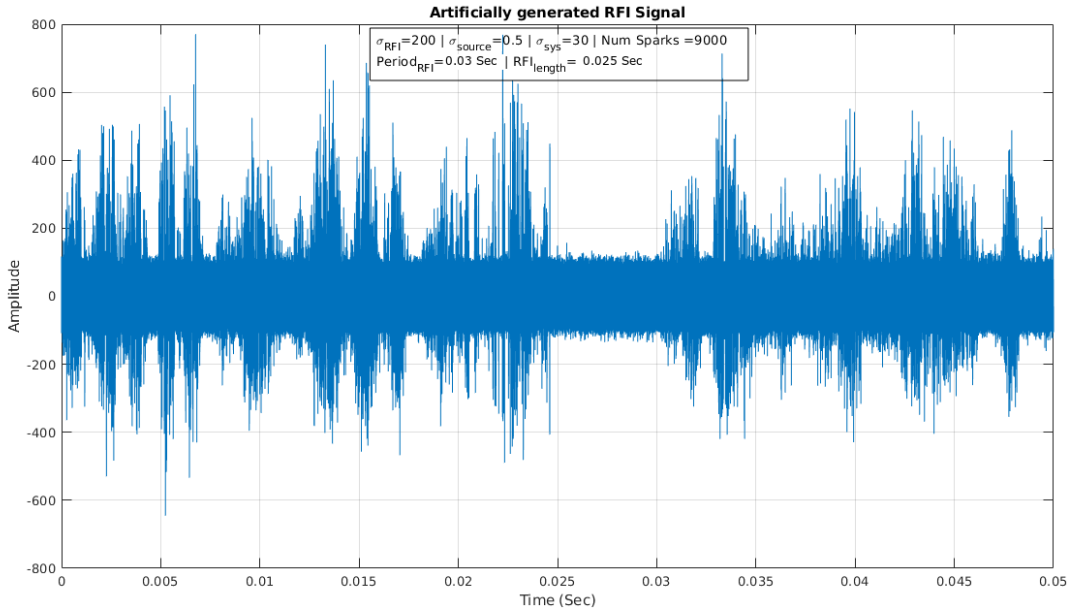


(a) Envelope of sparks

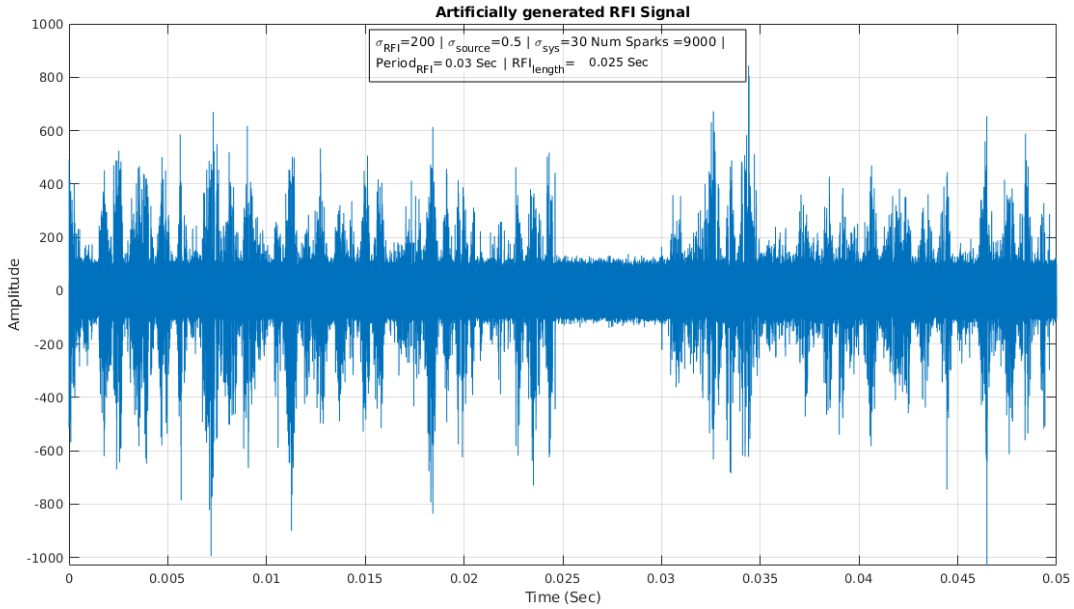


(b) Stacked signal envelope

Figure 4.29: Signal envelopes



(a)



(b)

Figure 4.30: Simulated RFI signal with artificial envelope

4.4 Distribution of Signals in ADC Bits

The signals are generated based on the standard deviation of the distribution (mainly Gaussian) from which they are chosen. This section will describe an analysis of bit occupancy of the ADC. For an RFI signal generated with σ_{RFI} as high as 200 units an ADC with 8 bits will not suffice as the number of ADC bins will fall short to occupy the higher amplitudes of the signal. This is depicted in Figure 4.31. The higher amplitudes can be accommodated by coarsening the resolution while maintaining the same bit size or increasing the number of ADC bits [19]. Figure 4.31 depicts the distribution of an artificial signal generated with parameters as described in the figure using 8 bit quantization. Quantization with 8 bit ADC would mean that the generated amplitude values have 256 (2^8) bins to occupy. The black dashed lines represent the ADC and the $3\sigma_{tot}$ limits. It is clearly seen in the figure that the signal amplitudes will not be captured with an 8 bit quantization. In contrast to this when a 12 bit quantization is used as represented in Figure 4.32, all the amplitude values are well captured in the ADC bins.

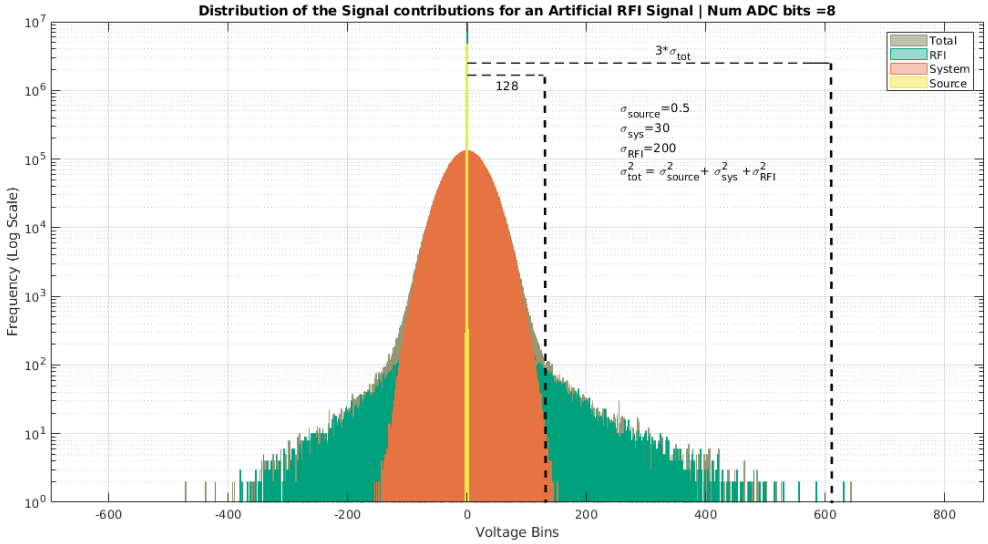


Figure 4.31: Distribution of signals with 8 bit quantization

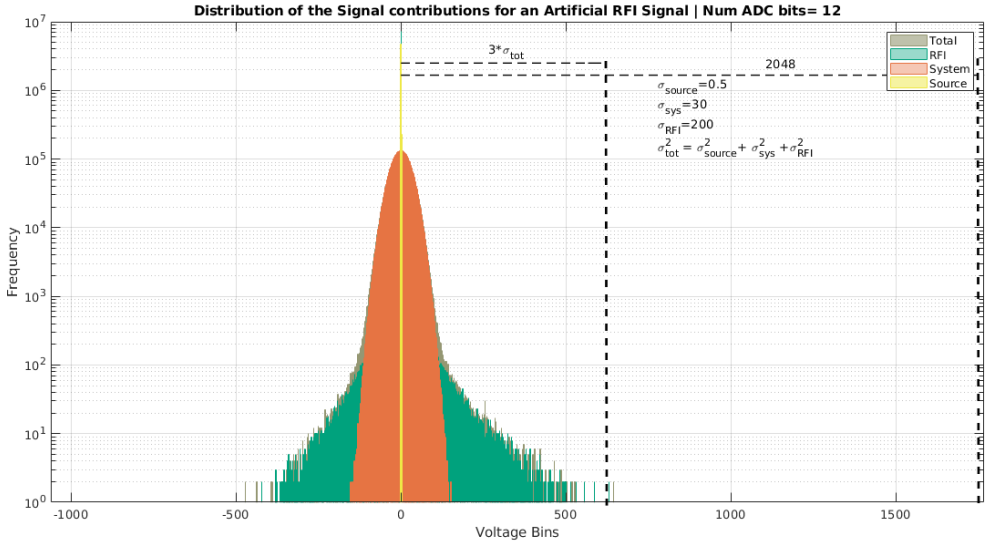


Figure 4.32: Distribution of signals with 12 bit quantization

Chapter 5

Tool Development and Test Results

5.1 Interferometer Simulation

The effect of the artificially generated RFI signals can be seen on cross correlation values as outlined in Section 3.5. The algorithms to generate the artificial RFI were embedded in an RFI generation pipeline. This pipeline can simulate a n element interferometer which will accept signals (artificially generated) with appropriate delay based on the GMRT antenna positions and output their cross correlation results. This would give user a freedom to choose a bespoke framework through which different RFI profiles can be generated and their effect can be studied. This pipeline is coded in MATLAB. The Figure 5.1 show the various simulation blocks which form a part of this simulation chain. The next sections briefly describe about these blocks and their corresponding functions.

5.1.1 Signal Generation Block

The signal generation block would take the following inputs from the user:

- Total time and sampling time.
- σ_{source}, σ_{sys}, σ_{RFI}.
- X,Y,Z co-ordinates of an RFI source.
- Bandwidth and frequency of operation.
- RFI_{length}, $Period_{RFI}$ and RFI type.

The time domain signals are generated based on the generalised model as expressed in Section 4.3. This signal model takes the user defined parameters to generate the signal and pass them through a band pass filter if selected.

5.1.2 ADC Block

The quantization of the signals is of crucial importance and has to be done in accordance to the selected signal thresholds (Section 4.4). The number of bits for quantization is a user-tunable parameter.

5.1.3 Fourier Transform

After the quantization of signals a Short Time Fourier transform is obtained based on the selected number of FFT channels or the frequency resolution. While dealing with real signals

the effective number channels will be half as the FFT would output spectrum in negative frequency domain as well. However, when complex signals are passed through this pipeline the number of channels remain intact. User is also given the provision to choose the type of signals (real or complex) (s)he needed to pass through the chain. The channel selector provides the user the choice of selecting a particular frequency channel and viewing its variation with time. This is further charted graphically in the test results for better understanding.

5.1.4 Correlator

The Correlator block implements the same functionality as described in Section 3.5. The spectrum of signals is multiplied channel by channel to result in cross correlation products. The cross correlation products for each baseline are averaged in time and result in the accumulated output. See Figure 3.10.

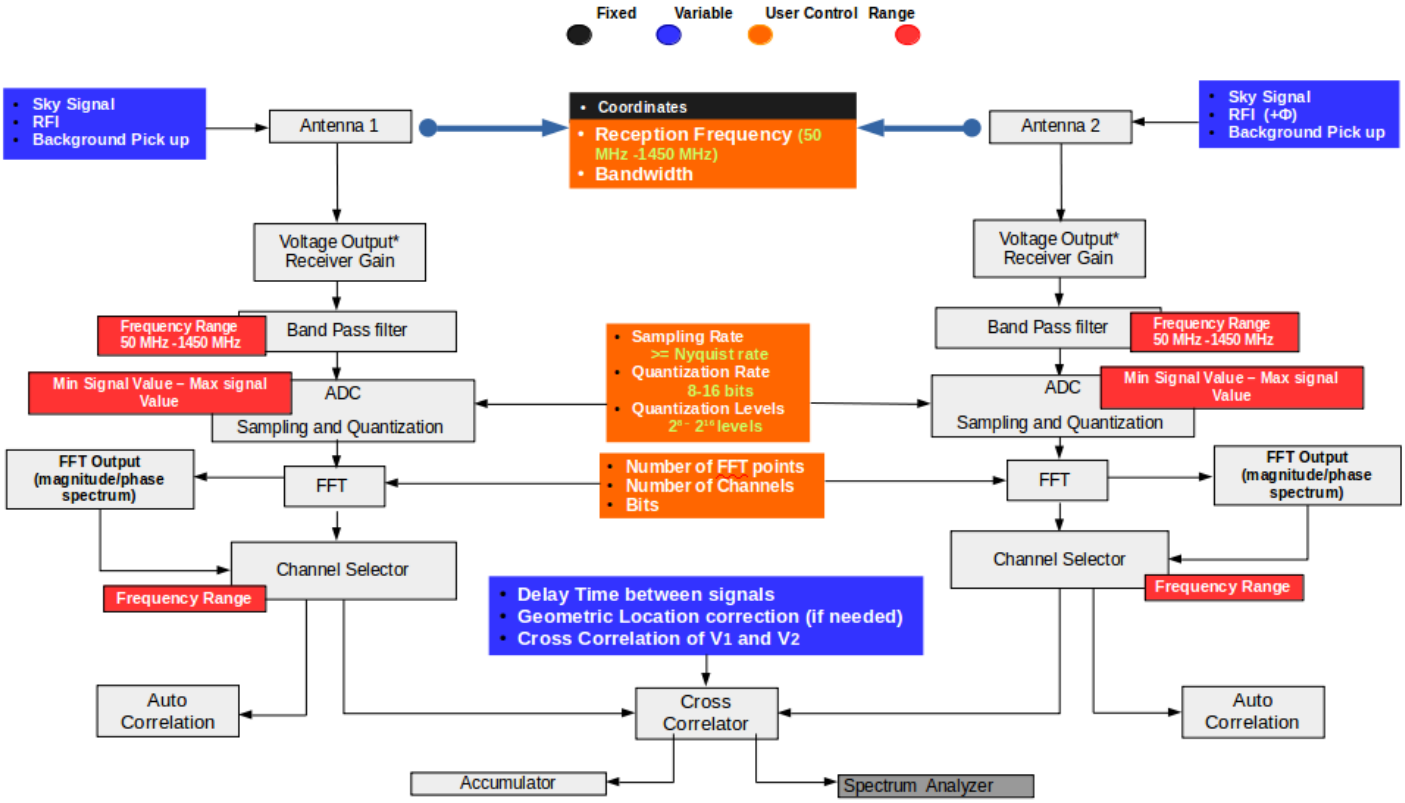


Figure 5.1: Simulation Blocks

5.2 Test Results from the Generated RFI

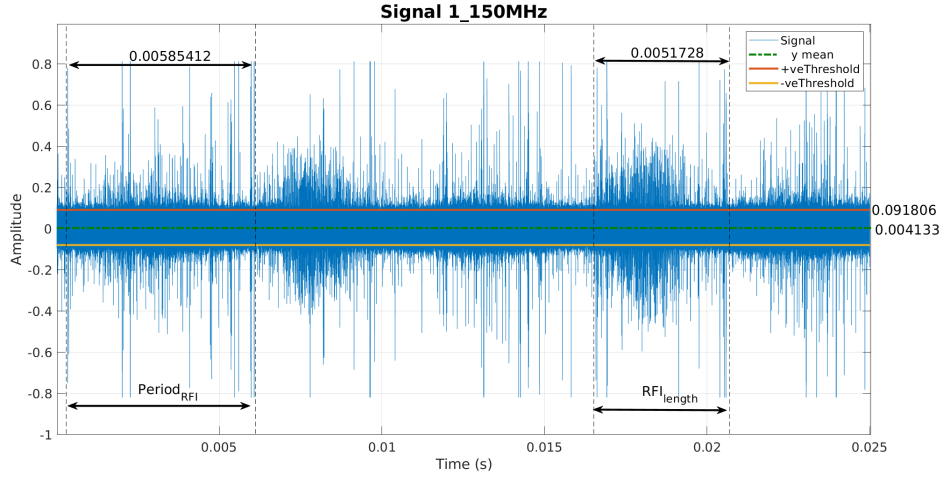
The results of the computational model of two element interferometer are produced in this section.

5.2.1 Results from No RFI Signal

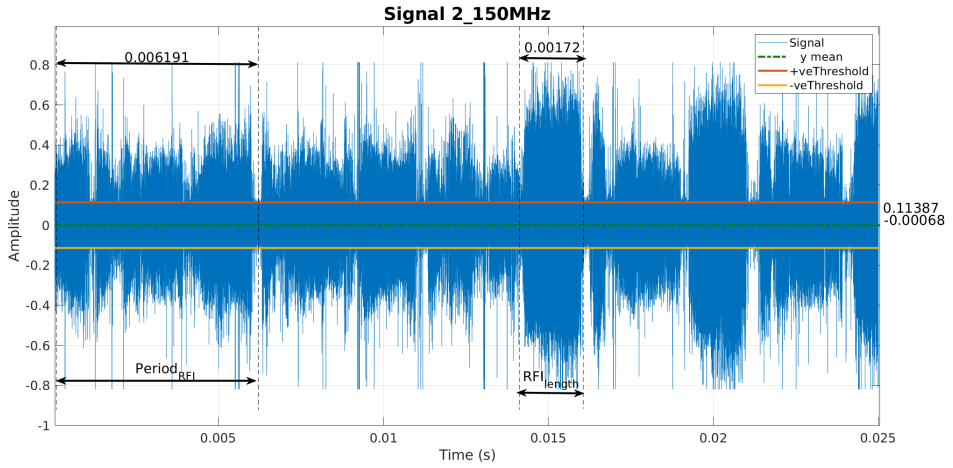
5.2.2 Results from the Cross Correlation of actual signals

Results for 150 MHz frequency band

The Figure 5.2 shows the actual signals recorded using one of the antennas of the GMRT antennae with a sampling rate of 200 MHz. The time length of each signal is 0.025 sec. Figures (a), (b) and (c) in Figure 5.3 show the auto correlation and cross correlation of signals mentioned here above. The auto correlation and cross correlation were calculated with 2048 FFT samples.

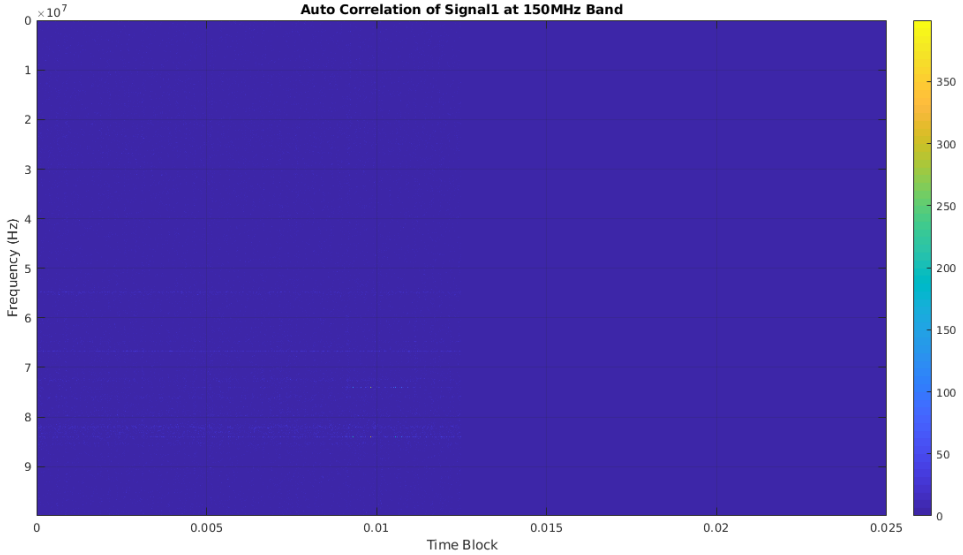


(a)

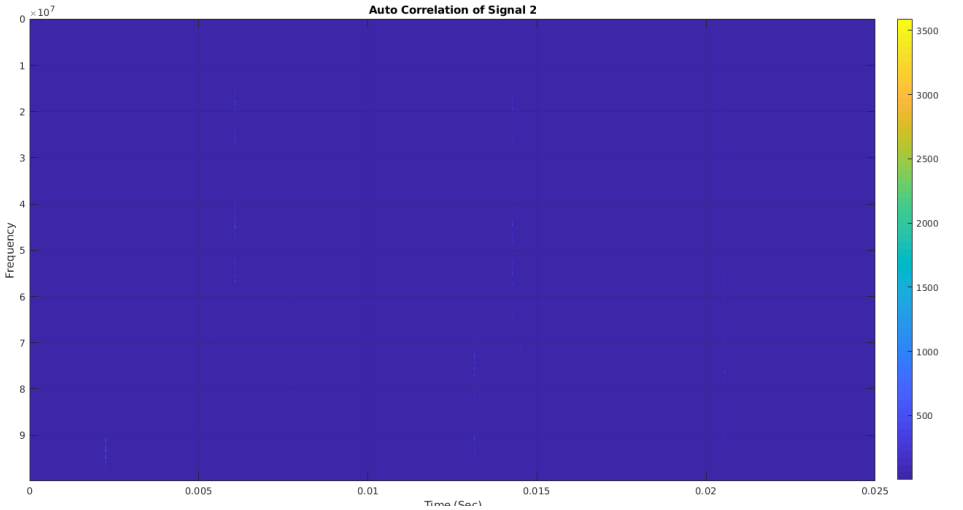


(b)

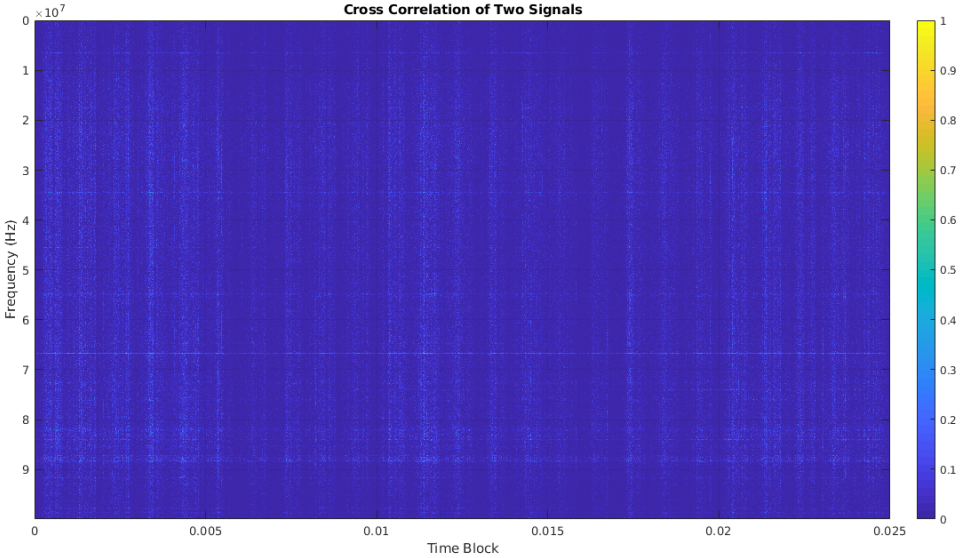
Figure 5.2: Actual signals for used for Auto Correlation Cross Correlation at 150 MHz band



(a) Auto Correlation of Signal 1



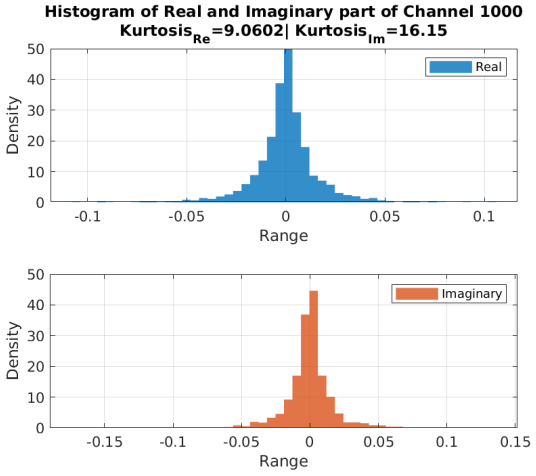
(b) Auto Correlation of signal 2



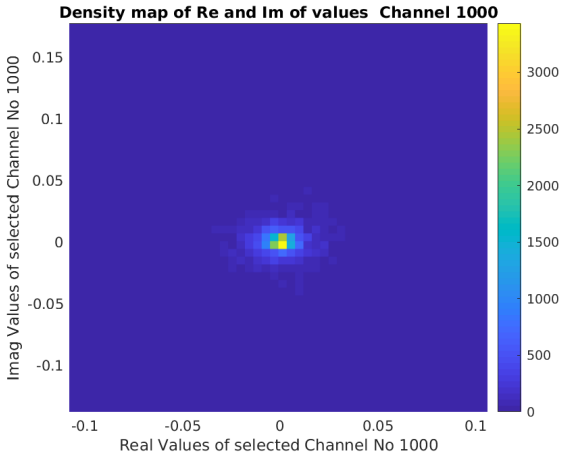
(c) Cross Correlation of signals

Figure 5.3: Auto Correlation and Cross Correlation of signals at 150 MHz with 2048 FFT samples.

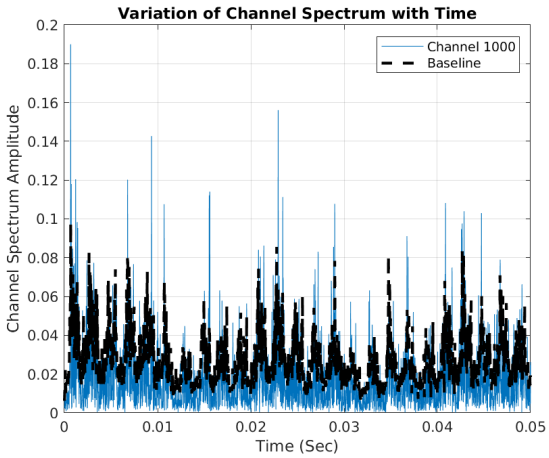
Plots shown in Figures 5.4 and 5.5 are extracted from cross correlation and auto correlation of the signals recorded at 150 MHz band. If there was no RFI in the signal, the signal would be purely Gaussain with only the source and the receiver component. Then the real and imaginary components of the dynamic spectra of an particular frequency channel would will follow a normal distribution and will be mesokurtic (Kurtosis = 3). However, the plots referenced below do not share the same story due to injection of RFI, the behaviour of the dynamic spectrum has changed from mesokurtic to leptokurtic (Kurtosis > 3). Figure 5.4(a) shows the histograms real and imaginary parts of the dynamic spectrum of Channel 1000. Figure 5.4(b) shows the 2D density map of these histograms. Figure 5.4(c) shows the variation of channel against the mean of the all the channels with time. The mean is represented as Baseline in the figure.



(a) Histograms of Re and Im components of Ch1000.



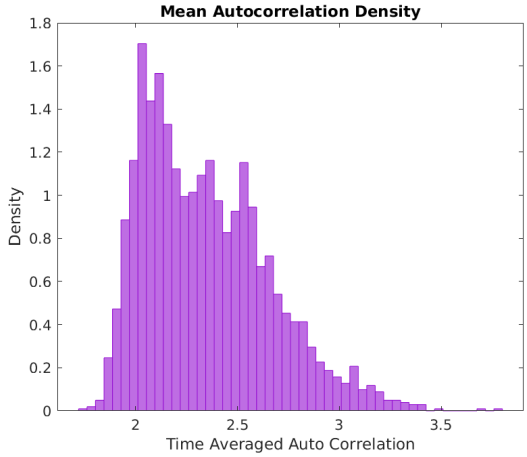
(b) A 2D density map of Re and Im components of Ch1000



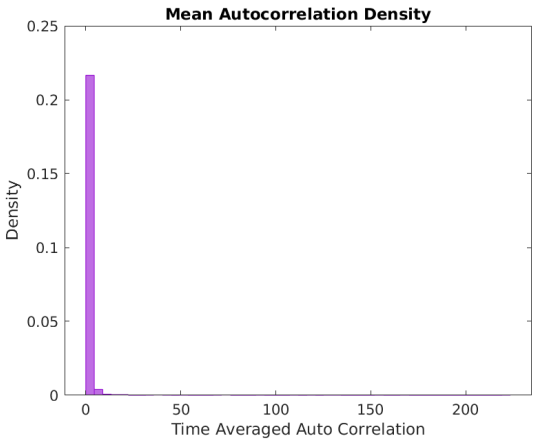
(c) Cross corr variation of Ch1000 against time

Figure 5.4: Cross Correlation Plots at 150MHz

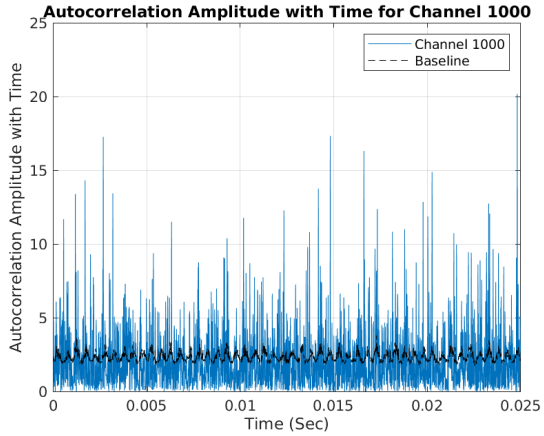
In the Figure 5.5 the auto correlation variation with time of channel 1000 is plotted. Figure 5.5(a) and 5.5(b) represents the histogram of the mean of the auto correlations of all the channels against time. The auto correlation is expected to follow a Rayleigh distribution as it represents the magnitude of a complex number. However, due to the presence of RFI, the histograms does not seem to follow this trend. Figures 5.5(c) and 5.5(d) depict the variation of auto correlation of channel 1000 of signal1 and signal2 with time respectively.



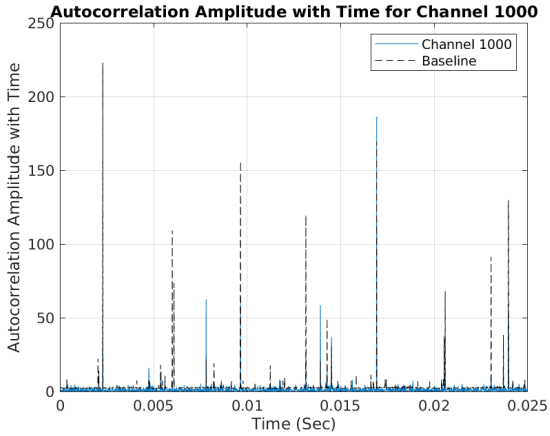
(a) Mean Auto correlation Density of Signal 1



(b) Mean Auto correlation Density of Signal 2



(c) Variation of Auto correlation of Channel 1000 of signal 1

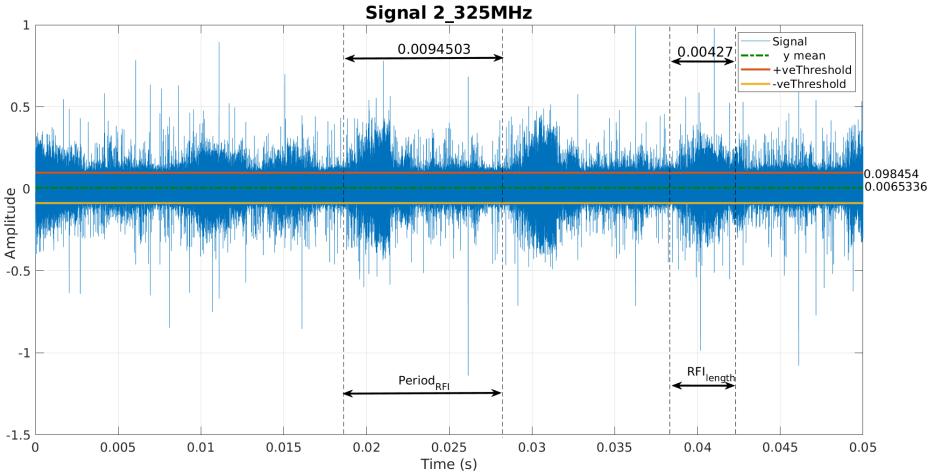


(d) Variation of Auto correlation of Channel 1000 of signal 2

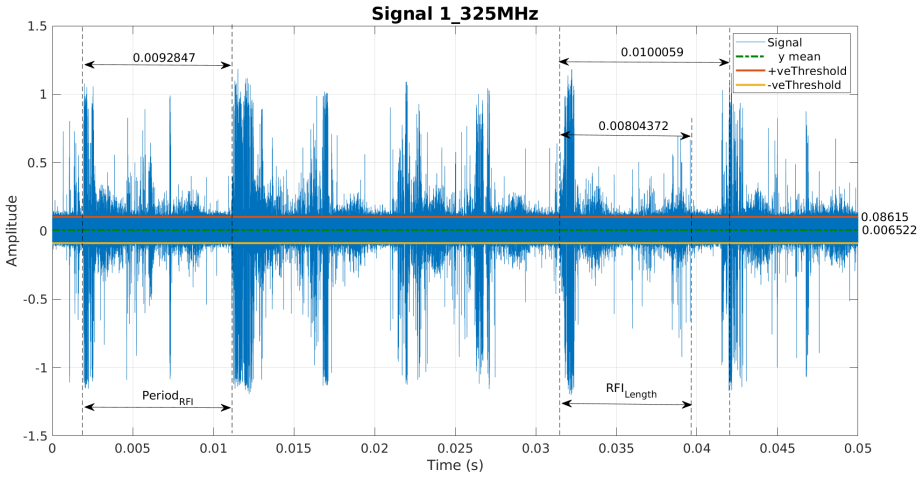
Figure 5.5: Auto Correlation Plots at 150MHz

Results for 325MHz

The Figures 5.6 to 5.9 show the signals recorded at 325 MHz band at GMRT and corresponding Auto correlation and Cross correlation plots as described in Section 5.2.2. The signals were recorded at a sampling rate of 200 MHz and the time duration of each signal is 0.05 sec.

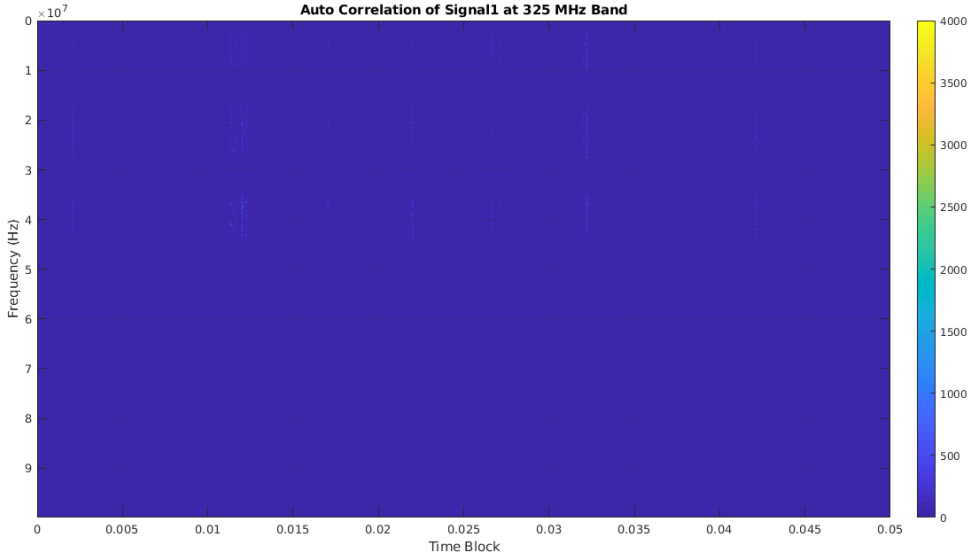


(a)

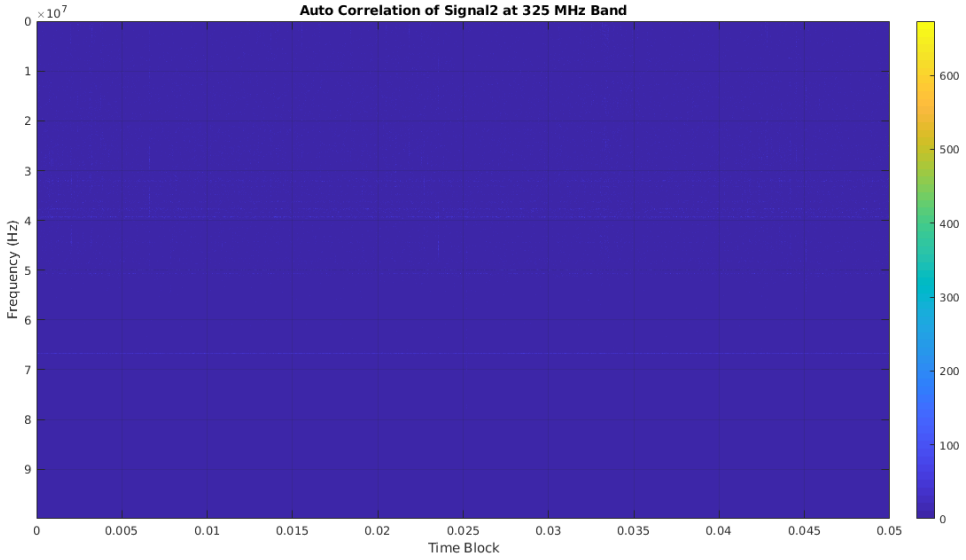


(b)

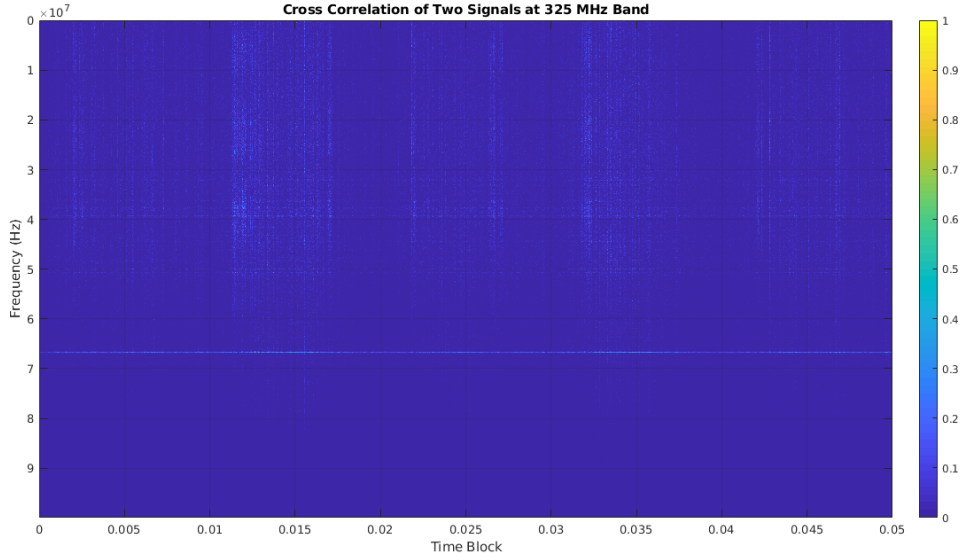
Figure 5.6: Actual signals for used for Auto Correlation Cross Correlation at 325 MHz band



(a) Auto Correlation of Signal 1

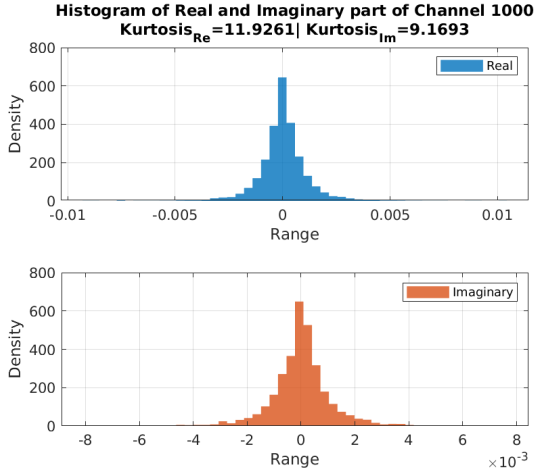


(b) Auto Correlation of signal 2

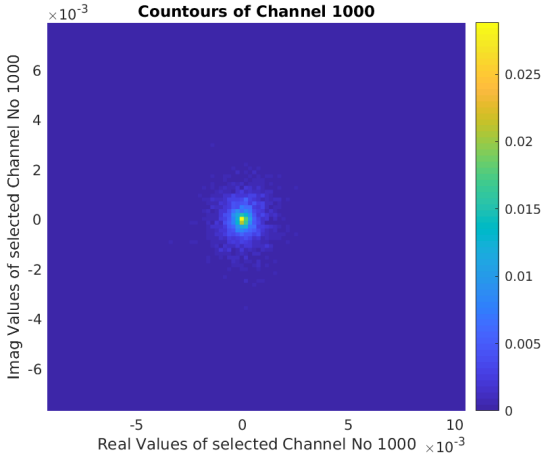


(c) Cross Correlation of signals

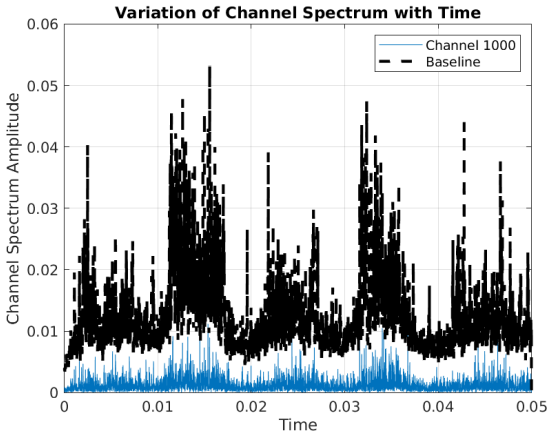
Figure 5.7: Auto Correlation and Cross Correlation of signals at 325 MHz band with 2048 FFT samples.



(a) Histograms of Re and Im components of Ch1000.

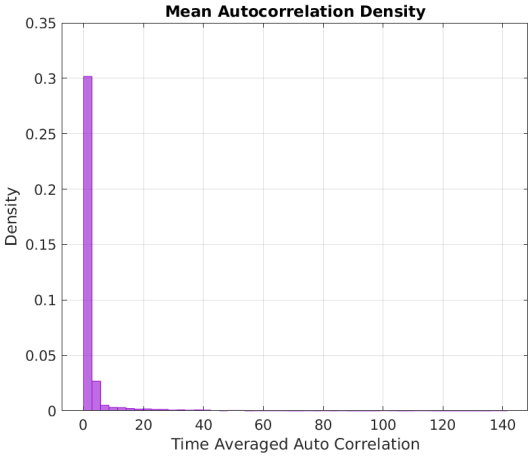


(b) A 2D density map of Re and Im components of Ch1000

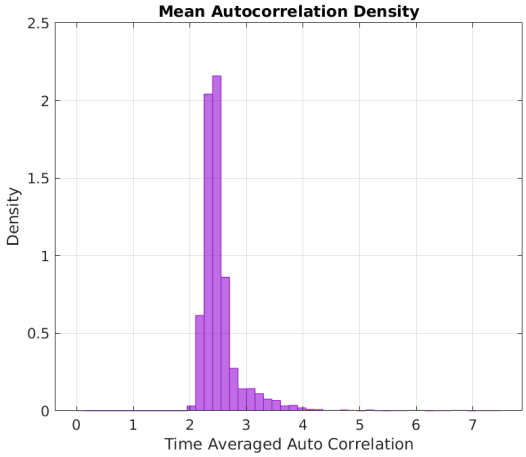


(c) Cross corr variation of Ch1000 against time

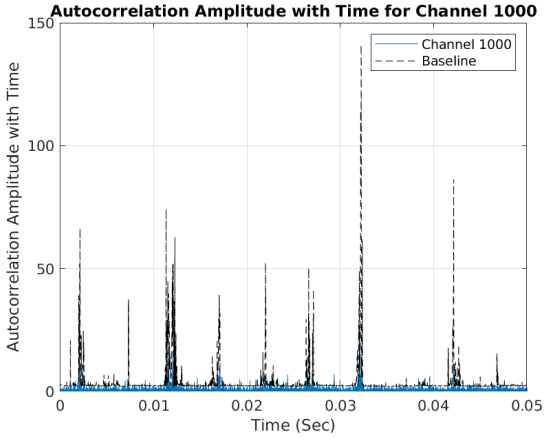
Figure 5.8: Auto Correlation and Cross Correlation Plots at 325MHz



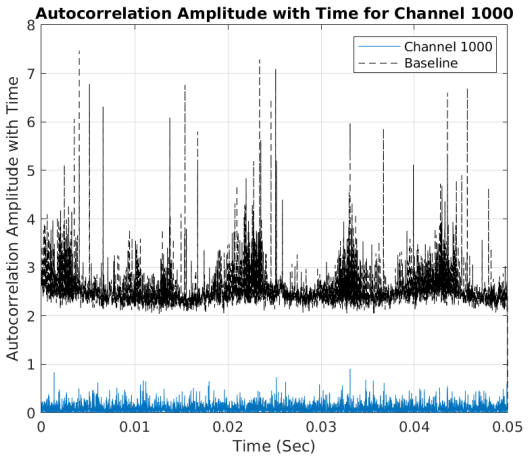
(a) Mean Auto correlation Density of Signal 1



(b) Mean Auto correlation Density of Signal 2



(c) Variation of Auto correlation of Channel 1000 of signal 1



(d) Variation of Auto correlation of Channel 1000 of signal 2

Figure 5.9: Auto Correlation Plots at 325MHz band

5.2.3 Artificially Generated RFI and Respective Cross Correlation Plots

The below plots from Figure 5.10 to 5.17 show artificial RFI signals generated using the techniques and algorithms described in Chapter 4. All these signals were generated with a constant **quantization** of **16 bits**. The chosen sampling frequency and time duration of the signals are 200 MHz and 0.05 sec. The parameters used for generating these signals are labelled in each of these plots.

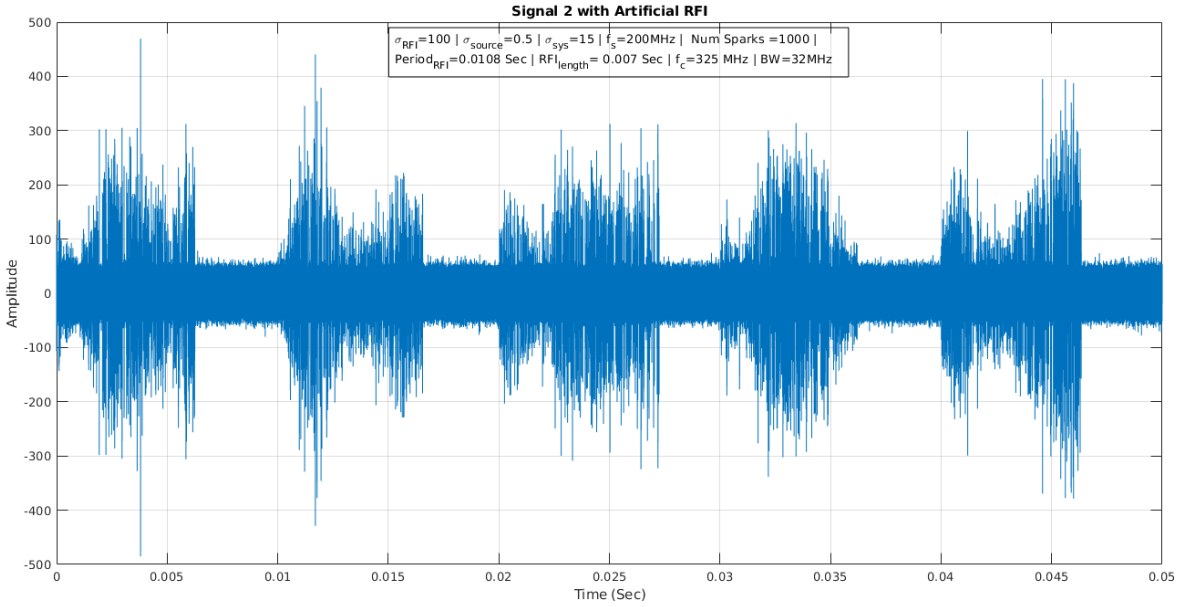


Figure 5.10: Artificial RFI signal 1

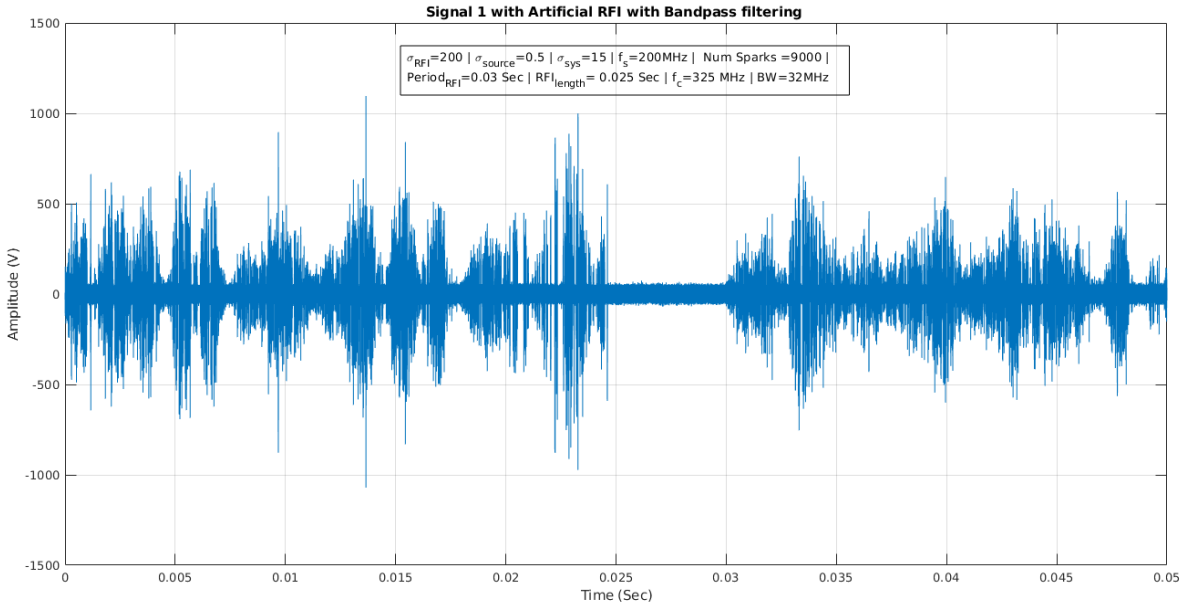


Figure 5.11: Artificial RFI signal 2

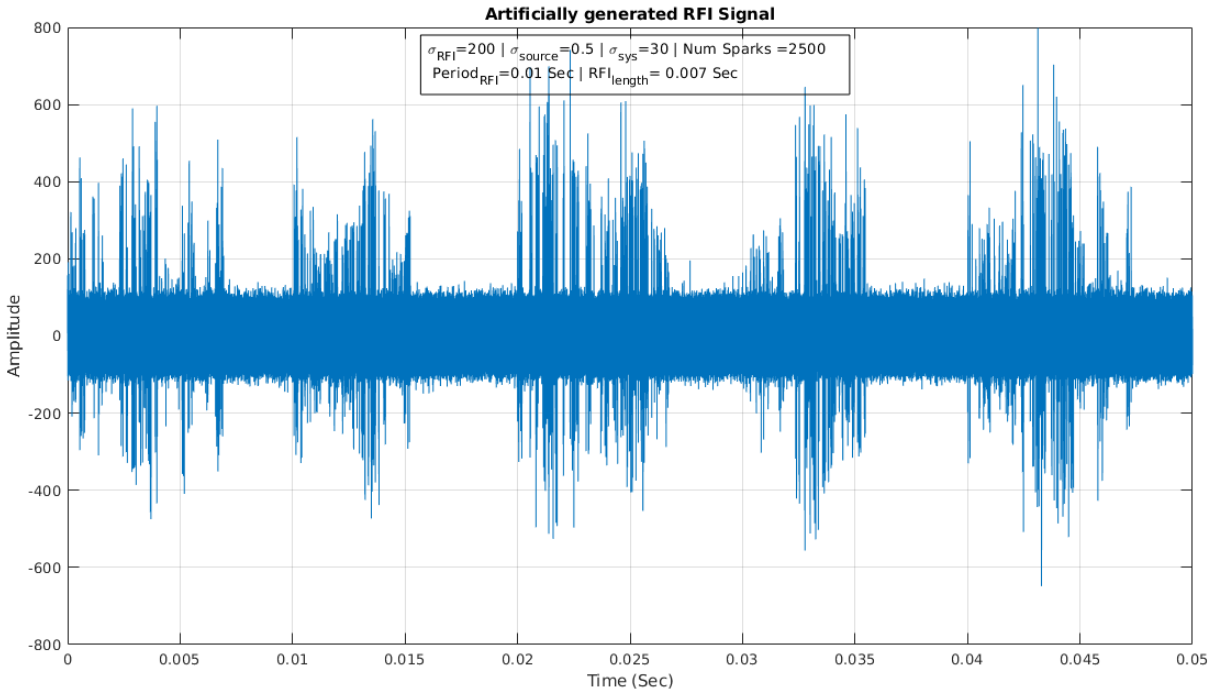


Figure 5.12: Artificial RFI signal 2

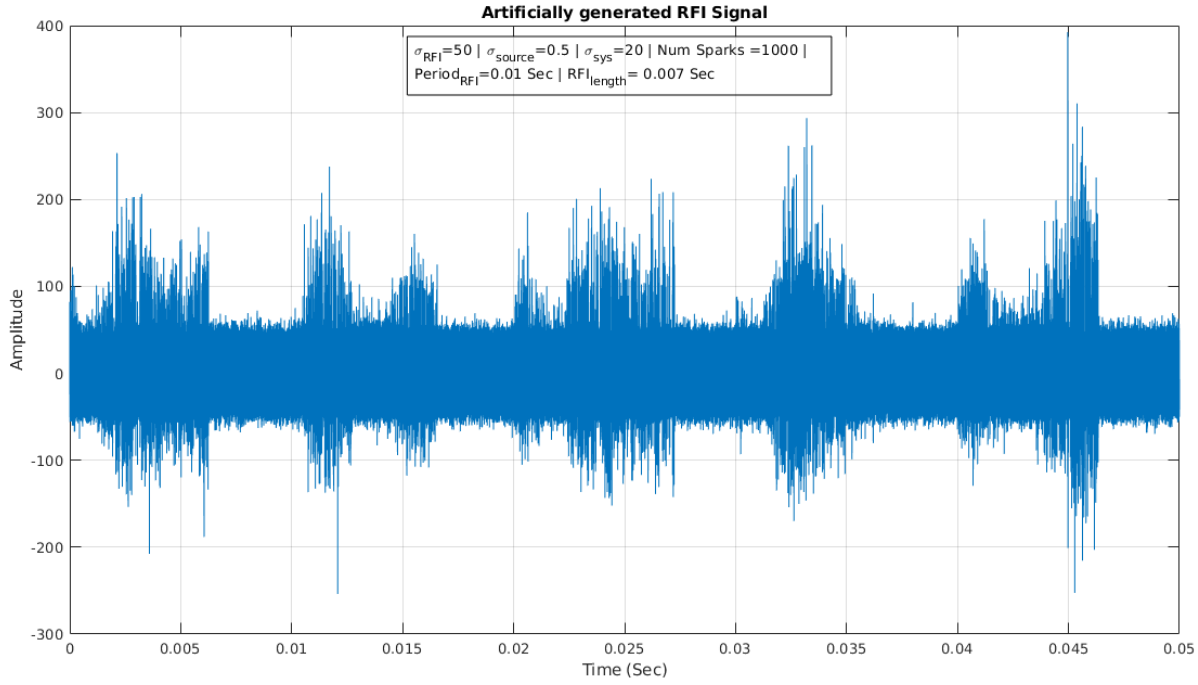


Figure 5.13: Artificial RFI signal 4

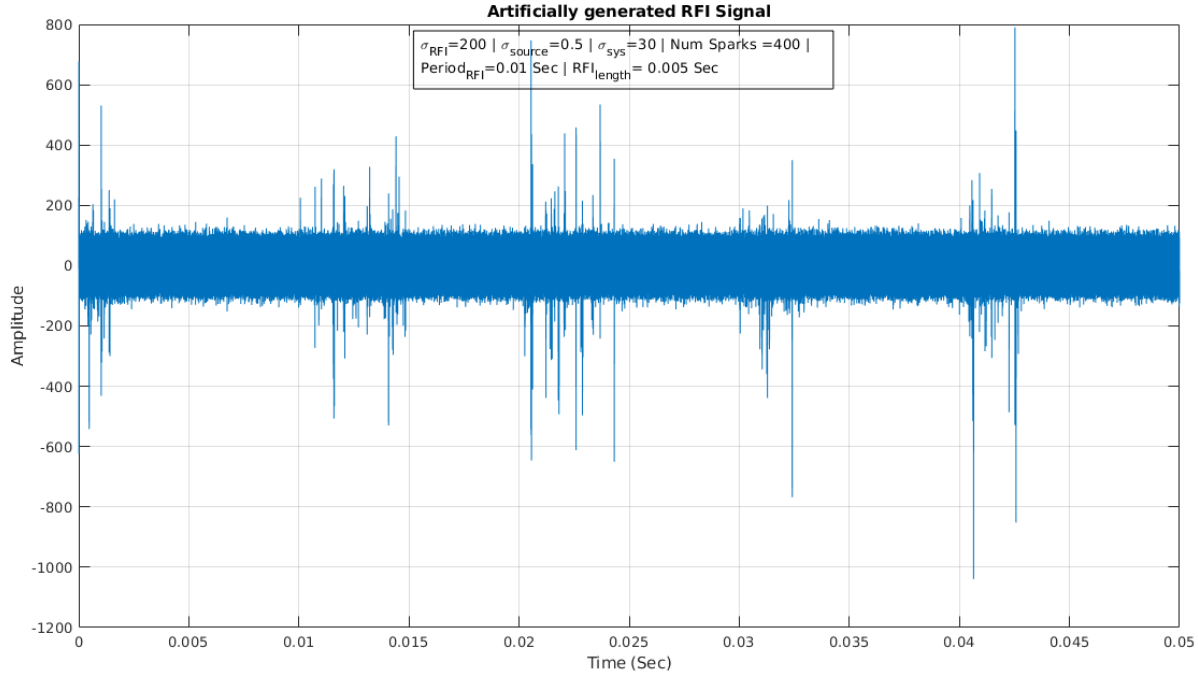


Figure 5.14: Artificial RFI signal 5

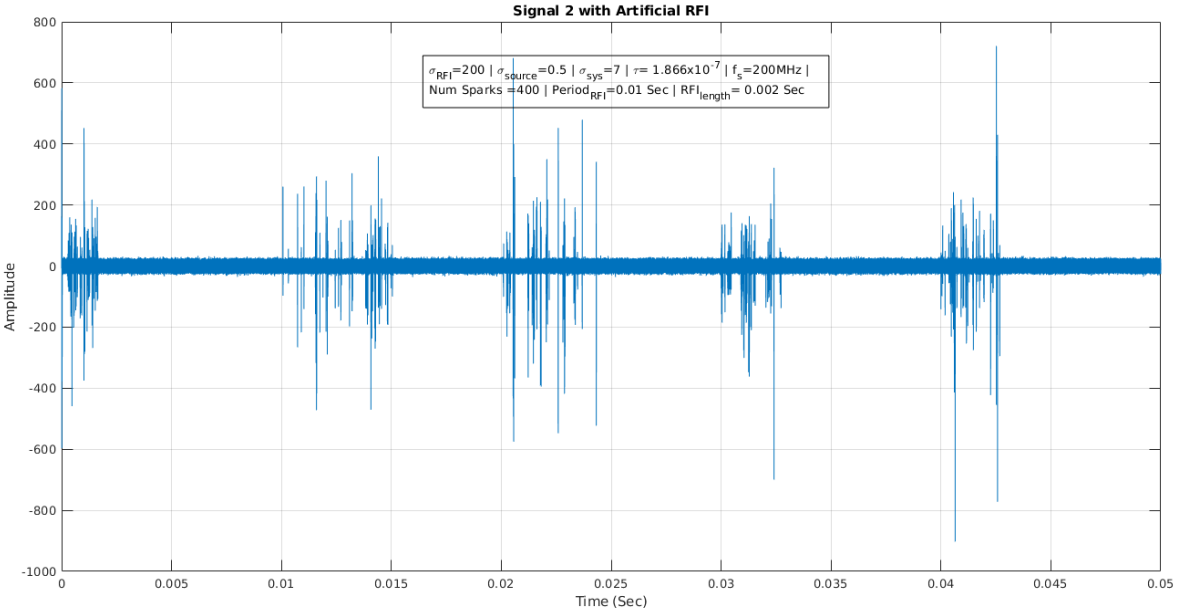


Figure 5.15: Artificial RFI signal 6

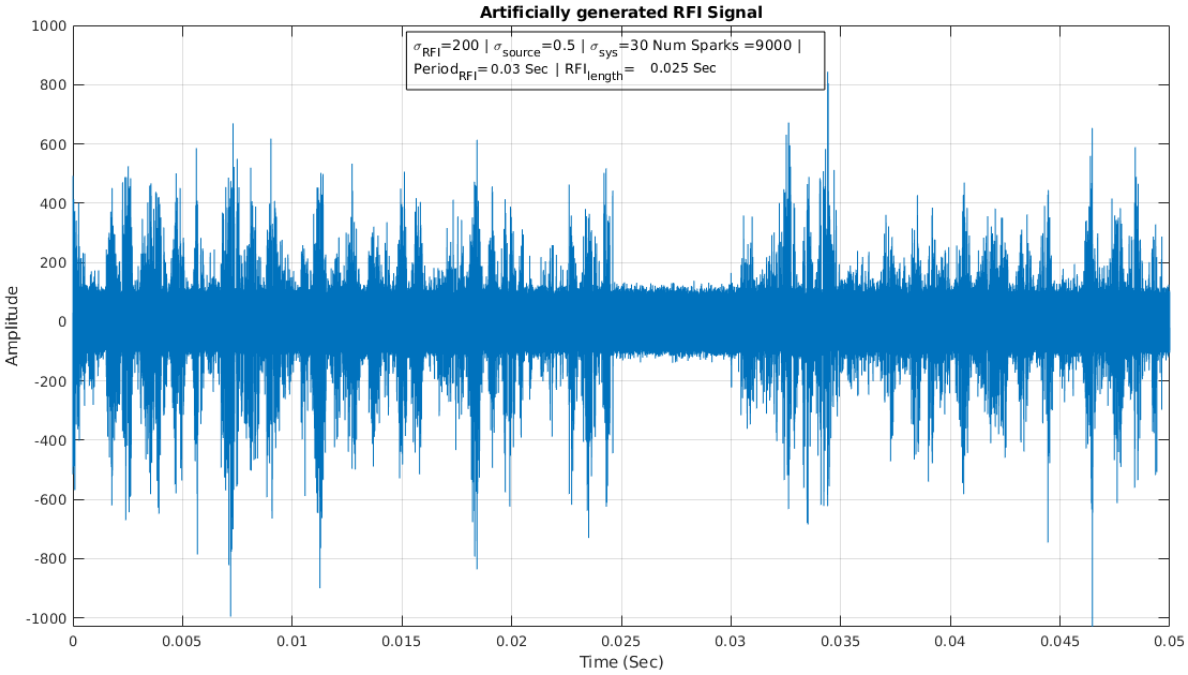


Figure 5.16: Artificial RFI signal 7

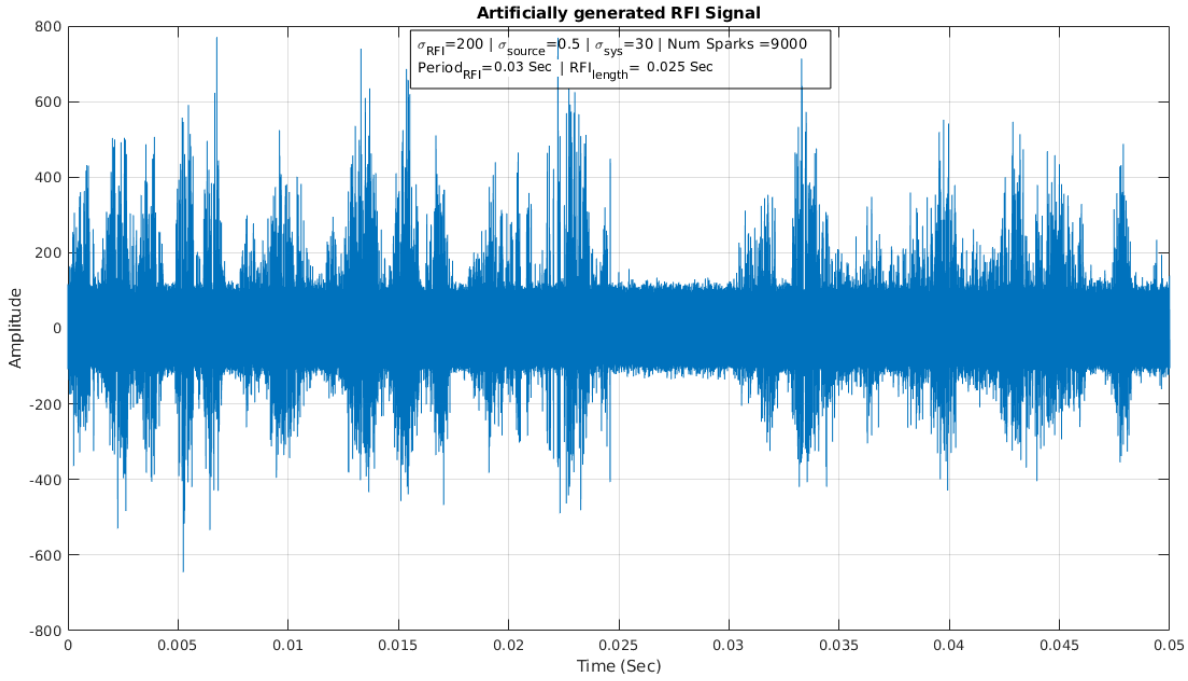
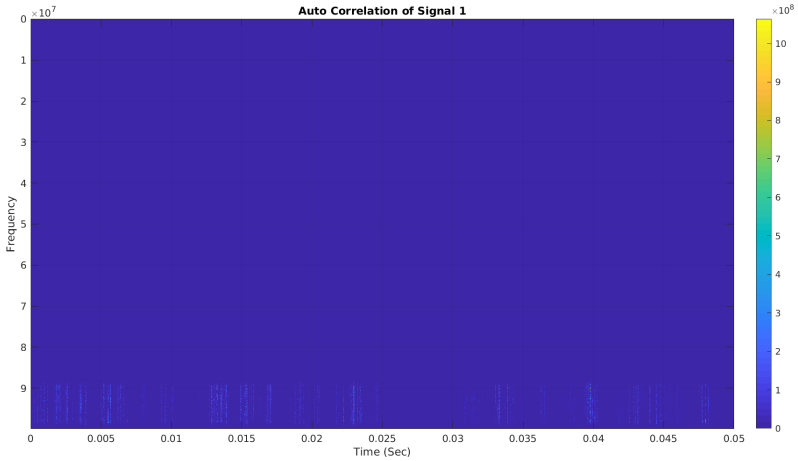


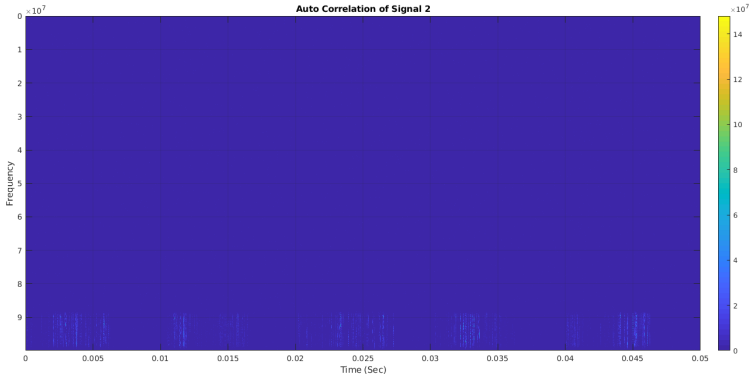
Figure 5.17: Artificial RFI signal 8

Cross Correlation and related plots for signal in Figure 5.10 and signal Figure 5.11

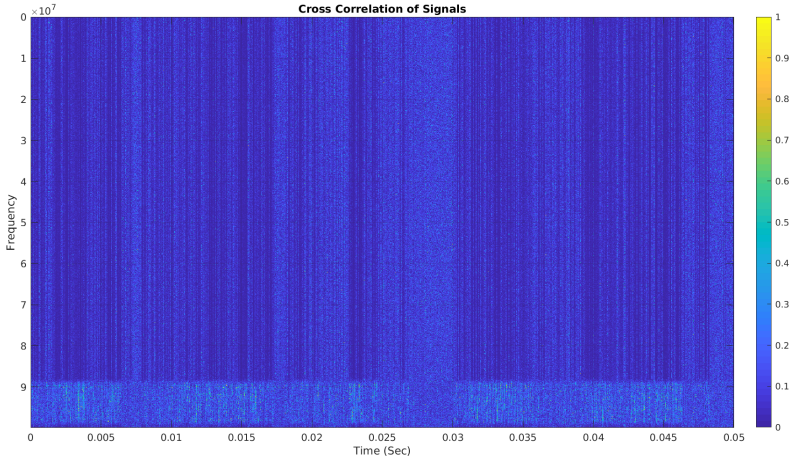
Here we produce the test results of one of the simulation tests. The test was done using signals in Figures 5.10 and 5.11. The cross-correlation and the auto-correlation plots of the same are depicted below.



(a) Auto Correlation of Signal 1

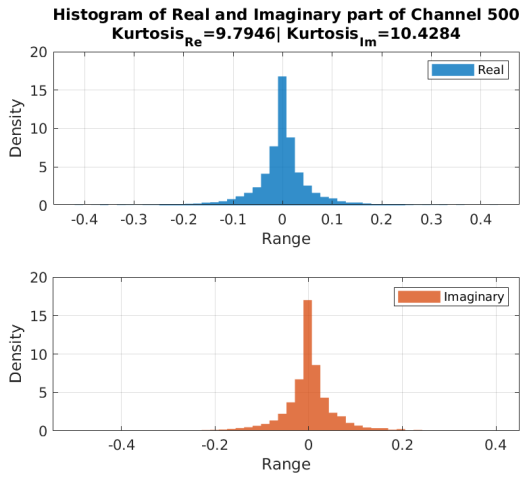


(b) Auto Correlation of Signal 2

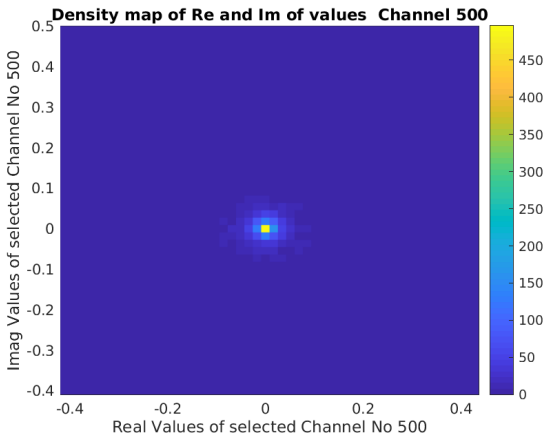


(c) Cross Correlation of signals

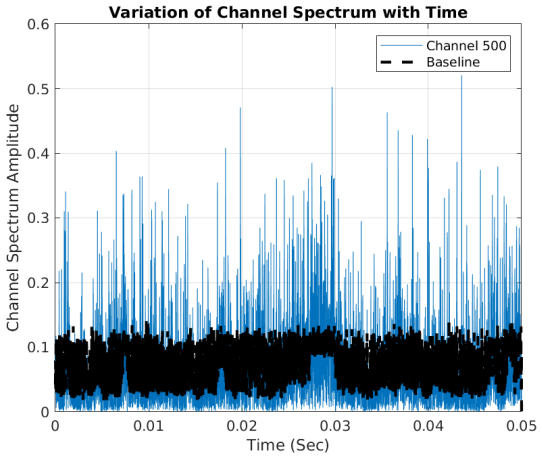
Figure 5.18: Auto Correlation and Cross Correlation of artificial signals with 2048 FFT samples.



(a) Histograms of Re and Im components of Ch500.

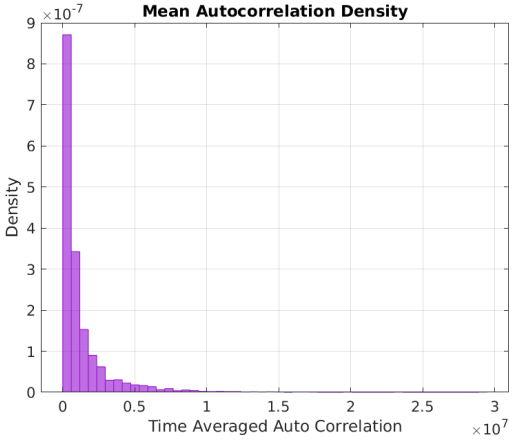


(b) A 2D density map of Re and Im components of Ch500.

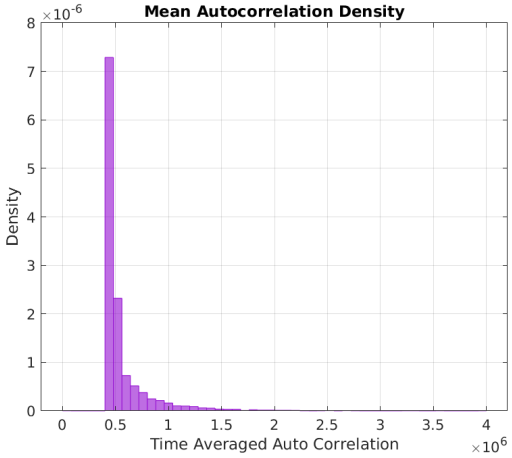


(c) Cross correlation variation of Ch500 against time

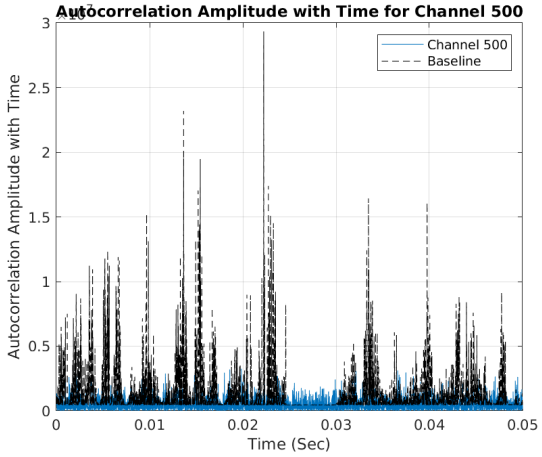
Figure 5.19: Cross Correlation Plots for a user generated signal



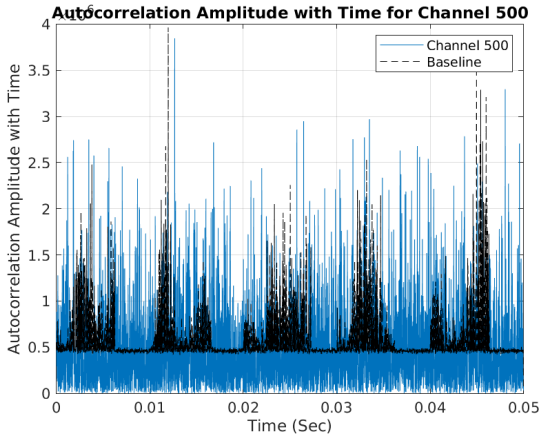
(a) Mean Auto correlation Density of Signal 1



(b) Mean Auto correlation Density of Signal 2



(c) Variation of Auto correlation of Channel 500 of signal 1



(d) Variation of Auto correlation of Channel 500 of signal 2

Figure 5.20: Auto Correlation Plots of user defined signal

Chapter 6

Conclusion and Future work

The presence of RFI poses a great challenge for astronomers and astrophysicists to extract interesting science from radio observations. Though the GMRT is among the leaders globally for developing an deploying efficient and innovative RFI excision algorithms, flourishing urban population in the nearby regions of the GMRT poses a risk for effective mitigation of the powerline RFI. The work presented here focuses on the statistical behavior of RFI and development of a tool to imitate and analyse the signals from powerline radiation. In Chapter 1 fundamentals of radio astronomy was presented which formed the foundation for the modeling tool. Next we outlined the two types of poweline RFI which corrupt the astronomical signals. The spark discharges are due to gaps in the conductor (electrodes) and corona discharges are due to ionization of air around the conductor. In Chapter 3 we detailed the basic components used to obtain the visibility and their functionality. Each of the component described forms an integral part of the simulation tool from which the visibilities of the artificial signals are obtained. In Chapter 4 we described the statistical behavior of the observed RFI at GMRT and using those statistics how various signals can be generated for further investigation. From a detailed examination of very high resolution data captured at GMRT, we concluded that the RFI signal appears in a series of discrete sparks and is not a continuous disturbance. This allowed us to estimate the inter-arrival times of these signals and attempt to identify RFI signals on spark to spark basis. An attempt was also made to understand their underlying behaviour and look for a formulaic pattern however, due to their unpredictable nature no concrete approach could be found that could quantify this nature. While this holds to our previous understanding of the randomness of this phenomenon at a macro level contrary to our expectations, we were unable to find an underlying functional form based on expectations from spark discharge models. Chapter 5 outlines the blocks of the simulation framework and present the results that are obtained from the simulation.

The observed RFI is completely stochastic in nature and hence applying general principles of prediction do not seem promising. An approach to time series forecasting would be limited to only few milliseconds of data up to which a similar pattern occurs. This forecasting would also not provide user the control of various parameters that the user can choose to generate the RFI and look for its properties in the visibility domain. This project presents the most comprehensive attempt yet to understand the properties of the RFI observed at GMRT and to model it with the objective of contributing toward eventually being able to efficiently identify and remove the corrupting RFI signals. As a part of this project, we examined in detail many high resolution data set recorded at GMRT to build a statistical description of the observed RFI. This work has led to the realization that individual RFI bunches show so much variation from one data set to another that it seems futile to look for a parametrized analytical description of sparks in a bunch. Another objective of this project was to build a simulation tool to quantify

the impact of RFI on the final observable of radio interferometry- the visibilities. To meet this objective a simulation framework modeling the entire GMRT signal chain was implemented, from the front end all the way to the output of the correlator.

This project started with an expedition to know the general properties of powerline RFI. During this journey many new ideas were left unexplored and we would like to layout what tasks can be taken up next which not definitely but seemingly would take us close to understanding this random phenomenon.

- Using a detailed Functional Data Analysis approach.
- Possible use Markov process to generate the powerline RFI.
- Understanding the cyclostationarity pattern of RFI and implementing the same in the signal generation.
- Optimise the simulation with parallel processing.
- Exploring potential machine learning algorithms for pattern generation, specifically Convolutional Neural Networks, Encoders and Adverserial Networks.
- Prediction of $Period_{RFI}$ and RFI_{length} from the data using Maximum Likelihood Method and EM algorithm.
- Extend this study to see the effect of RFI on image by taking a 2-D FFT, primarily applying the van CittertZernike theorem.

Bibliography

[1] Swarup, G., "Power-line Radio Frequency Interference at the GMRT," 2008.

[2] N. Rajaram, "Signals in Radio Astronomy," in *Low Frequency Radio Astronomy* (J. N. Chengalur, Y. Gupta, and K. Dwarakanath, eds.), ch. 1, Pune 411007, India: National Institute for Radio Astrophysics (NCRA-TIFR), 3 ed., 2007.

[3] P. K. Langat, *Power-line sparking noise characterisation in the SKA environment.* PhD thesis, University of Stellenbosch, Department of Electrical Electronic Engineering, 2011.

[4] J. D. Kraus, "Radio Astronomy Fundamentals," in *Radio Astronomy*, ch. 3, McGraw-Hill Inc, 2 ed., 1988.

[5] G. W. S. J. A. Richard Thompson, James M. Moran, "Van CittertZernike Theorem, Spatial Coherence, and Scattering," in *"Interferometry and Synthesis in Radio Astronomy"*, ch. 15, Springer Open, 3 ed., 2016.

[6] Griffin Foster, "van Cittert-Zernike Theorem," 2016. Lecture Slides, Rhodes University.

[7] G. W. S. J. A. Richard Thompson, James M. Moran, "Geometrical Relationships, Polarimetry, and the Measurement Equation," in *"Interferometry and Synthesis in Radio Astronomy"*, ch. 4, Springer Open, 3 ed., 2016.

[8] Peter Joseph Bevelacqua, "Antenna Theory." http://www.antenna-theory.com/basics/main.php.

[9] Steven W. Smith, "Introduction to Filters." http://www.dspguide.com.

[10] Steven W. Smith, "ADC and DAC." http://www.dspguide.com.

[11] Bhalchandra Pujari, "Fourier Transforms and Complex Analysis," 2018. Lecture Notes, Savitribai Phule Pune University, Centre for Modeling and Simulation.

[12] J. N. Chengalur, Y. Gupta, and K. Dwarakanath, eds., *Low Frequency Radio Astronomy.* Pune 411007, India: National Institute for Radio Astrophysics (NCRA-TIFR), 3 ed., 2007.

[13] Kaushal D. Buch et al, "Real-Time Implementation of MAD-Based RFI Excision on FPGA," *Journal of Astronomical Instrumentation*, vol. 08, 2019.

[14] Kaushal D. Buch and et al, "Implementing and Characterizing Real-Time Broadband RFI Excision System for the GMRT Wideband Backend," *IETE Technical Review*, 2018.

[15] NIST SEMANTECH, "Measures of skewness and kurtosis." https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm.

[16] Kaushal D. Buch and Anand D. Darji, "FPGA-based emulator for functional verification of radio frequency interference mitigation algorithms," *Annual IEEE India Conference (INDICON)*, 2013.

[17] Mihir Arjunwadkar, "Sampling via Transformations of the Uniform," 2018. Lecture Notes, Savitribai Phule Pune University, Centre for Modeling and Simulation.

[18] Aurore Delaigle and Peter Hall, "Defining Probability Density For a Distribution of Random Functions," *The Annals of Statistics*, vol. 38, 2010.

[19] S. Harshvardhan Reddy, Sandeep C Chaudhari, "ADC Input Power to Correlator Counts Study Experiment," 2017. Experiment notes at GMRT.