Final report of the AST398 group's effort at the Giant Metrewave Radio Telescope to remove RFI and investigate the epoch of reionization

Joshua Albert Nidhi Banavar Mark Kuiack Connie Lien
Hans Nguyen

26 August, 2011

Abstract

The primary goal of this project is to devise a standard procedure to locate and eliminate radio frequency interference (RFI), within the near vicinity of the Giant Metrewave Radio Telescope, within about ≤ 20 km from the central square, and ultimately make the procedure streamlined so that others may carry it out. The second aim is to continue the data collection required for CITA's research group which is to continue the study of the structural evolution of hydrogen during the epoch of reionization. The RFI is detected by applying singular value decomposition on short observation data, as well as on epoch of reionization data, and selecting the twenty largest eigenvalues. In two cases RFI is detected by directly monitoring the 150 MHz channel while in the field. Several positive sources of radio frequency interference were located and, in some cases, initial steps of removal were initiated. A prospective method to aid in the detection of RFI is also proposed.

1 Introduction

1.1 Science Behind Epoch of Reionization

The Epoch of Reionization (EoR) is the greatest frontier being explored observationally by cosmologists today. The large scale structure of the universe can be revealed by studying the distribution of the very first sources of ionizing radiation, (stars and galaxies). The anisotropic nature of thermal and radiative energy in the cosmic dark ages is also probed by measuring the density perturbations during the EoR.

Due to the fact that the dark age is the era that connects two important areas of study in cosmology and astrophysics, the study of the Large Scale Structure (LSS) in

the universe and the Cosmic Microwave Background (CMB), it will likely prove to be a fruitful repository of information. The study of the early stages of the EoR will reveal the conditions of the universe during the formation of the universes first luminous objects, which in turn should yield new information for CMB and LSS study [2]. It will allow for the study of the CMB at greater resolution than currently observable, and reveal the evolution of LSS which have been subject to the influence of complex non-linear gravitational forces at lower redshifts.

The EoR began near the end of the dark ages when the more dense regions in the universe collapsed into structures which accommodated stellar creation. It has been shown with measurements of the velocity/radius curves of nearby galaxies that dark matter must have a significant effect on maintaining the large scale gravitationally bound structures in the universe, such as galaxies. Regular matter, however, must have been present within the regions of high dark matter density for stars and galaxies to be created.

The ionizing radiation emitted from the earliest galaxies was absorbed by the surrounding neutral hydrogen that made up the inter galactic medium (IGM). This ionized gas has a much higher kinetic gas temperature and therefore a higher gas pressure. This pressure resists further gravitational collapse and acts as a self-regulating throttle to stifle the rapid creation of very large galaxies.

A result of this is that smaller dwarf galaxies, around 10^8 $M_\odot$[5], were likely responsible for most of the radiation throughout the EoR. This behaviour leads to the rapid ionization of the high density regions in the beginning, followed by the very gradual ionization of the lowest density voids between the galaxies. Early luminous matter galaxies at the greatest redshift will only be present in the regions of high density. Furthermore, the reionization will occur with a complex looking highly inhomogeneous and patchy geometry [1].

We are focused on observing in the 150 MHz (200 cm) band of the electromagnetic spectrum. This is the location of the 21 cm radiation line which has since been redshifted to ten times the emitted wavelength, $z \approx 9$. In this region of time and space the envelopes of ionized matter should still be small, restricted to the high density regions.This means that, at early times, or low frequencies, a high resolution large beam size is preferable [1]. The difference in radiation temperature of the CMB will be on the order of 10 mK.

With a lower resolution array like the Giant Metrewave Radio Telescope (GMRT), it is important to keep in mind that the frequency at which the observed root-mean-square temperature peaks will vary significantly. The peak position is determined by the difference between the brightness temperature of the neutral hydrogen and the CMB and by how well the beam size matches the patch size. It it therefore thought that only the planned Square Kilometre Array (SKA) will have the sensitivity and resolution required to observe the earliest stars at the end of the cosmic dark ages. The GMRT however has the benefit of a very large collecting area, 48000 m^2, which allows for statistical detection of small fluctuations on much shorter integration times for a given sensitivity level [1].

The GMRT is a 25 km diameter array of thirty, 45 m parabolic dishes, fourteen of which span over 1 km^2 of the central square.[4] The giant diameter increases the resolution which, for the observing frequencies we use, yields an angular resolution of 20 ” as well as increasing the brightness sensitivity. Since we are interested in the half way point of the EoR, which is believed to to be $8.1 \sim z \sim 9.2$, we observe at 150 MHz with a bandwidth of 16.7 MHz[3]. The antenna feeds each record two circularly polarized signals, one left and one right [4]. This data is then fed to the GMRT Software Backend (GSB) which records with a frequency and time resolution of 7.8 kHz and 0.25 s.

The method used to observe the patches of ionized hydrogen at the EoR require a pulsar in the line of sight. During the summer of 2011 the team predominantly used the pulsar B2045-16. The pulsars period is divided equally into 16 parts, each of which is assigned to a gate, one of which contains the pulsar flash. The pulse is used to equalize the signal gain from each antenna in real time during an observation [3]. Given that the period of fluctuation of the ionosphere is much longer than that of the pulsar during each of the pulsars on/off cycles, a new signal gain and phase calibration is made. Then, the gate in which the pulse appears is removed form the final data set.

1.2 Goals and Aims

The first goal of this project was to search for radio frequency interference (RFI) and eliminate it, and to create a near-automated procedure that continues this in the future. The procedure would identify RFI sites that would then be physically investigated by another arbitrary group of individuals.

The second goal was to continue data collection of the EoR so that the CITA group[1] may continue research into the structural evolution of hydrogen during the EoR.

Presented here first are the technical aspects of procuring data for RFI detection, see section 2.1. This is followed by the procedure to analyze the data to locate the candidate RFI source, see section 2.2, and how to deal with the source of RFI, see section 2.4. The specific RFI sources that were identified are presented, see section 3, along with discussion relating to the accuracy of the detection procedure, see section 2.5.

Following this is a discussion of the procedures involved with collecting EoR data, see section 4. A prospective method for better detecting RFI is shown in section 7. Finally, a conclusion will discuss future improvements and necessary objectives that need to be met in order to continue hunting for RFI.

[1] Ue-Li Pen et al.

2 Technical Aspects of Locating RFI

The process of mitigating RFI can be divided into three parts: detecting the RFI, physically locating the RFI, and removing the RFI.

2.1 Detecting the RFI

To detect a potential RFI source the array of telescopes at GMRT is used as a giant interferometer to make images of the near field radio emission. The setup used has the antenna array recorded on 150 MHz with the antennas pointed at the zenith and the 90 cm feed in focus. This configuration ensures the geography in focus is the horizon rather than the sky. A GMRT software backend then generates chimaps, which are visual maps created by the CITA research team. Chimaps have a domain of the real 40 km by 40 km area centred on the GMRT central square. The mode of a chimap is the eigenvalue rank when order by magnitude. For a visual example of a chimap see section 2.5.1.

2.2 Details Of Singular Value Decomposition

Singular value decomposition (SVD) is used to separate celestial and terrestrial static broadband radio sources. This is possible due to the fact that since terrestrial sources do not move as a function of time with respect to the array of radio telescopes, they are the dominant contributors to the largest eigenvalues. As the sky is rotating with respect to the array, the sources in the sky have less impact on the eigenvalues because they are not stationary. The data can be put into a matrix to which we apply SVD and obtain the twenty largest eigenvalues. Each mode is the data from one eigenvalue. [4]

Each mode can be represented as an image called a chimap. The chimap depicts multiple hyperbolas that intersect at a near-singular point. The hyperbolas are generated by the correlation of multiple antenna baselines. The brightest intersection point corresponds to a potential RFI source. The chimap is also a 2-D geographical representation of a 40 km by 40 km area centered on GMRT. In this way, the location of the brightest pixel on the chimap image corresponds to a geographical coordinate. A script extracts this coordinate in degrees which can then be used as a starting point for an RFI search.

2.3 Locating the RFI Source

After gathering the appropriate GPS coordinates, yagi antenna, AM radio receiver, ducky antenna and an ultrasound detector, a team travelles to the GPS coordinate. It is helpful to have a radio on in the car while travelling to the potential noise source so that other interference sources can be detected en route and a note of them made. Two potential RFI sources were detected using this method. After reaching the GPS coordinate found

in the chimap of the RFI source, it is useful to vary the antennas used. To detect the exact source of RFI in the field, the direction of the source is narrowed by first using the yagi, then switching to the ducky antenna for close range, and then the ultrasound detector. It is useful to switch the equipment as each has a respectively shorter radius that easily localizes the problematic equipment. The radio is always tuned to 150 MHz. It's important to note that the chimaps are reasonably accurate but not perfect, so what's picked up on the yagi may be 20 m, or greater, in practice, away from the GPS coordinate. The chimaps produced in June 2011 had an error bar of approximately 100 m as determined by Ue-Li Pen through investigation of multiple sources. In July 2011, mapping algorithms changed resulting in better accuracy as the lag sign of *rfiposlag.x* changed.

Examples of interference sources were transformers, insulators, and both high and low tension power lines.

2.4 Removing the RFI Source

If the problem lies outside of the central square, a regional wire-man is contacted and shown the faulty equipment. Which wireman to talk to varies depending on the region the RFI source is located.

On occasion, the wireman is able to fix the problem himself. However, if the problem deals with power lines, a letter needs to be sent to the MSEBs closest subdivision in a village called Otur. They will have to turn off power to that area so the interference issue can be fixed. This part of the procedure is rather slow.

Another option is working with the GMRT RFI team. However, due to bad coordination, this option was not explored. Future excursions with personnel from GMRT would be beneficial to RFI removal process.

2.5 Calibration of the Chimap Method

The chimap method utilizes singular value decomposition to make a consistent guess at the source of RFI within data. This method has some error associated with it, whether it comes from multi-path effects or rounding errors in software. The errors stemming from software are negligible when compared with the systematic error resulting from multi-path effects.

To calibrate the method for this unknown uncertainty, a controlled RFI emission is made via a device that sends out a relatively high powered signal, peaking at 150 MHz. The device oscillates from being on and off at a frequency of 1 Hz. The emission is made from the chimap locations of the candidate RFI source and a new chimap is made with this emission data. The location that the chimap method produces for the emission data has two pieces of information.

First, it ensures that the antennas and software are functioning properly. Clearly, if the emission source is not evident then something is not working properly. The emission is three orders of magnitude brighter than the average RFI source in the field, and thus blocks out all weak signals on the chimap. If there is no visible emission RFI on the chimap, then there could be a problem in the software; antenna setup or RFI emitter apparatus.

The second piece of information that the second chimap gives is the relative difference between the known GPS coordinate of the emission and the chimap coordinate of the emission. This difference could be due to multi-path effects. If the transmission was done close enough to the real GPS location of the candidate RFI site, then the same multi-path effects that affected the emission should affect the RFI source as well. Then, by subtracting this difference from the candidate RFI sources chimap, the coordinate should approximate its true GPS location.

One possible flaw in this idea is that if the chimap coordinate of the candidate RFI source is far from the actual GPS coordinate, then the multi-path effects could be much different for the RFI source and the emission. Thus, one could not use the above method to correct for multi-path effects.

2.5.1 Qualitative Discussion of Chimaps

By looking at the chimaps of all positively identified RFI sources, one can see that even the chimaps that best predicted the actual location of RFI source were not without error. Thus, there is an element of uncertainty that can be discussed. Ideally one would like to mathematically relate the features of the chimap to the accuracy of the prediction. This would be an impractical task since the chimaps depend on so many variables. Rather, an intuitive approach is preferred.

In examining Figure 1, one will see that the brightest pixel is amid an ill-defined arc and indeed the localization seems to span an area. This particular chimap gave the best starting point for finding the actual RFI source. This chimap was used to find the RFI source in section 3.1. The more localized the chimap arcs are, and the smaller the area of bright pixels, the more accurate the prediction.

When looking at all chimaps from a particular scan, one clearly sees some do not localize to a single bright pixel. Also, some modes of a scan localize to different spots on the chimap. These should correspond to different RFI sources. One can separate which well-localized modes belong to the same RFI source. The spread of these positions which correspond to the same candidate RFI source is then calculated. In the case of Figure 1 the spread of chimap positions of the various modes that localized to the same candidate RFI source was 1800 m. The real RFI source was actually found and the closest chimap mode was Figure 1. It clearly does not exhibit sharp features. Since the spread is so

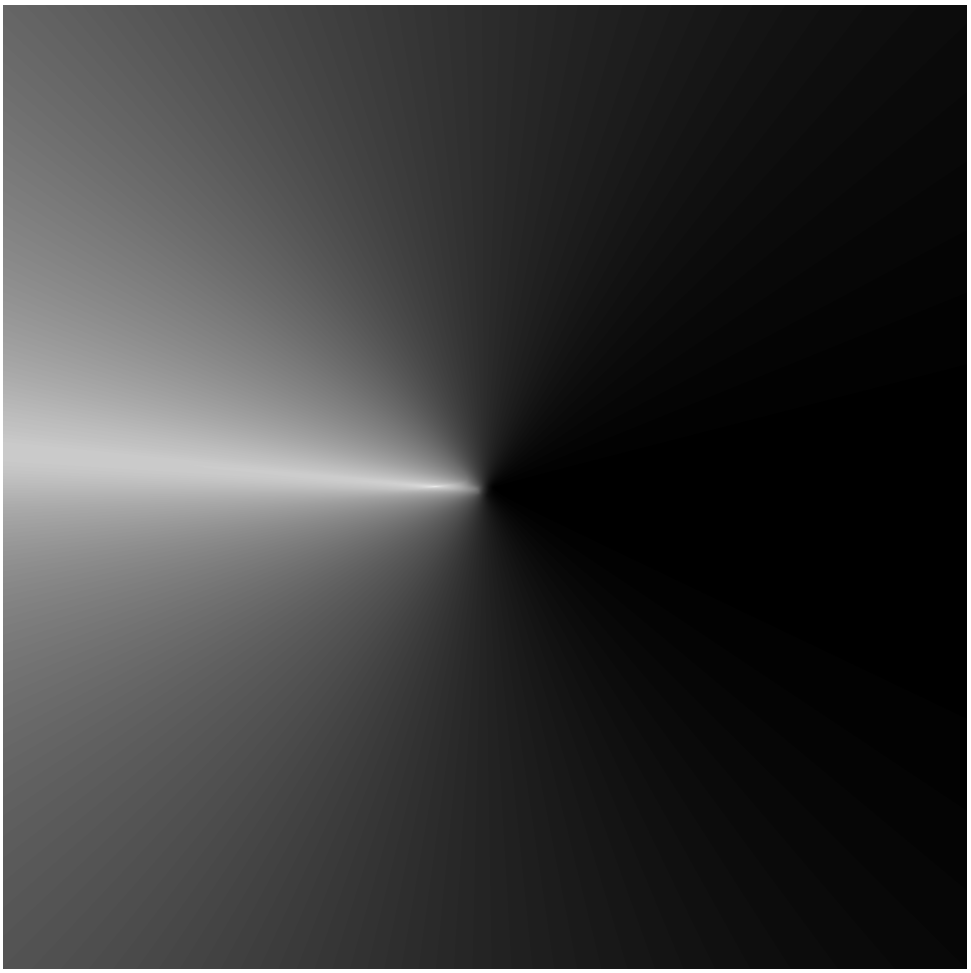

Figure 1: This is the chimap from data taken on June 24, scan 2, mode 17 chimap. It represents a 20 km by 20 km area that is centred on the GMRT central square. Note the ill-defined arcs around the brightest pixel. The brightest pixel is amid a number of similarly bright pixels and seems to span an area.

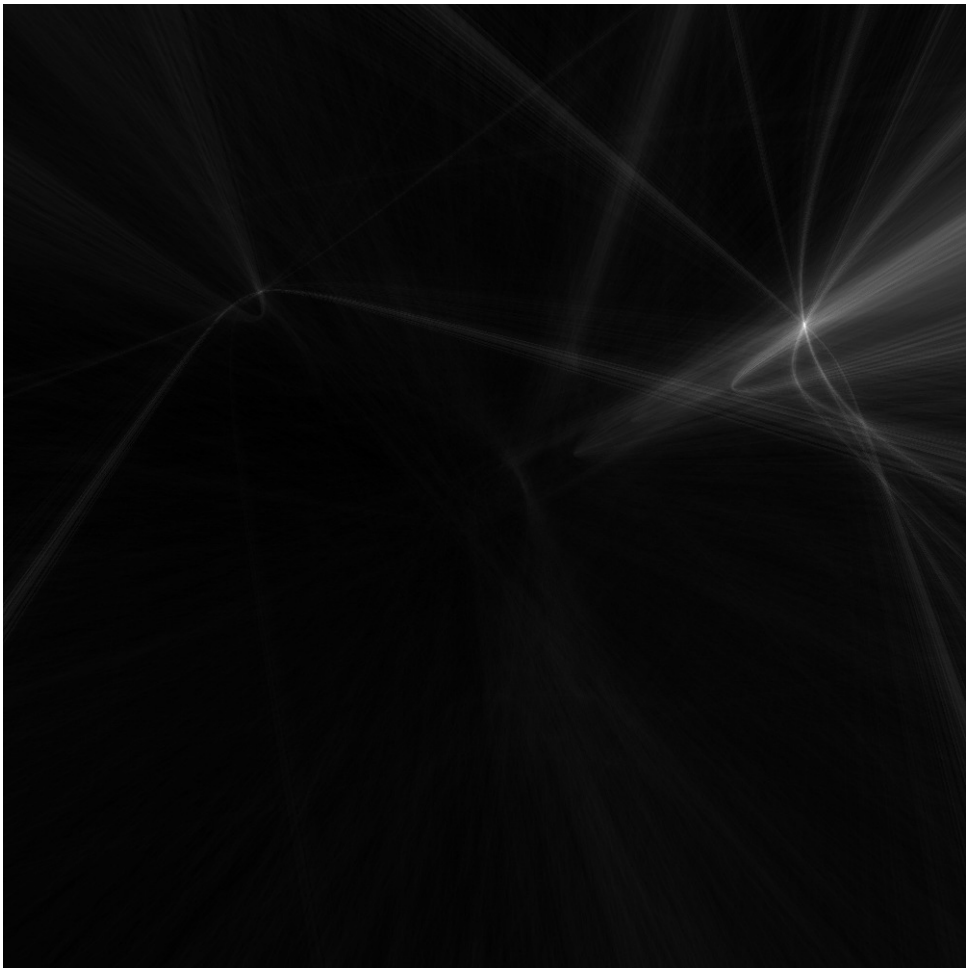

Figure 2: This figure is the mode 3 chimap of data taken on July 12, scan 2. There is a clear localization and well defined arcs.

large, had one started looking for the RFI source from any other chimap mode then one could have easily missed the actual RFI source. Such a large spread may be indicative of multi-path factors. Perhaps the fuzzy features of the arc are also indicative of multi-path and thereby a large spread.

Figure 2 corresponds to the discussion in section 3.3. The spread of chimap positions of the various modes that localized to the same candidate RFI source was 25 m.

The features of Figure 2 support the theory that a large spread implies a fuzzier arc, which may be due to multi-path factors. Here one sees the complimentary evidence. It shows well defined arcs and a clearly localized point. This is clearly in contrast with the preceding figure.

If fuzzy arcs do imply larger spreads, and possibly mutli-path factors, and tight arcs small ones then it makes logical sense to first pursue candidate RFI sites that have tight arcs as they have a better chance of being found quickly and not overlooked.

3 RFI Sources Identified

Most potential RFI sources were not looked into due to the overwhelming number of locations found. For a few of the candidate positions an attempt was made to locate the RFI source.

In all cases, to locate the RFI source the team started at the position identified by the chimap method and then tried to locate the RFI source with yagi antennas, which were designed for 146 MHz, receiving on a radio at 150 MHz. A source was not considered identified until a cause for radio emission was identified, such as a faulty ground connection or a broken insulator.

3.1 A Pomegranate Farm

During an RFI scan on June 24, done between 13:00 and 15:00, a candidate RFI source was located at $(+19.09234°, +74.01513°)$. It was identified via the chimap method. The starting search location was taken from the chimap of mode 17. Alternatively, there were modes $(2, 4, 12, 15, 19, 20)$ which also had reasonably well localized chimaps, giving a total of seven possible positions returned from the chimap method. The standard deviation of the positions across these seven modes is $(\pm 750\text{ m}, \pm 1600\text{ m})$, possibly suggesting that there is some geographic feature that causes the signal to reach the antennas along different paths thereby causing different positions in different modes. The actual source was found at position $(+19.09264°, +74.01472°)$ which is 2156 m from the average position of the cluster of seven predicted locations.

The actual source position, assuming there was only one in the area, was found only 330 m away from mode 17 chimap position. This was the closest chimap position to the actual location out of all the modes, though indeed the spread of chimap points does fill up a rather vast area and the assumption of only one RFI source within that area could be unjustified. It may be that there are several RFI sources in the area, however the various chimap modes all seemed to localize to the same location with similar arc features. Since it is thought that the sharpness of the chimap features should imply something about the accuracy of prediction, the qualitative properties of mode 17 chimap should be examined; see Figure 1 and section 2.5.1 for a discussion on it and other chimaps.

The source was found to be coming from a steel pole on a power line. From the best ground visual observations, it seemed that the only possible source of RFI was from an faulty insulator, though this has not been conclusively identified by an informed authority. The next step that must be taken in regards to this particular RFI source is to contact and convince the local power authority to identify and replace the malfunctioning part.

Figure 3: One of the faulty meters in the DGB.

3.2 Diesel Generator Building

An RFI source was localized to the central square in the diesel generator buildings (DGB). There are two buildings which both contain generators. The third scan taken on July 7 gave the coordinate $(+19.094397°, +74.049324°)$, near the DGB. After thoroughly examining the area, a loud hum was detected on the radio and found emitting from both buildings in the vicinity. The noise localized to three electric meter boxes, two in one building, and one in the other. Figure 3 shows one of the meters in the building with two meters. The coordinate of the Diesel Generator Building is $(+19.094167°, +74.049056°)$. There are many frequencies coming from the meters at which the signal strength peaks. To remove this interference a Faraday cage was constructed around the meter itself, and the cables leading out from the back. A copper mesh was fitted around the meter box and this did, temporarily, reduce the signal strength by approximately a half. However, the next time the meter was inspected, the signal strength had returned to its previous level. The Faraday cage was not very well constructed with various holes visible to the naked eye.

The insulation of the cables at the back of the meter posed an even greater problem. Covering it with a copper mesh made no difference to the signal strength. The ultrasound detector at a very low gain setting of ≈ 2 detected the most interference while pointed at the cable. This leads the team to believe that the improperly insulated cables are more

of a problem than the meter box itself. It is important to note that all of these attempts at removing the interference have only been attempted on the one electric meter. The two in the other building have yet to be tested.

The electrical department at GMRT is aware of the problems stemming from the DGB and have also suggested a few other ideas to better insulate the cables such as using zero-copper insulation. As of now, this has not been implemented. Nissim Kanekar, a scientist at NCRA, is also well aware of the problems at both the DGB and the power station hub.

3.3 A Chicken Farm

Mode 2 from a scan on July 12 at 17:15 suggested that there was an RFI source located at $(+19.137867°, +74.145737°)$. Upon visiting the site it was observed that there were actually three RFI sources in the area within 50 metres of each other. Other good chimap modes that localized near the actual RFI sources were modes (3,4,5,7,10,16,17, and 18).

The spread of these points is $(\pm 25\,\mathrm{m}, \pm 0\,\mathrm{m})$ and the distance of the furthest actual RFI source in the area to the average position of the modes is 110 m. This, when compared with the spread and distance from the average of the modes' position mentioned in Section 3.1, is quite small and does not immediately imply multipath effects in the geography. Furthermore, on July 17 during EoR observations of pulsar B2045-16 between 01:30 and 08:00, it was observed, after analysis, that mode 3 produced a chimap which localized to $(+19.138138°, +74.145164°)$, only 80 m away from the average position from those on 12 July. The fact that the same RFI sources were found in data during different times of the day on different dates indicates that these sources may be some that would show up in multiple observers' data. This makes the removal of these sources a high priority. The three actual RFI sources have positions $(+19.137729°, +74.146060°)$, $(+19.137410°, +74.145890°)$, and $(+19.137120°, +74.146510°)$. The wireman inspected these sites. The main problems, agreed by both the AST398 team and the wireman, in all three cases, were bent insulators and improperly grounded poles attached to 33 kV power lines. Unfortunately, the wireman was unable to fix the bent insulators as the power line had too high a voltage. The MSEB has to shut down the power to this line in order for him to straighten out the insulator. Attempts were made to contact the MSEB and the appropriate people at GMRT, however no progress was made. The wireman re-grounded the wires which were spiralling off the power poles. This did not make a noticeable difference in terms of decreasing the interference implying that the main source is the insulators. The ultrasound detector could detect the arc of the interference at low gain.

It is suggested that a letter be sent to the MSEB which asks them to shut down the relevant power line. The wireman of the Rajuri district knows what needs to be shut

down and will be able to fix the faulty insulators once that has been done.

3.4 Power Station Hub

A possible RFI noise source was first detected at the power station hub at the GMRT entrance by walking around the central square and following an erratic noise source with the yagi antennas. Upon further investigation, a beeping sound was heard with the radio at approximately 1 Hz when at the doorway of the building. It was detected through the use of the yagi antenna The signal strength increased upon entering the building. The periodic noise was not audible far from the building.

Using the ducky antenna the actual RFI source position was found at GPS coordinate $(+19.097139°, +74.048139°)$. The problem was the middle two electric meter boxes maintained by the MSEB - see Figure 4 The reason as to why the boxes are radiating has not yet been determined. It is suspected that if MSEB changes the boxes, they will no longer emit any interference as the rightmost meter box was the only one not giving off any interference. Currently, it is the only one that had been recently replaced. The major cable at the back of the power station hub behind the electric meter boxes which links to each node has also been found to give off interference. The proper authorities have been notified and mentioned that it will take time to fix as the back wall of the hub needs to be removed to gain sufficient access to the cable to insulate it. There have also been problems detected with the grounding in the building. The grounding cables need to be rechecked and re-grounded and then the major cable must be checked for interference again. If it is still a problem, then it needs to be insulated as soon as possible.

3.5 Near E03

The July 12, mode 2, RFI scans localized to $(+19.127288°, +74.077141°)$. Upon visiting this location a broadcast was made from a point as near as possible to the coordinate location, $(+19.125938°, +74.077654°)$, approximately 20 m away. Unfortunately the GPS location was in the middle of a grass field and therefore a broadcast from the exact locations was impossible. After the broadcast, a bank of transformers emitting radio noise at $(+19.126883°, +74.077457°)$ was discovered by circling the original location with a yagi. When broadcasting from this location the generated source localized on the July 13, scan 18, chimap to $(+19.125988°, +74.077713°)$. This is one chimap pixel length east of the original candidate source location. The bank of transformers can therefore be considered positively identified as the noise source to within experiment uncertainty, 20 m.

No attempt has been made to repair the transformer apparatus, nor contact to anyone to have it repaired.

Figure 4: One of the faulty meters in the powerstation hub.

3.6 Girish Bhor, near Ranjani village

From scan 6 on July 7, mode 3 a very probable source was revealed at (+19.054707°, +74.051704°). A team travelled to this location and broadcast using the noise generator. Using the yagi it was discovered that a low voltage power line was a strong RFI source at (+19.055364°, +74.051738°). From beneath this pole another broadcast was made. This localized on the chimapfrom scan 1 on July 12 to (+19.054707°, +74.051994°). Given the pixel size of 20 m on the chimap this confirms this as the original source to within experimental uncertainty.

No attempt has been made to repair the transformer apparatus or contact anyone to have it repaired.

3.7 Telephone Poles

During an RFI scan on July 7 a source was observed to be located at (+19.03368°, +74.08000°). The location was taken from the first mode of the third scan of the day. Upon the first field excursion, the source of the RFI was localized to two telephone phones situated at (+19.03370°, +74.07993°) and (+19.03343°, +74.08033°). The next step is to contact the local power authorities so they may repair the problematic poles.

4 EoR Observations (Cycle 20)

4.1 Observational Procedure

During Cycle 20 of GMRT's observational time, the team performed drift scans of pulsar B2045-16 with the following default settings for the array:

L01 = 218 MHz
L04 = 62 MHz
L05 = 16 MHz(138 MHz → 167 MHz)
Baseband bandwidth = 16 MHz
IF bandwidth = 16 MHz
Noise source ON, (low cal, 25% duty cycle)
Power equalization (3C flux calibrator close to source, 300 counts)
ALCOFF

On occasion, different settings were used when problems occurred. Difficulties were, for the majority, concerned with time constraints. In those cases, power equalization was not performed and ALC was left *ON* as this step often took half an hour. The pulsar is observed with multiple scans with each scan lasting 14 minutes followed by a 2 minute interval before the next scan to allow for antenna rotation. Scans are taken at LST ranges that range from LST 1900 to LST 2400 with the convention that a given LST range will

have the same scan number each night, but a different pointing. The pointings are either on target and pointed directly at the pulsar, or 1.6 degrees north or south. These three pointings cycle through after each scan and are different for the subsequent day. This information is given to the operators in the control room through the use of a command file which directs the antennas at each specified time by the user to match the LST range convention.

The start and stop times of each scan are extracted from the command file for the current observation and are converted into seconds. Afterwards, they need to be input into the EoR correlator script. For each observation, a few parameters must be changed within the correlator script such as the date, start and stop times as well as where the data is to be stored on the nodes. The convention for data storage between different observations is that odd numbered scans are stored in one drive while the even numbered scans are saved into another such as */mnt/EoRa* and */mnt/EoRb* respectively so long as there is enough disk space.

In order to run the correlator properly, the period of the pulsar must be obtained for the current observation as it is changes slightly. The information is obtained by running a script which prompts the user for the pulsar name, the date of the observation and the time of the observation. The time used is the time halfway through the observation as the EoR correlator does not adjust the pulsar period throughout the observation. This is chosen as a rough average as the pulsar period changes slightly over the course of the observation. The script outputs the period in milliseconds, the right ascension in hours and the declination in degrees. The coordinates are used later after the observation has been completed and is ready for analysis. The period is converted into seconds and input into *rtcorr.c* at the beginning of the script.

Since the EoR correlator requires different inputs from the standard GMRT correlator, a switch in the correlator room must be turned to the side that reads *Node 1* as opposed to *1ppm* before observations are run. Once the switch has been flipped, two more scripts must be run prior to the observation to equalize the gains of the antennas. These steps were originally in the correlator script but were malfunctioning for Cycle 20 and have yet to be fixed.

The correlator script can now be run and data taken. Data is recorded at a rate of roughly 110 Mb/s on average. A higher value of 120Mb/s is typically expected however, the write speed depends on both the disk write speed as well as concurrent CPU usage. If a lower value that is less than 90 Mb/s is encountered, it is likely that there is another process that is running and interfering with the observation and that process should be terminated.

4.2 Other Scans Taken

While at the GMRT, the AST398 group also took pulsar observations in tracking mode and spider scans for polarization beam mapping. The Pulsar Tracking scans focused particularly on B2217+47 as this was the pulsar observed in Cycle 18. This was done so the GMRT-EoR group could measure the changes in the ionosphere (by comparing the polarization angles with the noise source), compared to one year ago.

Polarization Beam Mapping (Spider) scans were also taken while observing pulsar B0531+21. In this set up, half of the antennas were set in drift mode while the other half were set to track the pulsar. This was done to keep track of the off-center beam polarization of the antennas.

4.3 Fringestop and Analysis

After data has been taken by the array, it must be converted into a format that is more easily usable. This is done by a process known as *fringestopping* the data. This process essentially orients all the baselines from the antennas to the same position in the sky. As the GMRT is an interferometer, this is done by accounting for the path difference between antenna pairs. The scripts involved in the process require the right ascension and declination obtained previously during the observation. A script written by the AST398 group converts the coordinates into radians before they are input into the pipeline. A script then runs and will fringestop and subsequently produce analysis files for later use. All the odd numbered scans and even numbered scans are done separately. After the data files have been converted, more scripts are run on the files that provide other important information. These steps collectively comprise the initial analysis portion.

Among the files that are produced, one that can quantify that the analysis has been performed properly and that observations were successful is an image of the sky as seen in Figure 5. The source should appear as a bright source at the center of the image. However, the analysis is not always performed correctly. The software tries to detect which gate out of 16 that the pulsar appears in for a given scan but may be unable to single out the proper gate. For this cycle, manual determination of the gate by eye was done for the majority of scans. To determine the proper gate that will produce the most accurate image of the sky, additional images that visualize the pulsar over time are first produced by a secondary program. It takes in as parameters, a data file, the length of time to plot, the number of antennas, the number of frequencies and the number of gates. For the latter, the three values used for this cycle were 64, 128 and 16. The script will generate images for both left and right feeds of each antenna of each baseline. One type of image it produces are *lag* images which depicts the lag of the data from left to right where the center of which corresponds to zero lag, while vertically it represents time, with settings chosen so that it is formatted into 16 gates each with 20 pixels which

correspond to a timestamp. The pulsar should appear at the zero lag line in a given gate. It will appear as a vertical contrasting line within a gate. It is often easier to detect the pulsar in images that correspond to the arm antennas as there is generally less RFI. The script also produces *lags* images which are similar to the *lag* images in what they depict save that the information is combined into 1 pixel. In this way, it is sometimes easier to detect the pulsar as it increases the contrast of the pulsar compared to the sky and noise background

Once the proper gate has been determined, the data can be recalibrated and a new image of the sky produced. Within an observation, the pulsar should be found in the same gate for all the scans. Therefore, even if the pulsar is hard to find in one scan but is easily identifiable in another scan, all scans for the observation can be redone with the same gate. The pulsar sometimes drifts between two gates in one observation session. This can be observed when looking at *lag* images and the pulsar seems to be stretched across two gates. It is worth mentioning that the pulsar, B2045-16 observed in EoR mode during July 2011 has been difficult to find. While it is not impossible, it is significantly more arduous than previous cycles. One possible reason for this difficulty is that after the switch of the number of antennas from 60 to 64, some programs may not have been updated. Furthermore, the pulsar was often not in transit when observed and was closer to setting when data was taken. This may contribute to the difficulty in detecting it.

5 Noise Masking of Data

Upon performing a Fourier transformation to produce an image of the sky, visibility maps are also produced as seen in Figure 6. They provide an image of the baselines over time. However, there is often noise that can be attributed to bad antennas, frequencies or timestamps. These appear as areas with significantly different values from the bulk of the map. They can be masked by determining what data is input into each pixel by the imaging program. This process is made easier if the imaging program prints the antennas, frequencies and timestamps of problematic pixels as discerned manually by eye. With this information, one can mask the specific antennas, frequencies or times to improve the dynamic range of the image, where dynamic range in this case, is the ratio between the smallest and largest root mean squared values of the brightness of the pixels. Achieving a value over 5000 corresponds to a well calibrated image. This procedure was performed on data from December 2007 of Cycle 18. Antenna 26 was observed from this data to be malfunctioning for multiple days and Dec 18 required no noise masking. See Table 1.

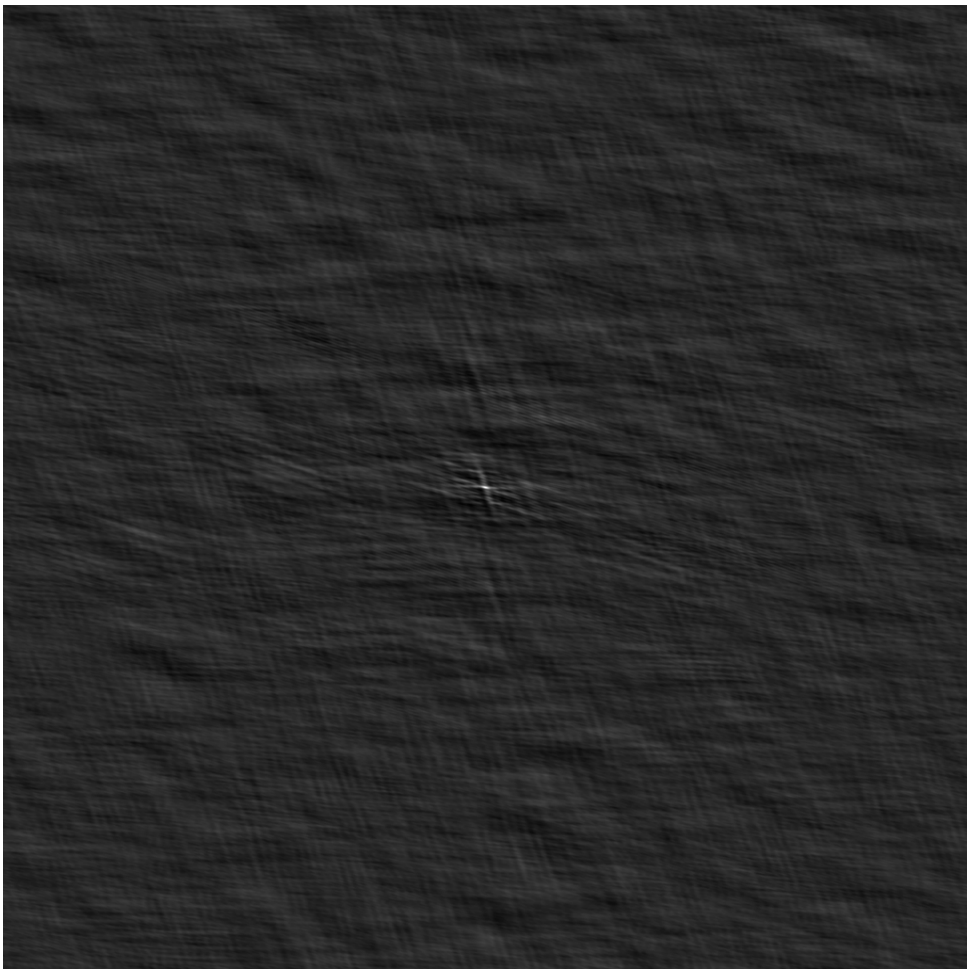

Figure 5: B2217+47, Aug 02, 2011 scan 22, rmap.pgm. This image depicts the pulsar at the center as a white dot. The data was calibrated using gate 7. This source was used in preference to B2045-16 as during the observation time, the latter would have already set. The software was able to detect the proper gate and no recalibration was done.

Date	Antennas(Full)	Antennas(Partial)	Timestamps	Frequencies	Dynamic Range
Dec 10 2007	26	-	310-346	-	6881.41
Dec 11 2007	26, 3, 5 11 ,18	-	430-448	2-6, 23-28	6272.555
Dec 14 2007	18, 24	17, 20 at 250-270	430-448	-	6635.093
Dec 16 2007	26, 24, 27	-	400-425, 440-448		5903.241
Dec 17 2007	8, 25, 27	18 at 170-174	135-156, 270-290, 300-366, 445-449		5519.115
Dec 18 2007	-	-	-	-	6308.905

Table 1: This table displays the data that was masked from the visibility maps of Dec 17 2007. Where Antennas(Full) represents the antennas that were masked over all timestamps, Antennas(Partial) represents antennas that were masked over a specific timestamps.

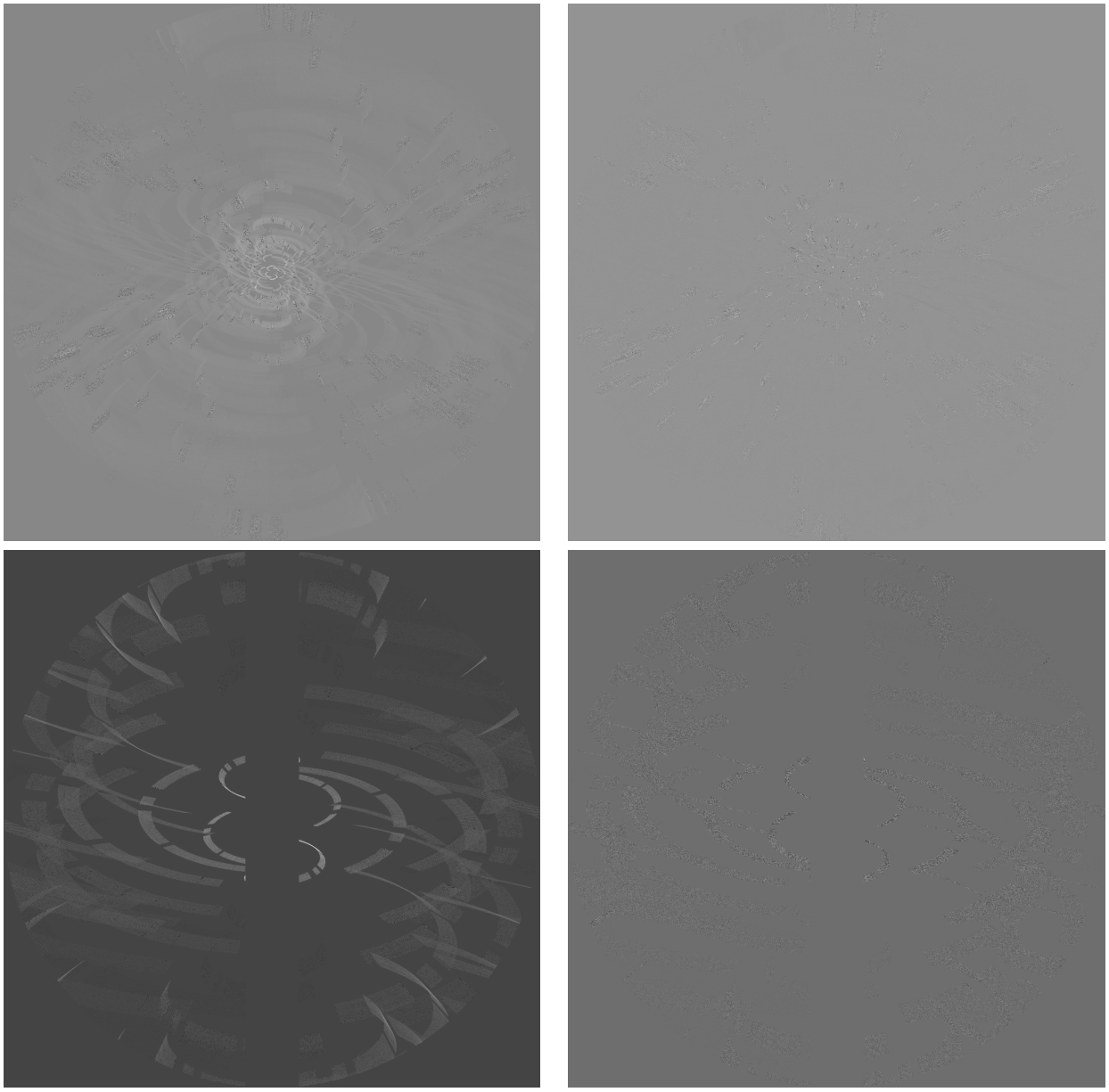

Figure 6: The images on the right are the real portions while the left corresponds to the imaginary visibility maps made from data from Dec 17th 2007. The top two correspond to the visibility maps prior to applying the maskings, while the bottom two have been masked. As referenced in Table 1, the maskings determined for this data set were very specific as they were narrowed down to specific timestamps intervals relative to specific antennas.

6 Calibration Noise Malfunctions

The data collected from looking at the amplitude plots generated by the noises sources built in at each antenna can be used to detect consistently malfunctioning antennas. The plots display the amplitude of the noise source with respect to time. In general, the plot should look slightly sinuisoidal with a very small amplitude and so any Looking at data from 2011, antennas C1, C11 and E2 have been consistently providing poor data since February. Antenna E2 has also malfunctioned in December 2007. Antennas C0, C2, C3, C8, C12 and W3 have also been malfunctioning, though less consistently, since February 2011.

7 Potential RFI Sites

In this section we propose a method method for determining what is considered a probable source of RFI. One of the aims of this project was to create a near-automated procedure that can be carried out by someone in the future to further eliminate RFI. There are several obstacles involved with such an operation. First, elimination of RFI requires people to address the problem. This is probably the hardest obstacle to overcome as there are no guarantees with regard to human action. Second, it would be more time efficient if there was a mechanism in place to output the most probable RFI candidates to search for. Although this would require extensive software intelligence, it should still be considered. While the team was able to identify several RFI sources, see Section 3, many potential locations were not investigated due to time constraints. They should be categorized by the likelihood of an RFI source being in the vicinity.

7.1 Method to Determine Probable RFI

It has been proposed that perhaps one of the best methods of discovering RFI sources was to constantly be observing on 150 MHz while driving around the country side.[2] This is obviously inefficient especially when there is a method for pointing one in the right direction to start with - the chimap method. Perhaps this method could be made more efficient by knowing where to drive around in the countryside. The procedures discussed in Section 2 are very effective once the RFI is within a detectable range.

Given a highly populated map of predicted RFI points, via the chimap method, map these to an $U \in \mathbf{R^2}$. For each $x \in U$ compute the standard deviation , $\sigma(x)$, of points within a radius r about x. Clearly, $\sigma(x) \leq r$ will only be equal when all the points lie on the boundary of the neighbourhood. The radius r is called the proximity radius. Now,

[2]During a talk on July 15 hosted by NCRA and given by Ue-Li Pen, Joshua Albert, and Nidhi Banavar on RFI elimination efforts at the GMRT.

for each x calculate the Gaussian probability distribution centred on x with width $\sigma(x)$ and map this to $V \in \mathbf{R^2}$. Call V the RFI probability density map.

A graph of V would visually show where the most likely location of finding RFI would be via the method mentioned above. Driving a vehicle through denser areas with a radio and antenna receiving on 150 MHz would have a higher probability of discovering an RFI site. Even without driving simply noticing that there is a tight bright point on the density map would suggest that that place should be looked into first. This is because the density map takes into account chimaps from scans done at different times and thus will over time show more strongly, as static RFI sources appear brighter.

One thing that this method requires is an initial estimate of the proximity radius. The proximity radius should give a lower bound on the distance between different RFI sites and an upper bound on the acceptable spread of the points surrounding x. There are clear cases with exception where the distance between different RFI sites is small. An example is in Section 3.3. In these cases so long as an RFI source is not overlooked it is fine to have a proximity radius greater than the minimum distance between RFI sites. The upper bound should take into account the discussion in section 2.5.1. If a large spread corresponds to a fuzzy chimap, and thus a large search area for the RFI, then the upper bound on acceptability of the spread of points should be low enough to avoid unclear chimaps. This will be an experimentally derived value.

$\sigma(x)$ as the width of the Gaussian will then be a model of a standard probability distribution not taking into account the topology of the region, or multipath effect and diffraction.

8 Suggestions

8.1 RFI Suggestions

One thing that can improve the ease of access for the GMRT-EoR group and other observers is to change from having a hardware switch to a software switch as suggested by N.K. This would allow RFI tests as well as other observations to be done remotely and more reliably. There would be less chance for accidental error and would make an automated periodic RFI search possible. Optimally, the team should update software and use the same correlator as GMRT if it does not negatively affect EoR observations.

Another limitation on detecting RFI is the amount of time available to physically detect RFI sources in the field. A suggestion would be for GMRT employees to carry radios during patrols or excursions around GMRT. This ensures that RFI detection is active on a daily basis. Finally, after an RFI test, the data files need to be formatted by separate scripts before SVD can be performed on the data. This process could be streamlined by having the the formatting processes included in the SVD script.

8.2 EoR Suggestions

Seeing the pulsar in recent EoR data has been a consistent problem for the AST398 group. An easy way to locate the pulsar is to use a process called pulsar folding. This can only be used, however, if the data has been run through the Pulsar Backend or observed in standard, non EoR mode. As our current method of observing does not allow for this, we have been unable to fold the pulsar. There are two possible solutions to this. Firstly, find a way to take EoR observations which can be run through the backend and then folded. Unfortunately, this would require a lot of time and effort. The second possibility is to pull the pulsar profile from the EoR data which would then allow the pulsar to be easily found.

9 Conclusion

The team has developed a solid procedure for detecting RFI in the nearby area surrounding the GMRT by observing at 150 MHz and producing chimaps with the GSB. The team has also developed a strategy for finding the RFI source in the field by searching with yagi antennas, ducky antennas and ultrasound. There are some concerns that multipath effects are causing large error in the creation of the chimaps. Methods to eliminate this error or reduce it should be developed and put into action. Possible future work in this area includes finding people who can regularly go into the field to locate RFI. The next steps should be to remove the identified sources of RFI and continue looking for new ones using the methods discussed in this report. In particular a method should be developed that can be automated, or remotely run.

References

[1] I. Iliev, G. Mellema, U. Pen, and P. Shapiro. Character and detectability of the dark ages and the epoch of reionization: the view from the simulations. *arXiv:0712.1356v1*, Dec 2007.

[2] Juhan Kim and Ue-Li Pen. Redshifted 21-cm in the dark ages. *arXiv:0908.1973v1*, Aug 2009.

[3] G. Paciga, T. Chang, Y. Gupta, R. Nityanada, J. Ogegova, U. Pan, J. Peterson, J. Roy, and K. Sigurdson. The gmrt epoch of reionization experiment: A new upper limit on neutral hydrogen power pectrum and z 8.6. *arXiv:1006.1351v2*, Dec 2010.

[4] Ue-Li Pen, T. Chang, J. Peterson, J. Roy, Y. Gupta, and K. Bandura. The gmrt search for reionization. *arXiv:0804.2501v1*, Apr 2008.

[5] Paul R. Shapiro, Ilian T. Iliev, Garrelt Mellema, Ue-Li Pen, and Hugh Merz. The theory and simulation of the 21-cm. *arXiv:0806.30901v1*, Jun 2008.