Measuring the 21 cm Power Spectrum from the Epoch of Reionization with the Giant Metrewave Radio Telescope

by

Gregory Paciga

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Astronomy & Astrophysics
University of Toronto

Abstract

Measuring the 21 cm Power Spectrum from the Epoch of Reionization with the Giant Metrewave Radio Telescope

Gregory Paciga

Doctor of Philosophy

Graduate Department of Astronomy & Astrophysics

University of Toronto

2013

The Epoch of Reionization (EoR) is the transitional period in the universe's evolution which starts when the first luminous sources begin to ionize the intergalactic medium for the first time since recombination, and ends when the most of the hydrogen is ionized by about a redshift of 6. Observations of the 21 cm emission from hyperfine splitting of the hydrogen atom can carry a wealth of cosmological information from this epoch since the redshifted line can probe the entire volume. The GMRT-EoR experiment is an ongoing effort to make a statistical detection of the power spectrum of 21 cm neutral hydrogen emission due to the patchwork of neutral and ionized regions present during the transition. In this work we detail approximately five years of observations at the GMRT, comprising over 900 hours, and an in-depth analysis of about 50 hours which have lead to the first upper limits on the 21 cm power spectrum in the range $8.1 < z < 9.2$. This includes a concentrated radio frequency interference (RFI) mitigation campaign around the GMRT area, a novel method for removing broadband RFI with a singular value decomposition, and calibration with a pulsar as both a phase and polarization calibrator. Preliminary results from 2011 showed a 2σ upper limit to the power spectrum of $(70\,\mathrm{mK})^2$. However, we find that foreground removal strategies tend to reduce the cosmological signal significantly, and modeling this signal loss is crucial for interpretation of power spectrum measurements. Using a simulated signal to estimate the transfer function of

the real 21 cm signal through the foreground removal procedure, we are able to find the optimal level of foreground removal and correct for the signal loss. Using this correction, we report a 2σ upper limit of $(248\,\mathrm{mK})^2$ at $k = 0.5\,h\,\mathrm{Mpc}^{-1}$.

Acknowledgements

This research could not have been completed without the work of my supervisor, Ue-Li Pen, and the many collaborators who have helped build the GMRT-EoR project, from theory, making observations, analysing the data, and preparing it for publication. These have included Tzu-Ching Chang, J. Richard Shaw, Julia Odegova, Kris Sigurdson, Chris Hirata, Jeffrey B. Peterson, Yashwant Gupta, Jayanta Roy, and Rajaram Nityananda. Special thanks as well to Kiyoshi Masui and Eric Switzer, and my office mates Sergei Ossokine and Nick Tacik, with whom I had many useful discussions.

Woodsworth College and the University of Toronto's Curriculum Renewal Initiative Fund made the summer undergraduate research course in 2011 possible. The work of Hans Nguyen, Joshua G. Albert, Nidhi Banavar, Connie Lien, and Mark Kuiack as research assistants hired as part of this course was useful in compiling the discussions in Chapter 3.

Financial support was provided by the Natural Science and Engineering Research Council of Canada and the Ontario Ministry of Training, Colleges and Universities. The GMRT is run by the National Centre for Radio Astrophysics of the Tata Institute of Fundamental Research. The computations were performed on the Sunnyvale cluster at the Canadian Institute for Theoretical Astrophysics, which is funded by the Canada Foundation for Innovation, the Ontario Innovation Trust, and the Ontario Research Fund. This research has made use of the database of published pulse profiles maintained by the European Pulsar Network[1], the SIMBAD database[2] operated at CDS, Strasbourg, France, and NASA's Astrophysics Data System[3].

Finally, I would like to thank my supervisory committee, Chris Matzner and Mike Reid, as well as Bob Abraham. Without their support this thesis would not have been completed.

[1] `http://www.jb.man.ac.uk/research/pulsar/Resources/epn/`
[2] `http://simbad.u-strasbg.fr/simbad/`
[3] `http://adswww.harvard.edu/`

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The earliest and latest stages of the universe, the time of recombination from which the Cosmic Microwave Background (CMB) originates and the era of stars and galaxies in which we live today, have been extensively studied and are well understood. The transition from an entirely neutral universe to one with luminous sources that ionize the intergalactic medium, however, is an era of the universe's history that we have not yet observed directly and as such is still relatively unconstrained. We will start with a review of what is known about this so-called "Epoch of Reionization" (EoR), the observations we expect to be able to make, and will give a brief introduction to the Giant Metrewave Radio Telescope (GMRT) and its use as a reionization instrument.

1.1 The Epoch of Reionization

In the years immediately after the big bang, the universe was filled with a hot plasma opaque to radiation. Once it had expanded and cooled to about 3000 K at redshift $z \approx 1090$ electrons could combine with nuclei to create long-lived neutral atoms for the first time, and the background radiation could propagate freely, creating what is now the cosmic microwave background (CMB). This is an epoch that has been studied thoroughly, most notably by the COBE, WMAP, and now Planck satellites.

The universe continued to expand and cool, and the density perturbations which were already in place could grow. This period is known as the "Dark Ages", since there were not yet any luminous sources. The first stars likely formed beginning around $z \approx 30$ (see, e.g., Barkana & Loeb, 2001) at local peaks in the matter density, and would have begun ionizing the neutral hydrogen around them, marking the beginning of the Epoch of Reionization (EoR). Furlanetto et al. (2006a) provides a thorough review of our current understanding of this epoch.

Over time the ionized bubbles grew and began to overlap, creating a patchwork of ionized and neutral cells. Eventually, by $z \sim 6$, the ionization of the broadly distributed material was complete, leaving only rare pockets of neutral gas, marking the end of the EoR. This patchy topology is supported by theory and simulations (e.g. Furlanetto et al. 2004; Iliev et al. 2006b; McQuinn et al. 2007; Zahn et al. 2007; Friedrich et al. 2011; Su et al. 2011; Griffen et al. 2013; see Trac & Gnedin 2011 for a review). The exact properties, however, are poorly constrained.

In addition to directly determining the redshift of recombination, CMB data has been used to estimate the redshift at which reionization took place. This is done by measuring the optical depth to Thomson scattering, which gives an estimate of the column density of free electrons to the surface of last scatter. Knowing how the density of the universe scales with time and assuming an instantaneous reionization event, Komatsu et al. (2011) estimated that it would have occurred at $z = 10.4$ using WMAP 7-year data. There is also some evidence from the CMB that indicates that reionization was an extended process (Dunkley et al., 2009). More recently, the Planck collaboration has fit the data allowing for an extended, but still relatively quick, reionization history with $\Delta z = 0.5$ to estimate the redshift at which the universe would have been 50% ionized to be $z \approx 11.4$ (Planck Collaboration et al., 2013).

Observations of individual luminous sources at high redshift can also probe the EoR, giving us substantial information on the neutral content of the intergalactic medium

(IGM) at more recent epochs. In particular, there are observations of quasar spectra at $z \approx 6$ (e.g., Becker et al., 2001; Willott et al., 2007) and high redshift gamma-ray bursts (e.g., Greiner et al., 2009) which can constrain the ionization fraction. Neutral hydrogen along the line of sight readily absorbs Lyman-α (Ly-α) photons on the red side of the rest wavelength, allowing us to infer the neutral fraction $x_{\rm HI}$ as a function of redshift (e.g., Fan et al., 2002). This can produce a Ly-α forest of absorption lines as $x_{\rm HI}$ changes with z, but the absorption completely saturates by $x_{\rm HI} \gtrsim 10^{-4}$ (Fan et al., 2006a), resulting in a Gunn-Peterson trough (Gunn & Peterson, 1965). Inferring an accurate *global* ionization fraction from such a spectrum, however, is difficult, being highly dependent on the properties of the model used.

It is generally accepted that reionization was complete by a redshift of $z \approx 6$ (Djorgovski et al., 2001; Becker et al., 2001; Fan et al., 2006b; Willott et al., 2007; Greiner et al., 2009) though there is some evidence that the actual HI fraction may still have been quite high (McGreer et al., 2011; Schroeder et al., 2013). The time evolution at higher redshift is still poorly constrained. Current observations are consistent with, for example, both a rapid evolution at $z \approx 6$ (Fan et al., 2006b) and a smooth gradual reionization that completes by about the same time (Becker et al., 2007). However, neither CMB nor quasar observations allow detailed examination of the reionization era itself.

1.2 The 21 cm line from neutral hydrogen

One of the most promising ways of probing the EoR is by observing the 21 cm emission line from neutral hydrogen. This spectral line arises from the slight energy difference ("hyperfine splitting") between the parallel and anti-parallel alignments of the ground state electron and proton spin axes in the atom. The case where the spin of the electron and proton are aligned in the same direction has a slightly higher energy level than the anti-aligned case by 5.9×10^{-6} eV. This energy difference corresponds to a rest wavelength

photon of about 21.106 cm, or a frequency of 1420.4057 MHz.

The transition between these two states is forbidden, with an Einstein coefficient of $2.85 \times 10^{-15}\,\mathrm{s}^{-1}$ (Furlanetto et al., 2006a) meaning that a hydrogen atom in the excited state would make this transition to emit a 21 cm photo only once every 12 billion years. Yet, the large number of hydrogen atoms in astronomical contexts, and the fraction of those that are in the excited state, combine to make an observable line. The relative population of atoms in the ground and excited states is quantified with the excitation or "spin" temperature T_s, defined by

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} \exp\left(\frac{E_{21\mathrm{cm}}}{k_\mathrm{b} T_\mathrm{s}}\right) \tag{1.1}$$

where n_1 and n_0 are the number densities of hydrogen in the excited and ground states, respectively, g_i is the statistical weight of each state, and $E_{21\mathrm{cm}} = 5.9 \times 10^{-6}\,\mathrm{eV}$ is the energy difference between the two states (Furlanetto et al., 2006a). Since the higher energy level is a triplet state ($g_1 = 3$), as long as T_s is much greater than the equivalent temperature difference between the two states $T_* = h\nu/k_\mathrm{b} \approx 68\,\mathrm{mK}$, then three out of four atoms will be in the excited state.

In the Rayleigh-Jeans tail of interest in radio astronomy, a useful quantity is the brightness temperature, defined as

$$T_\mathrm{b} = \frac{I_\nu c^2}{2 k_\mathrm{b} \nu^2} \tag{1.2}$$

where I_ν is the intensity at frequency ν. For an incident radiation field with temperature T_R on a cloud with uniform excitation temperature T_s, the emergent brightness according to radiative transfer is

$$T_\mathrm{b}(\nu) = T_\mathrm{s}\left(1 - e^{-\tau_\nu}\right) + T_\mathrm{R}(\nu) e^{-\tau_\nu} \tag{1.3}$$

where τ_ν is the optical depth. A precise calculation of τ_ν must include a correction for stimulated emission, the line profile, expansion of the IGM with the Hubble flow, the neutral fraction, and the column density. For 21 cm radiation passing through the diffuse

IGM, the result is given by Furlanetto et al. (2006a) to be

$$\tau_{21\mathrm{cm}} \approx 0.0092(1+\delta)(1+z)^{3/2}\frac{x_{\mathrm{HI}}}{T_{\mathrm{s}}}\left[\frac{H(z)/(1+z)}{\mathrm{d}v_{\parallel}/\mathrm{d}r_{\parallel}}\right] \tag{1.4}$$

where $(1+\delta)$ is the local overdensity, and the quantity in square brackets accounts for both the motion due to the expansion of the universe and peculiar velocities in the gas.

Taking the CMB as the background radiation field, so $T_{\mathrm{R}} = T_{\mathrm{CMB}} \approx 2.7(1+z)\,\mathrm{K}$, the variation of brightness temperature of the neutral hydrogen against the CMB can be shown to be

$$\begin{aligned}\delta T_{\mathrm{b}} &= \frac{T_{\mathrm{s}} - T_{\mathrm{CMB}}(z)}{1+z}\left(1 - e^{-\tau_{21\mathrm{cm}}}\right) \\ &\approx 31\,\mathrm{mK}\left(1 - \frac{T_{\mathrm{CMB}}}{T_{\mathrm{s}}}\right)(1+\delta)\left(\frac{1+z}{10}\right)^{1/2}\left(\frac{\Omega_{\mathrm{b}}}{0.0482}\right)\left(\frac{0.309}{\Omega_{\mathrm{m}}}\right)^{1/2}\left(\frac{0.678}{h}\right)\end{aligned} \tag{1.5}$$

where cosmological parameters have been adjusted from Furlanetto et al. (2006a) to reflect the best fit values from Planck Collaboration et al. (2013). If $T_{\mathrm{s}} > T_{\mathrm{CMB}}$ then the hydrogen shows up in emission, and if not it will be in absorption.

The spin temperature is influenced by several factors which evolve with time. These are the background CMB temperature, the gas kinetic temperature, and the temperature of the background Lyman-α radiation field. The coupling between 21 cm and each of these is poorly constrained by the lack of information on the high redshift universe. Figure 1.1, reproduced from Pritchard & Loeb (2010), shows a fiducial evolution of the 21 cm brightness temperature relative to the CMB temperature, as well as several alternate models.

In the early universe collisional coupling, primarily between H-H and H-e$^-$ pairs, is the main process that sets the spin temperature (Purcell & Field, 1956; Field, 1959). This couples the spin temperature to the gas kinetic temperature, which cools faster than the CMB temperature. When the density becomes too low for collisions to be efficient, the spin temperature decouples from the kinetic temperature (point A in Figure 1.1) and interactions with CMB photons begin to dominate, warming T_{s} back up to T_{CMB}. If there

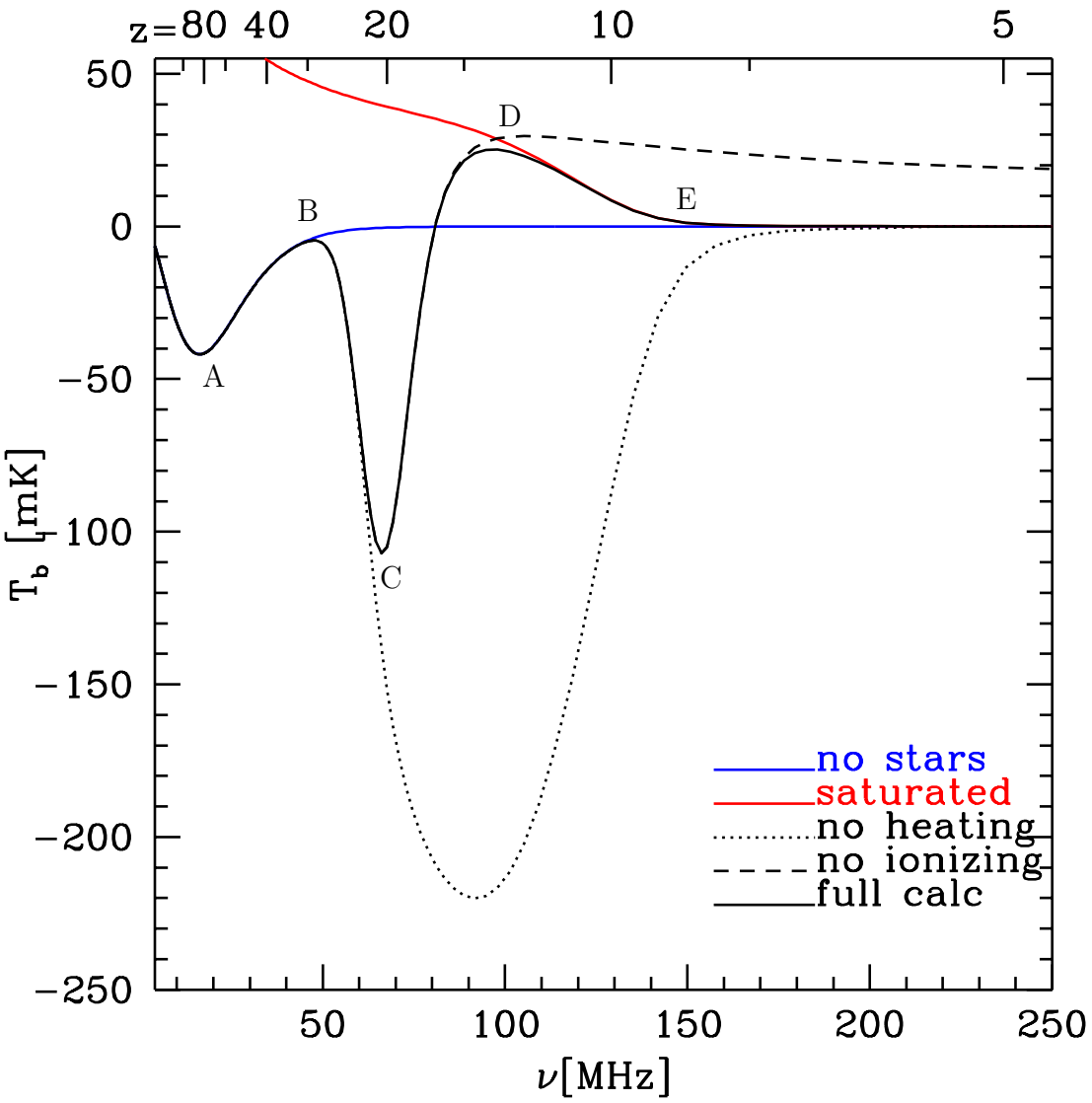


Figure 1.1: The evolution of the global 21 cm brightness temperature relative to the CMB temperature. The black line is a fiducial model which has all the major features: collisional decoupling (A), Lyman-α becoming effective (B), X-ray heating beginning (C), heating saturating (D), and the end of reionization (E). The other lines each show a different scenario with different physics dominating the evolution. Figure reproduced from Pritchard & Loeb (2010).

is also a UV background, the excited spin state can be populated by the Wouthuysen effect (Wouthuysen, 1952; Field, 1959) whereby an electron in the singlet state absorbs a Lyman series photon and de-excites into the triplet (excited) state. Since each Ly-α photon is scattered many times before redshifting out of resonance, Ly-α pumping of the hyperfine transition requires only a small fraction of the flux required for ionization. This begins with the formation of the first stars around $z \approx 30$ and recouples the spin temperature to the gas kinetic temperature to create a strong absorption feature (B to C in Figure 1.1). Eventually, X-rays from quasars or supernovae heat the gas again until $T_{\rm s} \gg T_{\rm CMB}$ (C to D in Figure 1.1), saturating the 21 cm emission. As the gas starts to ionize, the temperature excess decreases until reionization is complete (E in Figure 1.1). In contrast, if there was no significant source of heat, the gas kinetic temperature could have been as low as its adiabatic expansion value. For $T_{\rm s} < T_{\rm CMB}$, the brightness temperature can become arbitrarily large with 21 cm in absorption.

1.3 Observational history and prospects

The 21 cm line was first detected in the lab in 1951 (Ewen & Purcell, 1951; Muller & Oort, 1951; Pawsey, 1951) and within a few years was used to observe spiral structure in the Milky Way (van de Hulst et al., 1954) and the rotation curve of M31 (van de Hulst et al., 1957). Being the first spectral line in the radio regime, attempts at detecting a "21 cm forest" in radio galaxy spectra were made well before quasars and the analogous Lyman-α forest were known (Furlanetto et al., 2006a). Since then direct detection of 21 cm has almost exclusively been observed inside local galaxies out to only about $z \sim 0.2$ (Zwaan et al., 2001; Catinella et al., 2008; Verheijen et al., 2007).

At higher redshift, there is not enough neutral hydrogen in individual galaxies to be observable with the sensitivity currently available. Stacking individual galaxies can be an effective way of making statistical measurements (e.g., Lah et al., 2007; Delhaize et al.,

2013). Similarly, intensity mapping over wide fields, where many galaxies contribute to the flux of each pixel, is a viable way to measure cosmological properties out to $z \approx 1$, such as baryon acoustic oscillations (e.g., Masui et al., 2010; Chang et al., 2010; Masui et al., 2013).

Despite being difficult to observe at low redshift, the 21 cm spectral line has the potential to be a useful probe of the dark ages and reionization. The redshifted 21 cm line can probe all three dimensions of a survey volume, offering many more independent Fourier modes of information about high redshift cosmology than the CMB (Loeb & Zaldarriaga, 2004; Furlanetto et al., 2006a). It is also a direct probe of the IGM, which contains a larger fraction of the universe's baryons than can be studied from individual luminous objects alone (Furlanetto et al., 2006a).

In the 1970s, Sunyaev and Zel'dovich proposed that large-scale structure forms in a "top-down" scenario (beginning with Zel'Dovich, 1970), with the collapse of cluster-sized masses that then fragment into small-scale structures such as galaxies. In this scenario, the early universe would contain $\sim 10^{14}\,\mathrm{M}_{\odot}$ primordial "pancakes" of neutral gas with brightness temperatures as high as 10 K (Sunyaev & Zeldovich, 1975). Many observational programs were run with this model in mind, most at $3 < z < 5$, but none were successful. Only Bebbington (1986) attempted a detection at $z > 6$, and found no structure down to 5 K in brightness temperature at $z = 8.4$.

Hogan & Rees (1979) made the first predictions of 21 cm signal in the context of the currently favoured "bottom-up" structure formation models. It was over a decade before Scott & Rees (1990) suggested the possibility of constraining the matter power spectrum with a statistical measurement, given that the 21 cm signal was far too faint to be detected at high redshifts at the time. Today the 21 cm power spectrum, resulting from a combination of the patchy ionized and neutral medium and the underlying mass power spectrum, is generally considered one of the most promising signals from reionization (Scott & Rees, 1990; Zaldarriaga et al., 2004) and much attention has been paid in the

literature toward designing suitable experiments to detect it (e.g., Morales & Hewitt, 2004; Morales, 2005; Bowman et al., 2006; Harker et al., 2010).

In principle this technique could be used to study the universe during the dark ages well beyond $z = 6$ (Zaldarriaga et al., 2004). In the future, 21 cm tomography during the EoR can be used to add strong constraints to cosmological parameters (e.g., McQuinn et al., 2006; Cooray et al., 2008; Mao et al., 2008; Furlanetto et al., 2009; Masui et al., 2010; Pandolfi et al., 2011) and deduce the nature of the first ionizing sources themselves (Iliev et al., 2012; Kovetz & Kamionkowski, 2013; Datta et al., 2012; Majumdar et al., 2012), including possible exotic reionization scenarios (e.g., Furlanetto et al., 2006b; Mack & Wesley, 2008; Haiman, 2011). However, there are many observational challenges that are currently difficult to overcome. Unfortunately temperature fluctuations in HI are usually predicted to be on the order of tens of millikelvin, many orders of magnitude less than foregrounds from galactic and extragalactic sources at the relevant frequencies (Oh & Mack, 2003; de Oliveira-Costa et al., 2008a), although this relies on estimates of the IGM temperature which are not well constrained. It has been suggested that with less X-ray heating from supernovae, a cold IGM can result in a variance in temperature up to -400 mK against the CMB (Kim & Pen, 2009). If this were the case, the EoR signal would be well within the sensitivity of current generation telescopes.

These foregrounds are currently one of the largest obstacles to detecting the 21 cm signal and several schemes have been developed to address the problem (e.g., Petrovic & Oh, 2011; Liu & Tegmark, 2011; Chapman et al., 2012; Dillon et al., 2012; Parsons et al., 2012b). The most promising approach makes use of the assumption that galactic foregrounds, e.g. synchrotron emission, will be spectrally smooth, while the HI signal is expected to vary on ≈ 1 MHz scales (when ionized bubbles have sizes on the order of a megaparsec). In modelling foreground subtraction often a parametric fit is used, e.g. a 3rd order polynomial in Jelić et al. (2008), although other techniques have been suggested (e.g., Harker et al., 2009). Complicating this, Ali et al. (2008) have reported

observations that foregrounds are not as smooth as theorized which would make fitting foregrounds difficult.

While still very challenging, study of the $z > 6$ universe and the EoR is beginning to become technologically feasible. Some interesting results have already been found from studies of the global 21 cm signal as a function of redshfit. Bowman et al. (2008) estimated an upper limit to the contribution of HI to the redshifted 21 cm brightness temperature of 450 mK. Bowman & Rogers (2010) have put a lower limit on the duration of the EoR of $\Delta z > 0.06$, while Zahn et al. (2012) have used measurements of the kinetic Sunyaev–Zel'dovich effect with the South Pole Telescope to suggest an upper limit on the transition from a neutral fraction of 0.99 to 0.20 of $\Delta z < 4.4$.

1.4 Power spectrum forecasts

There have been numerous simulations of reionization scenarios aiming to make predictions of the 21 cm power spectrum and determine how observations of the power spectrum could be interpreted. Trac & Gnedin (2011) provides a good review of reionization simulations, the broad features of which tend to point to a relatively consistent picture (Iliev et al., 2006a). The most basic (and because of their relative computational simplicity, the largest) simulations evolve the dark matter distribution with an N-body code, while more advanced codes include hydrodynamics. There are also many schemes for solving radiative transfer. The difficulties in accurately simulating the reionization history and topology are primarily from the large range of scales necessary to cover the sources and sinks of radiation to the large ionized bubbles that form as a result.

Simulations have often focused on constraining particular properties of the EoR that shape the ionization topology. For example, the number of ionizing photons which escape their host galaxy to ionize the IGM (the "escape fraction") and the number of times ionized atoms recombine influence both when and how quickly reionization can proceed.

Estimates currently suggest that a few to tens of ionizing photons must be produced per hydrogen atom to ionize the medium (Trac & Gnedin, 2011). As a consequence of this, it is generally not known whether reionization occurred first in the densest regions and progressed outwards ("inside-out"; e.g., Friedrich et al., 2011), or if most ionizing photos escape the dense regions and ionize the low density regions first ("outside-in"; e.g., Miralda-Escudé et al., 2000). Choudhury et al. (2009) suggested that the situation may be more subtle, with the true topology following a combination of the two trends depending on factors such as distribution of ionizing sources and the recombination rate.

Typically simulations for estimating power spectra span $\sim 100\,\mathrm{Mpc}/h$ boxes, which is enough for many independent ionization cells. A typical example, from Jelić et al. (2008), is shown in Figure 1.2 and includes both the brightness temperature of the neutral hydrogen as a function of redshift and position, and the brightness temperature along a single line of sight as a function of frequency. The signal is typically on the order of tens of mK, in line with equation 1.5, and declines steadily as the neutral fraction drops. The size of ionized bubbles at these redshifts tends to be on the order of megaparsecs, or about 1 MHz wide in frequency space. Similar results were shown in Iliev et al. (2008).

These simulations can be used to forecast the performance of EoR experiments aiming to measure the power spectrum. Jelić et al. (2008) used simulations described in Thomas et al. (2009) in the context of reionization observations at the Low Frequency Array (LOFAR), and Iliev et al. (2008) has done similarly for the current GMRT project. These are shown in Figures 1.3 and 1.4, respectively. These curves will be used later as fiducial EoR power spectra when presenting our power spectrum measurements at the GMRT.

Several groups are making progress towards measuring the 21 cm power spectrum. The most prominent of these is the Low Frequency Array (LOFAR[1]; Kassim et al., 2004; Röttgering et al., 2006) which names the EoR as one of its Key Science Projects

[1] `http://www.lofar.org/`

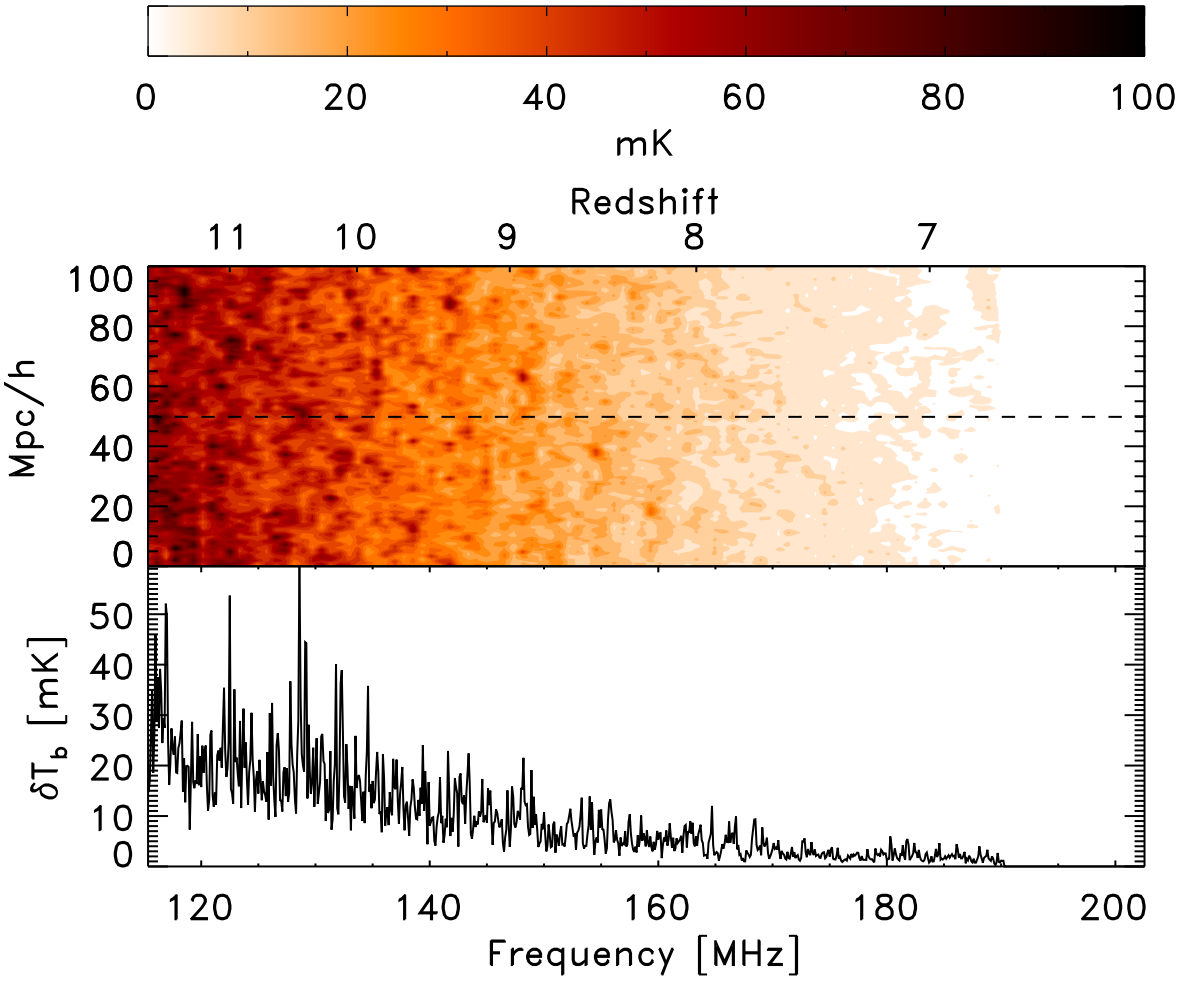


Figure 1.2: Reionization simulation used in power spectrum forecasts of Jelić et al. (2008), showing both the topology as a function of time (top panel) and the brightness temperature (bottom panel) along the line of sight indicated by the dashed line. The simulation itself is described in Thomas et al. (2009). The redshift range $6 < z < 12$ covers a physical distance larger than the $100\,h^{-1}$ Mpc size of the simulation box, resulting in some repetition of structure.

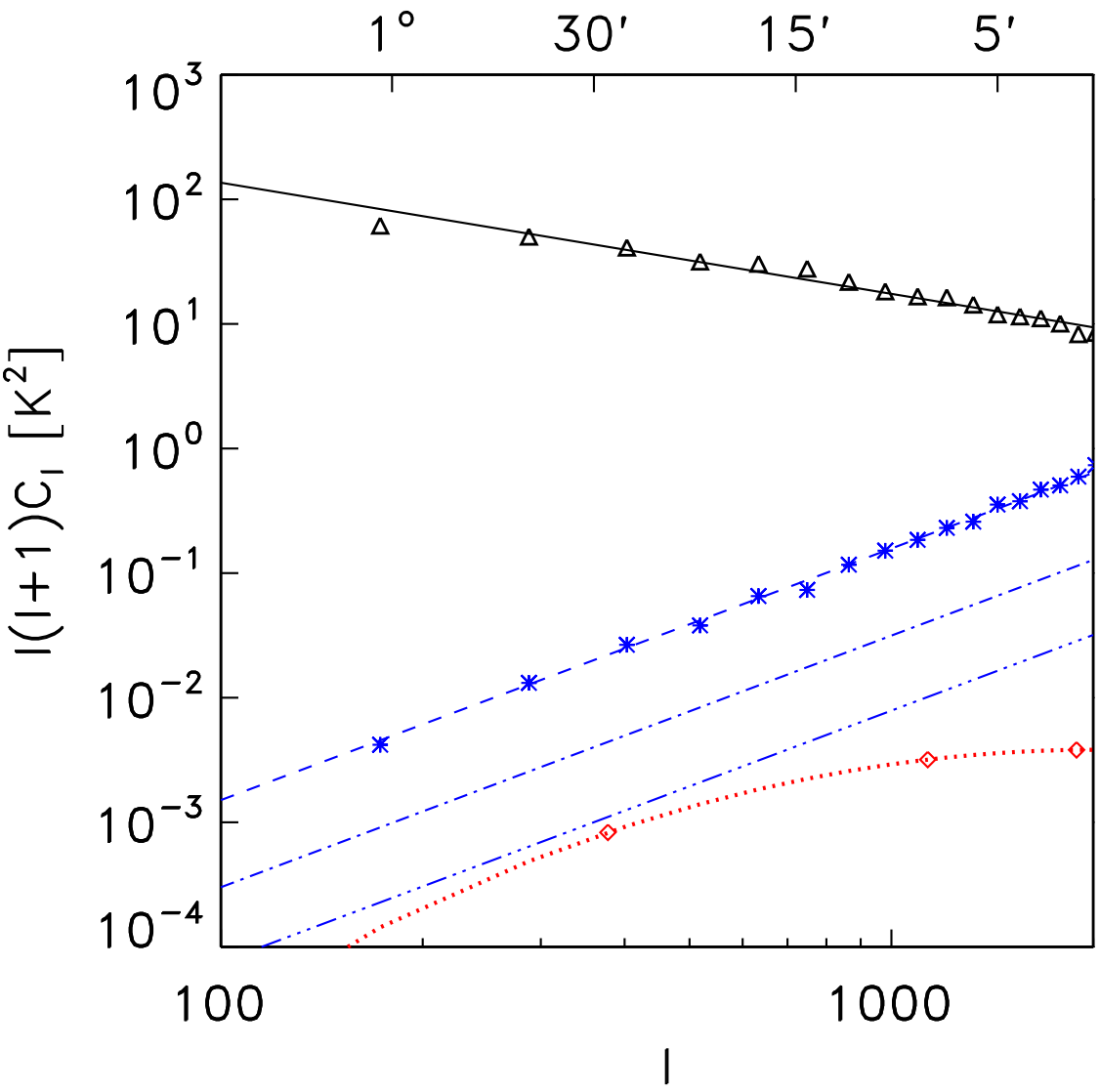


Figure 1.3: 2D power spectrum forecasts for LOFAR from Jelić et al. (2008), in terms of the multipole moment ℓ, which at $z = 8.6$ is related to the perpendicular wavenumber by $k_\perp \approx \ell/6608\,h\,\mathrm{Mpc}^{-1}$. The black line shows the estimated foregrounds. The three blue lines are the estimated noise levels after, from top to bottom, one year (400 hours) of integration with a single beam, one year with fives beams, and four years with five beams. The red line at the bottom is the simulated power spectrum. Though the signal-to-noise is less than one, Jelić et al. (2008) expect to be able to make a statistical detection by estimating the excess in the noise due to the EoR signal.

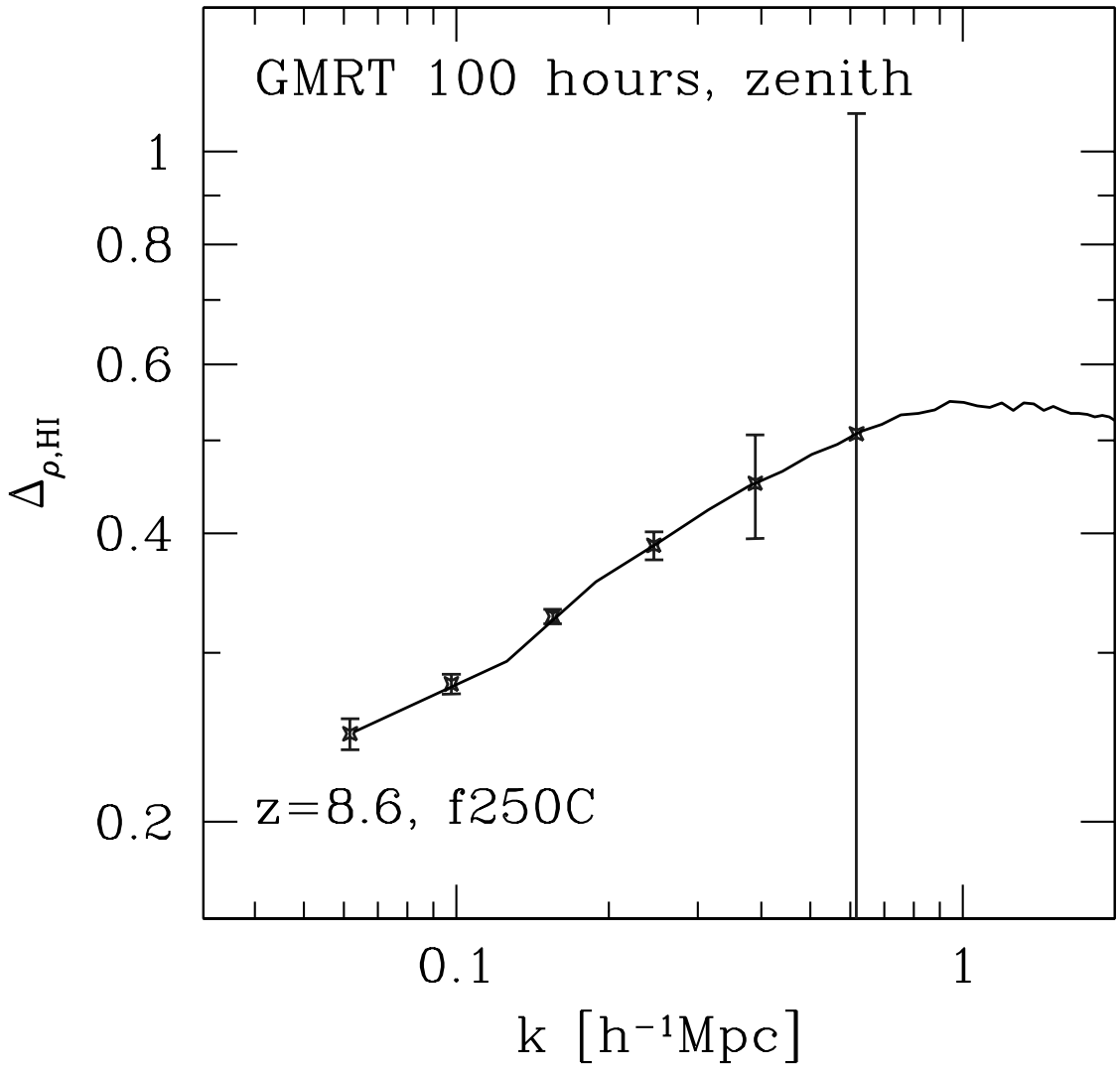


Figure 1.4: Forecasts from Iliev et al. (2008) for the GMRT-EoR observations, presented in terms of the neutral hydrogen density $\Delta_{\rho,\mathrm{HI}}$, which is related to the power spectrum by a factor of δT_b^2, or approximately $9 \times 10^{-4}\,\mathrm{K}$. The error bars assume 100 hours of integration, which is about twice the amount that will be included in our analysis in Chapters 5 and 6.

(Zaroubi & Silk, 2005; Harker et al., 2010; Brentjens et al., 2011) and is currently being commissioned in the Netherlands. LOFAR is a dipole array which features a large field of view and high brightness sensitivity. Zaroubi et al. (2012) estimate that LOFAR will have the potential to overcome the low signal-to-noise to directly image the neutral hydrogen, though a power spectrum measurement will no doubt come first.

While LOFAR is a multipurpose instrument, there are several dedicated EoR projects underway. The 21 cm Array (a.k.a. PaST[2]; Peterson et al., 2004) has been under construction in China but has not had any recent results. The Precision Array for Probing the Epoch of Reionization (PAPER[3]) had reported a limit of approximately 5 K with a 310 mK noise level (Parsons et al., 2010), but has more recently claimed an upper limit of only 52 mK (Parsons et al., 2013) using some novel foreground suppression techniques which we will address in Chapter 6. The Murchinson Widefield Array (MWA[4]; Lonsdale et al., 2009) is expected to be able to detect both the amplitude and slope of the power spectrum with a signal-to-noise ratio > 10 (Beardsley et al., 2013). For more on the EoR case see, e.g., Morales et al. (2006a), Bowman et al. (2006), and Lidz et al. (2008).

Future generations of telescopes, in particular the Square Kilometre Array (SKA[5]), should be capable of achieving even higher sensitivities, including direct imaging of the ionization structures (Carilli et al., 2004) but is not yet under construction and will not be in operation for at least another decade (Rawlings & Schilizzi, 2011). Nonetheless, many of the current generation of EoR experiments are often billed as "path-finder" projects for such future endeavours, and results here will help inform this next generation of EoR science.

[2] http://web.phys.cmu.edu/~past/
[3] http://astro.berkeley.edu/~dbacker/eor/
[4] http://www.mwatelescope.org/
[5] http://www.skatelescope.org/

1.5 Interferometry

Most 21 cm experiments are interferometers. At the low frequencies required to probe the high redshifts of the EoR, interferometers are a natural way to get the brightness sensitivity and angular resolution required, and they measure the power spectrum directly. In this section we briefly describe the basics of interferometry as it applies to EoR experiments. Much more thorough treatments can be found in standard texts such as Taylor et al. (1999) and Thompson et al. (2001).

The most basic interferometer is two radio telescopes separated by some distance b and pointed towards the same target at an angle θ above the horizon. This is illustrated in Figure 1.5. An incoming plane wave from a distant source will arrive at each antenna at different times, due to the path length difference. This difference is called the geometric delay, τ_g. The antenna feeds record the voltage of the incoming electric field, and these are multiplied together by the correlator. The signals suffer additional delays τ_1 and τ_2 between the feeds and the correlator, but these are known and corrected for before correlating.

With two antennas, the situation is exactly analogous to Young's double slit experiment. If the vector between the two antennas is projected onto a plane perpendicular to the direction to the source, they are separated by a baseline vector $\mathbf{u} = (u, v)$ with u measured along the east-west direction and v along north-south, both in units of the observing wavelength. The length of this vector $|\mathbf{u}|$ is equivalent to the separation of the two slits in Young's experiment, which sets the length scale of the interference fringes. The correlation of the signals from two antennas is called a "visibility". Mathematically, it can be expressed as

$$\mathcal{V}_{ij}(u, v) = G_{ij} \int\int \frac{A(l, m) I(l, m)}{\sqrt{1 - l^2 - m^2}} e^{i2\pi(ul+vm)} \mathrm{d}l\mathrm{d}m. \tag{1.6}$$

where $I(l, m)$ is the intensity on the sky, $A(l, m)$ is the primary beam of the antennas, and G_{ij} is a factor accounting for the gain of the antennas. The coordinates l and m are

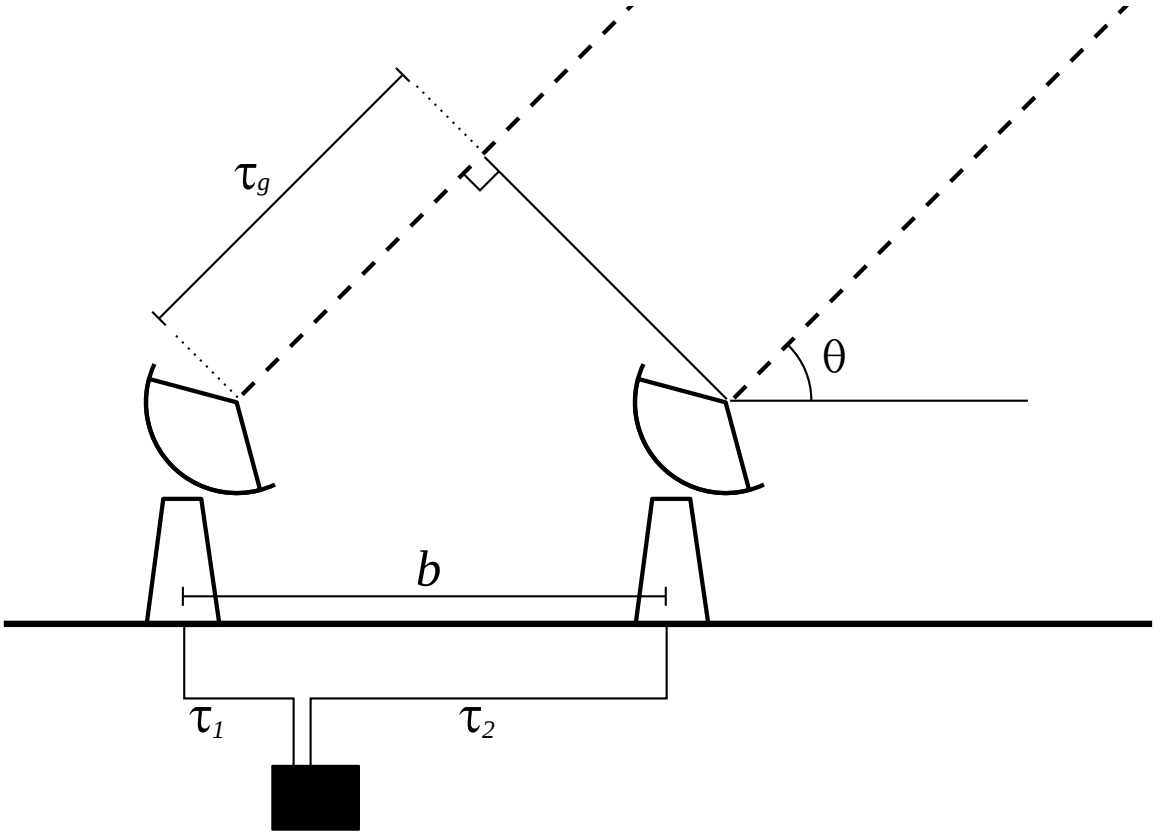


Figure 1.5: Schematic of a basic two-element interferometer observing a source at an elevation of θ from the horizon. The geometric delay τ_g is due to the path difference to each receiver, and depends on both the position of the source on the sky and the distance between the receivers, b. Additional delays between the receivers and the correlator due to the layout of wiring and electronics, τ_1 and τ_2, are different for each antenna but known and fixed.

the cosines of the angles from the pointing centre, measured along the same axes as u and v. If the order of the antennas is switched, this is equivalent to changing the signs of u and v, giving the complex conjugate of the visibility.

A single correlation forms an interference pattern that samples a single Fourier mode on the sky. The angular size of this mode corresponds to $1/|\mathbf{u}|$. A single mode corresponds to a sine wave on the sky, with an orientation determined by the direction of the vector between the two antennas. By changing the baseline length and orientation, one can sample different Fourier modes and build up the complete Fourier transform of the sky.

In practice, interferometers have multiple antennas, with the baselines between every pair sampling as many different modes as possible. As θ changes with the rotation of the Earth, the projected baseline length changes, and two fixed antennas sweep out a series of points in (u, v) space. The baseline also changes as a function of frequency, meaning a single baseline measures different Fourier modes across the bandwidth.

For a point source on the sky, the maximum possible geometric delay for a given baseline is equal to the light travel time between the two antennas. As the sky rotates, this delay, or "lag", changes such that the path difference oscillates between constructive and destructive interference, creating characteristic "fringes". This is corrected for by applying a phase correction given the right ascension and declination of the target in a process called "fringestopping". This is effectively just a correction for the geometric delay, but we can in principle fringestop to locations other than the pointing centre.

To make an image of the sky, we take a Fourier transform of the visibilities. To do a discrete transform, the visibilities need to be gridded. The coarseness of the grid cells, Δu, sets the field of view. Equivalently, this can be written in terms of the number of grid cells N_x (assuming a square grid where $N_x = N_y$) and $u_{\max}$, the maximum $|\mathbf{u}|$ that the grid supports. The field of view of an image is

$$\text{F.O.V.} = \frac{1}{\Delta u} = \frac{N_x}{2}\frac{1}{u_{\max}} \text{ radians.} \tag{1.7}$$

This can, of course, be different from the actual field of view of the telescope (the "primary

beam") which is limited by the size of the dishes. The approximate angular resolution (neglecting incomplete coverage of the (u, v) plane) is

$$\Delta\theta = \frac{1.22}{u_{\rm cut}} \text{ radians} \tag{1.8}$$

where $u_{\rm cut}$ is the maximum $|\mathbf{u}|$ that was included in the transform, and necessarily $u_{\rm cut} \leq u_{\rm max}$. By adjusting $u_{\rm cut}$, $u_{\rm max}$, and N_x, one can tune the spatial scales to which an interferometer is sensitive. Any signal power that only exists outside the range of modes probed by the smallest and largest baselines will be not be detected.

Although we have used only 2D coordinates (u, v) to specify the location of antennas in a plane, the rotation of the Earth turns this plane over time with respect to the line of sight. For long integrations, baselines actually move in a three dimensional (u, v, w) space. If the w term is ignored by setting $w = 0$, this effectively blurs the image. For a thorough explanation of w term issues, see for example Cornwell et al. (2008) and, for a more general overview, Thompson et al. (2001). When the w coordinate is included, the data becomes very sparse and correcting for the frequency dependence of both the beams and foregrounds becomes more difficult. Techniques for projecting visibilities with non-zero w onto the (u, v) plane have been developed in the context of CMB studies and could be applied to EoR data as well (Myers et al., 2003; Hobson & Maisinger, 2002). Such corrections have not been applied to the data present here except where noted in Chapter 4.

Since each baseline of an interferometer measures a Fourier mode directly, imaging is not required to estimate the underlying power spectrum. While the power is most naturally expressed as a function of baseline length $|\mathbf{u}|$, since a given baseline length probes a single angular scale on the sky, it is more common to use units of multipole moment $\ell = 2\pi|\mathbf{u}|$. In two dimensions, the power spectrum C_ℓ is often given in terms of $\ell^2 C_\ell/2\pi$, which is the contribution from a range of Fourier modes per $\ln \ell$, and the integral of which gives the RMS intensity fluctuations of the map. Given that we are interested in scales at high redshift, around $z \approx 8.6$, we can convert the multipole moment

ℓ to a physical distance. Since u and v are derived from angular scales on the 2D sky, this is the 2D power perpendicular to the line of sight, which we express in terms of the wavenumber $k_\perp \approx (\ell/6608)\, h\,\mathrm{Mpc}^{-1}$. Information along the line of sight, $k_\parallel$, is found along the frequency direction. While we will continue to use ℓ for the 2D power, $k_\perp$ is useful when discussing the 3D power as a function of $k = \sqrt{k_\perp + k_\parallel}$. As in the 2D case, we will express the power spectrum as $k^3 P(k)/2\pi^2$, the power per logarithmic range in k.

1.6 GMRT as an EoR instrument

This research uses the Giant Metrewave Radio Telescope (GMRT; Swarup et al., 1991; Ananthakrishnan, 1995) near Pune, India, to make measurements of the power spectrum of the neutral hydrogen signal with the hope of characterizing the structure in the range $8.1 < z < 9.2$. We will discuss more of the technical details of observing redshifted 21 cm emission with the GMRT in the next chapter.

Among the currently active EoR experiments, GMRT is unique in that it was designed and built with the goal of detecting much brighter EoR signatures from "primordial pancakes" before that model fell out of favour. Somewhat serendipitously, despite the significantly different EoR structure that is expected in more modern theories, the design of GMRT still lends itself well to detecting the signal.

Compared to modern EoR experiments which are currently coming online and producing data, GMRT has several design strengths that are difficult to match. These include:

1. *The largest primary dishes.* This results in the smallest field of view, which is a great advantage for an initial detection. The non-coplanar w terms decrease with the field of view, and the ionosphere is less problematic. In the compact core, GMRT has a similar point source sensitivity to the most sensitive competitor, LOFAR. The

details depend on sky brightness, elevation of source, integration time, and other factors. Most importantly, large dishes minimize the impact of side-lobes, because of the high forward gain.

2. *Precise, over-engineered infrastructure.* The GMRT is designed to work up to 1.4 GHz, which is an order of magnitude higher in frequency than the target EoR frequencies. This makes most science specifications trivially achieved, without need to worry about details like pointing, tracking, or surface accuracy to the same degree as other experiments.

3. *Rigid, steerable dishes.* Being able to track a target source means that the GMRT primary beam does not change with time, whereas dipole arrays like LOFAR, PAPER, and MWA have beams which change significantly depending on position on the sky. A constant beam pattern makes analysis much simpler. Beam patterns and polarization leakage do not vary with elevation of GMRT antennas, and we can track targets down to elevations of 18 degrees.

Of course, being dedicated EoR instruments, groups such as PAPER and MWA have an advantage in being much more flexible in approaching the EoR signal with different techniques. Both PAPER and MWA have emphasized the importance of antenna layout in maximizing sensitivity to the EoR signal (Parsons et al., 2012a; Beardsley et al., 2013), a freedom which we do not have at the GMRT. In these respects one of the major advantages of the GMRT is simply that we were the first group to have science data with which to begin to understand some of the unforeseen challenges to detecting EoR. Of course, the various experiments are complementary to each other; the significantly different systematics provide a useful cross-check for a detection in other instruments.

1.7 Structure of this dissertation

This chapter has laid out the background, theoretical, and observational basics of the 21 cm signal from reionization.

Two papers, Paciga et al. (2011) and Paciga et al. (2013), have been published based on the material in this dissertation. The next three chapters on the data and calibration form a common foundation for both papers. In Chapter 2, we outline the observations that have been made at the GMRT from 2007 onward, which form the basis of this research. Chapter 3 discusses the impact of radio frequency interference on our data, and the efforts made to mitigate it both by identifying the physical sources and techniques for masking it. This contains some material that was published as part of Paciga et al. (2011) but has been expanded significantly to include more background and some newer information. Chapter 4 then goes over the calibration of the data and some diagnostics on assessing the quality thereof.

The bulk of Paciga et al. (2011), including the main power spectra results, is included in Chapter 5. In Paciga et al. (2013) we applied a second type of foreground subtraction and quantified the amount of signal loss that results. This forms our Chapter 6.

The component programs of the RFI and calibration pipelines described in Chapters 3 and 4 were in place before this thesis work began, while the analysis itself is new. The programs used in the analysis presented in Chapters 5 and 6 were written by myself unless otherwise noted. For the purpose of providing a complete reference, in some places the names of programs used to perform certain tasks, such as the steps of the RFI removal and polarization leakage calibration, will be given in a monospaced font in parentheses (`example_program`). When necessary we will use the WMAP7 maximum likelihood parameters $\Omega_{\mathrm{M}} = 0.271$, $\Omega_{\Lambda} = 0.729$, and $H_o = 70.3\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Komatsu et al., 2011). All distances are in comoving coordinates.

Chapter 2

Observations

The first large time allocation for the EoR experiment at the GMRT came in the winter of 2007. In this chapter we start by detailing the technical details of GMRT and introduce the general observational strategy. We then briefly give the details of the three target fields that have been used, followed by a summary of the observations of those fields made since 2007.

2.1 Technical details of the GMRT

The GMRT is a radio interferometer located 120 km east of Mumbai, India, the technical specifications of which have been thoroughly documented in Chengalur et al. (2007). The array consists of 30 antennas with separations ranging from approximately 100 m to 26 km. Fourteen of these are arranged in a dense central core of one square kilometre, which allows the high brightness sensitivity required to search for the dim EoR signal (Pen et al., 2009). The remaining 16 are arranged in three long arms extending northeast, northwest, and south from the central square. These arms provide the long baselines and high angular resolution which are useful for imaging compact sources. Coordinates relative to the centrally-located antenna C02 are given in Table 2.1. A map of the (u, v) coverage over a typical eight hour observing run can be seen in Figure 2.1.

		60 Input Number		64 Input Number		Relative Location (m)	
	Antenna	L	R	L	R	Δx (East)	Δy (North)
Central Square	C00	1	29	1	33	687.88	−21.19
	C01	2	30	2	34	326.43	−42.67
	C02	3	31	3	35	0.00	0.00
	C03	4	32	4	36	−372.72	141.29
	C04	5	33	5	37	−565.94	130.54
	C05	6	34	6	38	67.82	−260.80
	C06	7	35	7	39	−31.44	−233.29
	C08	8	36	8	40	280.67	−423.40
	C09	9	37	9	41	41.92	−160.39
	C10	10	38	10	42	−164.88	−621.34
	C11	11	39	11	43	−603.28	−340.09
	C12	12	40	12	44	174.85	−672.16
	C13	13	41	13	45	−639.53	−1182.34
	C14	14	42	14	46	−473.71	−664.85
West Arm	W01	15	43	15	47	−1591.94	625.39
	W02	16	44	16	48	−3099.41	1501.18
	W03	17	45	17	49	−5199.90	3066.16
	W04	18	46	18	50	−7039.03	5359.52
	W05	19	47	19	51	−8103.13	8267.57
	W06	20	48	20	52	−11245.60	9430.02
East Arm	E02	21	49	21	53	2814.55	1008.62
	E03	22	50	22	54	4576.00	2043.81
	E04	23	51	23	55	7780.69	3070.58
	E05	24	52	24	56	10200.00	3535.84
	E06	25	53	25	57	12073.46	4804.91
South Arm	S01	26	54	26	58	633.92	−2967.61
	S02	27	55	27	59	−367.30	−4525.73
	S03	28	56	28	60	333.03	−6774.02
	S04	59	57	29	61	947.68	−9496.90
	S06	60	58	30	62	−369.04	−14153.59

Table 2.1: GMRT antenna names, position relative to antenna C02 in metres, and ordering in both the 60-input and 64-input formats used by the EoR correlator. Note that there are no antennas C07, E01, or S05. In the 64-input case, positions 31, 32, 63, and 64 are empty.

Each antenna is a paraboloid reflector with a feed turret mounted at the primary focus supported by four evenly spaced support legs. The turret can rotate through four positions to allow observing at frequencies from 150 MHz to 1420 MHz. Figure 2.3 shows the different feeds. The diameter of each dish is 45 m, giving GMRT a nominal collecting area of 47 713 m^2. However, the feed was designed with a tapered response close to the edge of the dish to minimize sidelobes and avoid contamination from the ground, so the effective collecting area is slightly smaller. The antenna dish has a mesh design which has the advantage of being cheaper to build and having low wind resistance. Being designed for frequencies up to 1420 MHz, the wire spacing increases from 10 mm in the centre of the dish to 20 mm at the edges, making it very smooth for 2 m wavelengths at 150 MHz.

The antenna feeds are pairs of dipoles sensitive to orthogonal linear polarization. The signals are passed through a hybrid coupler which combines these to provide a pair of left (L) and right (R) circularly polarized signals. With two polarizations from 30 antennas, there are a total of 60 signals to process, which combine into 1830 cross- and auto-correlations. For a pair of antennas, there are four possible correlations: LL, RR, LR, and RL. These can be converted to the standard Stokes parameters as follows:

$$\begin{aligned} I &= LL + RR \\ Q &= RL + LR \\ U &= -i(RL - LR) \\ V &= LL - RR \end{aligned} \tag{2.1}$$

which have units of intensity.

A noise source at the antenna adds a noise signal of known power to both the L and R feeds with a 25% duty cycle and a period of 1 Hz. This can be extracted later for the purposes of calibrating the relative antenna gains. The noise sources tend to be unreliable and poorly monitored, however. A visual inspection of the LR auto-correlation of antennas can indicate which antennas have working noise sources. One quarter to one

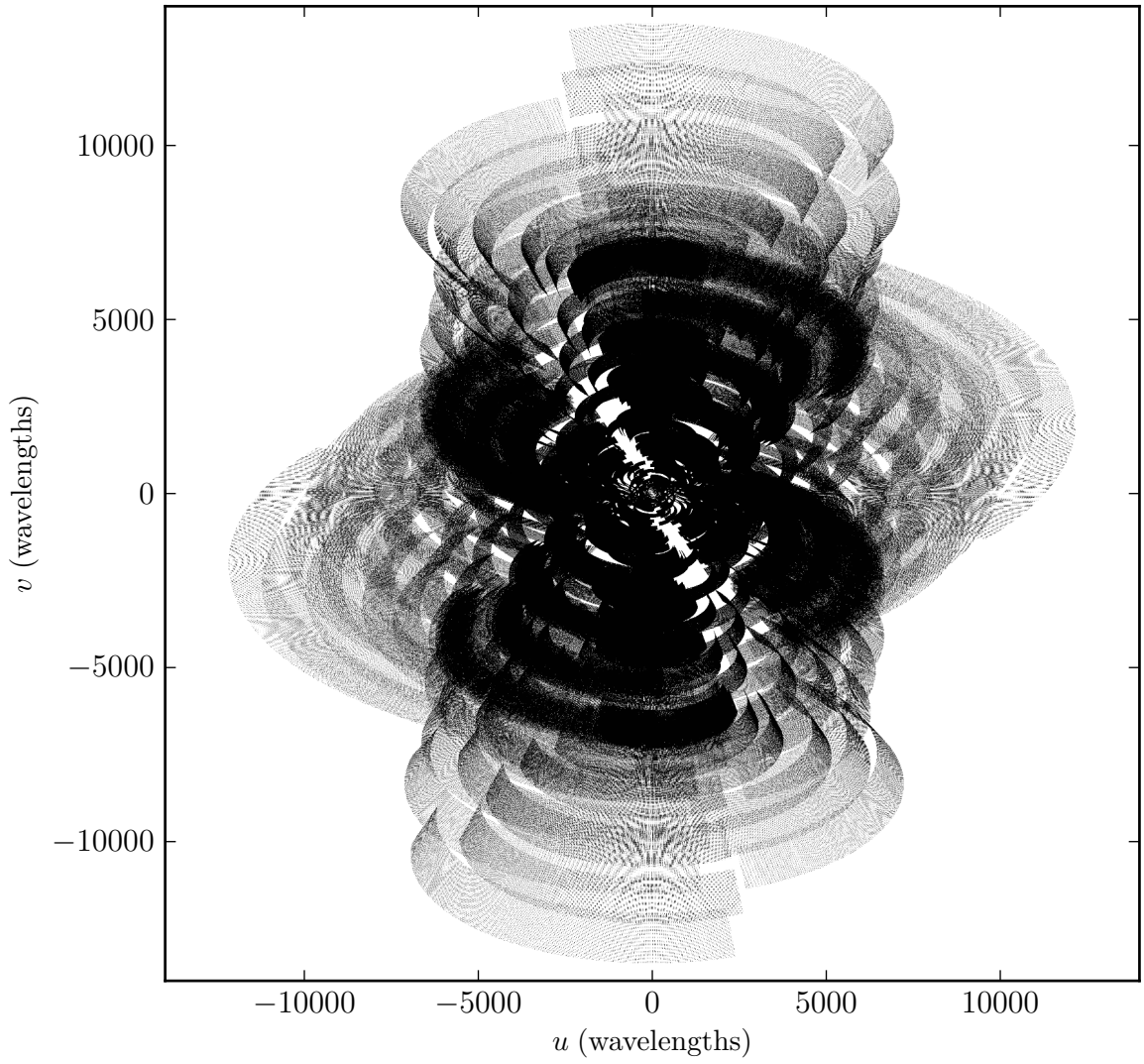


Figure 2.1: Coverage in the (u, v) plane for GMRT during an eight hour observation on December 14, 2007. Two dots appear for every two antenna baseline at a single time and frequency, and these move in the (u, v) plane as the Earth rotates. The GMRT has good coverage for $|\mathbf{u}| \lesssim 5000$. The Fourier transform of this plane, shown in Figure 2.2, is the synthesized beam.

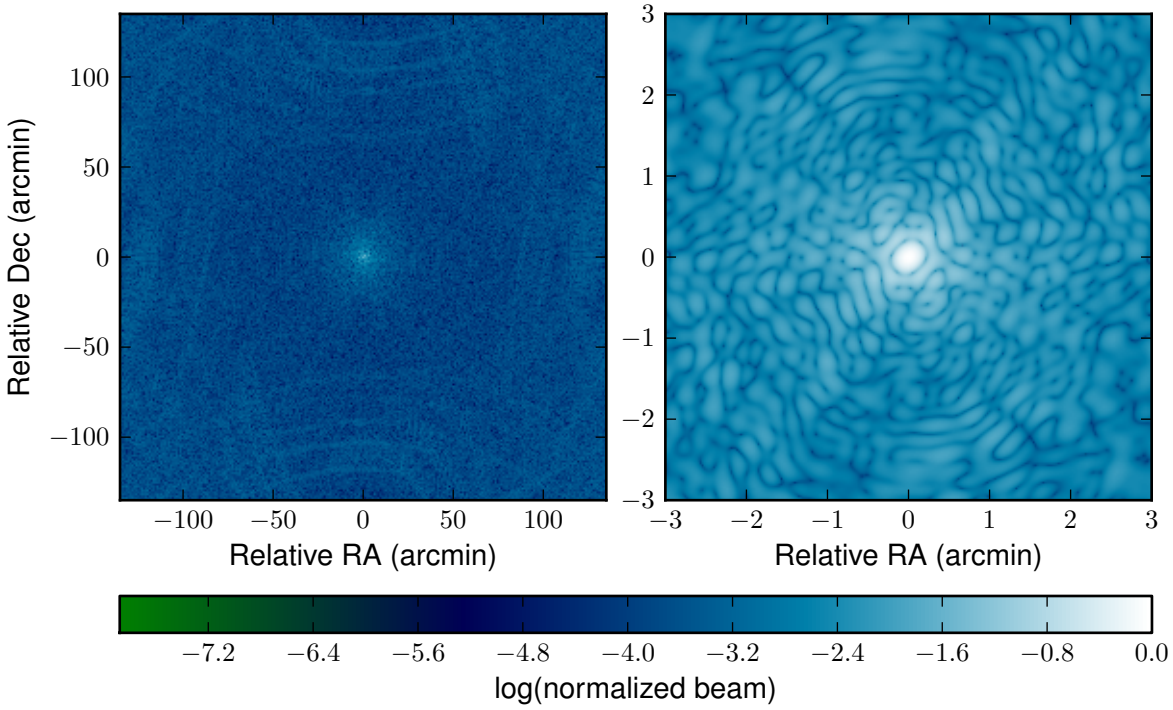


Figure 2.2: The synthesized beam profile, derived from the Fourier transform of the (u, v) plane shown in Figure 2.1, with a width of approximately 20 arcseconds using all baselines. While the synthesized beam is mostly important for imaging, it influences the flux calibration in how well it resolves substructure of sources and in how bright sources outside the field contribute through side-lobes.

Figure 2.3: Close up photo of the GMRT antenna primary focus. Clockwise from the top are the 150 MHz, dual 610/233 MHz, 327 MHz, and 1420 MHz feeds. *Photo by Mark Kuiack.*

half of the noise malfunctioning at any time is typical. A thorough check of all antennas was done in 2007 and again in 2011, with results shown in Table 2.2. Noise sources which have not worked consistently over the course of this project were reported to GMRT staff in 2011 to be repaired.

The antenna inputs are processed through the GMRT Software Backend (GSB; Roy et al., 2010), a 48 node cluster which was built in part for this project and has now replaced the hardware correlator for all observing at the GMRT. Sixteen nodes are dedicated acquisition nodes, using commercial analog-to-digital converter (ADC) sampling boards to feed four antenna inputs to each node. The remaining 32 nodes are used for computation, data storage, and some post-processing. The GMRT-EoR project uses a dedicated software correlator which records the visibilities in a plain binary format without any headers or meta-data. This EoR correlator was developed for the GSB before the standard GMRT software correlator was in place, and was necessary to record visibilities "gated" on the period of the pulsar at a high resolution, which was not possible with the hardware correlator in place at the time. By gating, we mean that the incoming data is folded on the period of the target pulsar, and a separate visibility is recorded for each subset of the period, or "gate". This is necessary to isolate the pulsar pulse from the rest of the data for calibration, as will be described in Chapter 4. The correlator itself was written by Ue-Li Pen, while the fringestopping component was written by Tzu-Ching Chang and Chris Hirata. Our file format has been modified slightly over the years to accommodate changes to the GMRT hardware configuration and to enable remote observations. Most significantly, data recorded in 2009 and onward has 64 antenna inputs instead of 60, four of which are empty. Table 2.1 shows the ordering of antennas used in each version of the binary format.

The hardware delays due to the different path lengths between the location of GSB and the individual antennas are accounted for by the correlator. However, regular maintenance and repairs of the connections can create changes in these lengths that are not

	Dec 2007					Feb 2011				June 2011	
Date	10	14	16	17	18	20	25	26	27	22	26
C00						I		X		I	I
C01						X	X	X	X	X	X
C02						I	I	X		I	I
C03						I	I	X		I	I
C04						I	X		X		
C05					X	I					
C06	X	X	X	X	X	X	X	X	X		
C08	X	X	X	X	X		I			I	I
C09			I				I				
C10											
C11					I	X	X	X	X	X	X
C12	X	X	X	X	X		I	I	I	X	I
C13	X	X	X	X	X						
C14	I		X	I	X		I				I
W01							I				I
W02							I	I	I	I	I
W03	I					X	I	I	I	X	I
W04	I							I		I	I
W05	I				I			I		I	I
W06						I	X	X			I
E02	X	X	X	X	X	X	X	X	X	X	X
E03				X				I			
E04						I		I	I		
E05						I					
E06	I							I			
S01	X	X	X	X	X	I					I
S02	X	X	X	X	X						I
S03					I			I			
S04								I			
S06	X										

Table 2.2: Antennas for which the noise source was not functioning during observations in 2007 and 2011. An "X" denotes no noise source visible at all in the LR auto-correlation of the antenna. An "I" denotes a source that was visible only irregularly (during some, but not all, scans from that day).

automatically detected. It is important that the delay table used by the EoR correlator (in nanoseconds) is regularly checked against the correct path lengths (in metres) maintained by the GMRT engineers to ensure the correlations are correct. These are updated by GMRT every few months, and can change the delays by tens to hundreds of nanoseconds. Plots of the pulsar lag, discussed in section 2.3.1, can be used to identify errors in the delay table.

The EoR correlator uses 2048 frequency channels over a $16.\bar{6}\,\mathrm{MHz}$ bandwidth from 156.0 to $139.\bar{3}\,\mathrm{MHz}$ for a frequency resolution of 8.1 kHz with a 0.25 s integration time. For 21 cm (1421 MHz) emission this frequency range probes a redshift range of $8.1 < z < 9.2$, which provides a cylindrical comoving survey volume of $(280\,\mathrm{Mpc}/h)^3$, with about equal dimensions in all three directions. The highest angular resolution at this observing frequency at GMRT is about 20 arcseconds.

The file sizes of the raw correlations are quite large, such that storage space on GSB is one of the biggest bottlenecks for large observing blocks. On each node, for a single scan, the disk space D required is

$$D = \left(\frac{N_\nu}{N_\mathrm{node}}\right)\left(\frac{N_\mathrm{in}(N_\mathrm{in}+1)}{2}\right)\left(\frac{\Delta t}{0.25\,\mathrm{s/visibility}}\right)\left(\frac{2\,\mathrm{bytes}}{\mathrm{visibility}}\right)\max\left(N_\mathrm{gate}, N_\mathrm{gate}\times P\right) \tag{2.2}$$

where $N_\nu = 2048$ is the number of frequency channels, $N_\mathrm{node} = 16$ is the number of storage nodes, $N_\mathrm{in} = 60$ or 64 is the number of antenna inputs, Δt is the length of the observation, $N_\mathrm{gate} = 16$ is the number of gates, and P is the pulsar period. The last factor reflects the fact that a maximum of N_gate gates are recorded, and if the period is longer than the integration time, empty gates will not be written to disk. For a pulsar with a 0.5 s period, this works out to approximately one gigabyte per minute per node (or a total of almost one terabyte per hour). With four terabytes per storage node dedicated to the EoR experiment, there is space for about 50 hours of observing. This can be further compressed by splitting the data into two streams: one with all 16 gates but rebinned to 128 frequency channels and 16 second integrations, and a second with

the full frequency and time resolution but averaged down to a single gate. Doing so provides a total reduction in size of about a factor of 15, to about 60 gigabytes per hour. This reduction is necessary to be able to move the data to the approximately 4 terabyte staging area before writing to magnetic tape and transporting back to the University of Toronto.

2.2 Sensitivity

Several system parameters useful for estimating the sensitivity are given in Chengalur et al. (2007). At 150 MHz the GMRT design specifications list a receiver temperature of 144 K, a ground temperature of 30 K due to the mesh and spillover from the ground, and a typical sky temperature of 308 K, giving a total system temperature of $T_{\rm sys} = 482\,{\rm K}$. Observationally, GMRT has reported that $T_{\rm sys}$ is closer to 582 K.

The gain G_A of a single antenna is 0.33 K/Jy. On a single baseline, the total noise can be estimated as

$$n = \frac{2k_{\rm b}T_{\rm sys}}{A\sqrt{\Delta t \Delta \nu}} = \frac{T_{\rm sys}}{G_A}\frac{1}{\sqrt{\Delta\nu\Delta t}} \tag{2.3}$$

where A is the collecting area of a single antenna, Δt is the integration time, and $\Delta\nu$ is the observing bandwidth. In an interferometer with many baselines, an additional factor of $N(N-1)/2$, where N is the number of antennas, appears in the square root. For the GMRT, $A \approx 1200\,{\rm m}^2$, $N = 30$, $\Delta\nu = 16.\bar{6}\,{\rm MHz}$, giving an RMS noise of about $40\,\mu$Jy/beam for a 10 hour integration. However, this assumes a bright point source, the Fourier transform of which is constant across the (u, v) plane, which is not true of the diffuse EoR signal.

For EoR observations, the relevant $\Delta\nu$ is not the total bandwidth but the frequency scales on which we expect the signal to reside; larger frequency scales will be treated as foregrounds and removed. This is typically on the order of 1 MHz. The relevant integration time, Δt, is the time it takes for the $\approx 20\lambda$ beam to move to an independent

point in (u, v) space due to the Earth's rotation. For the smallest baselines of about 50λ this takes roughly 90 minutes. These values give an estimated noise of 26 mJy, significantly higher than the expected EoR signal of less than 10 mJy. However, in a single night each of the three 50λ baselines will move over as many as six independent regions of the (u, v) plane, with sixteen 1 MHz wide frequency bins across the bandwidth. This gives approximately $3 \times 6 \times 16 = 288$ independent measurements, for a factor of 17 improvement on the noise in the power at this choice of scales.

A thorough estimate of the noise across the whole array is more complicated since these factors change with each baseline, choice of field, and timing of observations. Nonetheless, the primary sensitivity to the EoR signal comes from the compact central core, which is contained within about 1 km or roughly 30 dish diameters. This means there are about 30^2 resolution elements across the primary beam for these baselines, for a total signal-to-noise approximately 30 times better than any individual resolution element. This suggests that if we were only thermally limited we should be able to measure the total spatial power to a few per cent.

The raw data has a very high spectral resolution, giving us flexibility in which line-of-sight physical scales ($k_\parallel$) to measure. The noise is effectively fixed for a given spatial ($k_\perp$) scale, and the power is inversely proportional to the total k, suggesting that the best signal-to-noise will be achieved at the lowest possible $k_\parallel$. In practice, however, this is also where the foregrounds are strongest, and moving to higher $k_\parallel$ means more frequency channels fit across the bandwidth, reducing the noise in the power. Balancing between these two effects, our highest sensitivity is likely to be when the $k_\perp$ and $k_\parallel$ scales are comparable.

2.3 Observational strategies

2.3.1 Real-time pulsar calibration

All GMRT-EoR observations use a pulsar as a calibration source, which has several advantages over standard modes of calibration for radio astronomy. Pulsar pulses have very short duty cycles relative to their periods, and aside from possible pulsar wind nebulae (see, e.g., Gaensler & Slane, 2006) have effectively no flux otherwise. The EoR correlator folds over the period of the observed pulsar, and divides the pulsar period into 16 equal "gates", such that the pulse is entirely contained within a single gate. The 0.25 s dump time of the correlator is typically shorter than the period of the pulsar, so in practice each dump contains fewer than 16 full gates of data.

The period of the pulsar is much shorter than the time-scale at which ionospheric fluctuations dominate, so by comparing the "on" gate with the neighbouring "off" gates, everything in the field that is constant over the period of the pulsar can be removed. This includes sky sources and most radio frequency interference (RFI). This technique allows calibration of both phase and polarization in real time, with 100 per cent observing efficiency, since no additional time is needed for phase or flux calibration. It also allows for a calibration as a function of time, with an updated solution at every timestamp. The full calibration procedure will be discussed in more detail in Chapter 4.

Since the on-pulse flux of a pulsar is typically greater than all other sources in the field, the pulse gate can usually be identified automatically as the one with the highest flux recorded. Occasionally this technique fails—the pulsar may be too dim, too far from the pointing centre, or there may be too much RFI. In these cases it is often possible to identify the pulsar by eye. This is done by using a plot of the amplitude in lag space as a function of both time and gate. An example of such a plot is shown in lag space for a single baseline in Figure 2.5. If the pulsar is visible in a single gate but not at zero lag, this is an indication that the correlator delay tables may be incorrect. In principle

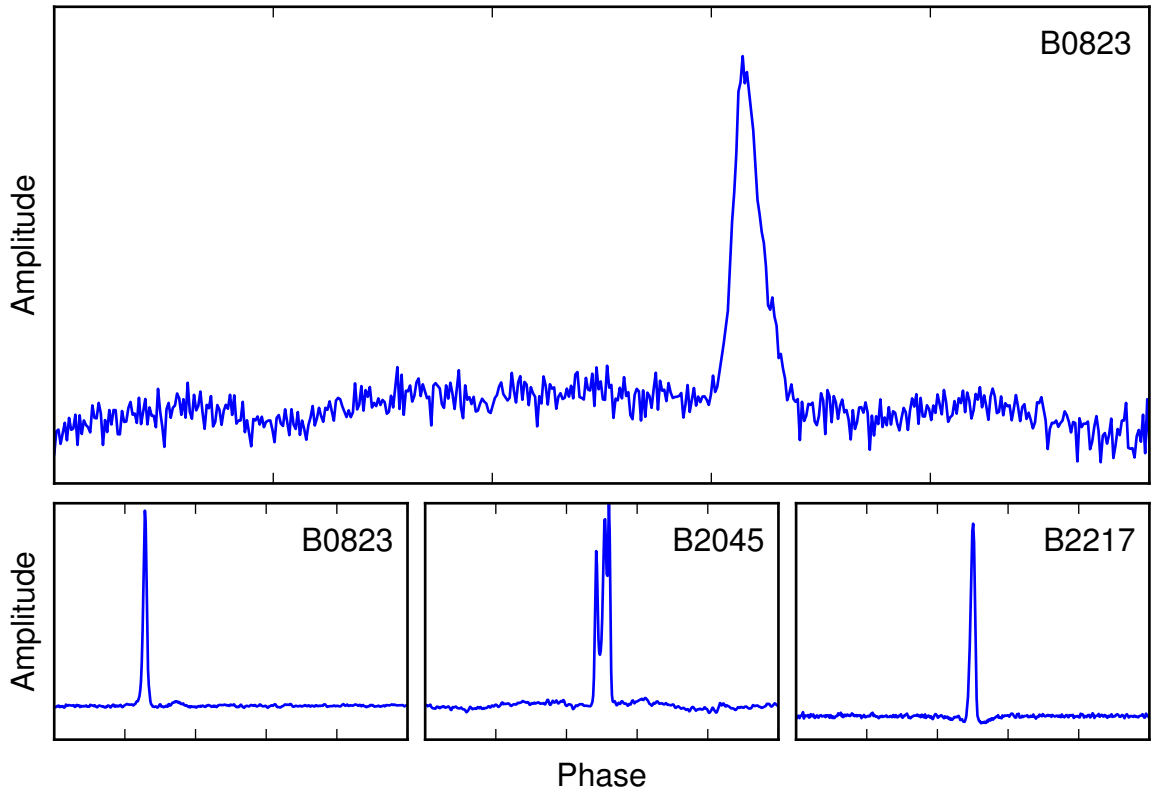


Figure 2.4: Pulse profiles for the observed pulsars, all on an arbitrary (uncalibrated) amplitude scale. The upper plot shows the profile of B0823+26 at 150 MHz, recorded using GMRT's C00 antenna during an EoR observation on March 2, 2010. The lower three plots show the profile for all three pulsars used in this project at 408 MHz using the 76 m Lovell Telescope, taken from Gould & Lyne (1998). Though there can be some substructure to the pulse, the width is significantly shorter than the total period in all cases.

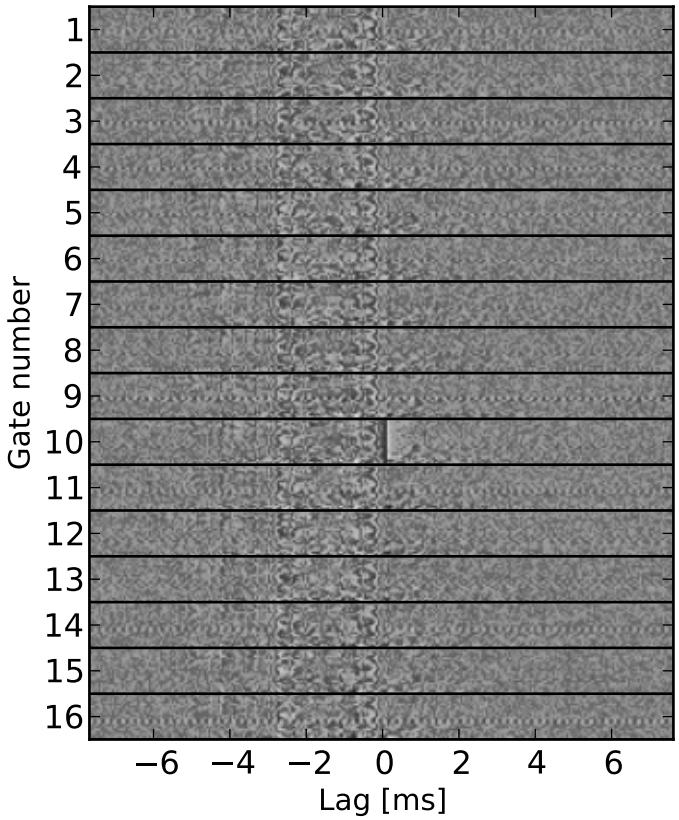


Figure 2.5: Lag space plot of a pulsar observation for the C00-C01 baseline. The sixteen rows are the sixteen gates spanning the pulsar period, with time increasing downwards in each, covering 20 minutes in one minute integrations. Each gate is integrated independently. The pulsar pulse is seen as the bright line at zero lag in the 10th gate. Sources of interference which are on only briefly show up as a horizontal line at the same vertical position in every gate.

delays can be tuned manually this way by changing the relative lag between the baseline antennas by an amount equal to the deviation from zero lag.

Errors in clock synchronization with the pulsar period sometimes cause the pulsar pulse to drift relative to the correlator timing and, for a time, straddle two gates. In these cases it is necessary to include more than one gate when identifying the pulsar signal. This is not ideal since the pulsar signal is diluted over multiple gates, but the correction is limited to only the portion of observations that require it. An alternative

approach is to dynamically select the pulsar gate at each timestamp rather than specifying it globally, but this has the downside that the incorrect gate will often be identified when, for example, RFI is brighter than the pulsar, distorting the pulsar signal even further. While this can be corrected by manually editing the pulsar gate location, this is not practical for large data sets.

Since the pulsar amplitude varies from pulse to pulse, the absolute system gain needs to be measured separately, first by calibrating the individual gain response of each antenna, while a sky radio source is used to set the absolute flux scale. For extended sources such as radio galaxies, this can only be done for the short baselines where its structure is not resolved. For antennas in longer baselines we determine the relative gain calibration using the pulsar itself. This procedure is described in Chapter 4.

2.3.2 Observational technique

Early observations made use of long (eight to ten hour) continuous observations with the antennas tracking the pulsar. The analysis in later chapters of this thesis use only continuous tracking data, but more recent observations have used different strategies.

In late 2009, drift scanning was used instead of tracking. In this mode, the antennas are positioned such that the target pulsar is at the eastern edge of the primary beam. They are locked into position while the pulsar drifts across the beam with the rotation of the Earth. When the pulsar reaches the opposite edge of the beam, the antennas are repositioned for the next scan. This has the advantage of keeping the array more stable over the length of one scan, reducing any possible interference from the pointing motors themselves, as well as filling out more of (u, v, w) space.

Beginning in 2010, a north/south scanning strategy was added to complement the east/west information added by drift scanning. For this, each scan was positioned at one of three declinations: equal to, half a primary beam width north, and half a primary beam width south of the pulsar declination. Mosaicing in this way is necessary to obtain

sub-primary beam information in order to compensate for the frequency dependence of u, v, and w.

More recently, beginning in late 2010, total observation lengths were limited to about two hours before and after transit, to minimize the w term, making more precise foreground subtractions possible. One additional benefit from this is that by limiting observations to the same LST range each day, we get deeper observations at the same (u, v) points than a similar number of hours spread over fewer long nights.

Typically, reducing the (u, v) coverage is not an advantage in radio interferometry, where the goal is usually high resolution images. However, deepening the coverage over a smaller (u, v) range should improve the power spectrum measurement, which does not rely on evenly sampling all angular information in the same way as imaging, assuming there is no preferred direction to the power spectrum. This strategy is now also being used by the PAPER EoR project, which has experimented with array configurations designed to maximize the redundancy in the (u, v) plane for a fixed number of baselines (Parsons et al., 2012a). While the GMRT antenna locations are fixed, using target fields near the equator in addition to observing in limited LST ranges also limits the (u, v) coverage and opens up more potential target fields that are far from the galactic plane.

2.4 Target fields

Three different pulsars have been the main target fields between 2007 and 2012, each selected with different motivations as described below. The relevant properties of each pulsar are given in Table 2.3, and the position of each field on the galactic plane is shown in Figure 2.6, along with the brightest radio sources on the sky for reference.

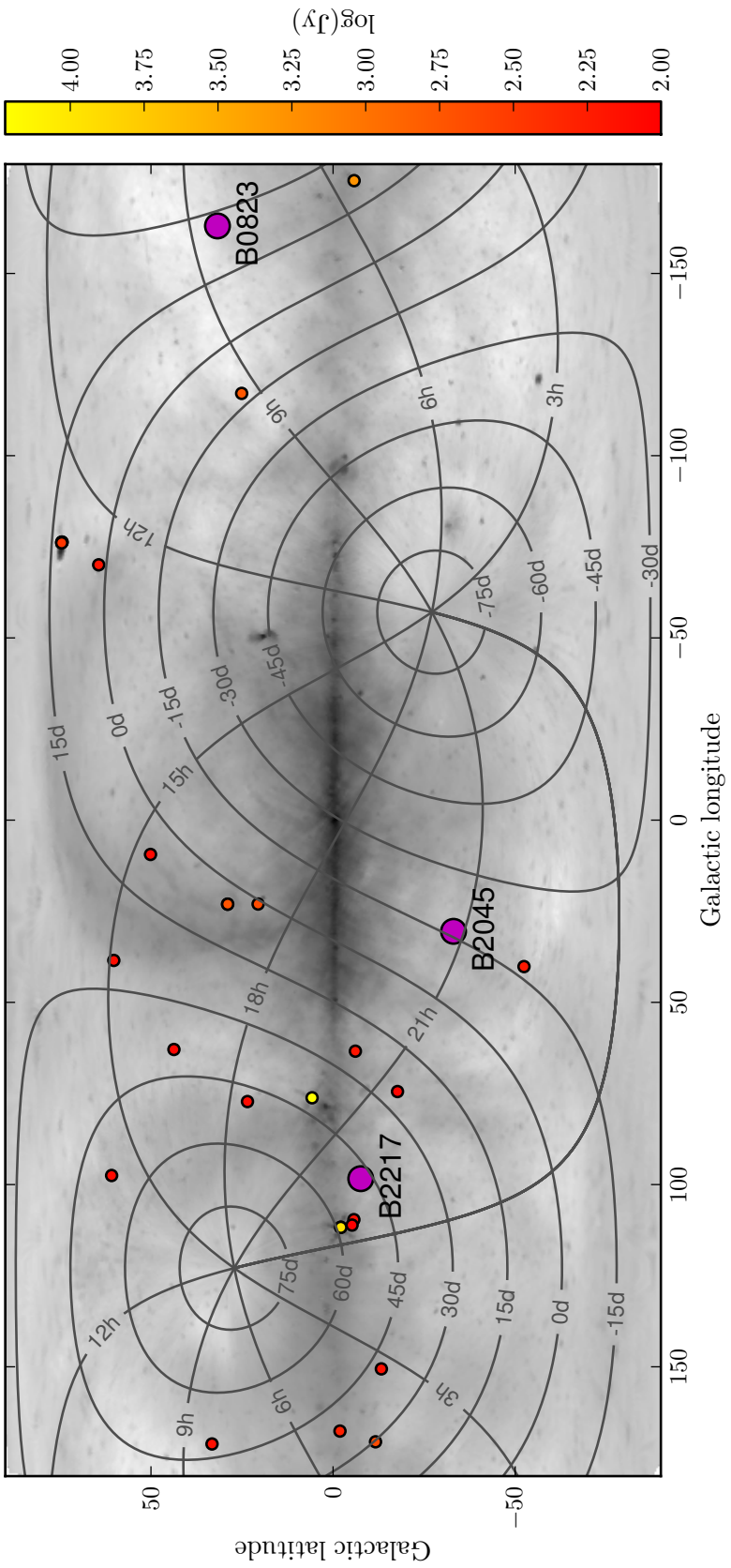


Figure 2.6: All sky map including sources brighter than 100 Jy at 74 MHz from the VLA Low-Frequency Sky Survey (Cohen et al., 2007), which covers 95% of the sky north of −30 degree declination. Purple circles mark each of the three pulsars observed. Note that B2217 is close to the two brightest sources on the sky, Cas A (left) and Cyg A (right). The background image shows 150 MHz diffuse emission on a log scale estimated with LFmap[a] from the Haslam 408 MHz maps (Haslam et al., 1982; Platania et al., 2003).

[a] `http://www.astro.umd.edu/~emilp/LFmap/`

Name		B0823+26	B2045-16	B2217+47
Right ascension	(J2000)	08h26m51.38	20h48m35.64	22h19m48.14
Declination	(J2000)	+26°37'23.79"	−16°16'44.55"	+47°54'53.93"
Galactic latitude	(deg)	196.96	30.51	98.38
Galactic longitude	(deg)	31.74	−33.08	−7.60
Period	(ms)	530.66051169	1961.572303613	538.4688219194
Dispersion measure	(cm^{-3} pc)	19.454	11.456	43.519
Flux at 400 MHz	(mJy)	73 ± 13	116 ± 26	111 ± 12
Flux at 1400 MHz	(mJy)	10 ± 2	13 ± 4	3 ± 0
Spectral index		−1.6 ± 0.2	−1.7 ± 0.3	−2.8 ± 0.3

Table 2.3: Parameters of observed pulsars, from the Australian Telescope National Facility Pulsar Catalogue (Manchester et al., 2005). Periods and dispersion measures are from Hobbs et al. (2004). Fluxes at 400 and 1400 MHz and spectral indices are from Lorimer et al. (1995) and Fomalont et al. (1992).

2.4.1 PSR B0823+26

This pulsar has been observed regularly during winter observations, since it is up during the night when radio frequency interference and ionospheric variations are thought to be less important. This is the main field discussed in the bulk of this thesis. Data from this field was first analyzed and published in Pen et al. (2009).

This pulsar has a period of about 0.53 seconds and an average flux of 350 mJy at 150 MHz (Hobbs et al., 2004). It is situated in a relatively cold part of the sky at a galactic latitude of 32° with few nearby bright sources, meaning foreground subtraction should be relatively easy. The on-pulse flux is about 6 Jy, brighter than all other sources in the field, making it a good calibrator.

Figure 2.7 shows the positions of the brightest sources in this field. The primary flux calibrator used for this field is 5C 7.245, a radio galaxy at $z \approx 1.6$ (Willott et al., 2001), located less than 14 arcminutes from the pointing centre.

2.4.2 PSR B2217+47

This pulsar was observed during the night in the summers of 2008 to 2010 as a complement to B0823, which is up only during the day at that time of year.

Though the fluxes measured at 400 MHz and 1400 MHz (Lorimer et al., 1995; Fomalont et al., 1992, respectively) imply that this should be even brighter than B0823, B2217 did not show up as clearly in the data, making it difficult to calibrate. Further, it is at a much lower galactic latitude than B0823 and much closer to some of the brightest radio sources on the sky, including both Cassiopeia A and Cygnus A, making foregrounds much more problematic than was anticipated when selecting the field.

As a result, as will be described further in section 2.5, most of the data on this field is not as useful as the B0823 field and has not been analyzed completely.

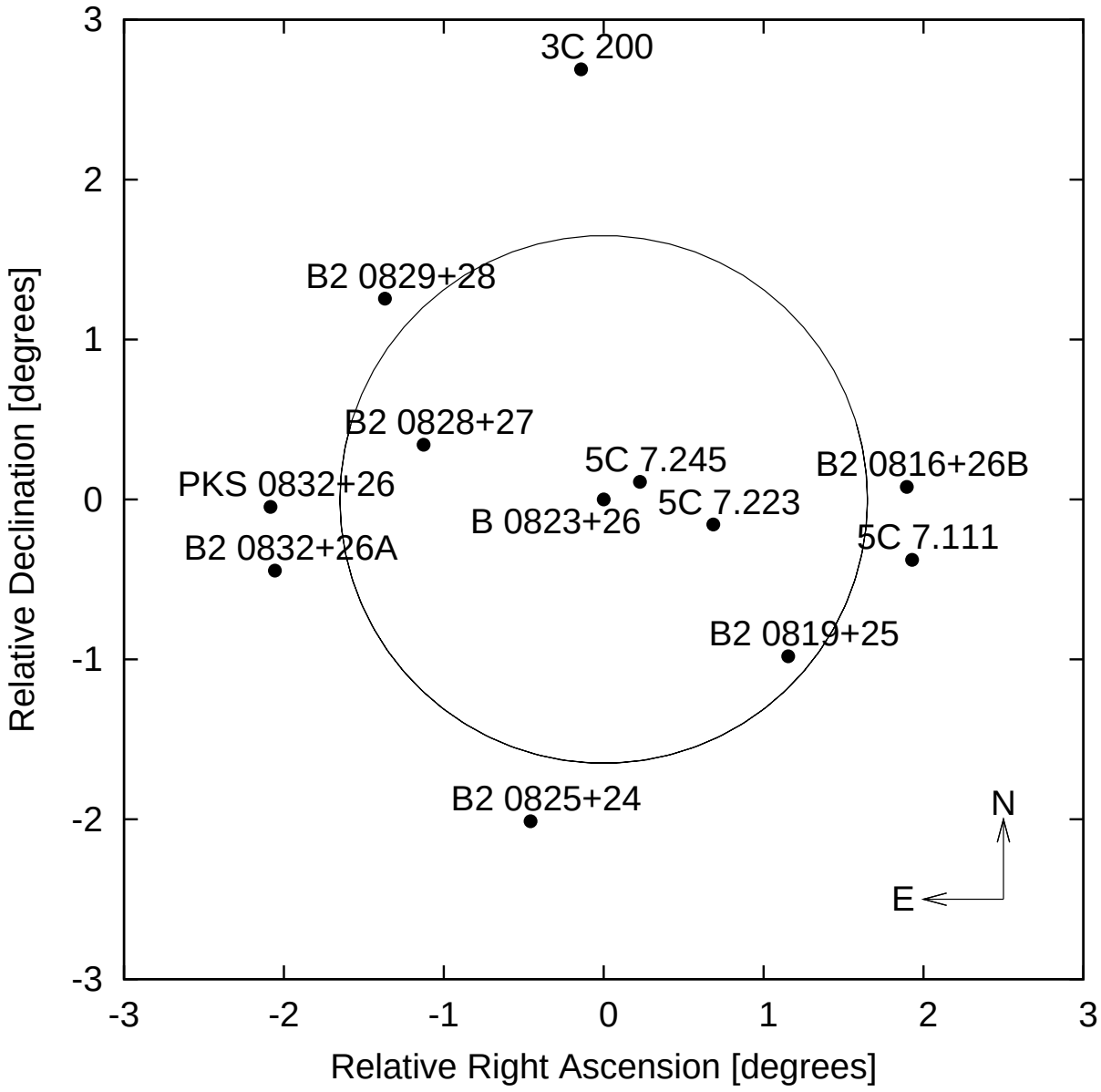


Figure 2.7: The 12 brightest sources in the B0823+26 field used for these observations. The pulsar is at the centre of the field. The circle denotes the half-power diameter of the main beam, within which there are few bright sources.

2.4.3 PSR B2045-16

After concluding that B2217 was not a productive field, summer observations focused on the new field B2045 starting in the summer of 2011. This pulsar is farther from the galactic plane but with similar right ascension to B2217, though it is much closer to the galactic bulge.

With a declination of -16 degrees, it is much closer to the equator than any previous source. This choice of field was motivated by the desire to increase the depth at every (u, v) point at the expense of reduced (u, v) coverage, as discussed previously.

There are six other pulsars with high galactic latitude and low declination that are at least as bright as our previous field PSR B0823+26 that we had identified as candidate fields. B2045 was chosen as our primary candidate because in the summer when it was to be first observed it transits at night, like B0823, when local RFI is lowest and the ionosphere is more stable.

Initial results from this field suggested the data will be useful in that the pulsar is reliably visible and nearby bright sources are not too problematic.

2.5 Summary of observing campaigns

Observations at GMRT have been ongoing since 2007, as part of the regular 5 month observing cycles scheduled by the GMRT Time Allocation Committee. These cycles run from April to September and October to March, with approximately one month of downtime between each for maintenance and repairs.

Due to the highly experimental nature of the EoR observations, each cycle saw slightly different observing techniques used on each of the target fields, with varying amounts of success. The methods and results of each are described here, and are summarized in Table 2.4.

Cycle	Dates	Pulsar	Hours	Observing mode
13	Winter 2007/2008	B0823+26	100	Continuous tracking
14	Summer 2008	B2217+47	50	Continuous tracking
15	Winter 2008/2009	B0823+26	160	Continuous tracking
16	Summer 2009	B2217+47	150	Continuous tracking
17	Winter 2009/2010	B0823+26	150	Drift scan
18	Summer 2010	B2217+47	150	Drift scan mosaic
19	Winter 2010/2011	B2045-16	16	Drift scan mosaic at transit
20	Summer 2011	B2045-16	32	Drift scan (3 degrees), mosaic, 3 hrs near transit
21	Winter 2011/2012	B0823+26	100	Drift scan (both 3 and 6 degrees), mosaic, 3 hrs near transit

Table 2.4: Summary of observing cycles at GMRT, including the time allocated and the main observation method used. See section 2.5 for more detail on the methods and results of each.

2.5.1 Cycle 13 (Winter 2007–2008)

This cycle marked the first large time allocation. B0823+26 was tracked throughout each night of observing for 11 nights of between 8 to 10 hours each. The analysis on six of these nights (December 10, 11, 14, 16, 17, and 18), totaling about 50 hours, is complete and comprises the principal data set used in later chapters of this thesis.

The days which did not produce useful data typically failed because of a problem with synchronization of the ADC board sampling clocks. During this cycle, each acquisition node had its clock reset at the beginning of each scan. While most nodes receive only L or only R polarizations, one receives both L and R from the S04 and S06 antennas. This common acquisition node is crucial for setting the correct timing between the L and R phases of all antennas, which meant that if both S04 and S06 were not functioning, phase calibration would fail. This was fixed in future cycles by having all nodes share a common trigger synchronization signal derived from a GPS time standard.

2.5.2 Cycle 14 (Summer 2008)

This is the first cycle in which B2217+47 was observed with a total allocation of 50 hours, spread over 5 nights. Of these, August 21 and 28 were successful and contain about 8 and 5 hours, respectively. Unfortunately the nights of August 20 and 23 were both dominated by severe RFI. The last night, August 22, contains only 3 hours of data. Given the limited size of this data set, these observations were useful mostly in checking the feasibility of using B2217+47 as an EoR field, but not yet for power spectrum estimates until additional days could be observed.

2.5.3 Cycle 15 (Winter 2008–2009)

Cycle 15 returned to the winter field B0823+26, using the same continuous tracking technique. In February 2009, we implemented a new RFI subtraction technique, whereby

6 antennas (E03, E05, W02, W04, S02, and S03) have their feeds pointed at the central square and are used as RFI reference antennas. This would allow the response to RFI on the other antennas to be solved and fully subtracted from the sky signal. This procedure was tested and shown to be viable, but it has not been used in the analysis in this work. Unfortunately, a software bug in the EoR software correlator corrupted the incoming raw data in every other frequency channel, making analysis of data from this cycle unrealistic. This bug was not present in earlier cycles.

2.5.4 Cycle 16 (Summer 2009)

This was the second cycle of observing B2217+47, and the last cycle to make use of continuous tracking. The same array configuration as Cycle 15, with a 6 antenna sub-array for RFI monitoring, was used. Unfortunately, because of the long time delay between recording and fully analyzing the observations due mostly to the difficulty of storing and transporting the large quantities of data necessary, the bug in the EoR software correlator which appeared in the previous cycle had not yet been discovered or repaired before this cycle ended. As such, no useful data was recorded during this cycle.

2.5.5 Cycle 17 (Winter 2009–2010)

Given the success with analysing earlier observations of this field, this cycle continued to use B0823+26. For the first time, draft scanning was used.

Data from February 8 was lost in transport from India to Toronto due to a failed tape. Of the nine remaining days, only two (February 5 and 25) were able to be calibrated successfully. With only two separate data sets, it is possible to make one cross-correlation to attempt to measure the power spectrum, but the noise on this measurement would be very high.

With the exception of February 26, most days do not show coherent phases, indicating that the polarization calibration will not work. Nonetheless, future analysis could include

this data without having the polarizations calibrated, with the caveat that this would introduce another source of error.

2.5.6 Cycle 18 (Summer 2010)

In addition to drift scanning, we began mosaicing in the north-south direction in this cycle. It was noted during this cycle that several nights were dominated by RFI, overwhelming the pulsar signal, making calibration difficult. Nonetheless, the pulsar was clearly visible on about three out of nine days (July 30, August 3, and September 1) and calibration was successful. On September 2nd, the calibration was marginally successful.

This was the last cycle to make use of the B2217+47 field. In addition to the difficulty of seeing the pulsar through strong RFI, it seemed the pulsar itself could not be identified reliably in lag space, even when RFI was not present. It was also determined that the foregrounds from the galactic plane and numerous nearby bright radio sources did not make it a likely productive field to access the dim EoR signal. In particular, sidelobe effects from Cas A were evident as a dominant source of noise.

2.5.7 Cycle 19 (Winter 2010–2011)

As in previous cycles, mosaic scanning was used with three pointings of $\pm 1/2$ of the primary beam width. In contrast to early cycles, the observations were limited to within approximately 2 hours of transit, which restricts the w range and thereby reduces the magnitude of the w-term corrections needed and enabling more precise foreground removal. Additionally, this is the first cycle that made use of a near equatorial field, B2045-16. As discussed, this limits the (u, v) coverage in the hopes of increasing sensitivity to the power spectrum. Limiting observations to the same LST range each day for the first time further deepens coverage with fewer hours. Since it is also up during the night in the summer but further from the galactic plane, it was hoped that this source could replace B2217 during the next cycle's summer observations.

It was unclear how either limiting the LST range or using an equatorial source would impact the later calibration steps, which were designed with full night tracking observations in mind. In order to pilot these new strategies, this cycle saw a smaller time allocation of only 16 hours compared to 150 hours in previous cycles.

Unfortunately, clock errors in the GSB hardware caused the correlator nodes to fall out of sync and limit the effective bandwidth to less than 2 MHz at the top end of the range. This was enough to confirm that B2045 was visible and would be a suitable candidate for longer observations, but is not enough bandwidth for EoR analysis.

2.5.8 Cycle 20 (Summer 2011)

With the preliminary success of the B2045 field, despite the technical problems, this cycle continued with the same observing strategy of short observing scans at a limited LST range near transit, mosaicing between three pointings north and south of the source declination. Unfortunately scheduling such small observing blocks proved difficult, and while most of the observations spanned the same 3 hour LST range, this range did not overlap with the observations from the previous cycle and begins approximately two hours after transit. Nonetheless, about six out of the thirteen days of observing (June 17, 18, 19, 21, 22, and 26) were of good quality.

This cycle unfortunately saw several disks fail, both on individual nodes of the GSB and on the manager nodes where data was staged before being written to tape. Approximately 30% of the data was affected. In cases where a GSB node failed, this usually resulted in a 1 MHz chunk of the total bandwidth being lost for some scans in a night. This could still provide useful data, but would require significant extra bookkeeping to track where the gaps are and provide dummy data to the analysis programs where needed. In many cases, because of disk space constraints and a tight observing schedule, this data had to be discarded in favour of new, complete, observations.

This cycle also saw the addition of a large team of undergraduate students working

on isolating and correcting local sources of radio frequency interference. For details, see Chapter 3. While this involved many additional hours allocated to our team during otherwise unused slots in the schedule, these observations focused on terrestrial interference sources and thus did not provide any astronomical data.

2.5.9 Cycle 21 (Winter 2011–2012)

Observations in this cycle moved back to our traditional winter field, B0823+26. This would be the first time mosaic drift scans were used on this target, hopefully providing a reference to quantify the benefits of the new observing technique compared to the tracking scans from 2007, and as a cross-check for results from the analysis of B2045-16 using these same techniques.

However, in contrast to the observations of B2045-16, we increased the mosaic sampling resolution. Experience with 21 cm intensity mapping at GBT by some members of our collaboration had shown that fine sampling of the beam is very important to allow to correct for the shift in frequency dependence of (u, v, w) coverage and improve foreground removal. Sensitivity to the cosmological signal is maximized when the radial wave number is comparable to the transverse, which for the shortest baselines is at about 3 MHz. This 2% change in frequency corresponds to approximately a 2 wavelength change in baseline, or one tenth of the primary beam aperture in the (u, v) plane. With this motivation, we sampled in the north/south direction in steps of one tenth of the primary beam.

For the first 12 nights of observations, spread from November 1st to January 8th, the drift scans were 3 degrees wide as normal. However, from January 21st to March 18th, for 18 nights this was extended to 6 degrees, to further probe the edges of the primary beam and allow better sidelobe rejection.

As in the previous cycle, scheduling so many small observing blocks over the same LST range each night proved difficult. In the first part of the cycle the LST range tended

to drift by up to an hour between observations. Typically, a single LST range from one 20 minute drift scan was observed about 3–6 times before the observation window moved past it. With 11 different declinations to cycle through, no single LST range was observed at every pointing. This improved somewhat in 2012 during the 6 degree scans, where the number of pointings was decreased again to five, and at least nine nights overlapped for four hours in LST.

Analysis of data from this cycle was being done primarily by Hans Nguyen, an undergraduate research assistant, and is not included in this thesis.

2.6 Alternative observing strategies

The observing strategies we have discussed are by no means the only approach. One limiting problem has been very bright sources in the far sidelobes, for example Cas A. Reducing our sensitivity to these types of foregrounds has been one of the primary motivators for some alternative strategies. Though these have not been a part of our normal observing campaign, we briefly discuss a few here for completeness.

In 2011 we explored the possibility of using Earth occultation, taking advantage of the fact that when our target field is close to the horizon, half the sky is blocked by the Earth itself. By cross-correlating opposing observations from rise and set, sources far from the pointing centre should not correlate, reducing the response to celestial interference. The GMRT would be in a unique position to implement Earth occultation mapping since other EoR telescopes without steerable dishes have limitations observing close to the horizon. However, there are unavoidable effects such as observing through a thicker and more variable ionosphere near the horizon, with many hours separating the correlated fields. Preliminary tests showed that variability from day to day was higher during rise and set than for full nights, suggesting such a mode of observing would increase the noise in a power spectrum measurement more than, say, correlating fields at transit.

A similar strategy would have only correlated disjoint sections of the beam. During a long integration, the (u, v) position of a baseline drifts, and the visibilities of objects far from the pointing centre vary rapidly. We can split the data into four sets, separated by one quarter of the primary beam, and cross-correlate every other set. Objects more than two primary beams apart have their phases effectively randomized, while objects inside the primary beam are not strongly affected and would still correlate well. An alternative, but related, method, is to grid in the (u, v) plane with a resolution smaller than the primary beam, and cross-correlate only adjacent cells. Both methods severely limit the number of cross-correlations that are available, however, again degrading the power spectrum measurement by reducing the signal-to-noise ratio.

Finally, we had also suggested making an all-sky map to create a sky model and subtract this from the data. By mid 2012, it was determined that one of the dominant remaining obstacles was celestial structures far from the primary beam, particularly the galaxy itself, which contaminates the short baselines needed for the 21 cm signal. An all-sky template could be used to remove this residual structure. This could be accomplished in a relatively short amount of time by rotating the 150 MHz feed on each dish out of focus and toward the sky, taking advantage of the full ~ 100 degree field of view, which enables surveying the entire sky (and all (u, v) configurations) in four 24 hour scans, with a higher sensitivity than is present in the focused sidelobes. Both the all-sky data and the EoR data could then be phased onto different parts of the galaxy, the individual antenna gains solved, and then subtracted from the EoR data. This strategy was included in our proposal for Cycle 22, which was awarded approximately 100 hours, but was not implemented.

2.7 Conclusion

In this chapter we have laid out the basic technical details necessary for understanding the GMRT-EoR observations, and have described the history and progress of our observing campaign.

After cycle 21 the observational side of the GMRT-EoR project has been led by J. Richard Shaw and Liam Connor, using a new technique in which drift-scans are extended for a full 24 hours at constant declination to build up a complete annulus on the sky. This can be decomposed into spherical harmonics, and m-modes containing the highest signal-to-foreground ratio can be isolated (Shaw et al., 2013). The details of this analysis are outside the scope of this thesis, but will likely be the principal strategy used by the GMRT-EoR team in the future.

This project was allocated over 900 hours of observing over the last five years. Some significant obstacles to efficient observing have been technical difficulties with both the changing hardware environment at GMRT as the GSB system has been built and upgraded, as well as software issues with our custom EoR correlator. Additionally, as we change the observing strategy to adapt to our improved understanding of the challenges we are facing, older data can sometimes become obsolete. The remainder of this work focuses only on about 50 hours of observing from 2007, but builds an analysis framework that should be useful for improving these results with more data. Further refinements to the data analysis techniques described in this chapter will be discussed in Chapter 6.

Chapter 3

Radio Frequency Interference Mitigation

One of the limiting factors for measurement of the EoR signal is broadband radio frequency interference (RFI), which can dominate the astronomical signal at 150 MHz. Reducing the RFI is critical to obtain high sensitivity observations of the sky. After reviewing the RFI environment, we discuss two approaches to accomplish this: physically identifying RFI emitters and removing them, and a singular value decomposition (SVD) to model and remove broadband RFI modes from the data. The component programs described in this chapter to isolate RFI were written by Ue-Li Pen, and combined into an analysis pipeline by Julia Odegova and myself.

3.1 The RFI environment

One of the motivations for building the GMRT in rural India was that it promised very low levels of RFI compared to other sites. A survey conducted at the GMRT site before construction showed that most bands were relatively free of RFI, though it was present about 20 per cent of the time around 150 MHz (Swarup & Venkatasubramani, 1991). These measurements had a sensitivity level of $-150\,\mathrm{dB\,W/m^2}$ over a 10 MHz bandwidth,

which is significantly higher than astronomical signals. It was estimated that RFI as low as $-200\,\mathrm{dB\,W/m^2}$ should be considered harmful. Though it is expected that RFI is still lower in India than similar sites in more developed areas of the world, the amount of RFI will have changed substantially over the 20 years since those measurements were made.

LOFAR, which primarily uses an automated flagging routine to excise RFI, has reported that approximately 3% of data around 150 MHz is obscured by RFI, though some frequency channels can be tens of per cent (Offringa et al., 2013). This is comparable to the fraction of narrowband RFI identified at GMRT, in spite of RFI measurements in the Netherlands showing signals as high as $-120\,\mathrm{dB\,W/m^2/Hz}$ at these frequencies (Bentum et al., 2010). Several reports have been produced for the SKA candidate sites in South Africa and Australia, the former of which is also being used for PAPER. While the RFI environment changes significantly for the remote stations given the very long baselines, the core sites have reported fluxes around -200 to $-180\,\mathrm{dB\,W/m^2/Hz}$ in Australia and slightly less in South Africa (Millenaar, 2011). Similar to LOFAR, RFI occupies a few per cent of the spectrum around 150 MHz both sites (SKA Expert Panel on Radio Frequency Interference, 2011). Neither LOFAR nor SKA expects these levels of RFI to be problematic.

RFI can come in numerous flavours. The Indian Ministry of Communications and Information Technology[1] maintains the National Frequency Allocation Plan (NFAP), which specifies the allowed usage of the entire wireless spectrum. The frequencies used for this project (139–156 MHz) are allocated for a wide range of activities: amateurs, satellite communications, car rallies and sports activities, wide-area paging, the remote control of construction equipment, cordless telephones, film shooting, and radionavigation. Only the 151.5–153 MHz range is restricted in the GMRT area, within a radius of 400 km.[2] In addition to regulated usage, there is also unintentional interference from sources such as

[1] `www.wpc.dot.gov.in`

[2] NFAP Channeling Plans; `www.wpc.dot.gov.in/DocFiles/NFAP/NPCHP.doc`. Retrieved May 31, 2013.

power lines, which we will see has a much larger contribution to the RFI environment.

Regardless of the source, RFI shows up in the data in one of only a few ways, and different mitigation strategies are useful depending on the particular character of the interference. For interference which lasts only for a short burst of time, or which exists over a very narrow frequency range, simply flagging the data is sufficient, as there is usually enough unaffected data to remain scientifically useful. Routines to flag such data can be automated effectively or done by hand if the volume of data is small.

Broadband interference poses an additional challenge. Though the affected data can be flagged and removed, this is usually undesirable since large fractions of the data can be lost. Some sources may contaminate the entire band for long periods of time. In section 3.4, we will discuss a singular value decomposition (SVD) that can be used to remove broadband RFI from the data while leaving the astronomical signal in place. This has the added benefit that SVD modes can be used to localise RFI sources in real space, opening up possibilities for preventative RFI mitigation programs. While SVD based RFI flagging has been implemented elsewhere (e.g., Offringa, 2010) we make use of code written specifically for the GMRT-EoR project.

3.2 Identification of physical RFI sources

Flagging data contaminated with RFI during the data analysis carries with it the risk of also removing the cosmological signal. Ideally one would prevent RFI from occurring during the observation at all by identifying and correcting the physical sources. We can take advantage of the fact that with many widely spaced antennas, the GMRT is able to make images of the near-field RFI environment, and create maps of bright sources near the GMRT itself.

3.2.1 Singular value decomposition of RFI modes

RFI sources can be identified by performing a singular value decomposition (SVD) on the raw visibilities, before fringestopping onto the sky, and identifying the modes likely due to RFI.

We first treat the visibilities as a matrix $\mathcal{V}$ with each row being the visibilities at every correlation and frequency at a single time. If we denote the number of timestamps as n_t and the number of visibilities at each time n_v, $\mathcal{V}$ is an $n_t \times n_v$ matrix. Similar to an eigenvalue decomposition, an SVD factors $\mathcal{V}$ into

$$\mathcal{V} = \mathbf{USW}^\dagger \tag{3.1}$$

where $\mathbf{S}$ is a diagonal matrix of the singular values. The matrix $\mathbf{U}$ is an $n_t \times n_t$ matrix containing the left singular vectors, and W is an $n_v \times n_v$ matrix containing the right singular vectors. In practice, we calculate the eigenvectors and eigenvalues of the $n_t \times n_t$ matrix $\mathcal{V}\mathcal{V}^\dagger$ (`svd_uvec`), which correspond to the left (time-wise) singular vectors of V and the square of its singular values, respectively.

RFI sources are stationary with respect to the antennas, while the sky rotates, meaning that they are relatively constant in time and will contribute to larger eigenvalues. The 50 largest eigenvalues and their corresponding eigenvectors are saved to disk. For the one hour scans from 2007, this is less than 1 per cent of the total number of modes. These time-wise eigenvectors can be used to examine the behaviour of a particular mode as a function of time. Examples of two such modes are shown in Figure 3.1.

3.2.2 Mapping RFI modes in real space

A second program (`svd_vvec`) computes and saves the right singular vectors of $\mathcal{V}$, which gives the amplitude of the eigenmodes for every correlated antenna pair and frequency. This is necessary for seeing the response of each antenna, from which we can get geographical information.

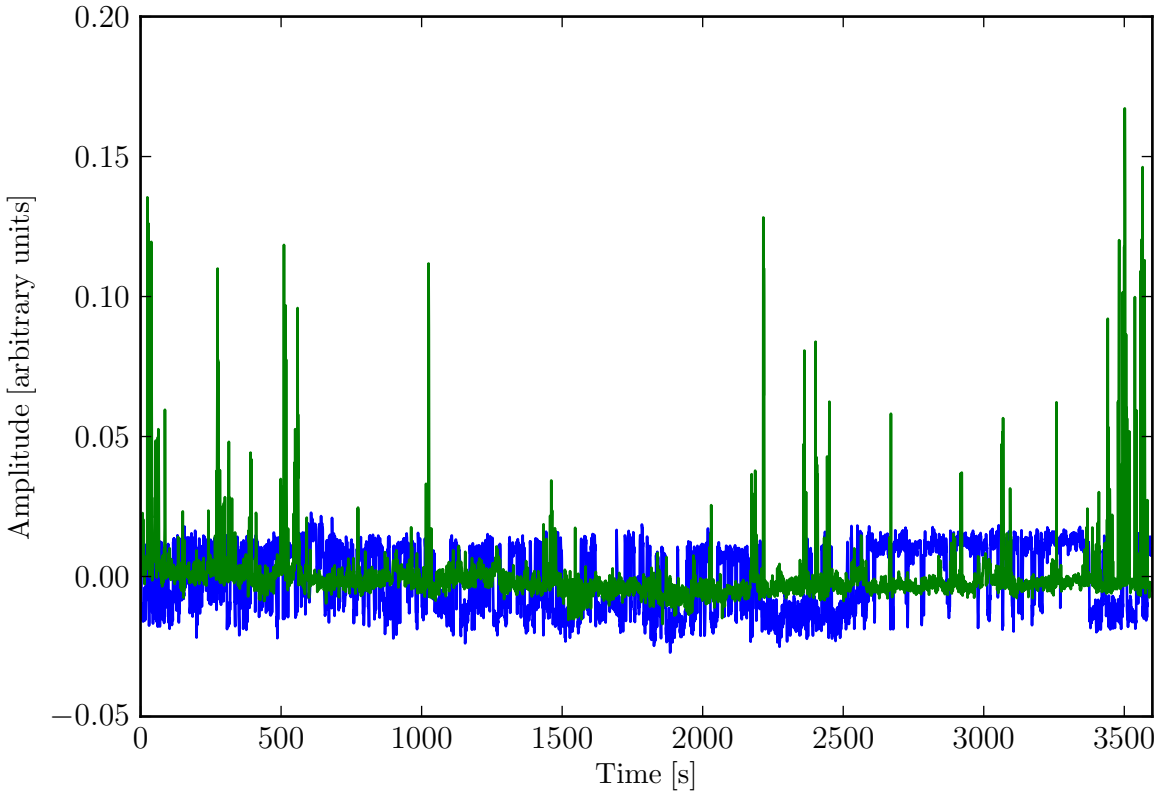


Figure 3.1: The eigenvectors of two candidate sources. Since these are the time-wise singular vectors of the visibility matrix, they indicate the behaviour of each source as a function of time. The blue line is from the strongest source seen in this particular data set and is discussed under the heading "Mode #1" in section 3.3.1. The green line comes from Mode #9. Though the latter has a higher maximum amplitude, it is intermittent. The first source is present at a lower amplitude but is always on, and thus corresponds to a larger singular value.

First, we define a physical area around our reference position which we are interested in exploring. For a pair of antennas, we calculate the delay (or lag) in arrival times of radiation from a hypothetical source at all positions in this area. Having already identified candidate RFI eigenmodes in the data, we reorganize the right singular vectors into a matrix of frequency and correlation, and take the Fourier transform along the frequency direction to get the amplitude of the mode as a function of lag and baseline (`rfimode`).

To create a map of the RFI mode, we step through every position in the area around GMRT and sum the amplitude of the RFI mode for all correlations of the GMRT antennas according to the lag appropriate for that baseline, calculated previously (`rfiposlag`). If an RFI source is located at a particular point, the amplitude of the RFI mode will be large for all baselines that saw the source, creating a bright pixel on the map.

Candidate sources detected by a single baseline appear as a hyperbola defined by the line from which the delay to each antenna is equal. When a source is detected by many baselines, the corresponding hyperbolas intersect at a single point in the image. An example of such a near-field image is shown in Figure 3.2. If a single SVD mode accurately models a single RFI source, there will be a single bright point. More generally, and especially as one looks at weaker modes, several RFI sources may mix to create multiple points. While the software pipeline automatically reports the brightest pixel for each RFI mode, each image needs to be checked visually to assess whether it may be a real single source, a combination of multiple sources, or noise. These images, with every pixel mapped to a real physical location, become the "RFI maps" which can be used to seek out the physical sources of interference.

The RFI maps are typically calculated to 20 m per pixel resolution. This is slightly more coarse than the accuracy of handheld GPS devices, and about twice the distance light travels in the 30 ns sampling time of the acquisition boards on the GSB. Maps of this resolution and 40 km on a side take less than a minute per mode to generate on

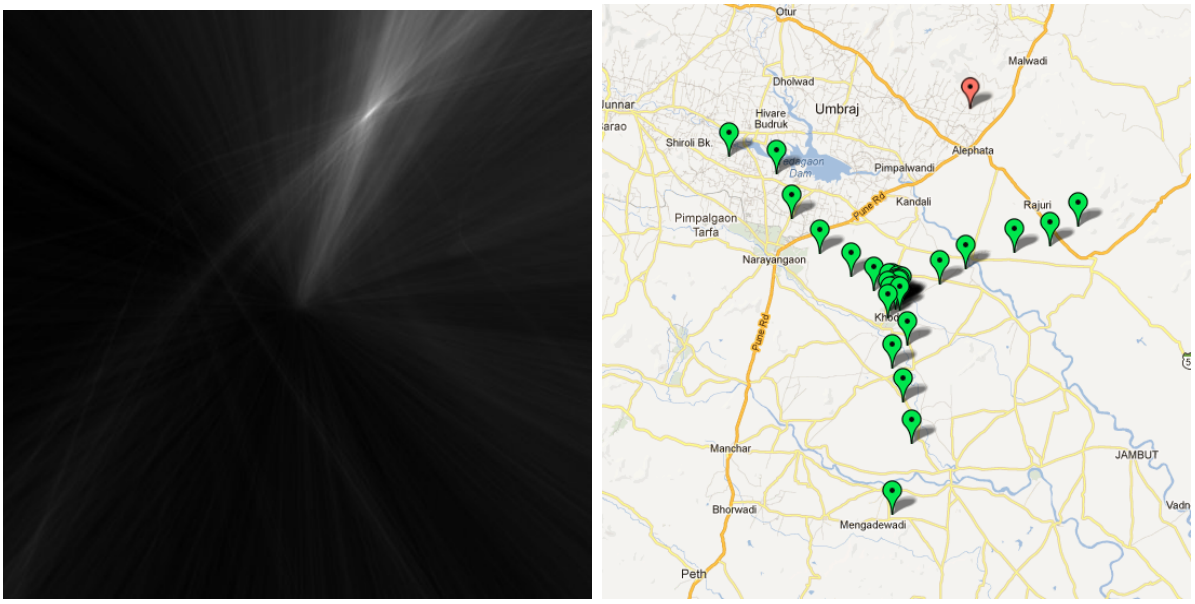


Figure 3.2: The left image shows near-field image, measuring about 40 km on each side and centred on GMRT. An RFI source identified through large SVD modes is clearly visible as a bright spot in the upper right. A map of GMRT covering the same area as in shown on the right. The green pointers indicate antenna locations, and the red pointer corresponds to the GPS position of the source identified on the left. This source was investigated and is described under the heading "Mode #14" in section 3.3.1. The map was generated with Google Maps. Map data © 2013 Google.

a single desktop computer. In practice, RFI sources are only located to within about 50–200 m of their actual location. This is likely due to such effects as reflection causing sources to be picked up at different delays, phase changes as the antenna feeds move when tracking a source on the sky, and mixing of several sources into a single SVD mode. Often there are cases where a single bright patch is evident but spread over a line of possible locations. This is particularly common for sources further from the central square, where the procedure of triangulating with light arrival times eventually becomes degenerate along the radial direction. Extending the maps further from GMRT will not localise sources that are further away because of this degeneracy. Nonetheless, even 200 m accuracy is often enough to pinpoint single unambiguously radiating sources.

3.2.3 Identifying physical RFI sources

Once a candidate source is chosen and a GPS coordinate identified, the challenge is to identify the physical source of radiation. Given the rural nature of the area around GMRT, navigating to arbitrary GPS coordinates itself can be challenging. Google Maps is extremely useful for providing satellite imagery, which includes the many dirt roads, footpaths, and landmarks that are not captured in the regular road maps. Printing satellite maps for the entire region is essential for navigating to these remote locations.

Once in the vicinity of the GPS coordinate, we use a commercially available handheld Yagi-Uda (or simply "yagi") antenna designed for 146 MHz with three elements spaced to optimize gain in a single direction. This is attached to any portable AM radio turned to the same frequency. Radio sources are heard as static. Since the yagi has directional sensitivity, the static will be louder when pointed towards an RFI source. While different types of sources may make different noises (e.g., "popping" or "sizzling"), an increase in volume alone is usually enough to indicate a candidate source. Care should be taken to avoid interference from other equipment, such as a handheld GPS device, which can itself be a weak RFI source. Figure 3.3 shows an example of the audio heard from a potential

RFI source as the yagi direction is varied.

Since directional uncertainty from the yagi alone is tens of degrees in either direction, sources usually need to be isolated by triangulation. As one approaches an RFI source it often begins to overwhelm the AM radio, at which point one can either suppress the signal with a squelch circuit on the radio or switch to a smaller antenna. A small rubber ducky antenna is useful when trying to differentiate between several closely spaced sources which the large size of the yagi would make impossible. At close range, the r^2 dependence is steep enough that an omnidirectional antenna such as this is useful for discriminating proximity to a radiating source.

For sources with a high enough duty cycle, it is relatively easy to simply follow the loudest source on the yagi to identify an RFI source. Unfortunately one of the biggest obstacles in identifying RFI is that sources rarely radiate constantly. Sources such as loose or faulty wiring may be very intermittent. GMRT engineers have recommended using an ultrasonic microphone to detect wiring faults, but our experience has been that this technique can identify a huge number of sources that are not radiating at 150 MHz at any significant level.

Additionally, the practice of "power shedding" in the rural region around GMRT, where villages may only be supplied with power for a fraction of the day, means that there is no guarantee that a real source is still being powered and radiating when the area is investigated. It has been suggested that this could be used to our advantage to limit RFI through clever scheduling of observations, but the power shedding schedules are irregular and the geographical areas which are without power at any time are much smaller than the range of the GMRT antennas. Despite this difficulty, this procedure has been successful and several of the brightest RFI sources have been identified and removed since 2008. Examples of such sources will be discussed in section 3.3.

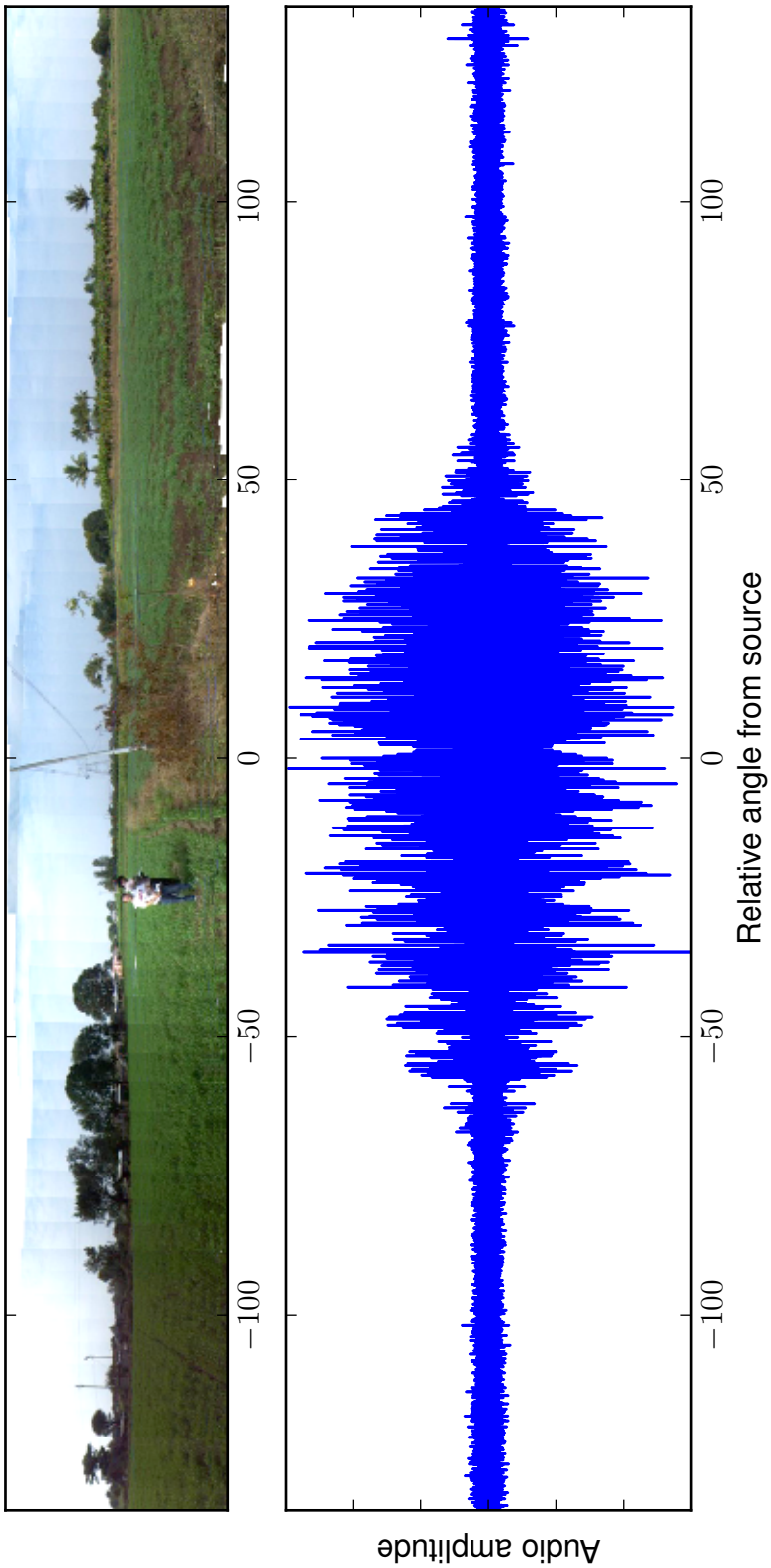


Figure 3.3: To demonstrate identifying an RFI source using a handheld radio, the audio output was recorded while sweeping the yagi antenna around 360 degrees near a candidate source. The panorama in the top panel was extracted from the same recording. The audio waveform in the lower panel shows the amplitude as a function of angle, covering the same range as the panorama, indicating where the yagi was pointing. While only 270 degrees are shown, the remaining 90 degrees had audio levels similar to the edges shown here. The audio gets significantly louder when within about 50 degrees of the source.

3.2.4 Calibration with an artificial RFI source

To better calibrate the RFI location programs, an artificial source can be put in the field, localized with an SVD, and then compared with the location of a candidate RFI source. The source used is a one watt noise transmitter which switches on and off with a period of 1 Hz. This strong, regular signal shows up easily in SVD modes after only a few integrations. This can be used to narrow down the position of a poorly localised source, or to confirm that the correct source has actually been identified. Occasional misidentifications are likely due to the intermittent nature of the sources; there is no guarantee that a radiating source identified when in the field is the same source that had been radiating when captured in the RFI maps.

Use of an artificial source can also confirm that the routines are working as expected. If well localised sources do not appear in any of the largest SVD modes, there may be a problem with the RFI mapping routine, the SVD, the software correlator, or GMRT itself.

Since the beginning of this effort, calibration has been improved to locate sources within about 100 m. Sources could in principle be located with a precision of 10 m, but the accuracy of available maps, GPS equipment, and other factors limit the real-world precision.

3.3 Known RFI sources

Candidate RFI sources identified so far include transformers, power line junctions, and loose wires in contact with power lines. Though RFI mapping and localisation efforts have been a regular feature during each of the GMRT-EoR observing cycles, most sources identified have come either from the first concentrated efforts in late 2008 or in an expanded program in the summer of 2011. The most note-worthy sources from both cycles are discussed here.

In February 2009, GMRT began a collaboration effort with the Maharashtra State Electricity Distribution Co. Ltd. (MSEDCL), which controls the power transmission lines in the area around GMRT, to assist us in removing local interference sources. A letter detailing known problems can be sent to the Maharashtra State Electricity Board (MSEB) in the hopes that they will choose to fix it, but in practice this route is indirect and slow. Occasionally we have made contact with local power line workers who are able to make repairs immediately.

3.3.1 Sources identified in 2008

Mode #1 (19.147676N, 74.034653E)

By far the brightest source ever identified, a candidate source at this position appeared in the maps of over 30 of the 50 largest singular values. It was conclusively identified as a long wire hanging on a 132 kV high tension power line running parallel to the state highway just north of GMRT. This source was bright enough to be audible from at least 2 km away with the yagi antenna, and was emitting in both ultraviolet and infrared.

A photo of the wire is shown in Figure 3.4a. The wire was reported to the MSEDCL, and they were able to shut down the power running through this line and remove the wire in February 2009.

Mode #9 (19.171797N, 74.085335E)

This source appeared to be due to wires hanging from an unused telephone pole, immediately below a 500 kV high tension line on the north side of the road at this coordinate. As far as was visible there was no power source associated with the telephone pole. It is possible that the high voltage lines were inducing the signal in the hanging wires, but this could not be confirmed. Neither of the high tension towers on either side of the field were audible sources.

The closest possible source to the exact candidate GPS coordinate was a transformer located behind a small house on the south side of the road. However, when this site was inspected there the power to this transformer was off, so we were unable to determine if it may have also been a source.

Despite uncertainty about the identification, the telephone pole was removed from the ground by GMRT staff in early 2009 and this position has not shown up as a candidate source in subsequent observations.

Mode #13 (19.037518N, 74.034271E)

A source in the area was transmitting very intermittently, audible only in very short bursts. This type of source is very difficult to localise because the signal is not audible for long enough to move the yagi and determine the direction of the source. A bank of transformers about 30 m north of the coordinate, on the north side of the road, was a potential source, but this was inconclusive.

Mode #14 (19.209597N, 74.096581E)

Similar to the source identified from mode #1, this was identified as a small wire hanging on a 500 kV power line, with one end wrapped around a rock as though it may have been thrown onto the wire purposefully. In this case, the identified source was a full 120 m east of the candidate GPS source, but was unambiguously radiating at a fairly high level. The wire was removed in February 2010.

Mode #15 (19.181877N, 74.120018E)

Two high tension transmission towers in the area of this coordinate were possible sources. The closet to the coordinate, approximately 20 m west, is a T-junction of two high voltage lines. The second, approximately 300 m northwest, was also radiating at similar levels. Aside from one small bird's nest on the tower to the northwest, neither tower showed any

obvious fault that would be causing the radiation. We hypothesized that something on the power line between them might be to blame, but a visual inspection showed nothing out of the ordinary.

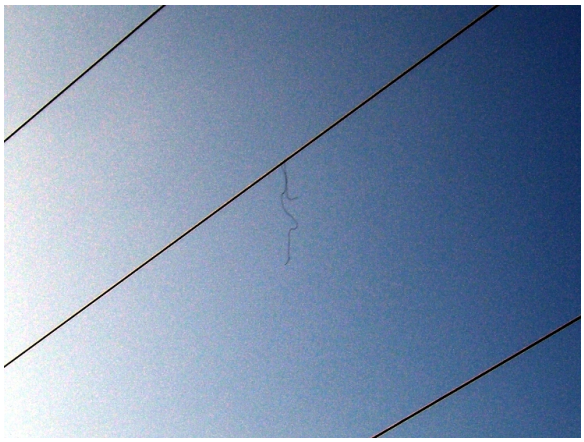
Mode #17 (19.088457N, 74.072952E)

This area showed several candidate sources, all of which were only radiating at very low levels and intermittently. The best candidate was a small pump located approximately 150 m west of the coordinate. However, several nearby transformers also seemed to be radiating; one 75 m north, and another 225 m west. The western transformer had one wiring glowing visibly, indicating a possible fault causing the radiation. Both had well localised signals but were only radiating at very low levels.

To help improve the fidelity of our candidate location, the artificial noise source was used from a location 10 m south of the pump. The resulting localisation of the artificial source was far from the original candidate coordinate, suggesting that the pump was not the original source. Sources in this area continued to show up in later years, but nothing conclusive has been identified as the cause.

Mode #33 (19.108976N, 74.075623E)

A transformer approximately 100 m northwest of the coordinate was unambiguously radiating at low levels. As is usual with this kind of source, it was unclear what the fault on the transformer may have been. These types of sources can be reported, and either GMRT or state staff can check the transformers for lose or worn out connections and repair them, but it is unknown whether this is effective at correcting these types of problems.



(a) Wire on 132 kV line

(b) Transformers

(c) Power line junction

(d) Generator

Figure 3.4: Photos of various RFI sources typical of those found within about 20 km of GMRT. The wire shown in (a) is identified as Mode #1 in section 3.3.1 and is the strongest source yet identified. The transformers and power line junctions shown in (b) and (c) are typically difficult to isolate due to having many components that could radiate. The generator in (d) represents a new type of source present, found in 2011, on the GMRT campus itself. *Photo (d) by Nidhi Banavar; all other photos by G. Paciga.*

3.3.2 Undergraduate RFI program in 2011

In 2011, a partnership between the GMRT-EoR team, Woodsworth College, and the Department of Astronomy, with financial support from the University of Toronto's Curriculum Renewal Initiative Fund (CRIF), designed an undergraduate research course which allowed five students—Joshua G. Albert, Nidhi Banavar, Mark Kuiack, Connie Lien, and Hans Nguyen—to visit GMRT for two months and take part in our observations and RFI mitigation efforts. With at least six active scientists on site at all times, this represents the most concentrated RFI removal effort ever at GMRT.

By having a continued presence at GMRT, it was hoped that the process of making RFI maps, going into the field to identify the sources, and reporting these sources both to MSEDCL and to the GMRT engineers could become a routine practice. Woodsworth College provided the infrastructure of an established study abroad program to which students could apply, get financial support, and earn credit towards their degrees. Funds from the CRIF were used to purchase equipment in the form of laptops, GPS devices, radios, antennas, and data storage space.

For the first few weeks, the students were trained on how to use GMRT to make both EoR observations and targeted RFI observations, run the SVD pipeline and make the RFI maps, and strategies for using the handheld radios to identify interference sources. Known uncorrected sources from previous cycles were used as test cases to confirm that the students could identify sources. Once established, the students were to continue trying to identify new sources under the supervision of different members of the GMRT-EoR collaboration.

This occurred in parallel with the observations of Cycle 20 for which 32 hours were allocated to the GMRT-EoR project. Additional time for RFI specific observing was requested during "white slots", during which no other observations were scheduled, under the umbrella of Director's Discretionary Time. Though an RFI observation need only last a few minutes, typically blocks of 2 hours were requested to allow for set up time

and coordinating the timing of artificial noise source transmissions. Given these longer blocks of time, it was also possible to experiment with alternative array configurations to improve the sensitivity to RFI, as will be discussed in section 3.5.

The undergraduate team was successful in identifying several RFI sources, including new types of sources which had been previously unconsidered, though none of the sources were as unambiguous as those found in earlier cycles. Several of the candidates are discussed here. Unless specified, most of these sources remain uncorrected.

Cell phone tower near Suzuki dealership (19,086991N, 73.965403E)

There is a large area on the west side of the Pune-Nashik Highway, south of Narayangaon, that had been identified as containing several possible RFI sources.

First visited in 2010, the most significant source was a cell phone tower near located about 260 m west of the highway. This source was used as a test case for the undergraduate group to get used to using the RFI equipment. Circling the tower with the yagi, equipment at the base of the tower was identified as the most likely source rather than anything mounted on the tower itself.

In addition to this tower, this field has one high tension power line and at least three lower voltage local power lines running north-south through it. In the area between the cell phone tower and the Suzuki car dealership about 350 m north, several power poles were identified as possible sources.

Since there were so many poles identified, we became suspicious that low level radiation was just a generic feature of all low voltage power poles, and we had picked out a few just by passing close to them. This would mean we could be guaranteed to identify sources in almost any area, especially where poles were so densely packed, whether or not they actually corresponded to candidates identified in the RFI maps. To check this, we tried observing other power poles with the yagi and found that even when immediately next to most poles, no radiation was heard on the radio.

Of the power poles that were radiating, the most likely cause was loose wiring. However, given that nothing was audible on the yagi from more than about 50 m away, it was not clear that any individual pole would have been an important RFI source.

Pomegranate farm (19.09234N, 74.01513E)

This candidate source was identified during a scan on June 24, 2011, with the telescopes pointed towards the zenith and the 150 MHz feed deliberately out of focus to increase sensitivity along the horizon. Seven SVD modes had candidate sources in the area, beginning with mode #2, with a scatter of approximately 750 m. It is unclear whether all of these candidates originate from the same source given their wide scatter. The coordinate from mode #17 was identified as the most well localised in the RFI maps. A steel power pole 330 m away from the GPS coordinate was identified as a potential interference source, possibly due to a faulty insulator, but the identification was inconclusive.

GMRT diesel generator building (19.094397N, 74.049324E)

This source was identified in a scan from July 7. This is notable because it was the first time that a potential RFI source was identified within the GMRT main campus itself. The source was investigated by Nidhi Banavar and Jeff Peterson, who heard a variable "crackling" noise while on the road just north of C03, but could not localise the intermittent source. Later, two diesel generator buildings nearby were found to emit a loud hum audible on the radio, which was localised to one of three electric meter boxes on the generators. One is pictured in Figure 3.4d.

To reduce the interference, GMRT engineers attempted to create a Faraday cage around the meters using a copper mesh, and to insulate the cables leading out from the back. The engineers reported that the ultrasonic microphone indicated that the cables were the dominant source of radiation, though it is not known how well this translates to emission at 150 MHz. The signal was observed to drop significantly after applying the

Faraday cage, but only temporarily. Though the GMRT staff reported an intention to improve the insulation around the generator cabling here and in similar locations on the GMRT campus, it is not known whether this was done or the extent to which it might have changed the RFI environment.

GMRT power station hub (19.097139N, 74.048139E)

After the above source was identified in the diesel generator building, Jeff Peterson, Nidhi Banavar, and Hans Nguyen decided to walk around the GMRT campus with a radio and see if any other sources could be found. The power station hub located at these coordinates was emitted a "beeping" sound with a frequency of about 1 Hz. Inside the building, the source was found to be two electric meter boxes. Unfortunately, these are maintained by, and would have to be fixed by, the MSEB. The cause of the emission is unknown, but a third box which had been replaced recently was not emitting. Since this source was discovered without the use of the SVD RFI maps, it is also unknown whether it may have impacted any GMRT observations. If it has, it may be below the threshold of the 50 largest singular values, or it may point to a possible hidden class of sources which are missed by the SVD mapping strategy.

Chicken farm (19.137867N, 74.145737E)

A scan from July 12 showed several candidate sources, beginning with mode #3 and including seven other modes with a scatter of 25 m, in the area around this coordinate. Three possible RFI sources were identified, all due to faulty insulators on 33 kV power lines. This same location was identified again during regular EoR observations on July 17, at a different time of day, indicating a persistent source. An MSEDCL representative inspected the site and found both bent insulators and improperly grounded wires. Though he was able to fix the ground connections, no improvement in radio emission was found. The power to the line would need to be shut down in order to fix the insulators,

so this could not be completed at the time.

Transformers near E03 (19.126883N, 74.077457E)

Several modes from July 12 and July 13 localised to an area around E03, in the middle of a grass field. A bank of transformers nearby was identified as a radiating source. Broadcasting with the artificial noise source from this location localised in the RFI maps just one pixel (20 m) away from the original candidate source, confirming that the transformers were the most likely cause.

Transformers near S03 (19.047148N, 74.056564E)

A faint source was detected from a transformer bank approximately 195 m southeast from the candidate location. When attempting to use the noise transmitter to improve the localisation, the transmitter localised to a position 120 m south of the original coordinate, and 188 m west of where it was actually located. Given the positional uncertainty this implies, it is not unreasonable to conclude that the transformer bank is the source detected in the original map.

Girish Bhor (19.055364N, 74.051738E)

Several scans identified candidate sources just west of the village of Girish Bhor. These include mode #3 from July 7, and modes #12 and #17 on July 18. A low voltage power line was identified as an RFI source. The noise transmitter, positioned underneath this line, localised to the same position as the candidate source in the maps, suggesting that it is the same source. However, the exact cause of the interference was not identified.

3.3.3 RFI from GMRT antennas

Ue-Li Pen and Jeff Peterson identified additional antenna-generated RFI on the shortest three baselines, namely the C05-C06-C09 triangle. This triangle has the highest EoR

sensitivity, but also the highest susceptibility to RFI, since it has the slowest fringe rotation rate, making it difficult to separate the sky from stationary RFI sources. Current data suggests that substantial antenna-generated RFI is present in the data. Since the dish meshes are not 100% opaque to radiation, it is feasible that the feeds are picking up radiation originating from below the dish, in the antenna bases.

These three antennas were investigated by Jeff Peterson, but a conclusive source was not identified. It is unknown whether this type of interference may be typical of all antennas, or if a problem from a single antenna was being picked up by the other antennas in these baselines.

On June 20, 2011, during regular RFI investigations in the area we identified a radio signal coming from the S03 antenna site, despite the fact the antenna was shut down and locked in position due to high winds. This has not been identified for any other antennas.

Nonetheless, the possibility that RFI is being generated from either the antenna electronics or the pointing motors motivated changing our observing strategy to use drift scanning, with the antennas in a fixed position, as discussed in Chapter 2.

3.4 Post-observation RFI removal

It is unfortunately impossible to remove all physical sources of RFI. Several approaches can be used to mitigate RFI further in post-observation analysis, which have been combined into a largely automated pipeline featuring a similar SVD process as used for physically locating the RFI.

Narrow line interference is first removed by masking points in each frequency bin with an intensity above a certain threshold. At this early stage, data above 8σ on a Gaussian scale in each frequency bin at every timestamp and correlation are flagged (`debias`, or `sigmacut` and `applymask` in earlier iterations). This mask can not be too aggressive or the SVD RFI subtraction which follows may fail due to there being too many gaps in

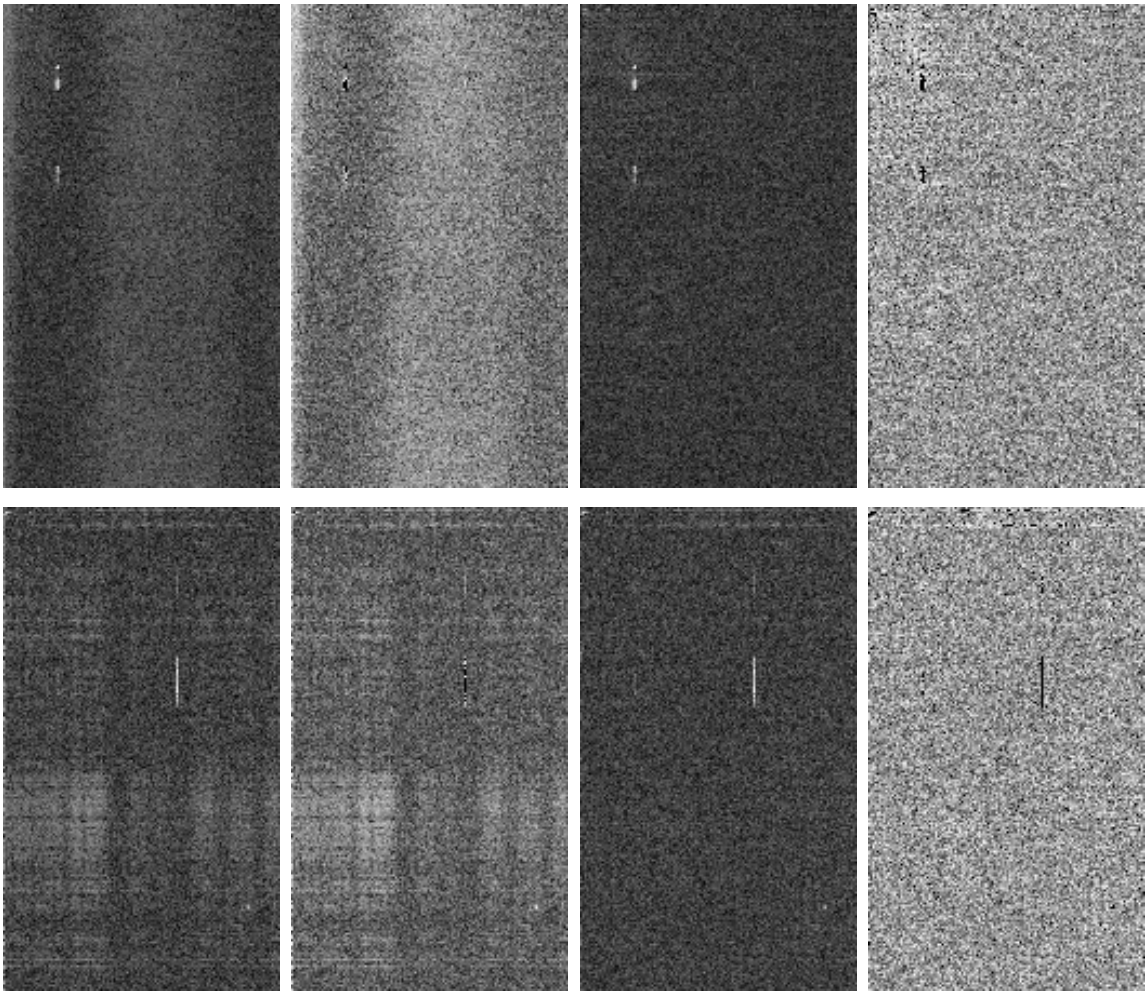
the data.

Singular value decomposition (SVD) is used to separate broadband radio sources on the ground from those in the sky. This is identical to the SVD that is performed to create RFI maps (`svd_uvec`, `svd_vvec`). As before, ground-based sources contribute most to the largest eigenvalues since they do not move as a function of time with respect to the array, while sky sources rotate. For observations of B0823+26 in 2007, one hour of data has 14396 time records, and each record has about 7.5 million entries corresponding to the number of frequency channels and baselines between the 60 antennas. Observations in drift scan mode from 2010 and later typically have 2000–5000 time records, depending on the declination of the source.

Once the modes are identified, the contribution of the largest modes are subtracted from the original data (`subsvd`), without the 8σ narrow line masks. After this, a new 3σ mask is applied. A sample of the data at a few intermediate stages showing the successful removal of both line and broadband RFI can be seen in Figure 3.5. The RFI patterns in (u, v) space both before and after the SVD are illustrated in Figure 3.6.

Since the SVD relies on the fact that RFI sources do not rotate with respect to the telescope, astronomical sources with a slow fringe rotation rate can be confused for RFI. An exactly north-south baseline does not rotate at all with respect to the sky, so we can expect any baseline with a low enough u coordinate (the east-west component of the baseline) may have its astronomical power reduced by the SVD RFI removal. In Chapter 5 we will use the power spectrum to quantify this effect and apply a cut to the visibilities to account for it (in particular, see Figures 5.3 and 5.4).

The pipeline concludes by fringestopping the data to the field centre (the location of the pulsar), and is rebinned from the native 2048 frequency channels and 0.25 second integrations to 64 frequency channels (0.25 MHz each) and 64 second integrations. In most recent versions of the pipeline the final narrow line masking, fringestopping, and rebinning is performed in a single step by `Fstop_mask_merge_rebin`.



(a) Input (b) 8σ mask (c) After SVD 3σ mask

Figure 3.5: Representative stages in the RFI removal pipeline. In each panel, the horizontal axis is frequency, covering approximately 1 MHz, and the vertical is time, increasing downward, and covering approximately one hour. The grey-scale is the visibility amplitude. The top row is the C00-C08 baseline and the bottom row the C00-W04 baseline, which are approximately 560 m and 9400 m long, respectively. Large bright patches indicate broadband RFI, and vertical lines indicate line RFI. From left to right the columns are the initial input data, after the initial 8σ mask, after removing the largest eigenvalues in the SVD, and finally the final 3σ mask removing the line RFI. After the SVD, broadband interference is no longer visible, and the last line mask leaves a nearly uniform image.

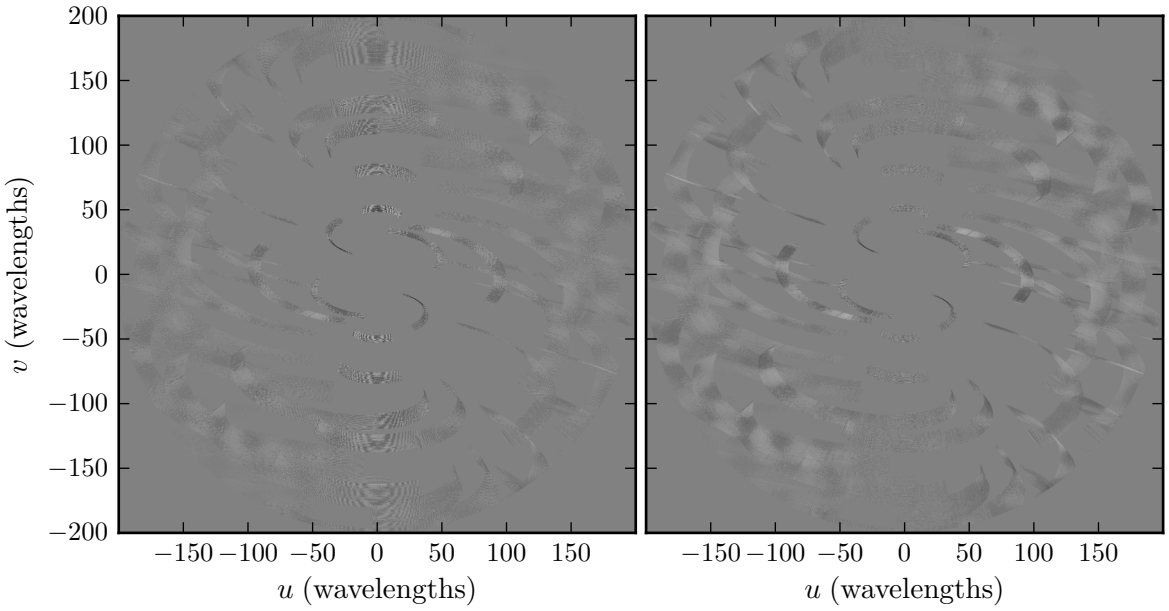


Figure 3.6: Raw visibilities with (u, v) distance $|\mathbf{u}| < 200$ before the SVD RFI removal step (left) and after the RFI removal (right). The structure in the central vertical strip (low $|u|$) is caused by RFI, and is significantly reduced by the SVD procedure. The effect of this process on the final power spectrum is considered in Chapter 5.

The current configuration takes approximately 11 hours to process one hour of data on the CITA Sunnyvale computing cluster, or approximately one hour of real time if the observations are broken up into 20 minute scans. During some observations, we were able to make use of the HP computing cluster at the National Centre for Radio Astronomy (NCRA) in Pune, India, near GMRT. On this cluster, processing one hour of data took approximately 20 hours but with the capacity to process four such data blocks in parallel. Unfortunately, the HP cluster was only available for two observing cycles, facilitating real time identification of RFI sources for only a small fraction of the total observing time allocated at GMRT.

While the Sunnyvale cluster has the capacity to process as many as 12 one hour scans at once, storage capacity, shared usage, and other bottlenecks reduce this significantly such that the actual throughput is comparable to the NCRA HP cluster. However, due to its location in Toronto, Canada, the large file sizes of the unreduced data, and the slow internet bandwidth between India and Canada, the RFI analysis is impossible to do in real time without a dedicated large cluster on site. Though this could in principle be accomplished with the GSB, the limited shared storage space and constraints on network and CPU usage imposed by active observing make this impractical except for short observations.

3.5 Increasing RFI sensitivity

Counterintuitively, increasing the sensitivity of the array to RFI sources in a controlled way has the potential to increase our sensitivity to the astronomical signal by improving RFI mitigation. There are two reasons for doing this: (1) to improve the localisation accuracy for dedicated RFI (non-astronomical) observations and (2) to make it easier to identify and remove RFI signals from the EoR data. Several different strategies to accomplish this have been tried:

1. While observing, a subset of arm antenna can be used as RFI monitoring antenna by directing them towards either the horizon or the central square. Since the arm antenna contribute mostly to baselines much larger than $|\mathbf{u}| \approx 500$ this should not impact the EoR sensitivity. In theory these can then be used to solve for the RFI response and subtract this from the data. This configuration was only used for the SVD RFI mapping and mode removal in 2009.

2. The feed turret at the primary focus can be rotated such that the 150 MHz is opposite the focus, effectively increasing the primary beam from 3 degrees to about 100 degrees. This decreases the sensitivity to the sky signal while becoming more sensitive to sources on the ground. This mode can not be used for EoR science observations, but is useful in short blocks to make RFI maps.

3. Once a candidate RFI source is identified, we can rotate each dish to point towards the GPS coordinate and take new data focused on the candidate location. This provides a more direct line of sight between the RFI source and the feed.

4. Even before candidate RFI positions are identified, the noise source can be used as a calibrator in the central square. This can be used to adjust the calibration for any of the above array configurations to compensate for effects such as changes in feed positions.

5. Antenna choice for the artificial source may also impact the sensitivity to different types of RFI. For example, using an omnidirectional antenna which broadcasts only with vertical or horizontal polarization may be useful for testing the sensitivity of our RFI algorithms to specific polarizations.

All of these techniques have been attempted at various times since 2008, however these observations have not been analyzed to the depth necessary to determine whether they allowed better RFI mitigation than previously possible. If the RFI environment continues

to be a significant limit on the power spectrum sensitivity, this could be pursued further in future observations.

3.6 Conclusion

Radio frequency interference is a significant obstacle to the detection of the EoR signal at GMRT. Despite the relatively radio-quiet environment, several very bright sources can be identified in GMRT data. The singular value decomposition and subsequent map making algorithms are successful in identifying the most egregious sources, and these can typically be corrected easily by the local power authority. While it is uncertain whether similar populations of interference sources will exist at similar facilities, it is reasonable to suggest that RFI map making should be a regular exercise to improve observations generally.

However, while the RFI hunting efforts during 2011 did turn up some interesting sources, the number of SVD modes that had well defined localisations, and the number of those localisations that resulted in concrete identifications, was significantly lower than in previous years. This may indicate that we have reached a kind of noise floor, where none of the brightest RFI modes identified alone are dominant sources. As such, it may be more useful going forward to focus on some of the other limiting factors to reaching the sensitivity necessary for detecting EoR.

One should also note that the RFI map making procedure is not sensitive to sources very close, or inside, the central square. Walking around the central square with a radio yielded at least two possible sources which did not have localisations in the RFI maps. This may also be a hint that there are also sources outside the central square to which we are not sensitive.

Even without removing physical sources, the SVD is successful at reducing broadband interference contamination in the visibilities. Although all the sources identified here may

be present in the raw data from 2007, they have been excised before fringestopping onto the sky and calibrating.

The fact that RFI map making succeeds in unambiguously identifying many RFI sources shows that the SVD modes removed from the 2007 data correspond to at least some real RFI contamination. The improvement the SVD RFI removal provides can be quantified in the context of the power spectrum after correcting for foreground contamination, which is much larger than the RFI. However, since the number of modes removed is arbitrary, it is possible that some signal from the sky is also mixed into these modes and will be removed along with the RFI. In Chapter 5, we will quantify both of these effects.

Chapter 4

Calibration

To get an accurate measurement of the sky signal, the response of the array needs to be calibrated, including the gains of the antennas, polarization leakage between the L and R feeds, and ionospheric variation. In this chapter we present a formalism for using a pulsar as a phase and polarization calibrator, discuss the implementation of this for GMRT data, and present some useful diagnostics and the results for the EoR data.

4.1 Polarization at 150 MHz

There are very few polarized sources at low frequencies. Recently Bernardi et al. (2013) reported only one polarized point source detected in a 2400 square degree survey at 189 MHz with MWA. Pulsars themselves tend to be highly polarized, with Hankins & Rankin (2008) reporting that B0823+26 has a polarization fraction as high as 40%, though the angle is variable. The ionosphere can also introduce polarization and change the apparent positions of sources. The large field of view of GMRT at 3 degrees means the ionosphere may vary over the width of the primary beam. However, this field is actually the smallest among EoR experiments, meaning ionosphere calibration will only be more difficult with other instruments.

The diffuse polarized emission is much higher than point sources. While some studies

(de Bruyn et al., 2006; de Oliveira-Costa et al., 2008b) have suggested the polarized signal at 150 MHz may be several to tens of kelvin, including structure on the arcminute scales relevant to EoR, Pen et al. (2009) reported the polarization of the B0823+26 field to be no more than 3 K. In their survey, Bernardi et al. (2013) reported a peak polarization of 13 K, but also that a large area of the survey was less than 1 K. Their survey area did not overlap any of our three fields, but does come within a few degrees of B2045-16, near an area of relatively high polarized emission. Without calibrating, leakage from Stokes Q or U into I at the level of a few per cent would be enough to create structure comparable in brightness to the EoR signal we are trying to measure, especially near the edge of the primary beam where the leakage will be the worst.

The polarization response can also vary with position in the primary beam due to instrumental effects. Pen et al. (2009) showed that GMRT has a slight "squint" even after polarization calibration, where the residual polarization increases with distance from the beam centre. However, a residual instrumental Stokes Q or U component does not impact the power spectrum measurement, as long as leakage into Stokes I has been minimized. The procedure we use to accomplish this is described in this chapter, though other implementations making use of similar input information are possible (e.g., van Straten, 2004).

4.2 Formalism

Though each antenna has a separate L and an R feed, it is possible for there to be leakage between the two, i.e., the L feed in practice will have a non-zero response to incoming right-handed polarized radiation. To get an accurate power spectrum measurement, we need to calibrate for the polarization response of the array and account for any leakage between the total intensity (Stokes I) and the polarized components (Stokes Q, U, and V). This leakage could change as a function of time, but we assume here that it is

constant during each night, i.e., on eight to ten hour timescales.

A general introduction to polarimetry can be found in Thompson et al. (2001). Incoming radiation $\vec{s}$ with orthogonal components (L, R), measured as voltages by the antenna feeds, is transformed by the instrumental response as

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = \begin{bmatrix} g_{L(L)} & g_{L(R)} \\ g_{R(L)} & g_{R(R)} \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix} \tag{4.1}$$

where $\vec{r} = (L', R')$ is the output measurement. The matrix that maps between the two is called a Jones matrix, $\mathbf{J}$. A relative rotation of the feeds and source adds a phase term to the diagonal elements, and differences from unit amplitude come from the gain of the antenna and electronics. The two off diagonal elements are the leakage terms between the two polarizations, e.g. $g_{R(L)}$ is the response of the right-handed feed to left-handed radiation.

For a correlation between the two antennas of a baseline, the Jones matrices combine as an outer product, giving the total response

$$\begin{bmatrix} (LL)' \\ (LR)' \\ (RL)' \\ (RR)' \end{bmatrix} = \begin{bmatrix} g^i_{L(L)}g^{j*}_{L(L)} & g^i_{L(L)}g^{j*}_{L(R)} & g^i_{L(R)}g^{j*}_{L(L)} & g^i_{L(R)}g^{j*}_{L(R)} \\ g^i_{L(L)}g^{j*}_{R(L)} & g^i_{L(L)}g^{j*}_{R(R)} & g^i_{L(R)}g^{j*}_{R(L)} & g^i_{L(R)}g^{j*}_{R(R)} \\ g^i_{R(L)}g^{j*}_{L(L)} & g^i_{R(L)}g^{j*}_{L(R)} & g^i_{R(R)}g^{j*}_{L(L)} & g^i_{R(R)}g^{j*}_{L(R)} \\ g^i_{R(L)}g^{j*}_{R(L)} & g^i_{R(L)}g^{j*}_{R(R)} & g^i_{R(R)}g^{j*}_{R(L)} & g^i_{R(R)}g^{j*}_{R(R)} \end{bmatrix} \begin{bmatrix} LL \\ LR \\ RL \\ RR \end{bmatrix} \tag{4.2}$$

where superscripts denote the antenna from which the gain factor g originates. From this it can be seen that every observed polarization includes components from all four, depending on the rate of leakage. The observed $(LL)'$ correlation will include some fraction of the true LR signal, determined by how much the L feed of the second antenna responds to R, and similarly for RL and RR. Since the total intensity, Stokes I, is a combination of LL and RR, these leakage terms need to be solved for and used to correct the visibilities.

For GMRT, the Jones matrix in equation 4.1 can be generalized to a 60×2 element gain matrix, $\mathbf{G}$, with one column representing the response of each of the 60 feeds to

left-handed polarization and the other the response to right-handed polarization. The visibilities can be written as the time averaged covariance of the incoming voltages,

$$\mathcal{V} = \left\langle \vec{r}\vec{r}^{\dagger} \right\rangle = \mathbf{G} \left\langle \vec{s}\vec{s}^{\dagger} \right\rangle \mathbf{G}^{\dagger}. \tag{4.3}$$

If we write the incoming signal as a 2×2 matrix in circular coordinates, the general form for arbitrary polarization in terms of the Stokes parameters is

$$\mathbf{P} = \begin{bmatrix} LL & RL \\ LR & RR \end{bmatrix} = \frac{1}{2} \begin{bmatrix} I+V & Q+iU \\ Q-iU & I-V \end{bmatrix} \tag{4.4}$$

From now on, we will ignore constant prefactors and take them to be absorbed into the gain solution. In the case of a purely unpolarized source, the true signal $\mathbf{P}_1$ is just the identity matrix times the Stokes I (which we omit for clarity), and the visibilities reduce to

$$\mathcal{V} = \mathbf{G}\mathbf{G}^{\dagger} = (\mathbf{U}\mathbf{S}\mathbf{W}^{\dagger})(\mathbf{W}\mathbf{S}^{\dagger}\mathbf{U}^{\dagger}) = \mathbf{U}\mathbf{S}^2\mathbf{U}^{\dagger} \tag{4.5}$$

where we have rewritten the gain matrix using a singular value decomposition. It is clear from the above that $\mathbf{U}$ is just the eigenvectors of $\mathcal{V}$, and $\mathbf{S}$ is the square root of the eigenvalues.

It may be tempting then to just diagonalize $\mathcal{V}$ and use $\mathbf{US}$ as the gain matrix itself. However, there is an ambiguity since any unitary matrix $\mathbf{W}$ will satisfy the rightmost equality but may not accurately reflect the true response of the system. If, for example, we set $\mathbf{W}$ to be the identity matrix, the above constraint is satisfied but we have effectively fixed the responses of the feeds to be perfectly orthogonal. To find the true $\mathbf{G}$, we need to solve for $\mathbf{W}$ by resolving the phase degeneracy.

There are four degrees of freedom in selecting $\mathbf{W}$, but one of these is a global phase that will always be unobservable. The remaining phases can be thought of as rotations between Stokes Q, U, and V. The space of possible matrices that satisfy the observational

constraints is spanned by the three Pauli matrices,

$$\sigma_{\mathbf{1}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_{\mathbf{2}} = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}, \quad \text{and} \quad \sigma_{\mathbf{3}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{4.6}$$

If we observe a polarized source with a known true signal $\mathbf{P}_2$, the visibilities are

$$\mathcal{V}_2 = \mathbf{G}\mathbf{P}_2\mathbf{G}^\dagger = \mathbf{U}\mathbf{S}\mathbf{W}^\dagger\mathbf{P}_2\mathbf{W}\mathbf{S}^\dagger\mathbf{U}^\dagger \tag{4.7}$$

which we can rewrite as

$$\mathbf{S}^{-1}\mathbf{U}^\dagger\mathcal{V}_2\mathbf{U}\mathbf{S}^{-1} = \mathbf{W}^\dagger\mathbf{P}_2\mathbf{W} \equiv \mathbf{M}. \tag{4.8}$$

The only unknown is $\mathbf{W}^\dagger\mathbf{P}_2\mathbf{W} \equiv \mathbf{M}$. If we diagonalize both $\mathbf{P}_2$ and $\mathbf{M}$, we see they share the same eigenvalues. Denoting the eigenvector matrices of each as $\mathbf{Q}_P$ and $\mathbf{Q}_M$ respectively, we can then rewrite $\mathbf{W}$ in terms of the known eigenvectors as

$$\mathbf{W} = \mathbf{Q}_P\mathbf{Q}_M^\dagger. \tag{4.9}$$

This effectively fixes two of the remaining phases. Each of $\mathbf{Q}_P$ and $\mathbf{Q}_M$ have two phases, making explicit the original four dimensional space of $\mathbf{W}$. One of these we have already said is an unobservable global phase, and two are degenerate with the phases of $\mathbf{U}$, leaving one phase remaining unknown.

We can use this expression for $\mathbf{W}$ to rewrite $\mathbf{G}$ as

$$\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{Q}_M e^{i\theta\sigma_{\mathbf{3}}}\mathbf{Q}_P^\dagger \tag{4.10}$$

where we have explicitly added the final phase degeneracy in terms of the third Pauli matrix, which relates the relative phases of the two components of $\mathbf{W}$. If we observe a third calibrator with response

$$\mathcal{V}_3 = \mathbf{G}\mathbf{P}_3\mathbf{G}^\dagger \tag{4.11}$$

we can use the matrices known so far to expand this into

$$\mathbf{Q}_M\mathbf{S}^{-1}\mathbf{U}^\dagger\mathcal{V}_3\mathbf{U}\mathbf{S}^{-1}\mathbf{Q}_M^\dagger = e^{i\theta\sigma_{\mathbf{3}}}\mathbf{Q}_P\mathbf{P}_3\mathbf{Q}_P^\dagger e^{-i\theta\sigma_{\mathbf{3}}} \tag{4.12}$$

where θ is the only unknown. If we choose $\mathbf{P}_2 = Q\sigma_\mathbf{1}$ and $\mathbf{P}_3 = U\sigma_\mathbf{2}$, then the right hand side can be evaluated as

$$\mathbf{Q}_M\mathbf{S}^{-1}\mathbf{U}^\dagger\mathcal{V}_3\mathbf{U}\mathbf{S}^{-1}\mathbf{Q}_M^\dagger = \begin{bmatrix} 0 & -iUe^{2i\theta} \\ iUe^{-2i\theta} & 0 \end{bmatrix}, \tag{4.13}$$

giving us θ in terms of the known matrices on the left hand side.

Together, observations of one unpolarized calibrator, $\mathbf{P}_1$, and two polarized calibrators, $\mathbf{P}_2$ and $\mathbf{P}_3$, completely determine the system gain $\mathbf{G}$, which can be used to correct for the polarization leakage.

4.2.1 Circular polarization

Our first "unpolarized" calibrator will actually have a circular component, which can be written in the form

$$\mathbf{P}_\mathrm{C} = \begin{bmatrix} I+V & 0 \\ 0 & I-V \end{bmatrix}. \tag{4.14}$$

In practice, it is difficult to separate the Stokes I and V components. Though we have not included such a correction in our analysis, we discuss here how the calibrators can be used to solve for the relative contribution from V.

First, we use the same basic approach as discussed in the previous section. An observation of our circularly polarized calibrator measures

$$\begin{aligned} \mathcal{V}_\mathrm{C} &= \mathbf{G}\mathbf{P}_\mathrm{C}\mathbf{G}^\dagger & (4.15)\\ &= \mathbf{U}\mathbf{S}\mathbf{W}^\dagger\mathbf{P}_\mathrm{C}\mathbf{W}\mathbf{S}\mathbf{U}^\dagger & (4.16)\\ &= \left(\mathbf{U}\mathbf{S}\mathbf{W}^\dagger\mathbf{P}_\mathrm{C}^{1/2}\mathbf{W}_\mathrm{C}^\dagger\right)\left(\mathbf{W}_\mathrm{C}\mathbf{P}_\mathrm{C}^{1/2}\mathbf{W}\mathbf{S}\mathbf{U}^\dagger\right) \equiv \mathbf{G}_\mathrm{C}\mathbf{G}_\mathrm{C}^\dagger & (4.17)\end{aligned}$$

where again we have noted that there is a degeneracy by inserting $\mathbf{W}_\mathrm{C}^\dagger\mathbf{W}_\mathrm{C}$, and used the matrix product in brackets to define $\mathbf{G}_\mathrm{C}$. This new estimator of the gain matrix can be used to correct the polarizations in observations of the second calibrator $\mathbf{P}_2$. We do this

by applying the inverse of $\mathbf{G}_\mathrm{C}$ to both sides of the visibility equation,

$$\mathbf{G}_\mathrm{C}^{-1}\mathcal{V}_2\mathbf{G}_\mathrm{C}^{-1\dagger} = \mathbf{G}_\mathrm{C}^{-1}\left(\mathbf{USW}^\dagger\right)\mathbf{P}_2\left(\mathbf{WSU}^\dagger\right)\mathbf{G}_\mathrm{C}^{-1\dagger}. \tag{4.18}$$

Using the expansion of $\mathbf{G}_\mathrm{C}$ defined in equation 4.17, this reduces to

$$\mathbf{M}_\mathrm{C} \equiv \mathbf{G}_\mathrm{C}^{-1}\mathcal{V}_2\mathbf{G}_\mathrm{C}^{-1\dagger} = \left(\mathbf{W}_\mathrm{C}\mathbf{P}_\mathrm{C}^{-1/2}\right)\mathbf{P}_2\left(\mathbf{P}_\mathrm{C}^{-1/2}\mathbf{W}_\mathrm{C}^\dagger\right). \tag{4.19}$$

Taking note that the product of the $\mathbf{P}$ matrices is antidiagonal, we can easily diagonalize it,

$$\mathbf{P}_\mathrm{C}^{-1/2}\mathbf{P}_2\mathbf{P}_\mathrm{C}^{-1/2} = \mathbf{U}_P\mathbf{S}_P\mathbf{U}_P^\dagger \tag{4.20}$$

and compare this to the diagonalization of $\mathbf{M}_\mathrm{C}$,

$$\mathbf{M}_\mathrm{C} = \mathbf{U}_M\mathbf{S}_M\mathbf{U}_M^\dagger = \mathbf{W}_\mathrm{C}\mathbf{U}_P\mathbf{S}_P\mathbf{U}_P^\dagger\mathbf{W}_\mathrm{C}^\dagger. \tag{4.21}$$

Since SVDs are unique, we must have $\mathbf{S}_M = \mathbf{S}_P$ and $\mathbf{W}_\mathrm{C} = \mathbf{U}_M\mathbf{U}_P^\dagger$. This allows us to rearrange equation 4.19 to find

$$\mathbf{P}_2 = \mathbf{P}_\mathrm{C}^{1/2}\mathbf{W}_\mathrm{C}^\dagger\mathbf{M}_\mathrm{C}\mathbf{W}_\mathrm{C}\mathbf{P}_\mathrm{C}^{1/2} \equiv \mathbf{P}_\mathrm{C}^{1/2}\mathbf{M}'_\mathrm{C}\mathbf{P}_\mathrm{C}^{1/2}. \tag{4.22}$$

We know $\mathbf{P}_2$ is traceless, and that the trace of a product is unchanged by cyclic permutations, so

$$\mathrm{trace}(\mathbf{M}'_\mathrm{C}\mathbf{P}_\mathrm{C}) = 0, \tag{4.23}$$

from which you can plug in $\mathbf{P}_\mathrm{C}$ and solve for V in terms of I to get

$$\frac{V}{I} = \frac{M'_{00} + M'_{11}}{M'_{11} - M'_{00}}. \tag{4.24}$$

Here, M'_{ij} are the elements of $\mathbf{M}'_\mathrm{C}$. With this correction, solving for $\mathbf{W}$ proceeds as before.

Unfortunately, by moving the $\mathbf{P}_\mathrm{C}^{1/2}$ terms to the left hand side of equation 4.22, it can be shown that $\mathbf{M}'_\mathrm{C}$ is traceless, suggesting that this procedure is formally singular. Another approach may be to change the order of calibrators, using $\mathbf{P}_2$ to form an estimator of the gain matrix to correct the leakage in $\mathcal{V}_\mathrm{C}$. However, since $\mathbf{P}_2$ is traceless, this

new estimator can not be written in the form $\mathbf{G}_2\mathbf{G}_2^\dagger$ as we did in equation 4.17. That is, we can not break $\mathbf{P}_2$ into the form $\mathbf{P}_2^{1/2}\mathbf{P}_2^{1/2\dagger}$. A solution must make use of a more complex decomposition where the $\mathbf{W}$ matrices holding phase degeneracies are members of SU(1,1), the group of matrices for which

$$\mathbf{W}^\dagger \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{4.25}$$

The procedure follows a similar pattern, where a gain matrix estimator is constructed, applied to the visibilities of the second calibrator, and this is used to solve for the relationship between I and V.

Since the circular polarization of B0823+26 is only about one per cent at 408 MHz (Gould & Lyne, 1998), with no obvious trend towards lower frequencies, we do not expect that this correction is an important one for this field. However, this could be implemented for further analysis of other fields if a large V component is expected to be problematic.

4.3 Implementation

The polarization calibration described above is implemented with a pulsar acting as all three calibrators by observing it over a range of parallactic angles. Since the data is gated on the period of the pulsar, the off-pulse gates can be subtracted from the gate with the pulse to get visibilities of the pulsar alone, without any sky or RFI contamination. These pulsar-only visibilities are used to calculate the calibration corrections. The component programs described in this section were written by Ue-Li Pen.

Before implementing the full gain solution to correct for the polarization leakage, we first calibrate LL and RR separately. The formalism above has assumed that everything is constant with time, while we know in practice the ionosphere causes variations. The ionosphere affects LL and RR identically, meaning they have the same time variability. Treating one polarization at a time allows us to go through the calibration with a scalar

instead of a 2×2 matrix, simplifying the calculation.

A principal component analysis (PCA) is used on each scan individually to find a preliminary bandpass and ionosphere calibration, and set LL and RR to have the same global phase (`pulsar_llrr`). First, the median visibility in time at every frequency and correlation is used to get the frequency eigenvectors, serving as a basic bandpass calibration for each timestamp. The median visibility in frequency is then used to get the time-wise eigenvectors, correcting for ionospheric variation over the night. The average gain for each antenna normalized to one. Finally, both the calibrated pulsar visibilities and the calibration model are saved to disk.

When starting a scan, the EoR correlator can update its software gain for each antenna. For observations before 2008, this was done at the beginning of each one-hour scan by measuring the total power from each antenna and setting the gains such that the powers from each were all equal. This step also resets the correlator clock, effectively randomizing the relative phases of the L and R inputs. In order to combine successive scans into a single data set, each scan must be renormalized to the same levels relatively to an arbitrary calibrator scan, usually chosen to be close to or during transit. The clock timing differences are calculated first (`triggerdiff`), and then these are applied together with a factor that renormalizes the software gains to match the calibration scan (`mergenorm`). From 2008 onward, the correlator scripts were changed such that the software gains and clock are reset only at the beginning of each night, making the renormalization and rephasing steps unnecessary when combining data from the same observing night.

Once all the scans are combined into a single data set with consistent normalization and phase, LL and RR are recalibrated separately again (`pulsar_llrr`) for an ionospheric and bandpass calibration that spans the entire night.

After applying the calibration from the unpolarized visibilities, the cross polarized (LR and RL) baselines are summed so that the remaining global phase can be found

(`sumlrraw`). This is also a useful point to check the quality of calibration so far, since the data condensed into a single baseline will readily show any catastrophic calibration errors.

Since the pulsar is observed over a range of parallactic angles, it can be decomposed into several signals: one which does not vary with time, one that varies with the parallactic angle ϕ, and one that varies against the parallactic angle (`projqu3`). With this decomposition the visibilities can be written as

$$\mathcal{V} = \mathcal{V}_I + \mathcal{V}_+ e^{2i\phi} + \mathcal{V}_- e^{-2i\phi} \tag{4.26}$$

and we define a frame for the linear Stokes parameters analogously to equation 2.1 but relative to the (unknown) pulsar polarization angle as

$$\mathcal{V}_Q = \mathcal{V}_+ + \mathcal{V}_- \tag{4.27}$$

$$\mathcal{V}_U = i(\mathcal{V}_+ + \mathcal{V}_-). \tag{4.28}$$

With this breakdown, $\mathcal{V}_I$ is the unpolarized calibrator (the visibilities of $\mathbf{P}_1$), and is used to solve for $\mathbf{U}$ and $\mathbf{S}$ in equation 4.5. The components $\mathcal{V}_Q$ and $\mathcal{V}_U$ are treated as two polarized calibrators. Unlike a real physical source, they will have no Stokes I component, but this does not affect the analysis. The first, $\mathcal{V}_Q$, has the polarization signal

$$\mathbf{P}_2 = \begin{bmatrix} 0 & Q \\ Q & 0 \end{bmatrix} \tag{4.29}$$

while the signal for $\mathcal{V}_U$ is

$$\mathbf{P}_3 = \begin{bmatrix} 0 & iU \\ -iU & 0 \end{bmatrix}. \tag{4.30}$$

Once the pulsar visibilities have been broken down into the three components, we can solve equations 4.8 through 4.13 to get the complete gain calibration (`polsol`). The final step (`unleak_data_noise`) is to break down $\mathbf{G}$ into the 2×2 submatrices $\mathbf{G}_i$ that give the gain response of each individual antenna i, and apply these to the measured

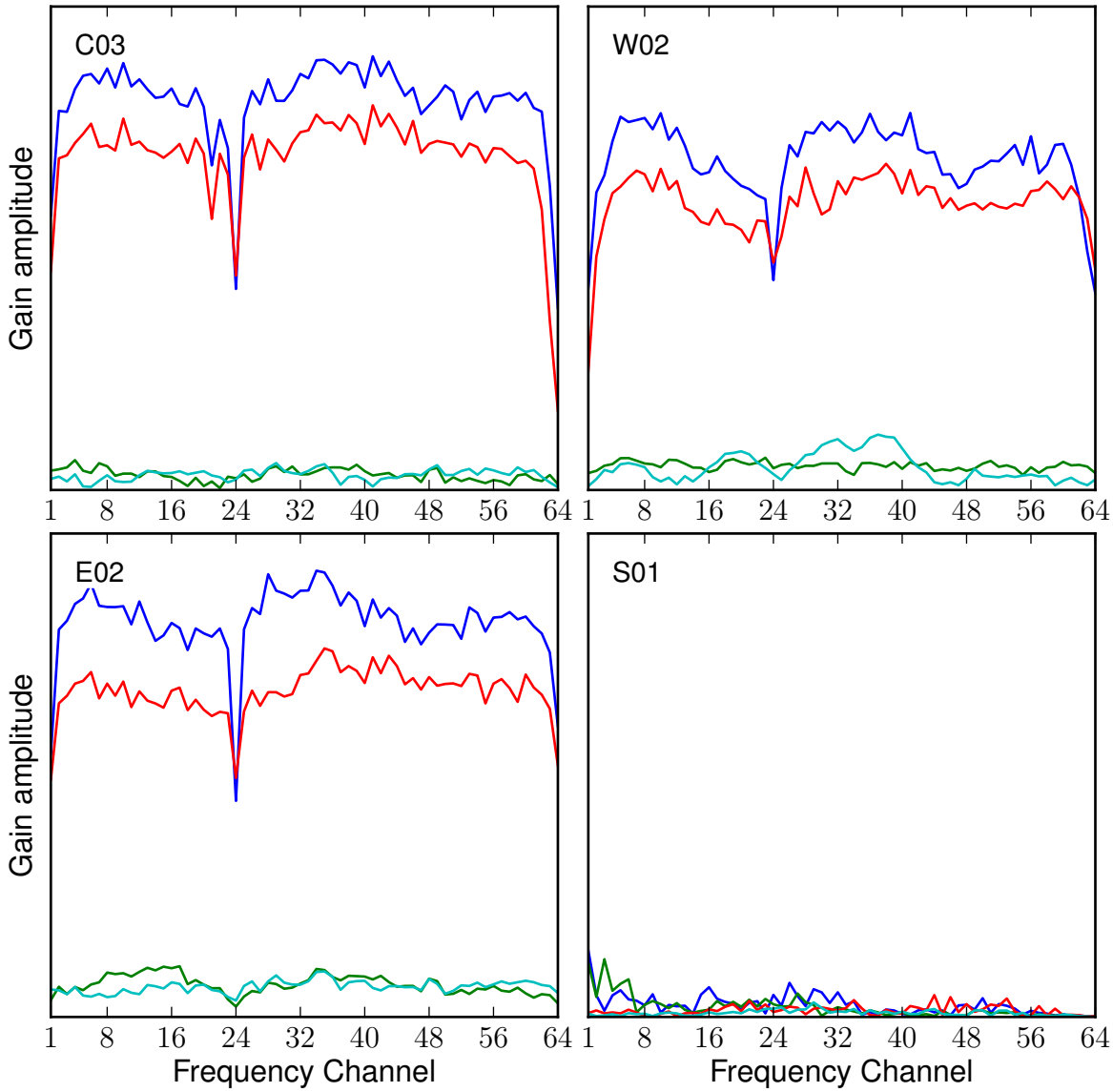


Figure 4.1: Gain solutions $\mathbf{G}_i$ from the night of December 10, 2007 for four antennas. The amplitude of the complex gain is shown as a function of frequency. In each plot the blue and red lines are the response of the left and right feeds to left and right polarization, respectively. The green and cyan lines are the response of the left and right feeds, respectively, to the opposite polarization. A perfect antenna with no polarization leakage would have both the green and cyan lines exactly zero. These levels of leakage are typical for all working antennas. Antenna S01 shows all four gains to be small and of comparable magnitude, indicating that it was not functioning on this night, and was masked from further analysis.

visibilities $\mathcal{V}_{ij}$ for each baseline as described by equation 4.2, where the $g^{(i)}_{p(q)}$ terms are taken from the $\mathbf{G}_i$ as appropriate. The solutions for $\mathbf{G}_i$ of four antennas as a function of frequency, with typical leakage terms of a few per cent, are shown in Figure 4.1.

This level of leakage is in line with results reported earlier by Pen et al. (2009) using a different subset of the GMRT-EoR data. With a modest polarization signal in this field of a few kelvin, this leakage would be enough to create structure in Stokes I comparable to the EoR signal if left uncorrected. Cutting the leakage by a modest factor of 10 would be enough to reduce the spurious structure to less than that of EoR, as long as the polarized emission is no higher than measured in this field.

Occasionally the calibration will fail. This is visually obvious in cases where the resulting image does not resemble the sky. One can use the covariance of the antennas after leakage correction to diagnose some problems. If the calibration was successful, the visibilities of the pulsar should be purely real. The actual result is easily assessed by creating an image of the visibilities sorted into a matrix by antenna, time, and frequency and broken into real and imaginary parts. Figure 4.2 shows such a matrix for the night of December 10. These figures for each night are produced automatically as part of the calibration pipeline (`prebin`).

It can also be useful to look at the phase of the LR correlations as a function of time and frequency, which can indicate whether the expected variation with parallactic angle is present, without which the decomposition defined in equation 4.26 may fail. As the target source transits, the parallactic angle changes quickly and a sharp transition can be seen in the phase angles. Two examples are shown in Figure 4.3.

4.4 Flux calibration

The flux scale of the field is set by calibrating relative to the pulsar in the centre of the field at each timestamp for every baseline, receiver, and frequency. Since the pulsar flux

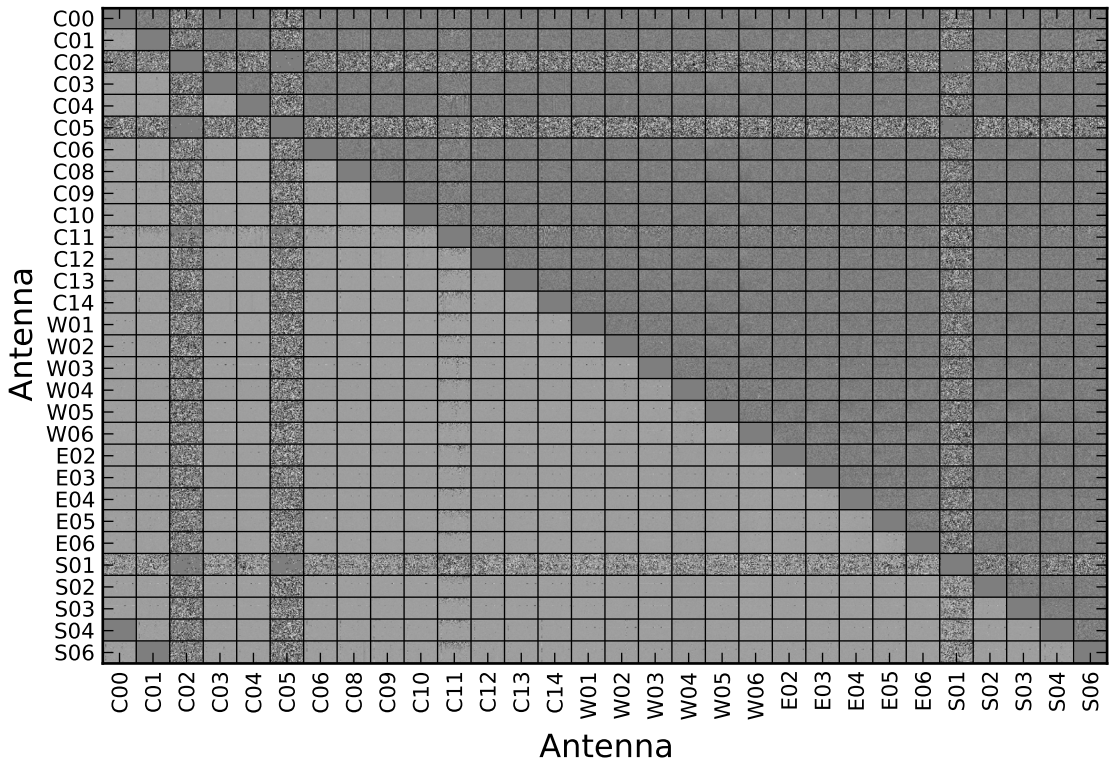


Figure 4.2: The covariance matrix of antennas after calibration can be a useful diagnosis tool, highlighting if particular antennas, frequencies, or timestamps are causing the calibration to fail. In each square, the frequency increases upward and time increases to the right. The upper right triangle is the imaginary part and the lower left triangle is the real part. These should be zero (black) and one (white) respectively, but are grey because of noise outside this range distorting the colour scale. In this example, C02, C05, and S01 were not functioning during this observation or could not be calibrated, as evidenced by the noise at all correlations.

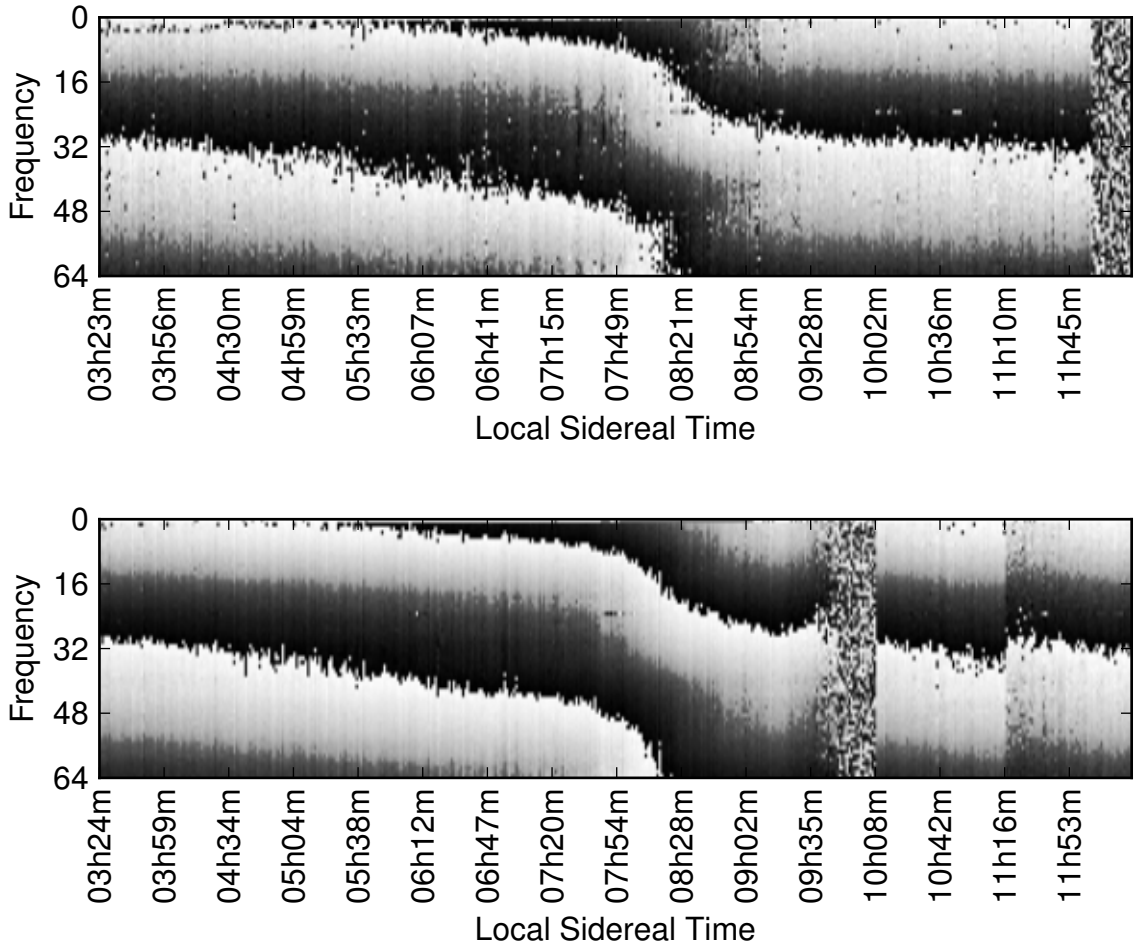


Figure 4.3: Phase of the LR correlation, summed over all baselines, as a function of frequency and local sidereal time for observations of B0823+26 demonstrating some typical errors. The top panel is from December 14, 2007. After 11h45m, the phases lose coherence, showing that the antennas were no longer on target and the observation had ended, even though the correlator continued recording. The bottom panel, from December 10, 2007, shows the sharp change in angle that occurs at transit (8h27) reversing slightly between 9h00 and 9h30. This is caused by the pulse moving slightly out of the gate so that the phase is no longer averaged over the entire pulse. Once the pulse moves completely out of the gate, there is no coherent phase until the next scan starts at about 10h. There is also a phase discontinuity at the transition between the last two scans, indicating a minor calibration error.

is known to be variable, to set the absolute flux scale we identify the source with the highest flux in the sky image and set this peak value equal to the known flux of that source. The exact value found depends on the angular resolution, determined by the maximum baseline length used. To calibrate the flux we require a point source so that the flux is not spread over multiple image pixels, and we also do not want to resolve any extended structure. This is achieved by limiting the maximum baseline length used while calibrating to $|\mathbf{u}| \leq 2000$, although this limit should be checked based on the substructure of any chosen calibrator. At much smaller $|\mathbf{u}|$, we become limited by confusion.

For the B0823+26 field, the brightest apparent source is the radio galaxy 5C 7.245, which is less than 14 arcminutes from the pointing centre. The flux of this source was measured as 215.1 ± 12.6 mJy at 1415 MHz by Willis et al. (1976) at the Westerbork Radio Telescope, and as 685 ± 47 mJy at 408 MHz in the 5C catalogue (Pearson & Kus, 1978) for a spectral index of 0.93 ± 0.07. Additionally, this galaxy has two components separated by 12 arcsec (Willott et al., 2001). GMRT is capable of resolving this structure with $|\mathbf{u}| \gtrsim 2600$, so our calibration limit of $|\mathbf{u}| \leq 2000$ is appropriate here. To achieve the best match with other bright sources in the field, we adopt a value of 1.6 Jy for 5C 7.245. Although the flux of this source changes by as much as 10 per cent across the band, this change is less than that due to the uncertainty in the spectral index. Pen et al. (2009) reported that this procedure gives an estimated error of approximately 10 per cent for the other bright sources in the field.

In the B2217+47 field, the brightest source is B3 2218+457, with a flux of 1.07 Jy at 356 MHz (Douglas et al., 1996) and 3.74 Jy at 74 MHz (Cohen et al., 2007), suggesting a 150 MHz flux of 2.15 Jy. This is consistent with the poorly determined spectral index of $\alpha = -0.4 \pm 0.4$ reported by Douglas et al. (1996). Since this source is approximately 53 arcmin south of the pointing centre, some correction should be applied for primary beam attenuation of about 20 per cent.

Finally, for the B2045-16 field, the brightest source is PKS 2048-14. The flux of this

source is reported in a number of sources: 2.5 Jy at 408 MHz (Ekers, 1969), 7.8 Jy at 160 MHz and 12 Jy at 80 MHz (Hunstead, 1972), and 11.88 Jy at 74 MHz (Cohen et al., 2007). From these we estimate a flux of 8 Jy at 150 MHz, but note that at 108 arcmin away from the pointing centre the source will be attenuated by the primary beam by about 39 per cent.

4.5 Ionosphere

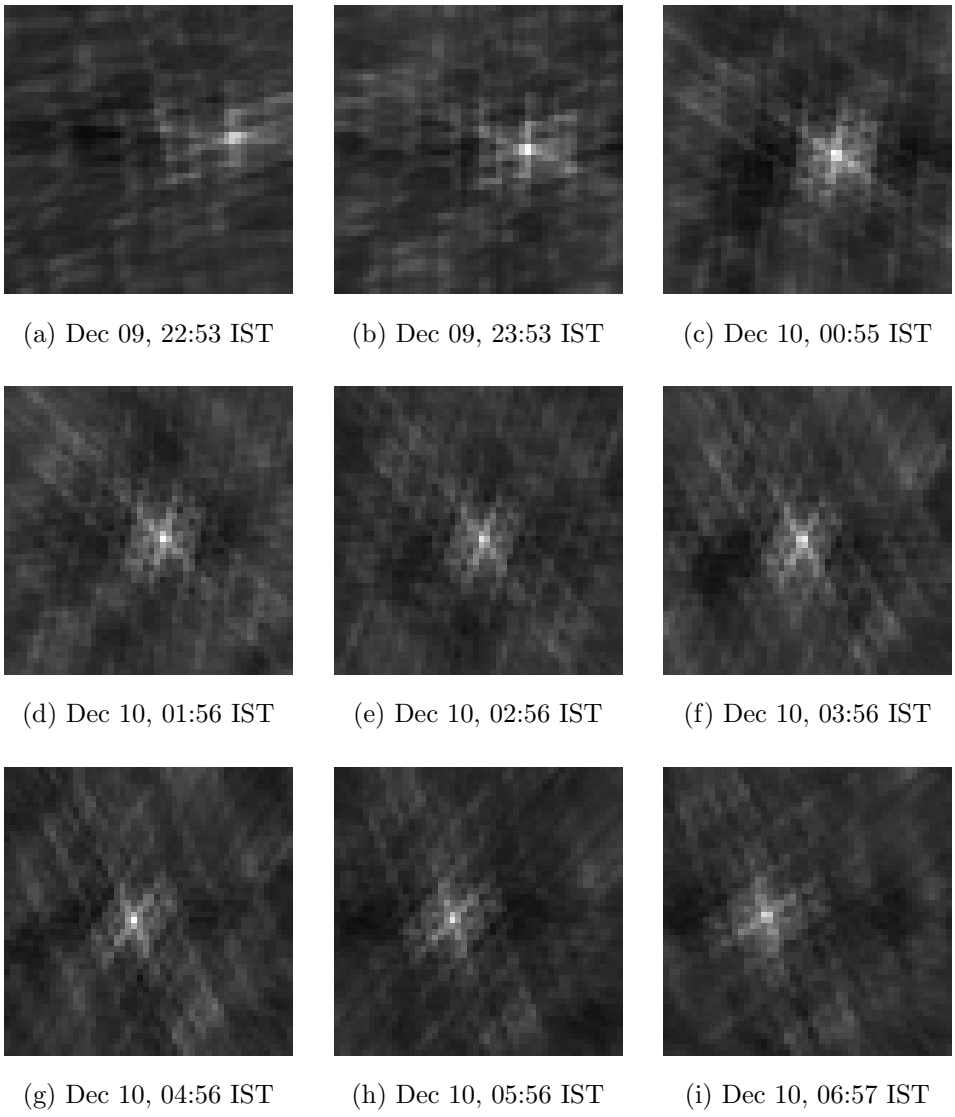
Our pulsar-based calibration automatically includes the ionosphere and any other atmospheric effects. At any instant only changes relative to the line of sight to the pulsar will remain. The most obvious effect the ionosphere can have is distorting the positions of sources, with sources farther from the pointing centre deflected more.

To test the magnitude of the effect after calibrating, we split one night's observations into a series of 10 minute long snapshots. A series of these, centred on 3C 200, is shown in Figure 4.4. Though it is clear that there is a systematic change in position as a function of sidereal time, the change is entirely consistent with the changes in the w term.

Each snapshot was imported into CASA[1], a standard radio astronomy data reduction software suite, and imaged with a resolution of 18 arcsec using the `widefield` imaging task. This performs a w term projection, and 10 000 iterations of the Cotton-Schwab CLEAN algorithm (Schwab, 1984). While we have found that the CLEAN algorithm is not useful for EoR science since it can distort the power of the diffuse signal, it is very useful for deconvolving point sources from the synthesized beam. Finally, the `imfit` task was used to fit a Gaussian profile to the eight brightest sources: 5C7.223, B2 0829+28, B2 0828+27, B2 0832+26A, B2 0816+26B, B2 0825+24, and 3C 200. These are scattered throughout the primary beam out to almost 3 degrees from the pulsar.

For these sources, the position between successive snapshots varies by less than

[1]Common Astronomy Software Applications; `http://casa.nrao.edu`. Version 3.1.0 was used for this work.



(a) Dec 09, 22:53 IST (b) Dec 09, 23:53 IST (c) Dec 10, 00:55 IST

(d) Dec 10, 01:56 IST (e) Dec 10, 02:56 IST (f) Dec 10, 03:56 IST

(g) Dec 10, 04:56 IST (h) Dec 10, 05:56 IST (i) Dec 10, 06:57 IST

Figure 4.4: Snapshots of 3C200, a bright radio source located approximately two degrees from the pointing centre, over the night of December 9–10, 2007, illustrating how the w term changes the apparent position of the source. Each image is 28 arcmin wide. The source rose at 22:26 IST, about thirty minutes before snapshot (a). Transit for this source was at approximately 03:48 IST, ten minutes before snapshot (f).

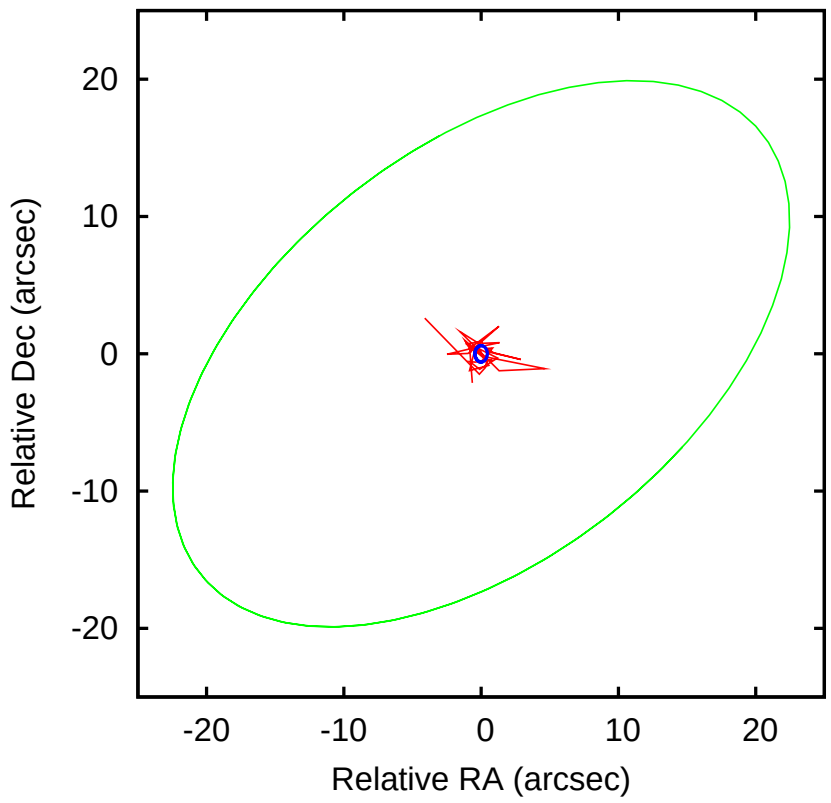


Figure 4.5: Position of 3C200 relative to the average in 10 minute intervals over 8 hours, shown in red. This source is almost three degrees from the pointing centre, where ionospheric distortions will be maximal. The blue circle is the average error on the positions and the green circle is the average size of the image, over all snapshots, which is effectively the same as the synthesized beam. The change in position with time resembles a random walk far smaller than the beam size. It is not even clear if the small residual drifts are larger than the thermal position noise and the phase variations in the primary beam sidelobes.

5 arcsec on average, and at most by 14 arcsec. The maximum deviation of any one snapshot from the average position is 25 arcsec, only slightly larger than the image resolution. The position as a function of time for 3C 200 is shown in figure 4.5, which illustrates that the error in the position is much smaller than the synthesized beam even at the edge of the field. 3C 200 is 3 degrees away from our calibration centre, with the primary beam half power point at about 1.6 degrees. This demonstrates that the isoplanatic patch extends beyond 3 degrees in radius. Ionospheric variation is responsible for only a tiny perturbation on the changing in position due to the w term seen in Figure 4.4. It should also be noted that this calibration does not depend on any special treatment (e.g., self-calibration) of 3C 200 or any other bright sources in the field other than the pulsar. This small position error is typical of all sources in the field of view.

4.6 Thermal noise

To demonstrate the level of thermal noise that would be present in their maps in the absence of foregrounds, Parsons et al. (2010) alternate the sign of successive integrations to suppress the slowly varying component and find a thermal noise limit of 310 mK in the first results from PAPER. We wish to find the comparable metric for GMRT. To do this, we alternate the sign of successive gates within 16 second integrations (`onegate_diff`). The four gates including and flanking the pulse are omitted, since the pulse would dominate any thermal signal, and could bleed into neighbouring gates. Since thermal noise scales with $\sqrt{t}$, the noise found with 12 gates should be reduced by a factor of $\sqrt{12/16}$ to give the noise with an integration over 16 gates. The power spectrum of the resulting visibilities can then be calculated, using the same method described generally in Chapter 1 and with more detail in Chapter 5, to give an estimate of the thermal noise.

Using six nights of observations from December 2007, comprising 50 hours, we estimated a thermal noise limit of approximately 25 mK, which is on the same order of the

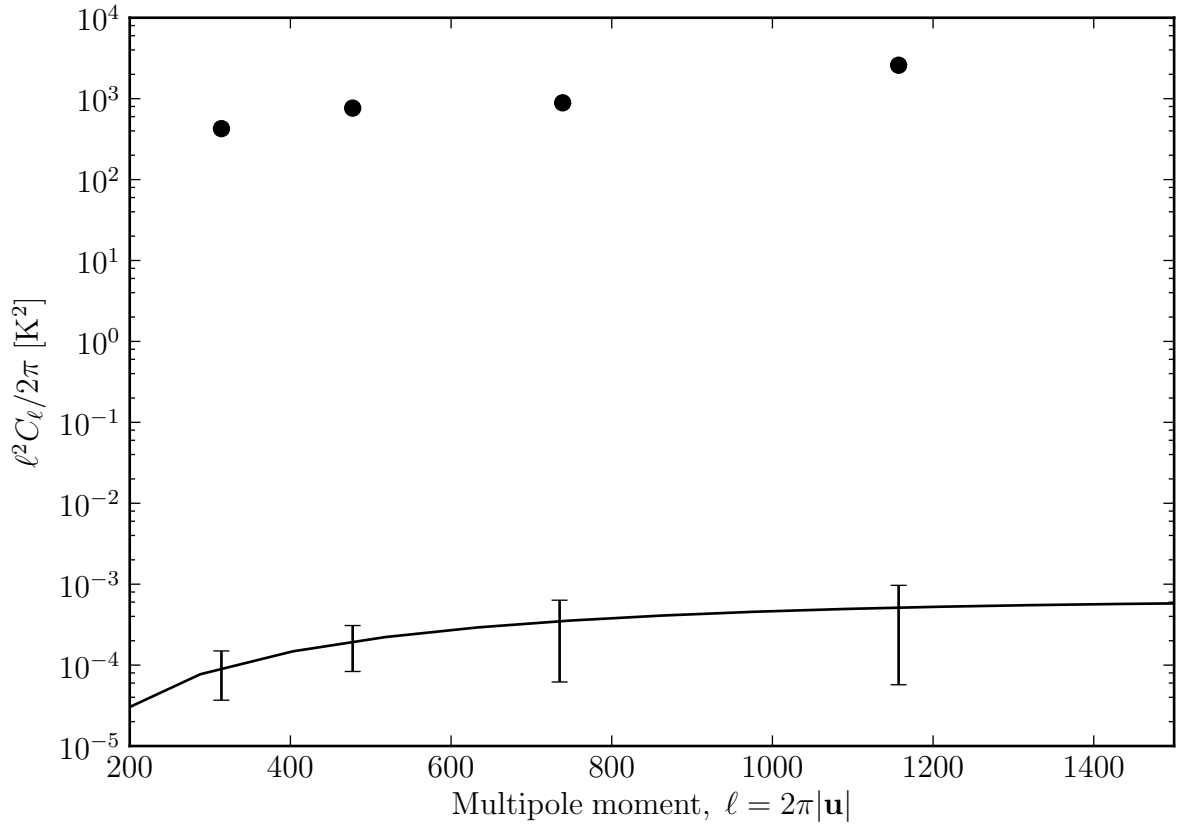


Figure 4.6: Forecast thermal error bars based on 50 hours of GMRT data from previous cycles. Forecasts assume an EoR signal from Jelić et al. (2008, solid line), 150 hours of observing time, and increased sensitivity from shorter integration blocks and a field closer to the equator. The circles above are the foregrounds measured in the PSR B0823+26 field in the same 50 hours. Bernardi et al. (2009) reported a power spectrum of foregrounds in the galactic plane at a level comparable to our measurement both here and in Pen et al. (2009). The error bars are approximately $(8\,\mathrm{mK})^2$ at the lowest ℓ.

expected EoR signal. This was used to motivate observing proposals from 2010 onward. If the total EoR data set could be extended to 150 hours, combined with changes in observing procedure designed to increase sensitivity, we estimated that the thermal limit could be reduced to 8 mK. Figure 4.6 shows these forecast errors in the context of the theoretical signal.

4.7 Eight-hour images

After RFI removal has been completed in each one hour scan, the polarization calibration described above allows us to combine these into images of an entire night, typically of about 8 hours, as can be seen in Fig. 4.7.

To correctly stack many hours together, one must consider the frequency and directional dependence of the measurement. One effect to consider is the change of the primary beam with frequency. For a given frequency, the primary beam is well defined and independent of (u, v). Deconvolution is difficult, but one can pick the frequency with the smallest beam and restrict the field of view at other frequencies to match. Computationally this could be done by convolving with the ratio of the beams, which is possible because of the commutativity of the primary beam operators.

However, this is a small effect compared to the effects of a w term. Typically, relative antenna positions are considered to be on a two dimensional plane with coordinates parametrized by (u, v) in units of the observing wavelength. Since baselines become non-coplanar over the course of the night, a third dimension, w, must be included, without which the image becomes blurred. Strategies for correcting this effect for GMRT by projecting visibilities onto the (u, v) plane have been partially developed, building on those used in CMB studies (Myers et al., 2003; Hobson & Maisinger, 2002), but these have non-trivial effects on the power spectrum and as such are not included here. In the interim we restrict ourselves to the short baselines for which w is small and such

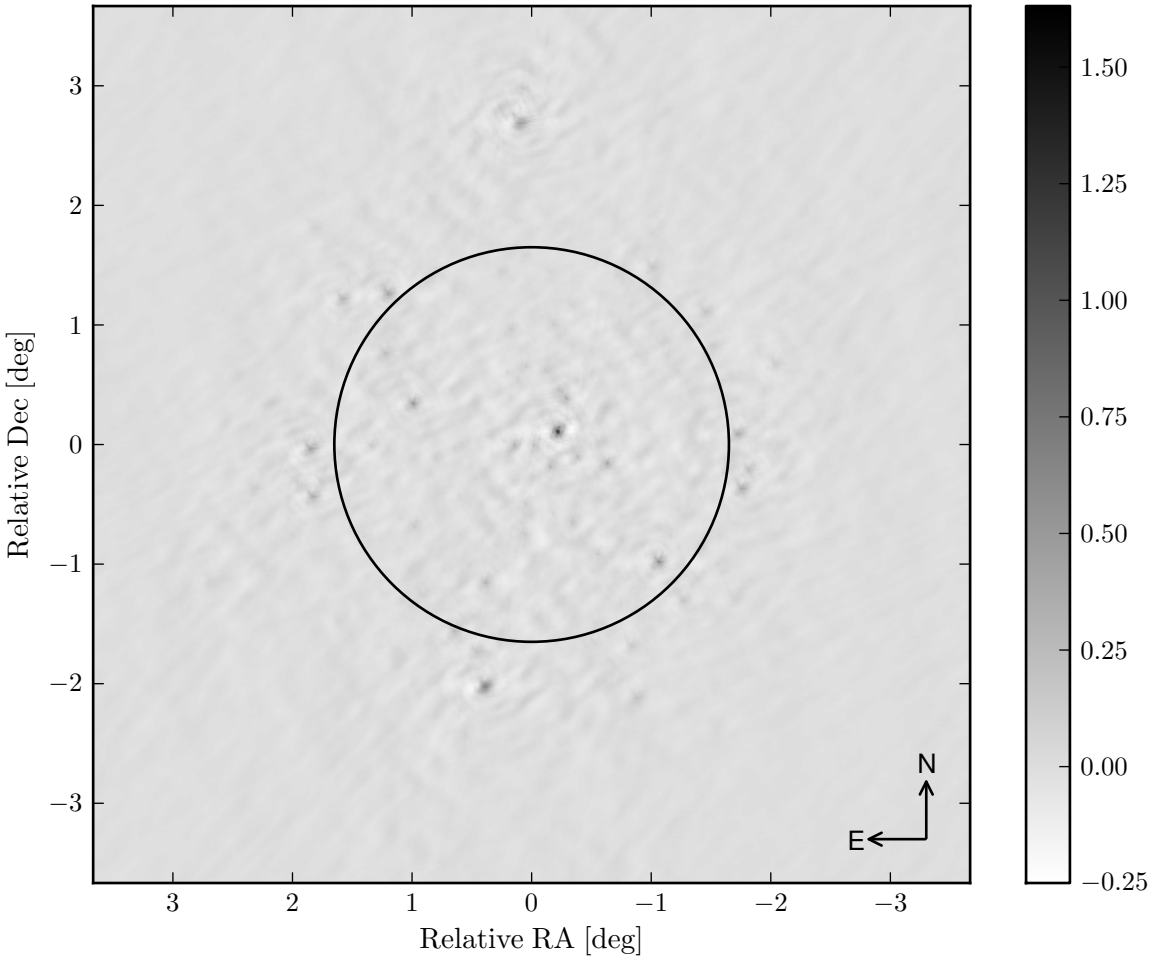


Figure 4.7: Dirty sky image made with 8 hours of data from 2007 December 10. The maximum (u, v) distance is 2000 with a grid size of approximately $\Delta u = 8$ giving a field of view of 7.3 degrees on each side. This same (u, v) distance cut is used to calibrate the peak flux of 5C 7.245 to 1.6 Jy for each day. The RMS of this image is 44 mJy within the primary beam.

corrections are not needed. The short baselines are the ones most sensitive to the EoR signal, so we retain most of the EoR sensitivity of the array. This is effectively the same as limiting the observations to within a few hours of transit, as was done explicitly in observations from late 2010 onwards.

4.8 Conclusion

Our procedure of using a pulsar at the centre of a field allows us to solve completely for the phase and polarization leakages, which are on the order of a few per cent. Since we effectively have a new measurement of the calibrator with every integration, we automatically account for variations in the ionosphere throughout the night, and the calibration works well even at the edge of the primary beam.

Given the low level of polarization in the B0824+26 field found in Pen et al. (2009), this calibration should be enough to reduce leakage from Stokes Q and U below the level of the EoR signal. We have not investigated either of the B2217+47 or B2045-16 fields here.

Having laid out the formalism and implementation of our calibration procedure, it is now possible to begin making estimates of the power spectrum.

Chapter 5

First Limits on the Power Spectrum

Once the observational data has been calibrated and as much RFI as possible is removed, the dominant remaining obstacle to detecting the 21 cm EoR signal is foregrounds. In this chapter we attempt a simple foreground removal strategy that operates in the frequency domain only, and use this to put preliminary limits on the power spectrum. Observations from 2007 December 10, 11, 14, 16, 17, and 18 are included in this analysis. Aside from a few additions, the content of this chapter was previously published in Paciga et al. (2011), Monthly Notices of the Royal Astronomical Society, Vol. 413, Issue 2, p. 1174–1183 (astro-ph/1006.1351).

5.1 Piecewise-linear foreground filter

Removing the foregrounds adequately is essential for detecting the EoR signal (see, e.g., Wang et al., 2006). Much work has been done on simulating foregrounds (e.g. Bowman et al., 2009; Jelić et al., 2010) and designing removal strategies (Morales et al., 2006b; Harker et al., 2009; Liu et al., 2009a, and others). For a review see Morales & Wyithe (2010) and references therein. The primary sources of foregrounds are synchrotron emission from the galaxy and extragalactic point sources. However, the precise sources which contribute to the foregrounds are not as important as the general form that they take as

a function of frequency.

For a simple method of removing foreground sources, we can take advantage of the fact that the fluxes of such sources do not vary greatly with frequency. A signal originating from the EoR will appear as additional variation on top of the foreground signal at megahertz scales. There are many possible approaches to modeling and subtracting the smooth component. For a preliminary estimate of the power spectrum, we use a simple piecewise-linear fit.

For each baseline and timestamp, there are 64 frequency channels spanning the $16.\bar{6}$ MHz bandwidth. We divide these into either 2, 8, or 32 bins (approximately 8, 2, and 0.5 MHz wide, respectively), find the median flux in each group, and extrapolate linearly between the medians at the midpoint of each bin. For the 8 MHz wide filter, there are only two bins and this is essentially a linear fit across the entire bandwidth. At 0.5 MHz, we are effectively smoothing over every other channel, which is the finest possible filter without removing all of the signal. This piecewise-linear filter is simple to implement and has the advantage over a polynomial fit of a having straightforward and local window function which is simple to interpret. We should note explicitly that this procedure may also remove some EoR signal, which we do not yet correct for. While more complex strategies for foreground removal are possible, and will be discussed in the next chapter along with a correction for signal loss, this provides a simple way of getting a preliminary estimate of the quality of the data.

Foregrounds which do not vary over the specified frequency range are removed, while features with variability at higher frequencies remain. This method results in an upper limit to the EoR signal since the measured power may still include residual foreground variation. By subtracting the mean flux over 2 MHz to remove galactic foregrounds, noise levels can be lowered to about 2 mJy for maps of an approximately 10 degree field of view. While bright point sources may play a significant role at some length scales (Datta et al., 2010; Liu et al., 2009b; Di Matteo et al., 2004), for the angular scales we are interested in

here the point sources are largely confusion limited and contribute in a similar manner to the diffuse background. We do not treat them separately here.

5.2 Sensitivity to physical scales

Since the observed frequency of the 21 cm emission corresponds to the redshift at which it was emitted, the frequency axis of our data contains all the information about the line-of-sight direction. Any foreground filter which operates by subtracting power at specific frequency scales necessarily changes the physical scales to which our power spectrum measurement is sensitive.

To first order, the piecewise-linear filter is the same as a boxcar average[1] which we can write as

$$\tilde{\mathcal{V}}(\nu) = \int \mathcal{V}(\nu')w(\nu - \nu')\mathrm{d}\nu' \tag{5.1}$$

where $\mathcal{V}(\nu)$ is the input visibilities at a given time and baseline, $\tilde{\mathcal{V}}(\nu)$ is the data after filtering, and the window function is

$$w(x) = \delta(x) - H(x, \Delta\nu). \tag{5.2}$$

Here, $\delta(x)$ is the Dirac delta function and $H(x, \Delta\nu)$ is a step function centred at x with width $\Delta\nu$ corresponding to our chosen filter, and of unit area.

In Fourier space, this window function is

$$w(k_\nu) = 1 - \frac{\sin(k_\nu \Delta\nu/2)}{k_\nu \Delta\nu/2}, \tag{5.3}$$

shown in Figure 5.1, where we have denoted the wave number as k_ν to make explicit that it is in units of inverse frequency. This can be rewritten in terms of the physical wavenumber $k_\parallel$ using the fact that in our redshift range $1\,\mathrm{MHz} \approx 11.6\,h^{-1}\mathrm{Mpc}$. When converted to units of $h\mathrm{Mpc}^{-1}$, we use the subscript "$\parallel$" to emphasize that the window

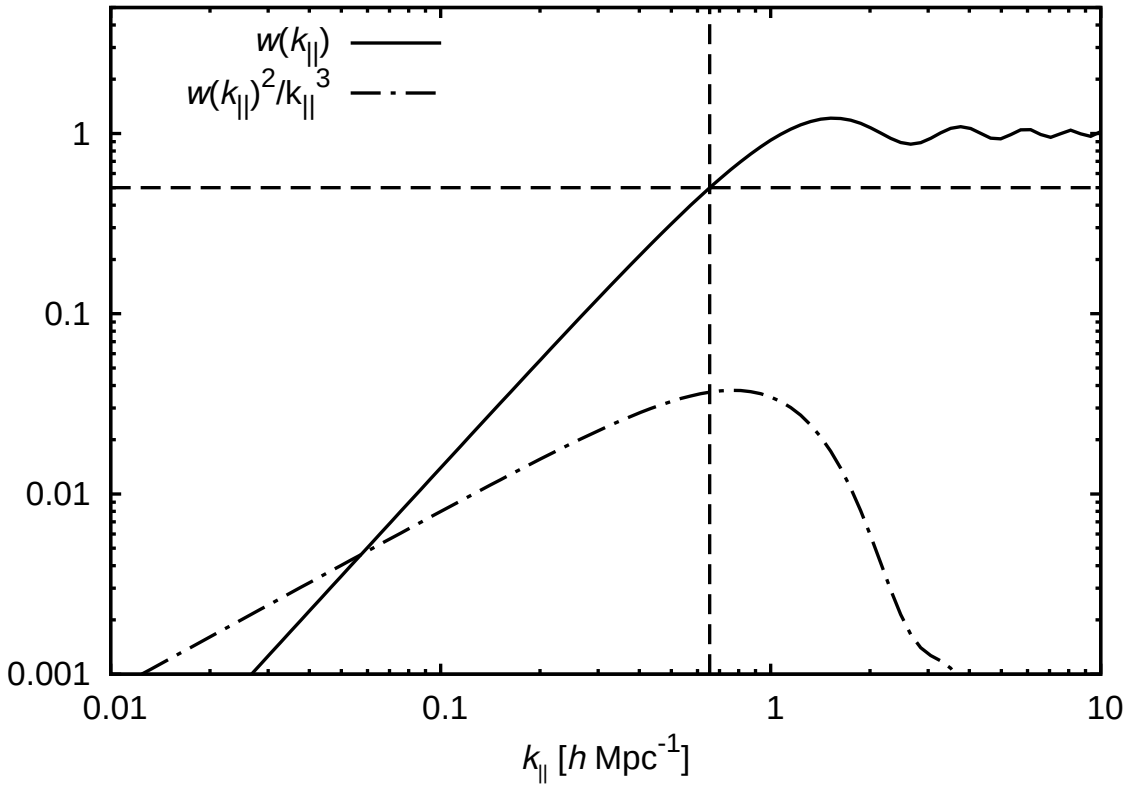


Figure 5.1: Fourier transform of our chosen window function for $\Delta\nu = 0.5\,\mathrm{MHz}$ (solid line). Dashed lines indicate where $w(k_\parallel) = 0.5$, which sets the minimum $k_\parallel$ to which we are sensitive. This value scales inversely with $\Delta\nu$. The peak of $w(k_\parallel)^2/k_\parallel^3$ (the dot-dashed line) indicates the $k_\parallel$ to which we are most sensitive under the assumption that the power spectrum is proportional to k^{-3}.

function acts on structure parallel to the line of sight only.

After filtering in this way, we denote the k_ν for which $w(k_\nu) = 0.5$ as corresponding to the minimum $k_\parallel$ along the line of sight to which we are sensitive, while smaller $k_\parallel$ are removed by the filter and will not contribute as strongly to the power spectrum. The three filters of 8 MHz, 2 MHz, and 0.5 MHz correspond to minimum $k_\parallel$ of 0.04, 0.16, and 0.65 $h\mathrm{Mpc}^{-1}$ respectively. Since the power spectrum is thought to be proportional to k^{-3} (Iliev et al., 2008), the line $w(k_\parallel)^2/k_\parallel^3$ indicates the $k_\parallel$ to which we are most sensitive.

The 0.5 MHz foreground filter reduces the peak and RMS values in the sky images by a factor of 50, as can be seen in Figure 5.2. The filter is most effective within the primary beam, where even point sources are removed without any visible residuals, while towards the edge of the beam the residuals grow more significant. Figure 5.3 demonstrates the dominance of RFI in the filtered dirty maps, and the improvement that the RFI removal via SVD provides. By using the same data both before and after applying the RFI removal, we find the maps in Figure 5.2 would have been a factor of four noisier.

5.3 Differences between nights

Since RFI and foreground signals are so much larger than the EoR signal, small errors in the subtraction of foregrounds could easily result in a spurious signal. To gauge how successful a power spectrum measurement might be, we are interested in the relative similarity of the different nights, which we can gauge by taking the differences of visibilities. This serves, in part, as a test of the importance of ionospheric fluctuations. Though the ionosphere is calibrated for along the line of sight to the pulsar in the centre of the field, the ionosphere could change across the field of view. Pairs of days for which this change away from the field of view is different will not subtract well, while days which do subtract well will show mostly noise. In principle issues such as RFI or antenna pointing

[1]More precisely it is a boxcar average with a variable width, and therefore a somewhat different curvature.

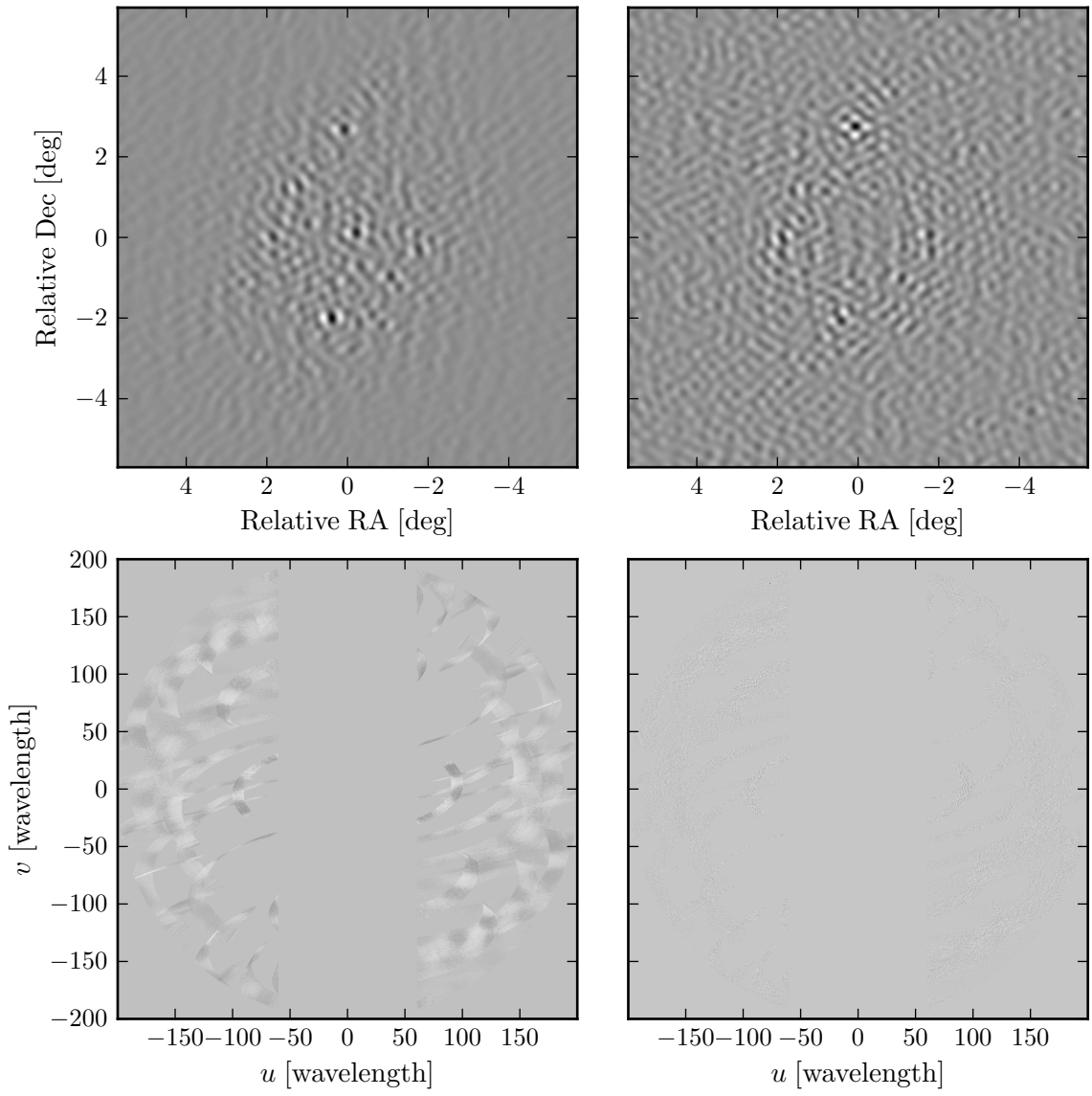


Figure 5.2: Data from Dec 10. The top row is the sky image with an 11.4 degree field of view, with $|\mathbf{u}| < 200$ and $|u| > 60$ binned in the (u, v) plane by $\Delta u = 5$, while the bottom row is the visibilities in the same range with $\Delta u = 0.4$ to show structure. The left column is before any foreground subtraction. In this image the dominant source is B2 0825+24 (a.k.a. 4C 24.17) just south of the full width at half maximum (FWHM) of the primary beam, with a peak value of 2.2 Jy. RMS within the beam is 343 mJy. The right column is after a 0.5 MHz subtraction. The peak value of this image is 47 mJy with an RMS of 6.2 mJy, lower by a factor of about 50. If put on the same grey-scale as the image without foreground subtraction no features would be visible. The dominant source after this filter is 3C 200, well outside the beam.

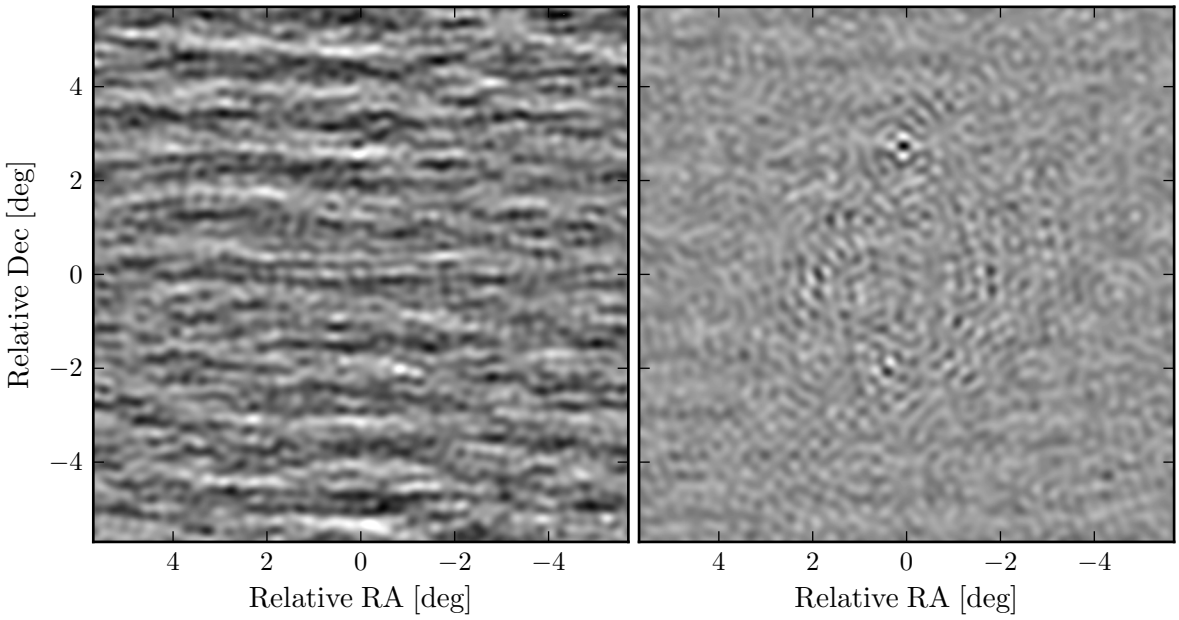


Figure 5.3: Dirty sky images with $|\mathbf{u}| < 200$ after the 0.5 MHz foreground subtraction. The left image is without any SVD RFI removal and the right image is with the RFI removal. As in Figure 5.2, the field of view is 11.4 degrees. RFI clearly dominates the map on the left, with peak value of 73 mJy and an RMS of 18 mJy. After RFI removal, the peak drops to 32 mJy with an RMS of 4.2 mJy.

errors which change from day to day will exhibit different signatures in the differences between days. Visibilities are only used when data exists at the same position on both days. The results of the subtractions are shown in Table 5.1.

It can be seen in Table 5.1 that December 11 gives consistently poor results in both the unfiltered and filtered images, and was thus excluded from all subsequent analysis. From these maps we can also conclude that RFI is not dominating the differences from day to day. If this had been the case, since RFI is typically isolated in (u, v) space, it would be visible across the whole sky image, including far outside the primary beam. However, it can be seen in Figure 5.2 that structure decreases rapidly away from the centre of the field, indicating an astronomical source. This is true of all subtraction pairs.

Maps of the subtracted visibilities were inspected visually for systematic patterns in the residuals, the nature of which can indicate different sources of error. For example, a pointing difference between two days will create closely spaced positive/negative peaks along the axis of the difference for every point source. Timing errors create a similar dipolar residual, oriented perpendicularly to the radial direction. When introducing a delay in the timestamp of the first day in a subtraction, the positive lobe of a point source residual will be on the westward side for a source to the north of the pointing centre and on the eastward side for a source to the south, with the separation of the lobes larger for sources further from the centre. A one minute delay (a single integration at this stage) is enough to make this effect visible. Neither timing nor pointing errors were detected.

5.3.1 A possible variable source

Comparing differences between days could also potentially be used to identify variable sources in the field, since the variability can stand out relative to residuals in the difference map more than the intrinsic flux of the source compared to other sources in the field.

Only one such candidate was found located approximately one degree southwest of

		Unfiltered		With 2 MHz filter	
Subtracted pair		Peak Flux	RMS	Peak Flux	RMS
Dec 10	Dec 11	1856.1	227.3	137.9	19.9
Dec 10	Dec 14	888.4	185.6	62.5	11.6
Dec 10	Dec 16	278.4	49.8	27.1	3.2
Dec 10	Dec 17	447.9	64.9	26.2	3.6
Dec 10	Dec 18	548.3	70.4	32.5	4.2
Dec 11	Dec 14	1037.8	145.1	256.1	56.5
Dec 11	Dec 16	2148.3	256.7	158.1	23.9
Dec 11	Dec 17	2528.2	318.5	106.1	21.1
Dec 11	Dec 18	2997.5	356.1	159.1	12.4
Dec 14	Dec 16	1221.5	193.9	50.2	7.0
Dec 14	Dec 17	1311.1	257.2	50.4	7.6
Dec 14	Dec 18	1433.5	233.0	67.1	4.9
Dec 16	Dec 17	202.0	78.9	23.8	2.8
Dec 16	Dec 18	224.2	31.2	24.4	3.4
Dec 17	Dec 18	206.9	60.6	24.6	3.2

Table 5.1: Peak flux and RMS of the difference of each two day pair available, in mJy, with a maximum (u, v) distance of 600, which corresponds to a maximum baseline of 1.2 km. All values are in mJy. To remove foregrounds, a 2 MHz linear filter was applied.

B0823+26, at J2000 8h21m41.6 +25°54'24". It appeared as a point source in the residual maps between 1 and $4\,\sigma$ above the RMS, implying a possibly daily variation of order 10 mJy. There is no source visible at this position on any of the maps. While pulsars are a natural candidate for a variable radio source, the ATNF Pulsar Catalogue lists no other pulsars in the field. There are several SDSS sources within an arcminute of this position, but the only nearby radio source in the NRAO VLA Sky Survey (Condon et al., 1998) is NVSS J082156+255833, over 5 arcminutes away. The NVSS source catalogue lists the flux as 2.5 mJy, which is below our RMS flux limit for individual nights.

It is unlikely that this candidate corresponds to the NVSS source and there is no other evidence to suggest that it is a real source. Since the variability is consistent with the noise in the maps, it will not impact the EoR measurement, and so we do not investigate it further.

5.4 Power Spectra

While in principle one could model and subtract the foregrounds from a single day and calculate the power spectrum from that data alone, any residual RFI and noise would be amplified in the auto-correlation of each visibility, making a clean measurement of the EoR spectrum very difficult to make directly. To avoid this, we use cross-correlations between as many pairs of days as possible, for which noise and, hopefully, RFI, are uncorrelated.

The power spectrum of sky structure can be determined directly from the visibilities (Zaldarriaga et al., 2004). Calculating the power spectrum from the visibilities instead of from the correlations in the sky image takes advantage of the fact that the visibilities have a nearly diagonal correlation matrix in the noise (White et al., 1999). To find the cross-power of two days, we grid the visibilities of each day with a cell spacing equal to the size of the beam ($\Delta u = 20$) and take the product of the two days in each cell. The noise

is calculated by taking the variance of the data within a single visibility before rebinning (i.e., at the full 0.25 second and 8 kHz resolution; `noisevar`, `covar2var`, and `mergevar`). When gridding, visibilities are weighted "naturally", by the inverse of this noise. We then find the weighted average of visibilities in annuli of (u, v) space which gives the power in units of Jy^2. Since the amount of data at large $|\mathbf{u}|$ decreases, we increase the width of each successive annuli by 60 per cent with increasing $|\mathbf{u}|$. The smallest bin width is equal to the size of the beam. This prevents large artificial variability in power due to sparse sampling.

Since we expect our sensitivity to the EoR signal to diminish rapidly with increasing baseline length, we look only at the first few points, averaged over all possible cross-correlations. Additionally, it is known that the RFI removal will introduce a loss of power at low $|u|$ because the SVD can not distinguish well between stationary RFI sources and the part of the sky signal that rotates slowly. To avoid this, we impose a limit of $|u| > 60$ when taking the cross-correlations, excluding baselines which are almost entirely along the north-south direction. This limit was determined by requiring that the power spectrum before and after the SVD differ by less than 1σ as shown in Figure 5.4. The part of the (u, v) plane that is lost with this cut was shown in Figure 3.6.

The power spectrum of the cross-correlation can be converted to units of K^2 using

$$\frac{l^2}{2\pi} C_\ell \Big|_{\ell=2\pi|\mathbf{u}|} = \left(\frac{|\mathbf{u}|}{11.9}\frac{3.3^\circ}{\theta_b}\right)^2 \left(\frac{150\,\mathrm{MHz}}{\nu}\right)^4 \left\langle \left|\frac{V(\mathbf{u})}{\mathrm{Jy}}\right|^2 \right\rangle \mathrm{K}^2 \tag{5.4}$$

where $\ell = 2\pi|\mathbf{u}|$, $\mathbf{u} = (u, v)$ is the visibility coordinate in units of wavelength, C_ℓ is the power measured in K^2, θ_b is the primary beam size, and ν is the wavelength (Pen et al., 2009). The quantity in the angle-brackets on the right is equal to the power in Jy^2 found above. This conversion is written in terms of the GMRT observations with a primary beam of $\theta_b = 3.3^\circ$ and $\nu = 150\,\mathrm{MHz}$. If gridding in the (u, v) plane is too fine, the data becomes noisy, while very coarse gridding requires the assumption that the data is constant across the whole cell. As mentioned, we use a gridding equal to the size of the beam. A fully optimized estimate would require a maximum likelihood code for

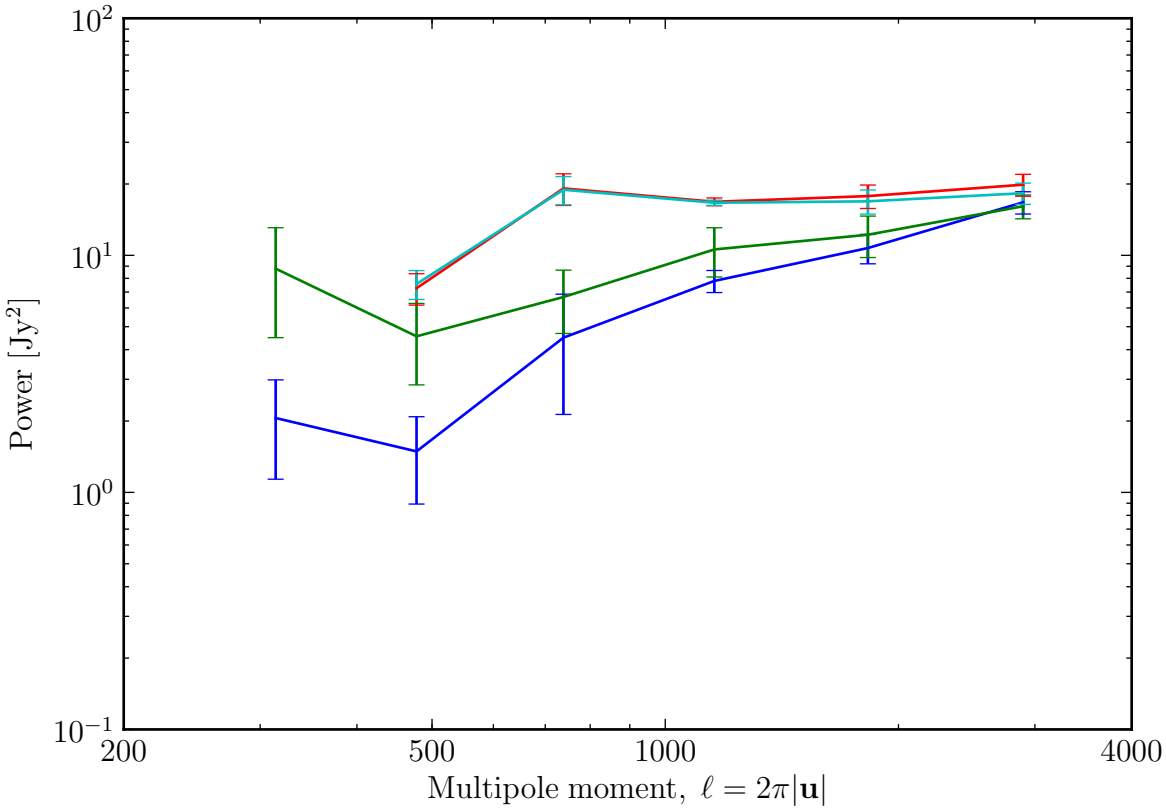


Figure 5.4: Comparison of the cross-power of December 10 with all other days under four different conditions. The green and blue lines are before and after the SVD RFI removal, respectively, including all u. It can be seen that power is lost in the SVD. Similarly, the cyan and red lines are before and after the SVD step, respectively, this time with a $|u| > 60$ limit imposed. In this case the two lines are almost identical, meaning the SVD had little effect on the total power after removing the low $|u|$ regions.

projecting the beam onto the (u, v) plane. This has been worked on in the context of the Cosmic Background Imager (CBI) gridder (Myers et al., 2003), but requires a much more complex analysis that we do not attempt here. Figure 5.5 shows the weighted-average of all cross-correlation pairs, excluding December 11. Bernardi et al. (2009) reported a power spectrum of foregrounds without subtraction in the galactic plane at a level comparable to our measurement both here and in Pen et al. (2009).

We have plotted the power spectra with 2σ bootstrap errors (Efron, 1979) which were derived as follows: Using five days of data, there are ten possible cross-correlations. From these, ten are randomly sampled, with replacement, resulting in a slightly different power spectrum. This is repeated 10^4 times, and the variance on this set of power spectra is calculated to give the error on the original. Formally this quantifies the error when taking independent samples of a statistical distribution. In our case this is not a rigorous error, but a suitable straightforward estimate given the main complications discussed earlier.

5.5 Constraints on the EoR

5.5.1 Cold vs. warm intergalactic medium

The brightness temperature of the 21 cm line relative to the CMB is determined by the ionization fraction of hydrogen and the spin temperature of the neutral population, which is in turn governed by the background radiation and the kinetic temperature of the gas (Purcell & Field, 1956; Field, 1959; Furlanetto et al., 2006a). Reionization requires a minimum expenditure of 13.6 eV of energy per hydrogen atom. In contrast, Ly-α photons, when absorbed by a neutral atom, are quickly re-emitted. Ly-α photons are said to undergo "resonant scattering" and each (multiply-scattered) photon can affect many atoms before cosmic redshifting makes them ineffective (Wouthuysen, 1952; Field, 1959; Chuzhoy & Shapiro, 2006). This means Ly-α pumping of the hyperfine transition requires only about 1 per cent of the UV flux required for ionization. Assuming a gradual

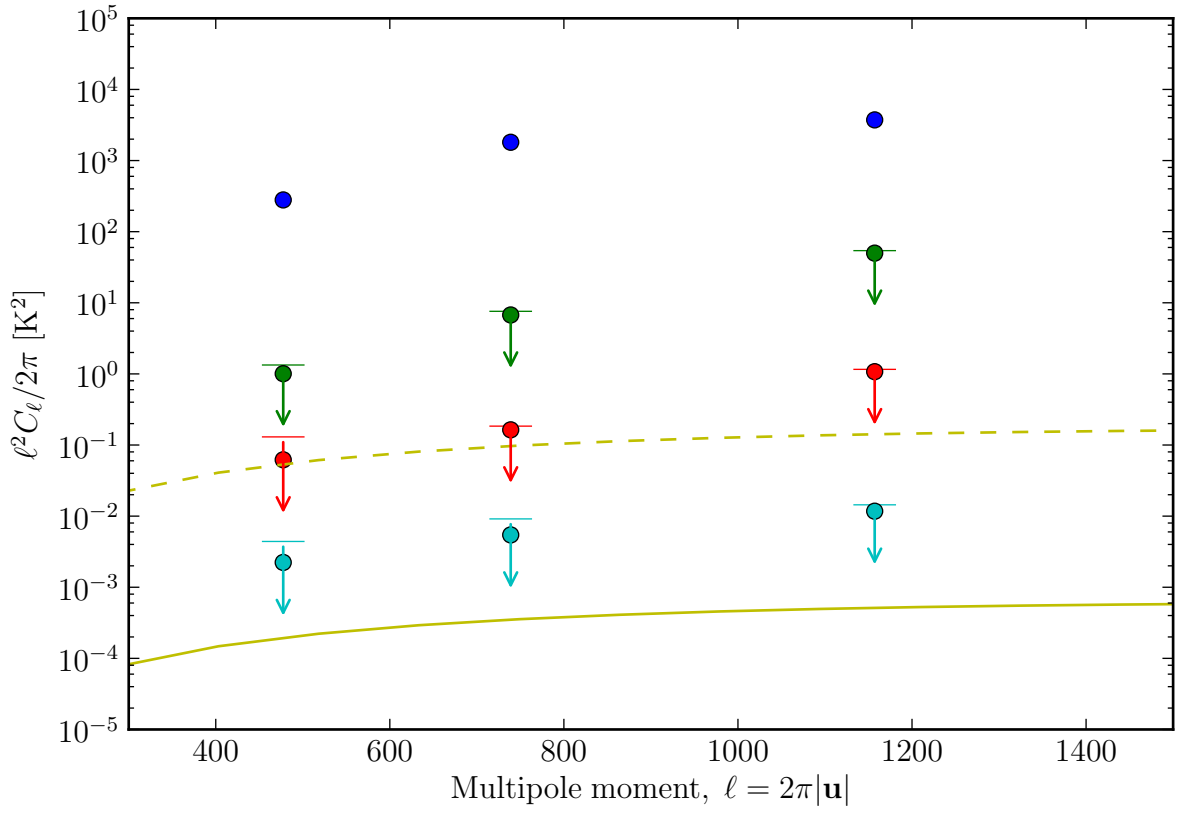


Figure 5.5: Average power spectrum in units of K^2 of all combinations of days, excluding December 11, as a function of the multipole moment ℓ. Each point is shown with a 2σ upper limit derived from a bootstrap error analysis, which is in most cases smaller than the size of the point. The points are logarithmically spaced as described in the text, from left to right covering the ranges $377 < \ell < 578$, $578 < \ell < 899$, and $899 < \ell < 1414$. Blue points are the power before subtracting foregrounds, green are after 8 MHz mean subtraction, red are after 2 MHz mean subtraction, and cyan are after 0.5 MHz subtraction. The solid yellow line is the theoretical EoR signal from Jelić et al. (2008), and the dashed yellow line is the theoretical EoR signal with a cold absorbing IGM as described in section 5.5.2.

increase in UV flux with time, well before flux levels for reionization are reached, Ly-α pumping will couple the spin temperature to the kinetic temperature of the gas. If there was no source of heat at that era other than a weak UV flux, the gas kinetic temperature must have been at its adiabatic expansion value of 1.7 K at $z = 8.5$, and neutral gas will produce a signal due to absorption of CMB photons in excess of stimulated emission and ionized structures would be seen as low brightness regions on the sky (Chen & Miralda-Escudé, 2004). In such a cold-gas model the brightness temperature of the neutral gas against the CMB can be as low as -500 mK.

A small fraction of the mass will have collapsed into minihaloes, which have a temperature higher than this adiabatic temperature (Shapiro et al., 2006). This fraction may even be substantially smaller due to non-perturbative velocity flows (Tseliakhovich & Hirata, 2010; Dalal et al., 2010). No study was found on the impact of these non-linear effects on the Ly-α pumped intergalactic medium (IGM) temperature. At linear order, the positive and negative over and under densities cancel, and the non-linear collapse fraction is still small, so we expect the realistic value to be similar to the adiabatic prediction.

Alternatively, X-rays from supernovae or quasars might have heated the IGM above the CMB temperature of ≈ 30 K before reionization (Madau et al., 1997; Chen & Miralda-Escudé, 2004). X-ray heating between 30 K and 10 000 K will result in a largely neutral but warm IGM, which would be seen in emission. In the limit that the spin temperature $T_{\rm s} \gg 30$ K, the volume emissivity becomes independent of temperature. Patchy X-ray heating can also result in large angular scale structure (Alvarez et al., 2010). The sky brightness temperature of the neutral gas has an asymptote at ~ 30 mK (Furlanetto et al., 2006a).

The cosmic luminosity of X-rays at $z \sim 9$ is not known and difficult to estimate (see e.g., Dijkstra et al., 2004; Salvaterra et al., 2007; Guo et al., 2009). If the rate of core collapse supernovae at high redshift matches that of today, the X-ray output at $z \sim 9$

would have been sufficient to raise the IGM temperature above the CMB temperature at the onset of reionization. However, the mechanism of and factors affecting core-collapse are not known and most numerical models appear to generically not result in supernovae at all (Mezzacappa, 2005). It is possible that at high z, core collapse supernovae were not as abundant as today, and the IGM was still in absorption during the EoR. We therefore consider two limits: one where the IGM is still cold and $T_{\rm b} = -500\,{\rm mK}$, and another where it is heated above CMB and $T_{\rm b} = 30\,{\rm mK}$. A general parametrization of these scenarios was recently proposed by Pritchard & Loeb (2010). Since the actual situation at high z is not clear, we will interpret possible constraints using both the cold (unheated by X-rays) and warm IGM cases.

5.5.2 Comparison to simulations

The results of Figure 5.5 can be compared to simulated results from the Low Frequency Array (LOFAR) EoR project in Jelić et al. (2008), which assumes $T_{\rm s} \gg T_{\rm CMB}$. At low ℓ, their simulated EoR signal is approximately $(10\,{\rm mK})^2$, while our lowest point with a similar 0.5 MHz bandwidth filter is $(50\,{\rm mK})^2$ with a 2σ upper limit of $(70\,{\rm mK})^2$. These results are comparable to the sensitivities LOFAR would expect after 400 hours of their EoR project. We have also considered the case where reheating of the IGM does not occur, so the spin temperature remains coupled to the kinetic temperature of the gas (Ciardi & Madau, 2003). In this case, the IGM cools adiabatically after decoupling from the CMB at $z \approx 150$. The temperature fluctuations scale with $(1 + T_{\rm CMB}/T_{\rm s})$, and the power scales with the same factor squared. Using $T_{\rm k} = T_{\rm CMB}(1+z)/150$ at $z = 8.6$, the power becomes approximately 275 times larger. This line is shown in Figure 5.5, and is comparable to the data. Scaling up the warm IGM power spectrum from Iliev et al. (2008) or Jelić et al. (2008) in this way is a reasonable approximation to the expected signal in a cold IGM. For more detailed studies of the signal in such an absorption regime, we point the reader to Baek et al. (2009) and Baek et al. (2010).

The power spectrum of reionization is intrinsically three dimensional (Morales & Hewitt, 2004). The strongest constraints on the 3D power spectrum for the $\Delta\nu = 2$ and 0.5 MHz foreground filter cases are shown in Figure 5.6. This uses $k^2 = k_\parallel^2 + k_\perp^2$, where $k_\parallel$ is given by the windowing function of the filter and $k_\perp \approx \ell/6\,h^{-1}$Gpc.

5.5.3 Comparison to a toy model

For a physical interpretation, it is useful to compare to a simple reionization model. We consider an idealized case in which ionized bubbles are of uniform scale and non-overlapping. Then for a given k there will be a characteristic bubble radius R at which the power is maximized. The Fourier transform of the brightness temperature profile for perfectly ionized bubble with radius R is given by

$$T(k) = 2\pi \int_0^R e^{ikr\cos\theta} \sin\theta \,\mathrm{d}\theta\, r^2 \,\mathrm{d}r. \tag{5.5}$$

By requiring that the universe is 50 per cent ionized, one can show the power is given by

$$\frac{k^3}{2\pi^2} P(k) = \frac{3T_\mathrm{b}^2}{2\pi} (kR)^3 \left[\frac{\cos kR}{(kR)^2} - \frac{\sin kR}{(kR)^3}\right]^2 \tag{5.6}$$

and is maximized when $kR \approx 2.5$. In this case, $k^3/(2\pi^2)P(k) \approx T_\mathrm{b}^2/5$, where T_b is the brightness temperature ≈ 30 mK in an X-ray heated IGM or almost -500 mK in a cold absorbing IGM. This signal would be more than an order of magnitude larger than the predictions by e.g. Iliev et al. (2008) or Jelić et al. (2008). In Figure 5.6 we have included the power spectrum from this model with k chosen to maximize the power in the range of interest. The data currently imposes a limit on the size of bubbles in this single-scale model. Our upper limits with a 0.5 MHz foreground subtraction rule out bubbles with diameters from 2.2 to 12.4 h^{-1}Mpc in the redshift range $8.1 < z < 9.2$. The cold IGM constraint is applicable even in the case of simulations, since only UV photons are included which themselves do not heat the IGM.

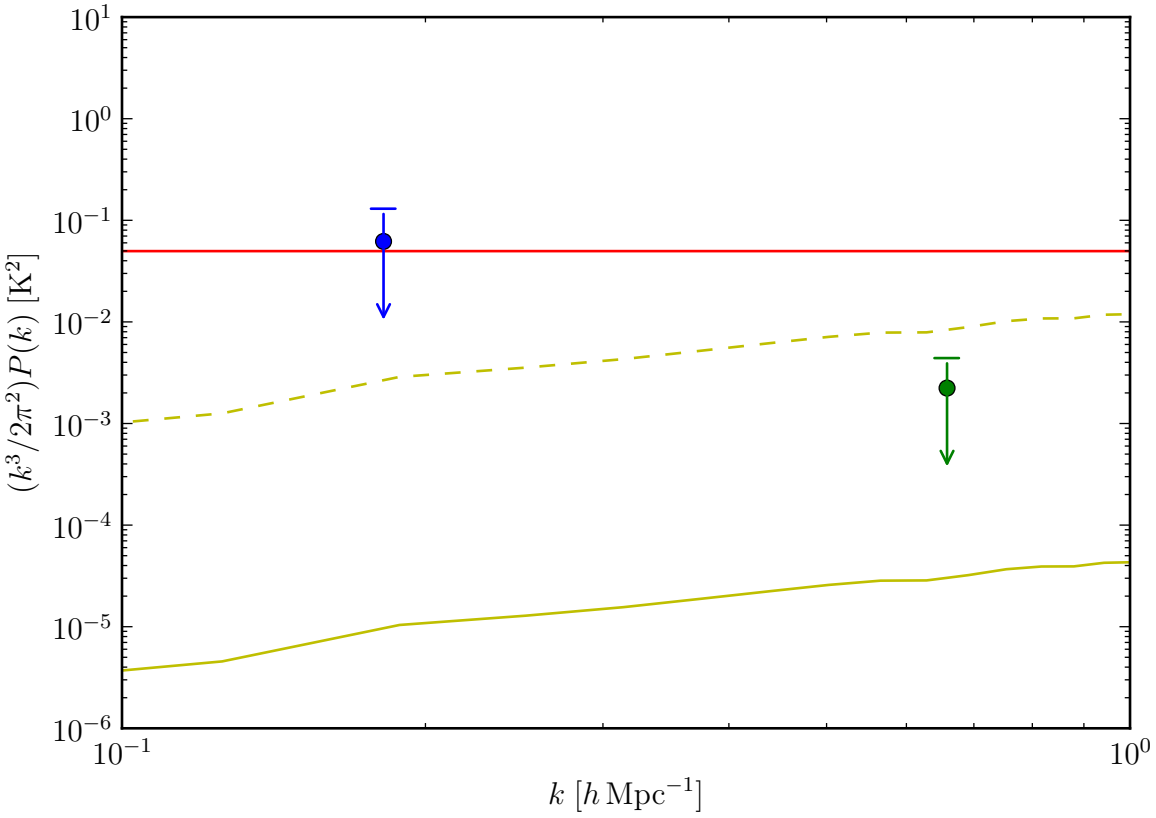


Figure 5.6: 3D power spectrum for the same data shown in Figure 5.5, using $k^2 = k_\parallel^2 + k_\perp^2$. This is dominated by $k_\parallel$, so the bin width in $k_\perp$ does not influence the horizontal position of the limits. The strongest constraints from the 2 MHz and 0.5 MHz filters are shown (blue and green, respectively). Upper limits are 2σ bootstrap errors. Three possible signals are shown. The solid yellow line is the prediction from Iliev et al. (2008) and the dashed yellow line is the same for a cold IGM. The red line comes from the single-scale bubble model as described in the text for a cold IGM, using bubble radii of $R = 2.5/k$ to show the maximum power at all k. For the two points shown, the bubble diameters which achieve this maximum power are 27 and 7.4 h^{-1}Mpc respectively. Only the 0.5 MHz point imposes a limit on the diameter. For a warm IGM case, this signal would be reduced by the same factor as in the two dashed lines.

5.6 Conclusion

The data analysis has been completed on six days from December 2007 with a noise level of approximately 2 mJy on most nights. The SVD removal strategies for broadband RFI used lower noise by a factor of four in temperature, or sixteen in power, which flagging alone cannot achieve. We have also tested for ionospheric variations and found that our pulsar calibration is sufficient for dealing with these effects. After RFI removal and foreground subtraction, we have measured a power spectrum which represents a preliminary upper limit on the 21 cm brightness temperature fluctuations during the EoR. These results can be used to constrain assumptions about the state of the IGM at these times, particularly in the case of a Ly-α pumped, but cold, IGM.

The previous best limit on 21 cm signal at comparable redshift was by Bebbington (1986), who reported no features down to 5 K at $z = 8.4$. When this analysis was first published, Parsons et al. (2010) had reported a similar limit of about 5 K using PAPER. The upper limit we present here is approximately 70 mK on the variance in 21 cm brightness temperature at $z = 8.6$, almost two orders of magnitude better than these previous limits. Residual foreground contamination and RFI may still be contributing to this power, but the EoR signal remaining can not be larger than this.

These results, however, do not include any corrections for possible signal loss due to the foreground subtraction itself, which will be shown in Chapter 6 to be a significant factor in setting robust limits on the EoR power spectrum. The $(70\,\mathrm{mK})^2$ limit we have reported here was an initial estimate that was later revised with more advanced techniques discussed in the next chapter.

Chapter 6

Quantification of Signal Loss with an SVD Foreground Filter

In this chapter we explore the use of a singular value decomposition to remove foregrounds with fewer assumptions about the foreground structure using the same data from December 2007. Using this method we also quantify, for the first time, the signal loss due to the foreground filter and present new power spectra adjusted for this loss, providing a revised measurement of a 2σ upper limit at $(248\,\mathrm{mK})^2$ for $k = 0.50\,h\,\mathrm{Mpc}^{-1}$. While this revised limit is larger than previously reported in Paciga et al. (2011), we believe it to be more robust and still represents the best current constraints on reionization at $z \approx 8.6$. These results have been previously published in Paciga et al. (2013), Monthly Notices of the Royal Astronomical Society, Vol. 433, Issue 1, p. 639–647 (astro-ph/1301.5906).

6.1 Data analysis refinements

After the preliminary analysis shown in Chapter 5, several aspects were identified that would improve the power spectrum result. These refinements, which were not part of the analysis in earlier chapters, are discussed here.

6.1.1 Manual flagging of data

In addition to automated flagging of line interference in visibilities and the SVD RFI removal pipeline for broadband interference, we have also added manual flagging of faulty antennas, timestamps, and frequency ranges that are exceptionally noisy or that visually appear to have RFI left after the automated procedures. These were identified primarily by eye in the visibility maps such as those shown in Figure 5.2.

Since the imaginary part of the pulsar-only visibilities should have no structure at all after a successful calibration, these are particularly illustrative for identifying problematic data ranges, and in quantifying the improvement from manual masking. A useful metric is the pulsar calibration dynamic range, which we define as the ratio of the flux of the brightest source in the sky map to the RMS noise in the imaginary part of the map of the pulsar alone. Since it is possible to be overzealous when masking by eye, this provides a check to indicate whether a suggested mask actually provides an improvement to the image quality.

Approximately 15 per cent of the visibilities are flagged with this procedure, using the masks given in Table 6.1. This improves the dynamic range by as much as a factor of four to approximately 8 000 to 11 000, which is comparable to what might be expected based on the pulsar flux, integration time, and system temperature of GMRT, taking into account that the pulsar pulse is only present for 1/16th of the time. A large portion of the masks is due to antennas which were not functioning, and the rest cover less data than the typical RFI occupancy rate at 150 MHz found by several studies mentioned in Chapter 3, suggesting that most RFI has already been removed by this stage.

6.1.2 LST regridding and pulsar subtraction

As discussed in section 5.3, when cross-correlating it is important that the signal in both data sets is as similar as possible. While all days from December 2007 were recorded

Date	Antennas	Timestamps	Frequencies
Dec 10	S01	*	*
Dec 10	*	374–418	*
Dec 11	C02, C04, C11, W04, S01	*	*
Dec 11	*	*	2–6, 23–28
Dec 14	W04, E05	*	*
Dec 14	W03, W06	295–319	*
Dec 16	E05, S01, S02	*	*
Dec 16	*	463–467	*
Dec 17	C08, E06, S07	*	*
Dec 17	W04	189–195	*
Dec 17	*	146–173, 309–334, 345–425	*
Dec 18	n/a	n/a	n/a

Table 6.1: Antennas, frequency ranges, and timestamps that have been manually masked. An asterisk indicates that the mask applies to the entire range, e.g. the first row shows that antenna S01 was flagged for all times and frequencies. The timestamps range from 1 to 467 in one minute intervals from a local sidereal time of 03h27m to 11h10m. The frequency numbers start with 1 at 156 MHz and progress to lower frequencies up to channel 64. No masks were applied for December 18.

over an 8 day period, the exact start and end time of each night varied. It is possible that the tail ends of a night may be gridded into the same cell as data from a different local sidereal time (LST) on another night without the corresponding contribution from the same LST.

To improve the comparability of each of the five nights of observing, we limit each night to the same LST range, which still provides a common window of slightly under 8 hours. We have also regridded the visibilities in time by taking the average of visibilities located on either side of timestamps spaced exactly one minute apart, weighted by the time separation from the timestamp, such that each night shares the exact same timestamps.

It is also known that the flux of a pulsar can change significantly with time, which creates another source of variability from night to night. Since the visibilities are gated on the period of the pulsar, isolating the on-pulse gates gives us a perfect model of the array's response to the pulsar flux in visibility space. Using the flux of the pulsar in the full sky map (made using all gates), the pulsar visibilities are scaled appropriately and subtracted from the data before calculating power spectra.

The combined effect of imposing common sampling in LST between each pair and subtracting the pulsar in this way reduces the RMS noise in the difference between pairs of days by a factor of two on average.

6.1.3 Mean vs. median power

When calculating the power spectrum from the weighted mean visibility in annuli of (u, v) space, it was found that outliers tended to skew the mean power in each annulus despite the noise weighting. Since a median is much more robust to such outliers, from this point forward we calculate the power in each annulus (that is, at each angular scale ℓ) as the median value including visibilities from all frequencies. Rather than the standard deviation, which is also sensitive to outliers, the error in each annulus is estimated as the

median of the absolute deviations from the median power, again weighted by the noise.

Finally, as in Chapter 5, the total power spectrum is calculated from the bootstrapped average of the power spectra over all ten possible cross-correlation pairs of the five nights. This gives the 2D power perpendicular to the line of sight as a function of baseline length $|\mathbf{u}|$ or equivalently multipole moment $\ell = 2\pi|\mathbf{u}|$. Since we do not yet include line of sight information, this is the power as a function of ℓ with fixed $k_\parallel = 0$, which we denote $P(\ell|k_\parallel = 0)$.

6.2 Singular value decomposition

Foreground removal techniques typically rely on the fact that the foreground signal is expected to be much smoother in frequency (that is, has fewer degrees of freedom) than the reionization signal, which decorrelates on the order of one to a few megahertz (Bharadwaj & Ali, 2005). Observations of foregrounds around 150 MHz with GMRT have shown that the fluctuations in frequency are large enough to make polynomial fits insufficient to model them (Ali et al., 2008; Ghosh et al., 2012). Our group had previously used a piecewise-linear fit. Here, we instead use a singular value decomposition (SVD), which still isolates smooth foreground modes but does not make *a priori* assumptions that the foregrounds can be approximated with a particular function. A similar technique has been used by Chang et al. (2010) and Masui et al. (2013) to clean foregrounds for HI intensity mapping at $z \approx 0.8$, where the relative dominance of foregrounds over the 21 cm signal is comparable to $z \approx 8.6$. For reionization, Liu & Tegmark (2012) have developed a framework for using SVD modes of a frequency-frequency correlation matrix to clean foregrounds at MWA.

We perform an SVD for each baseline individually on the visibilities arranged in a matrix by time and frequency. This is functionally similar to the RFI removal technique used in Chapter 3, with the distinction that the visibilities have been fringestopped to

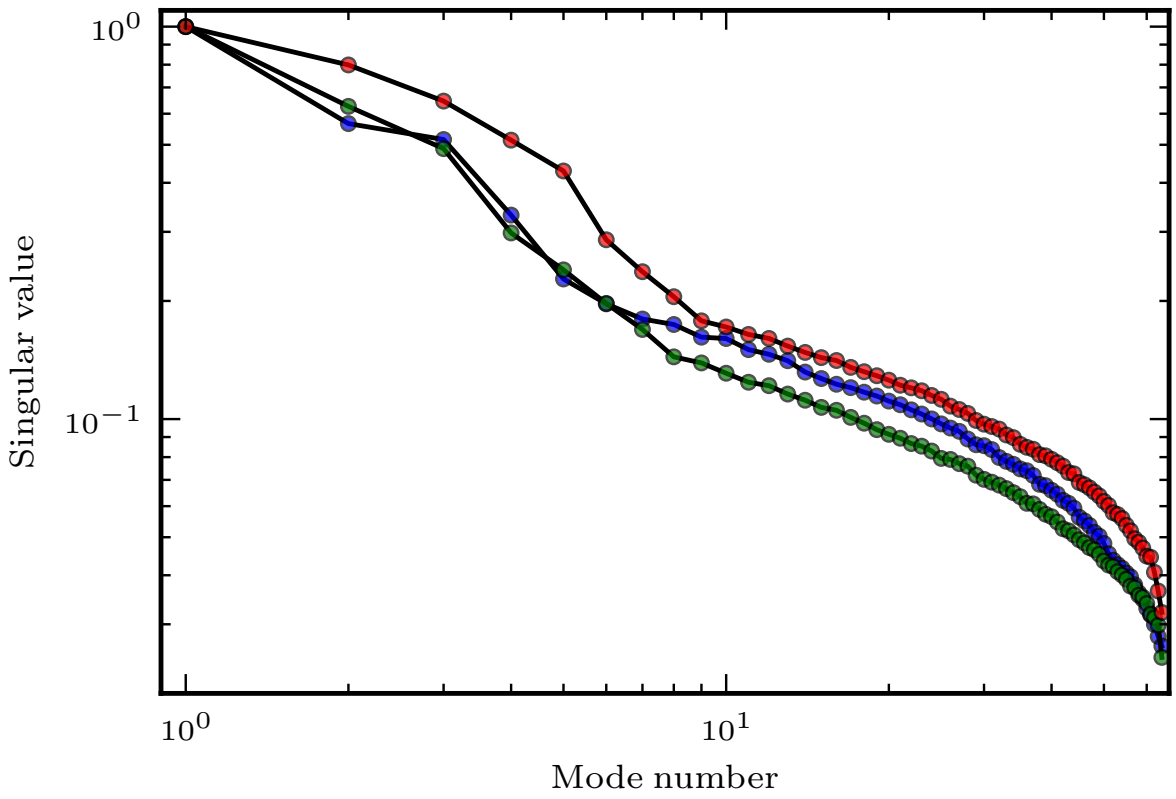


Figure 6.1: An example singular value spectrum for the three shortest baselines, each approximately 50–52 wavelengths at the zenith, with the largest singular value normalized to 1 for each.

the sky frame, such that now slowly varying (in both time and frequency) modes on the sky will contribute to the largest eigenvalues. The number of modes is limited by the 64 frequency channels. Figure 6.1 shows the singular values for the shortest baselines. The spectra of values on a given baseline is generally consistent from day to day, but occasionally there are large jumps in both amplitude and rate of decline with mode number, which are likely due to either RFI or calibration errors. In these cases, the noise on the baseline also becomes much larger, such that in the final calculation of the power spectrum their contribution is significantly down-weighted.

A sky image using 8 hours of data from a single night is shown in Figure 6.2, compared with the same data after the first eight SVD modes, shown individually in Figure 6.3, are

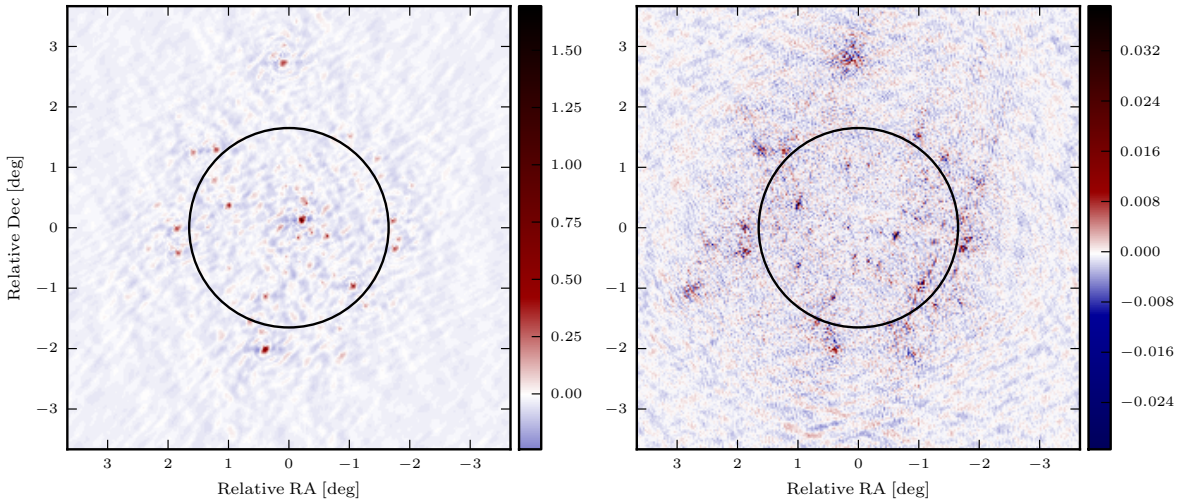


Figure 6.2: Sky image before and after an SVD foreground subtraction, for the night of December 10, 2007, using baselines up to 4 km. The colour scales are in units of Janskys and the black circle shows the full width half maximum (FWHM) of the primary beam. The RMS before any foreground removal (left) is 50 mJy. After removing eight SVD modes (right) the peak goes from 1.6 Jy to 39 mJy with an RMS of 2 mJy. Residual point sources can still be seen around the edge of the beam while those within about one degree of the centre are effectively removed.

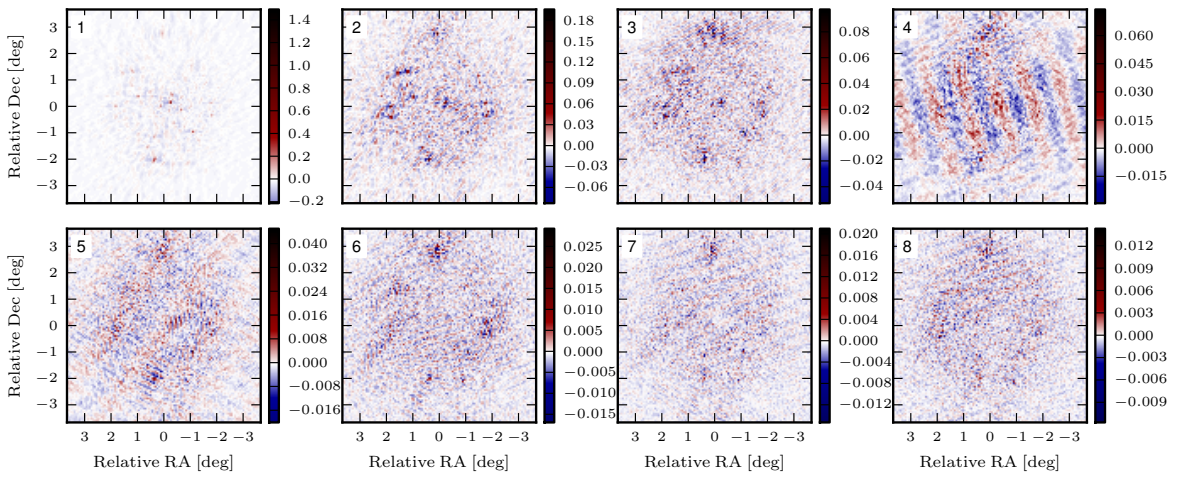


Figure 6.3: Sky images of the top eight SVD modes identified in the data from December 10, 2007, with all other modes set to zero. These modes are the ones subtracted between the two sky images in Figure 6.2. The colour scales are in units of Janskys.

removed. The overall flux is reduced substantially after only a few modes are removed. While the sources in the centre of the field are removed quite well, the dominant residuals are the point sources near the edge of the beam. This is generically true of any foreground subtraction used on this data set, as was also seen in Paciga et al. (2011) (Chapter 5). This is most likely due to beam edge effects, the worst residuals being close to the first null where the frequency dependence of the beam pattern is most significant. Though there are sophisticated schemes that may be able to model point sources while minimizing the impact on the 21 cm signal (e.g., Datta et al., 2010; Bernardi et al., 2011; Trott et al., 2012), at the angular scales we are interested in for this work ($\ell \lesssim 2000$) the point sources are confusion limited and contribute in the same way as the diffuse background.

Each night goes through the SVD foreground removal separately, and then the cross-correlations are used to arrive at a power spectrum using the same method described previously. The spectra for several numbers of SVD modes removed are shown in Figure 6.4.

6.3 Signal loss due to filters

A general problem with any foreground removal strategy is that it is impossible to completely separate the foregrounds from the signal, such that the foreground removal will likely remove some signal as well. Early work by Nityananda (2010) used a simple model of an SVD applied to a single visibility matrix to show that the signal loss could be calculated analytically. Our method of using an SVD for each baseline independently is more complex, and we wish to estimate the signal loss directly from the data itself. To quantify the signal loss, we aim to find the transfer function between the observed power $P_{\mathrm{SVD}}(\ell)$ and the real 21 cm power $P_{21\mathrm{cm}}(\ell)$. Since the real power is unknown, we use a simulated signal as a proxy. This is added to the data before the foreground subtraction and the resulting power spectrum after subtraction is compared to the input signal.

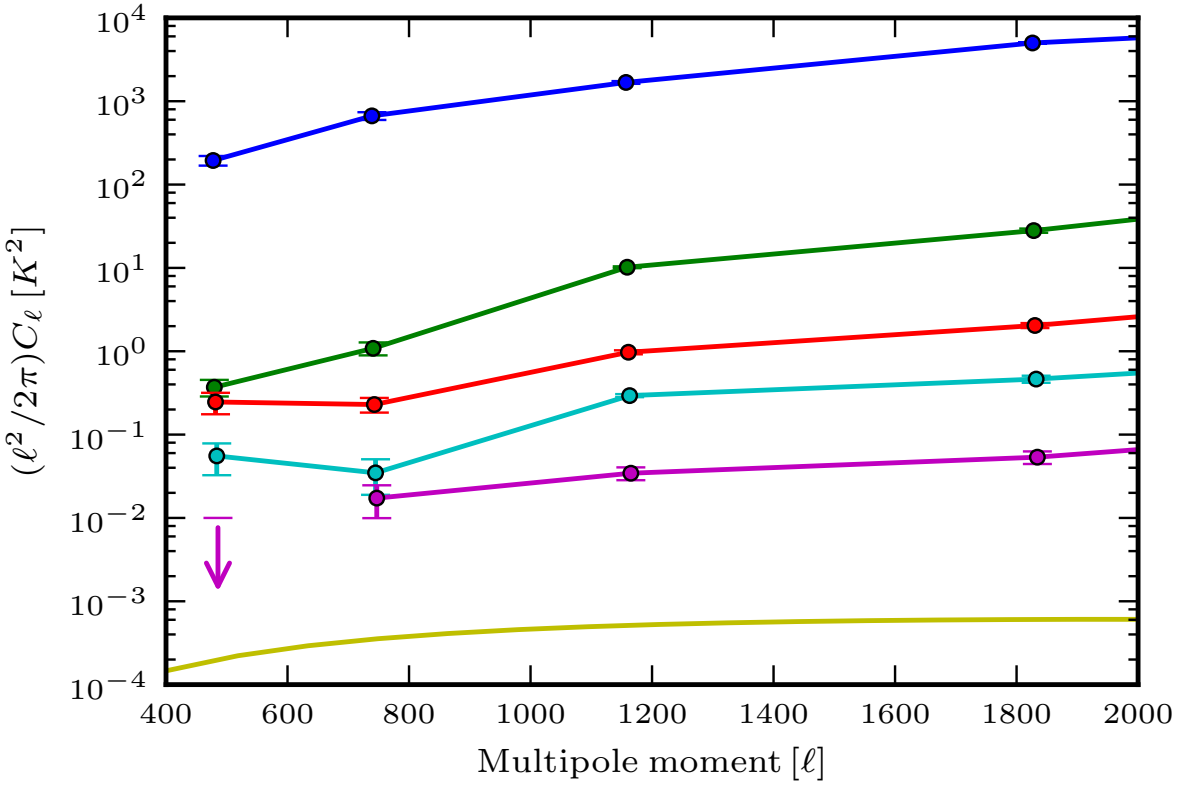


Figure 6.4: Power spectra before and after SVD mode removal. The blue line is before any modes are removed. The green, red, cyan, and purple lines are for 4, 8, 16, and 32 modes removed, respectively. The error bars are from a bootstrap analysis of all cross-correlated pairs from the 5 nights of data. The solid yellow line is the theoretical signal from Jelić et al. (2008).

6.3.1 Simulated signal

The simulated signal we use is a Gaussian random field with a matter overdensity power spectrum from CAMB[1] scaled to $z = 8.6$ using the linear-regime growth from $z = 1.5$, and with the amplitude calibrated to be similar to the expected 21 cm signal from EoR assuming the spin temperature is much greater than the CMB temperature. The code to produce the sky signal was written by J. Richard Shaw with input from Eric Switzer, while the code to produce GMRT-EoR observations of that signal was written by myself. Figure 6.5 is an image of the simulated signal as it would be seen by GMRT in the absence of any foregrounds or noise. The effect of the beam profile on the power spectrum is less than 3 per cent for scales in the range $40 < \ell < 2000$, and so has a relatively small effect on the result.

6.3.2 Transfer functions

Given a data set x which is the sum of the observed data and a simulated signal, the transfer function $T(x \to F(x); \ell)$ measures how much of the signal survives in the foreground filtered data set $F(x)$ as a function of ℓ. While $F(x)$ can stand in for any filter method applied to the visibilities x, for the SVD we must also specify that the modes removed are those calculated from the visibilities x themselves. While the transfer function measures the signal loss for a single set of data, the power is measured from the cross-correlations of those data sets, so the relationship can be written as

$$P_{\rm SVD}(\ell) = T(x \to F(x); \ell)^2 P_{\rm 21cm}(\ell). \tag{6.1}$$

Unless it is necessary to be explicit about the mapping T is measuring, we will shorten this notation to simply $T(\ell)$.

There are numerous ways one can estimate this function. The most direct way is to cross-correlate $F(x)$ with the injected signal, and normalize by the auto-power of that

[1] Code for Anisotropies in the Microwave Background; `http://camb.info`

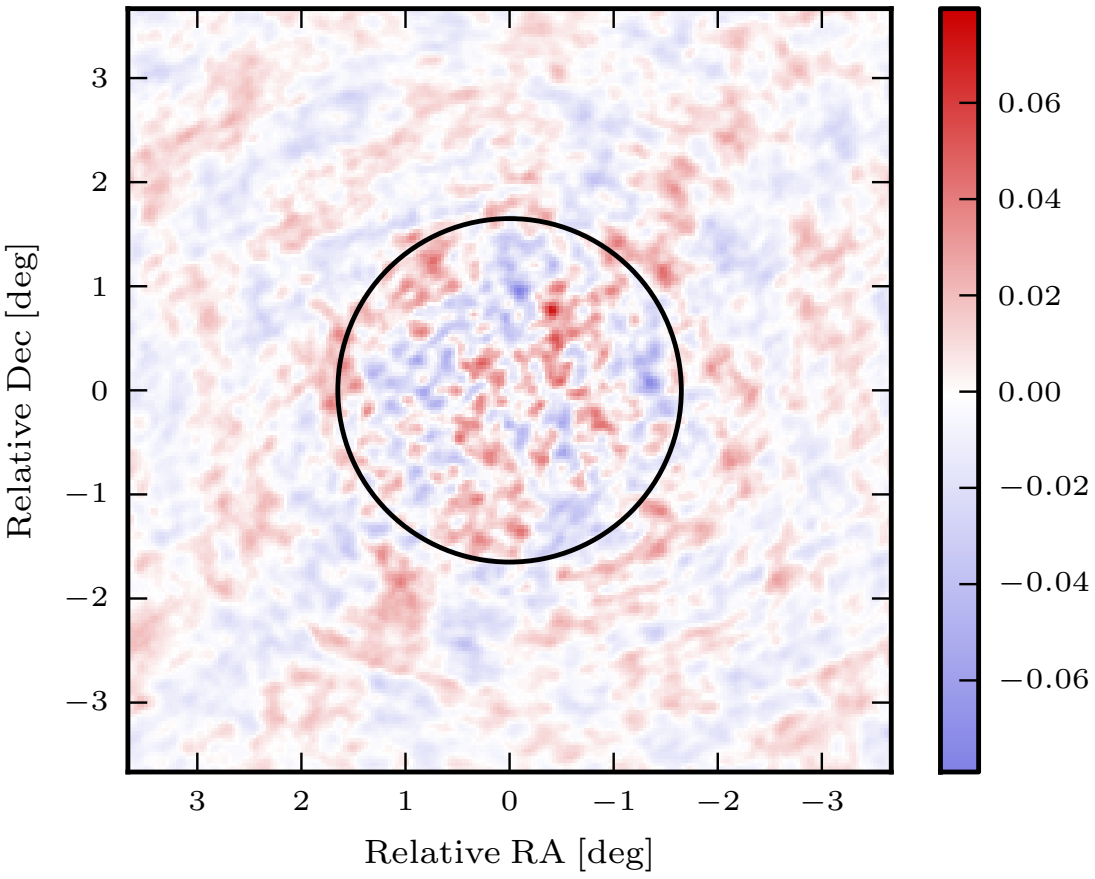


Figure 6.5: An image of the simulated signal as seen with GMRT in the absence of foregrounds or noise, using the same baselines and field of view as Figures 6.2 and 6.3. The colour scale is in units of mJy. The solid circle is the FWHM of the primary beam.

same signal. This is written as

$$T_0(\ell) = \frac{F(\text{data} + \text{signal}) \times \text{signal}}{\text{signal} \times \text{signal}}. \tag{6.2}$$

While conceptually simple, and used successfully by Masui et al. (2013) for data at $z \approx 0.8$, we find this estimator of the transfer function to be exceptionally noisy for realistic cases where $F(x)$ leaves residual foregrounds. In the case of the SVD, we would expect the function to become less noisy as more modes are removed and the residual foregrounds decrease, but we are still significantly limited in being able to measure the power.

An alternative is to subtract the original visibilities, under the same foreground filter, from the combined real and simulated visibilities before cross-correlating with the simulated signal. To distinguish it from the previous one, we denote this version of the transfer function T_1, which takes the form

$$T_1(\ell) = \frac{[F(\text{data} + \text{signal}) - F(\text{data})] \times \text{signal}}{\text{signal} \times \text{signal}}. \tag{6.3}$$

In addition to being much less noisy when residual foregrounds are present, this has the benefit that by subtracting the original data we remove the possibility of the real 21 cm signal in the data correlating with the simulated signal and biasing the result. If $F(x)$ left the signal untouched, this would in principle equal 1.0. However, deviations are possible even when $F(x) = x$. This is due to the fact that the cross-correlations with real data in the numerator introduce RFI masks, noise, and day-to-day variations which are not present in the pure signal in the denominator. Thus, the transfer function will also correct for these effects, which enter at a level of a few per cent.

Two additional transfer functions can be defined in terms of the power spectra themselves. The first method uses the power of the difference of the original data and the data plus simulation, and again compares this to the input power. This can be written as

$$T_2^2 = \frac{[F(\text{data} + \text{signal}) - F(\text{data})] \times [F(\text{data} + \text{signal}) - F(\text{data})]}{\text{signal} \times \text{signal}} \tag{6.4}$$

where we have included the square in the definition since this is already a measure of the change in power. Like T_1, with no filter this should reduce to exactly one, and should also be strictly positive. Similarly, we can take the difference of the powers by subtracting the power spectrum calculated in the usual way from the power spectrum calculated with the injected signal. Again normalizing by the input, we write this as

$$T_3^2 = \frac{[F(\text{data} + \text{signal}) \times F(\text{data} + \text{signal})] - [F(\text{data}) \times F(\text{data})]}{\text{signal} \times \text{signal}}. \tag{6.5}$$

In practice, the power spectrum is much more sensitive to small changes in the underlying data, making both T_2 and T_3 very noisy. For this reason, we will use T_1 as our estimate of the transfer function for the remainder of this analysis.

We carry out this process of estimating $T(\ell)$, averaging over 100 realizations of the simulated signal, after which both the mean and the standard deviation are well determined, and the error in the mean is small enough that it will not contribute significantly to the corrected power spectra later. Figure 6.6 shows $T_1(\ell)$ for a selection of SVD filters. While the transfer function in principle can depend non-linearly on the amplitude of the input signal, we find that the result does not change significantly within a factor of 10 of realistic signal temperatures. In the regimes where the transfer function does begin to depend on the input temperature, the two are anti-correlated; larger signals are more readily misidentified as foregrounds by the SVD, leading to a small value of $T(\ell)$.

The transfer function can be used to determine the best number of modes to remove, since as more modes are removed the 21 cm signal will eventually be reduced to a point where the additional correction to the signal outweighs the gain from reducing the foregrounds. Figure 6.7 shows that correcting for the transfer function after 32 modes are removed gives a weaker limit on the power than only removing 16 modes.

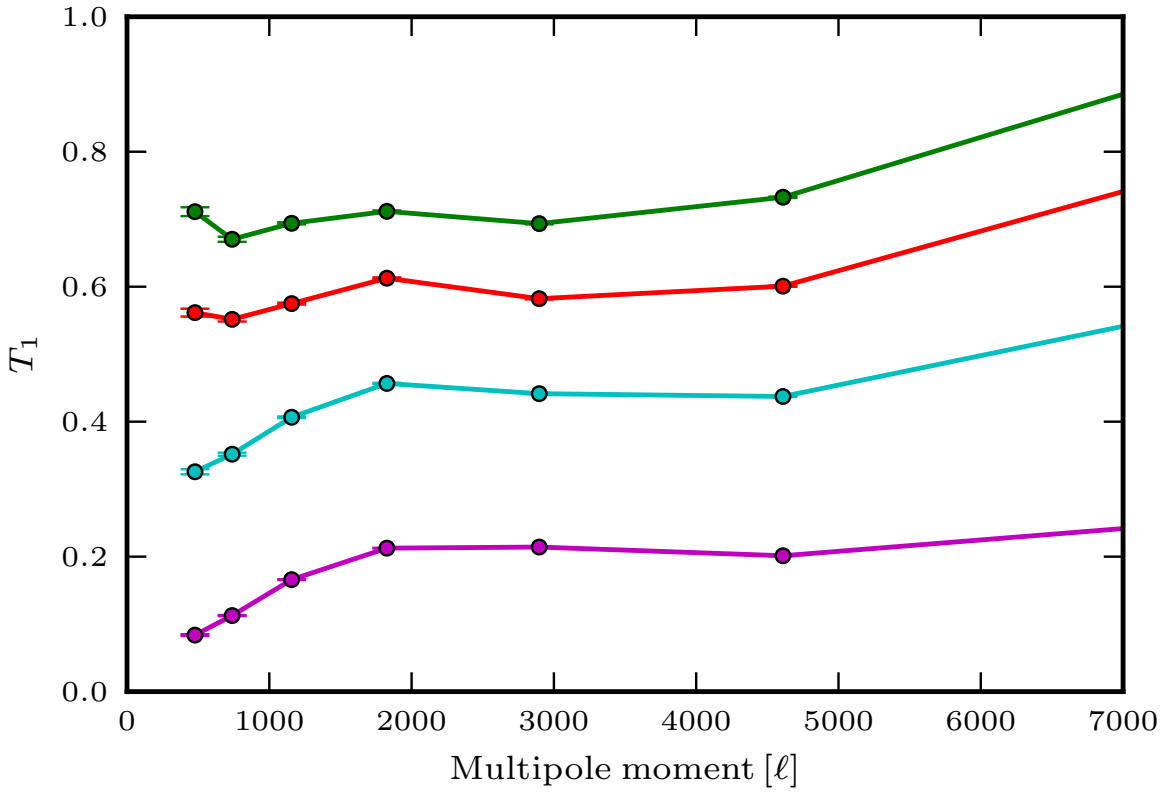


Figure 6.6: Transfer function T_1 with 4 (green), 8 (red), 16 (cyan), and 32 (purple) SVD modes removed, showing the fraction of the 21 cm signal that we estimate survives the SVD foreground removal. With only 4 modes removed, most of the 21 cm signal is expected to survive. However, when 32 modes are removed, about 20 per cent or less survives, depending on the angular scale.

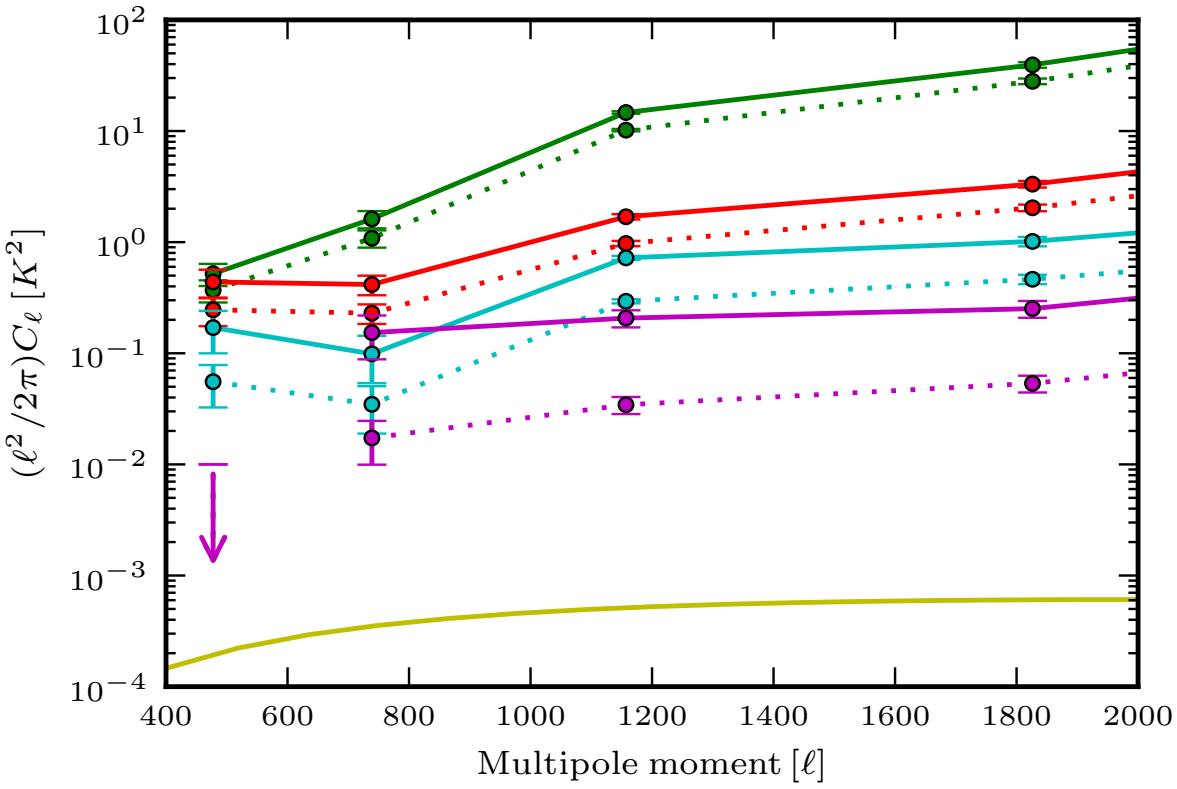


Figure 6.7: The T_1 corrected power spectra after SVD foreground removal. The colours represent 4 (green), 8 (red), 16 (cyan), and 32 (purple) SVD modes removed. The dotted lines are the uncorrected power spectra as in Figure 6.4, while the solid lines are the power spectra after correcting for the transfer function. For 32 modes removed, at low ℓ, the corrected power is larger than that for only 16 modes removed. As in Figure 6.4, the solid yellow line is the theoretical signal from Jelić et al. (2008).

6.4 Selection of line-of-sight length scales

The power calculated from annuli in visibility space only measures the 2D power perpendicular to the line of sight (that is, as a function of the multipole moment ℓ or wavenumber $k_\perp$). To find the full 3D power, we must also look at the line of sight, or frequency, direction and measure power as a function of $k_\parallel$. While certain forms of foreground filters will have a window function that naturally selects a $k_\parallel$, the SVD filter does not have a well defined behaviour along the line of sight. The gives us the flexibility of selecting the window function.

Hermite functions, having the benefit of zero mean and a simple Fourier transform, are well suited to select a narrow range of $k_\parallel$. In frequency space, we define a window

$$h(\nu) = \frac{1}{\sqrt{8\pi}\zeta}\left(1 - \frac{\nu^2}{\zeta^2}\right)\exp\left(1 - \frac{\nu^2}{2\zeta^2}\right) \tag{6.6}$$

where ζ is a parameter analogous to the standard deviation of a Gaussian distribution in units of MHz, which in this case specifies the location of the zeros. This is shown in Figure 6.8 for several ζ compared to the frequency bin size. This window has the Fourier transform

$$\tilde{h}(k_\parallel) = \frac{(k_\parallel r\zeta)^2}{2}\exp\left[1 - \frac{(k_\parallel r\zeta)^2}{2}\right]. \tag{6.7}$$

We have used the conversion factor $r \approx 11.6\,h^{-1}\mathrm{Mpc/MHz}$ such that $k_\parallel$ is in units of $h\,\mathrm{Mpc}^{-1}$. This peak in Fourier space, shown in the inset of Figure 6.8, occurs when $k_\parallel = \sqrt{2}/(r\zeta)$ and determines the $k_\parallel$ at which most power survives the Hermite window. The normalization has been chosen such that the maximum of $\tilde{h}(k_\parallel)$ is 1, thus preserving power at the chosen $k_\parallel$, though the shape of the window in Fourier space will also affect the sensitivity. The parameter ζ is inversely proportional to $k_\parallel$, with the range of possible values limited by the frequency resolution and bandwidth.

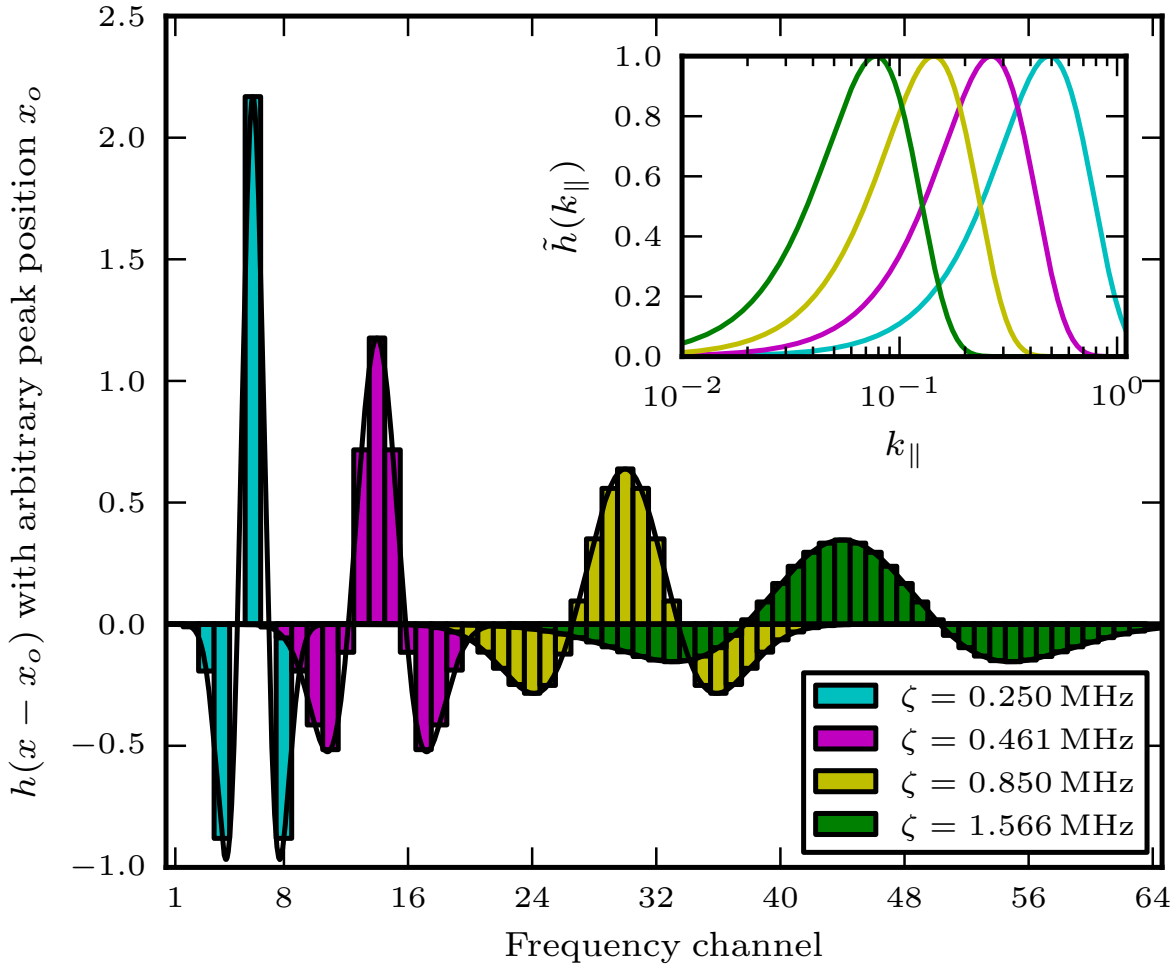


Figure 6.8: Hermite window in frequency space. Four examples are shown with different values of ζ increasing from left to right, with arbitrary horizontal offset. Bars indicate the value of the window in each frequency bin. In practice ζ should not be smaller than the frequency resolution, limiting the $k_\parallel$ available. The inset plot shows the Fourier transform.

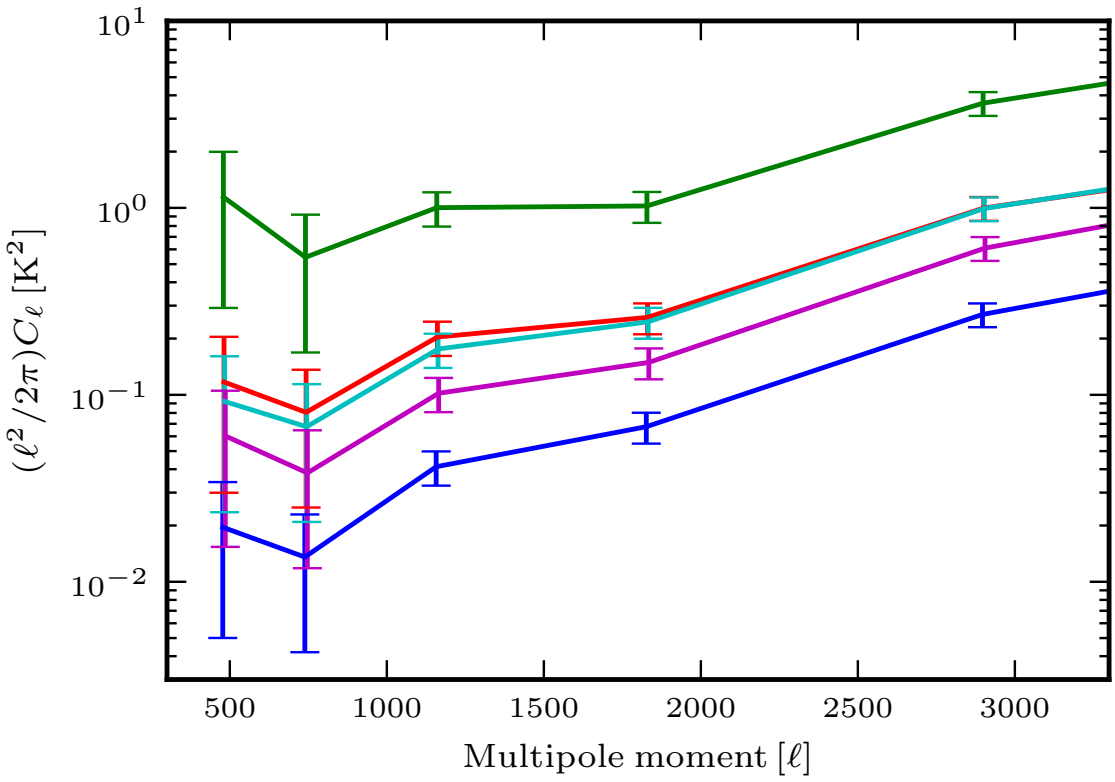


Figure 6.9: An example power spectrum at $\zeta = 0.25\,\mathrm{MHz}$ ($k_\parallel = 0.49\,h\,\mathrm{Mpc}^{-1}$) with 16 SVD modes removed (blue line), with the effect of each type of transfer function correction for T_1 shown. The green line corrects for both the Hermite window and the SVD subtraction, while the red line reintroduces the Hermite window, which agrees quite well with the semi-Hermite correction (cyan). The purple line uses the SVD only transfer function. Error bars include contributions from the transfer function and the bootstrap error from the raw power spectra. In this example, the three approaches agree quite well, though they can diverge by an order of magnitude for other selections of mode subtraction and $k_\parallel$.

6.4.1 Three approaches to the transfer function

By applying a Hermite window to the data, we can calculate the 2D power spectrum at a fixed $k_\parallel$. There is some complication, however, in how we can use a transfer function to correct for possible signal loss. The Hermite filter by design reduces power on most scales while leaving power only at a specific $k_\parallel$, and we want the transfer function to only adjust for signal lost at the same $k_\parallel$. Ideally one would apply the Hermite filter first to isolate the input power at the scales of interest and run the foreground filters on that data. If $H(x)$ represents the data set with a Hermite window applied, this would measure $T(H(x) \to F(H(x)); \ell)$. Unfortunately, the SVD actually depends strongly on information in the $k_\parallel$ direction, which means that $F(H(x))$ may have a much different effect on the power at the chosen length scale than $F(x)$. That is to say, the Hermite and SVD operations do not commute.

There are several possible approaches to get around this, which are as follows.

1. Assume that the transfer function is not strongly dependent on $k_\parallel$, and use $T(x \to F(x); \ell)$ from the $k_\parallel = 0$ case independent of the $k_\parallel$ selected by the Hermite window. We call this the "SVD only" approach. The $k_\parallel$ behaviour only enters in through calculation of the power spectra after the Hermite window. Any important behaviour of the SVD in the $k_\parallel$ direction will not be captured.

2. We can calculate a transfer function for the signal loss due to the total effect of both the Hermite window and the SVD, $T(x \to H(F(x)); \ell)$, and correct for both. We can then use an analytical form for the transfer function of the Hermite window alone to reintroduce the scale window and keep only the power at our selected $k_\parallel$. We call this the "full Hermite" approach.

   To find its analytical form, we start with the fact that the transfer function associated with the Hermite window measures the ratio of the windowed power to the

full power,

$$T_H^2(k_\perp) = \frac{\int P(k_\perp, k_\parallel)|\tilde{h}(k_\parallel)|^2 \mathrm{d}k_\parallel}{\int P(k_\perp, k_\parallel)\mathrm{d}k_\parallel}. \tag{6.8}$$

If we assume the power spectrum has the form

$$P(k_\parallel) \propto \frac{1}{{k_\parallel}^2 + {k_\perp}^2} \tag{6.9}$$

both the numerator and denominator of this can be evaluated analytically. The result is

$$T_H^2(k_\perp) = \frac{e^2 r\zeta k_\perp}{8\sqrt{\pi}} \left(1 - 2r^2\zeta^2 {k_\perp}^2 + 2\sqrt{\pi} r^3 \zeta^3 {k_\perp}^3 e^{r^2\zeta^2 {k_\perp}^2} \mathrm{erfc}[r\zeta k_\perp]\right) \tag{6.10}$$

where $\mathrm{erfc}[x] = 1 - \mathrm{erf}[x]$ is the complementary error function. Requiring the most steps, this method has more avenues to introduce errors or biases.

3. Apply the Hermite window first to the simulated signal. When added to the full data and passed through the SVD foreground removal, the larger amplitude of the foregrounds present in the data ensures that the SVD still has data at all $k_\parallel$ to operate on. However, since there is only a simulated signal at a specific $k_\parallel$, the cross correlation with the simulated signal when calculating the transfer function $T(\mathrm{data} + H(\mathrm{signal}) \to F(\mathrm{data} + H(\mathrm{signal})); \ell)$ only measures the effect of the SVD on that $k_\parallel$. We call this the "semi-Hermite" approach. This assumes that the SVD as applied to the $k_\parallel$ limited simulated signal is a suitable proxy for how the SVD affects the real signal, given that both the real signal and the $k_\parallel$ limited simulated signal are of significantly lower amplitude than the foregrounds.

6.5 Adjusted power spectra

Figure 6.9 shows a typical power spectrum for a particular choice of SVD filter, transfer function, and $k_\parallel$, without any correction and the resulting spectra after each of the above approaches. Differences in each approach illustrate the difficulty in finding an unbiased

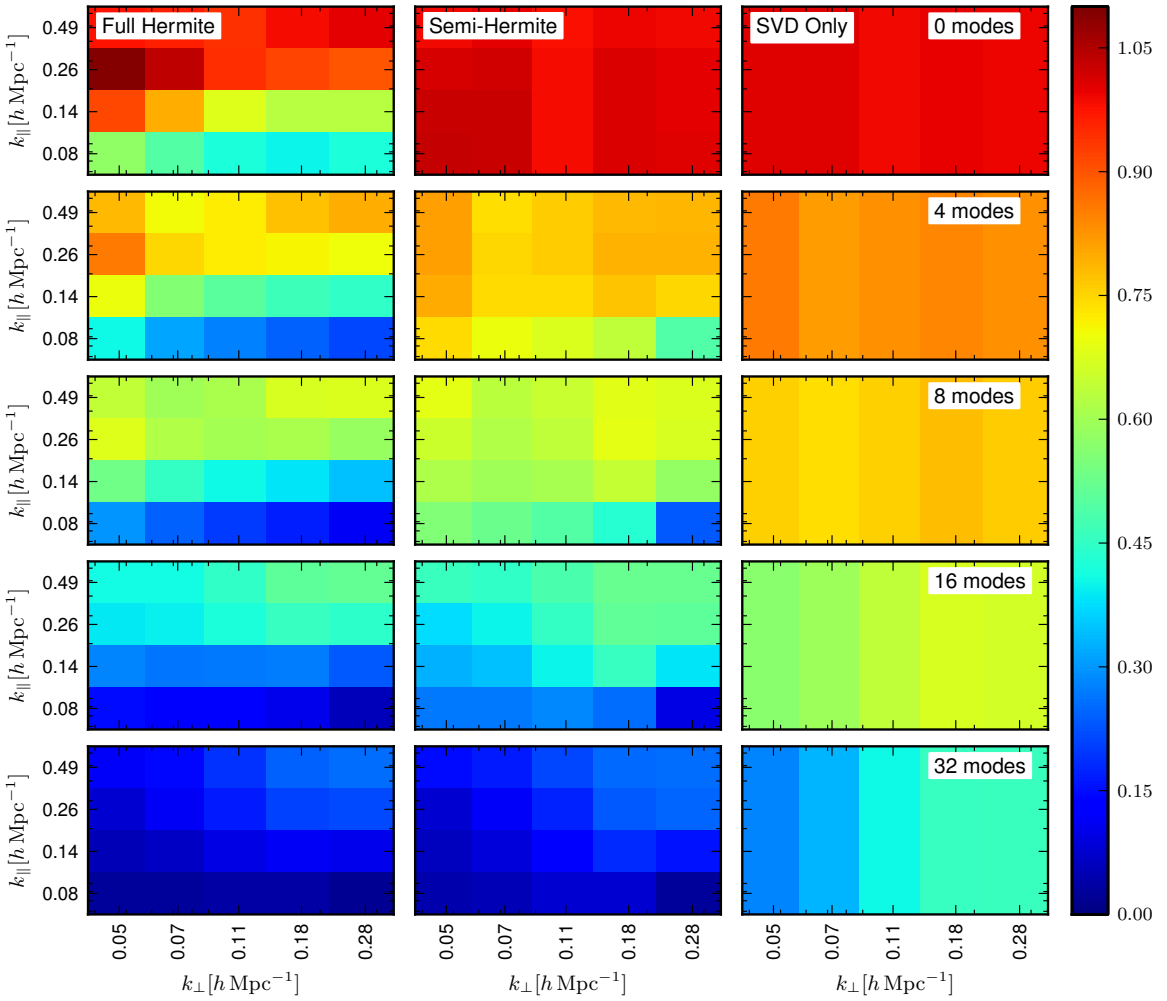


Figure 6.10: The transfer function T_1 as a function of both $k_\perp$ and $k_\parallel$ calculated using the full Hermite (left column), semi-Hermite (middle column), and SVD only (right column) methods. From top to bottom, 0, 4, 8, 16, and 32 SVD modes are removed. Slight deviations from unity with zero modes removed illustrate the additional effects from RFI masking, noise, and the beam that are captured by the transfer function.

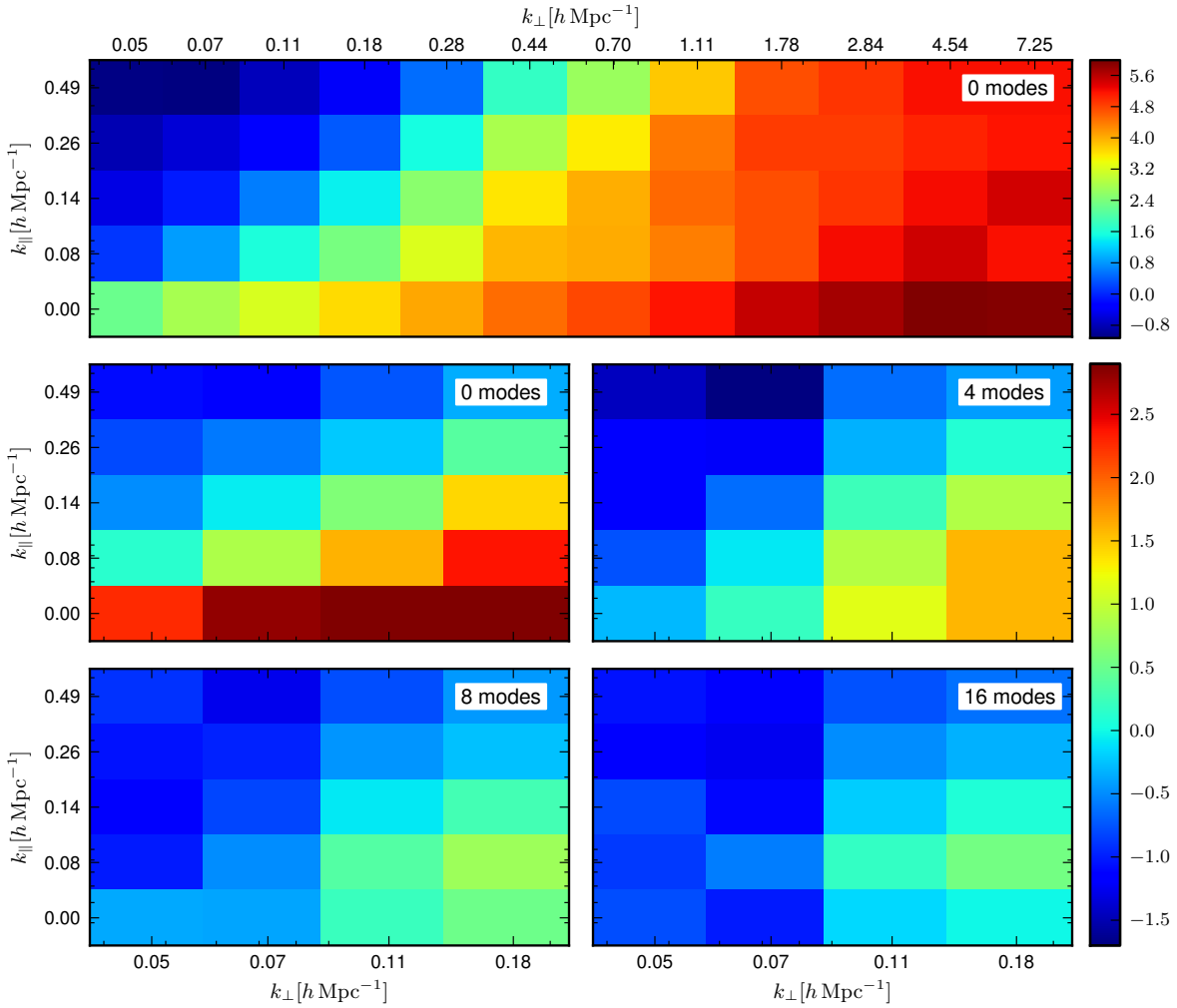


Figure 6.11: Power spectra as a function of both $k_\perp$ and $k_\parallel$ corrected with T_1 calculated using the semi-Hermite approach. The topmost plot shows the entire $k_\perp$ range without any foreground removal. The four smaller plots show 0, 4, 8, and 16 SVD modes removed on the same colour scale for only the lowest few $k_\perp$ bins. The colour scales are in units of $\log(\mathrm{K}^2)$. Compared to the case with 0 modes removed, the SVD tends to reduce the overall power by 1–3 orders of magnitude. See also Figure 6.12, which shows the power as a function of the total k.

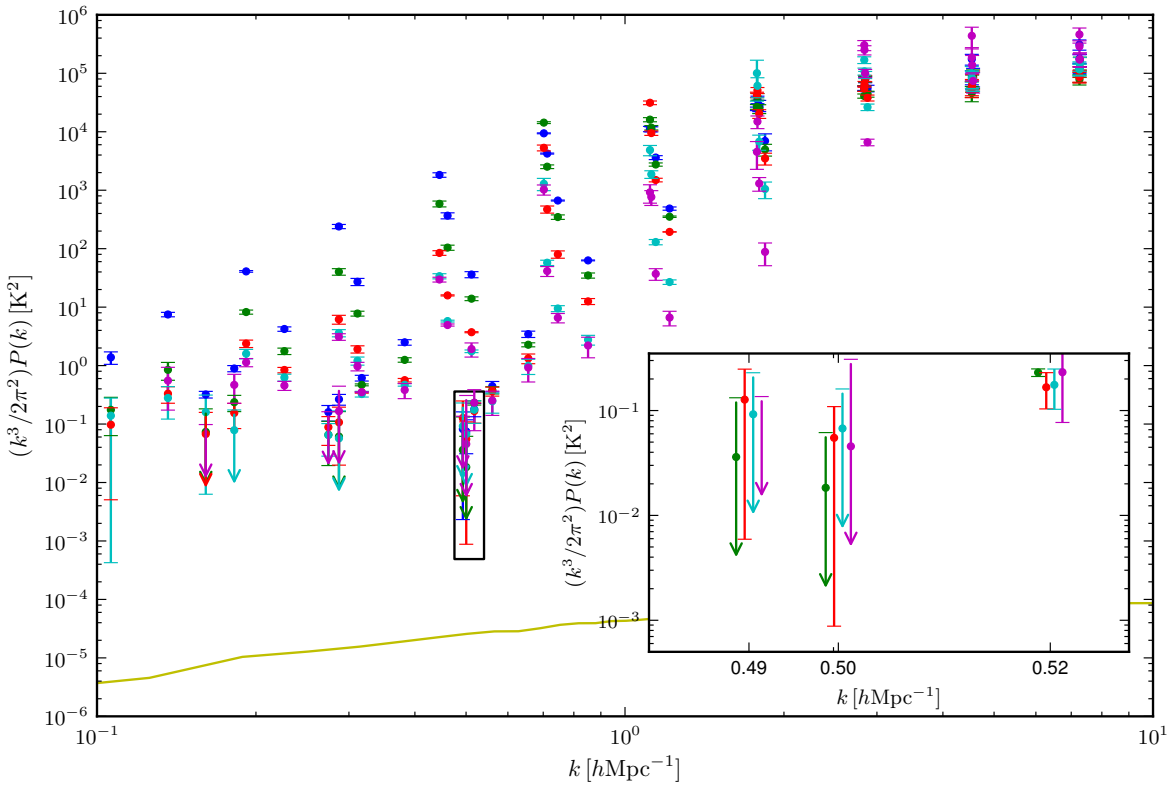


Figure 6.12: Power as a function of the total wavenumber $k = \sqrt{{k_\perp}^2 + {k_\parallel}^2}$. Each point represents a different $(k_\perp, k_\parallel)$ pair; there is no binning in k. Colours indicate the number of SVD modes removed; 0 (blue), 4 (green), 8 (red), 16 (cyan), and 32 (purple) are shown. The boxed region at $k \approx 0.5$ is shown inset, with nearby points each of the three marked k spread out slightly for clarity. The best limit at 2σ is $(248\,\mathrm{mK})^2$ at $0.50\,h\mathrm{Mpc}^{-1}$ achieved with 4 SVD modes removed. The solid line is the predicted 3D power spectrum from Iliev et al. (2008) assuming a 30 mK signal.

estimator that gives a robust result. Figure 6.10 shows the transfer function as a function of both $k_\perp$ and $k_\parallel$ for each of the three possible approaches.

We find that in both the full Hermite and semi-Hermite methods there is a $k_\parallel$ dependence which is not captured by the SVD only method, which is constant with $k_\parallel$ by definition. All three methods show deviations from unity on the order of a few per cent with zero SVD modes removed due to the additional effects from RFI masking, noise, and the beam that are captured by the transfer function. It is also notable, however, that the full Hermite approach finds T_1 deviating from one by tens of per cent in some regimes, especially at low $k_\perp$. This is likely indicative of a mismatch between the amount of power being removed by the combination of SVD and Hermite filters and the amount modelled by the analytic form. This suggests that in areas of $(k_\perp, k_\parallel)$ space where $T_1 > 1$, this method may overestimate the amount of signal present, in turn underestimating the 21 cm power by failing to fully correct for the signal loss. Nonetheless, the full and semi-Hermite approaches agree much better with more SVD modes removed. Since the semi-Hermite approach seems to capture both the $k_\parallel$ dependence and is relatively well behaved with $0 < T_1 < 1$, we use it as the canonical transfer function.

6.6 Sampling $(k_\perp, k_\parallel)$ space to get $P(k)$

Using the Hermite window to select a fixed $k_\parallel$ allows us to calculate $P(\ell|k_\parallel)$ and the associated transfer function at that $k_\parallel$. By repeating this for a series of $k_\parallel$, we can build up the full 3D power spectrum.

Figure 6.11 shows the power as a function of both $k_\perp$ and $k_\parallel$ using the semi-Hermite correction, given a series of different SVD mode subtractions. The power shows a pattern of lower values towards low $k_\perp$ and high $k_\parallel$. Figure 6.12 shows the same measurements as a function of the 3D wavenumber $k = \sqrt{{k_\perp}^2 + {k_\parallel}^2}$. Though the SVD is our primary mode of foreground removal, the Hermite function itself acts as a foreground filter removing

the large-scale structure in frequency space. This is reflected in the points where zero SVD modes have been removed. It is clear that our ability to remove foregrounds drops off quickly above about $k \approx 0.5\,h\,\mathrm{Mpc}^{-1}$. Our best limit at 2σ is $(248\,\mathrm{mK})^2$, achieved at $(k_\perp, k_\parallel) = (0.11, 0.49)\,h\,\mathrm{Mpc}^{-1}$, or a total k of $0.50\,h\,\mathrm{Mpc}^{-1}$, with 4 SVD modes removed. At this point, the semi-Hermite value of the transfer function was $T_1 = 0.74$, meaning an estimated 26 per cent of signal was removed by the SVD mode subtraction and Hermite window operating on each day in the cross-correlations. If instead 16 modes are removed, the limit changes to $(319\,\mathrm{mK})^2$ but 55 per cent of the signal is lost. Any residual foregrounds, though reduced by a much larger fraction than the signal, will also have been boosted by this correction, making this measurement an upper limit on the actual 21 cm signal.

6.7 Conclusion

Using a singular value decomposition as a foreground removal technique and a simulated signal to quantify the loss of a real 21 cm signal the SVD may cause, we have calculated an upper limit to the HI power spectrum at $z = 8.6$ of $(248\,\mathrm{mK})^2$ at $k = 0.50\,h\,\mathrm{Mpc}^{-1}$. The $k_\perp$ component was found using the median power in annuli of the (u, v) plane, while a Hermite window was used to sample the $k_\parallel$ direction. This is in contrast to the piecewise-linear foreground filter used in Chapter 5, which operates only in the frequency direction and carried with it an implicit $k_\parallel$ window.

This limit is dependent on the method one chooses to calculate the transfer function between the real 21 cm signal and the observed power. Both the $k_\perp$ and $k_\parallel$ behaviour of the foreground filter chosen needs to be taken into account. While the semi-Hermite method chosen uses a simulated signal with power in a limited $k_\parallel$ window, and may miss interactions between the SVD filter and the signal over larger $k_\parallel$ bands, we believe it to give the most reliable estimate of the transfer function and a suitably conservative

estimate on the final upper limit.

Had we instead used the full Hermite approach described, this limit would have been $(260\,\mathrm{mK})^2$. That this second approach gives a similar value suggests that this limit is a fairly robust one. The difference can likely be attributed in part to the simplifying assumptions necessary when deriving the analytical Hermite windowing function. We also consider the current result to be more robust than that reported previously in Paciga et al. (2011). While the previous limit was considerably lower, this can be accounted for by many factors; the different k scale, the change in foreground filter, several minor changes in the analysis pipeline detailed in section 6.1, and most significantly the fact that this is the first time a transfer function has been used to correct for signal lost in the foreground filter. Without such a correction, our best upper limits with the SVD foreground filter would have been incorrectly reported as low as $(50\,\mathrm{mK})^2$.

This limit still compares favourably to others established in the literature which are on the order of several kelvin (e.g., Bebbington, 1986; Ali et al., 2008; Parsons et al., 2010). Recently, after submission of the paper publishing these results, PAPER (Parsons et al., 2013) claimed an upper limit of $(52\,\mathrm{mK})^2$ at $k = 0.11\,h\,\mathrm{Mpc}^{-1}$ and $z = 7.7$. However, signal loss from their primary foreground filtering step (their section 3.4) has not been accounted for and so it is not clear how to compare their result to ours. LOFAR has begun publishing initial results from reionization observations, but have so far focused on much longer scales ($\ell \approx 7500$; Yatawatta et al., 2013).

In Chapter 5 we considered a model with a cold intergalactic medium (IGM), a neutral fraction of 0.5, and fully ionized bubbles with uniform radii. In such a model this current limit would constrain the brightness temperature of the neutral IGM to be at least 540 mK in absorption against the CMB. However, a value of the HI power spectrum of $(248\,\mathrm{mK})^2$ is almost an order of magnitude higher than what is generally considered physically plausible in most reionization models. In particular, this result does not constrain reionization models with a warm IGM where the spin temperature is

much greater than the CMB temperature.

The SVD procedure could be refined further by a baseline-by-baseline accounting of the optimal number of modes to subtract or by limiting the field of view on the sky to the innermost area of the beam where point source residuals are minimal, although it is not obvious what effect this would have on the signal at small angular scales. It is also possible to derive the SVD modes removed in other ways, e.g. by stacking all data and removing a common set of modes from each day, rather than calculating a set of modes separately on a day-by-day basis. Making a measurement at larger ℓ would require a more careful treatment of point sources but is also limited by the fact that the SVD is less effective for longer baselines. Regardless of the foreground removal technique used, it is likely that accurately correcting for the any resulting loss of the 21 cm signal, and disentangling the 21 cm signal from any residual foregrounds, will remain a significant challenge in measuring the true EoR power spectrum.

Chapter 7

Conclusion

7.1 Summary of results

We have now described the efforts of the GMRT-EoR experiment since 2007 and the various observing techniques we have used. While over 900 hours have been allocated to our group over five years, in this work we have analyzed only 50 hours of continuous tracking data completely. Of the remaining hours, a significant portion is potentially useful for further analysis. The overall observing efficiency, however, has been relatively low due to the experimental and ever-evolving nature of the project, leading to a large fraction of the data becoming obsolete.

A significant amount of effort has been put into improving the RFI environment around GMRT, and we believe this has been largely successful. Though new sources may arise, as of 2011 there were no individual sources that stood out in the data. It is likely that checking maps of the RFI environment a few times per cycle would be sufficient to spot any new egregious sources that may arise, while the process of investigating the lower level sources is unlikely to provide much further value.

Identifying physical sources via their SVD modes in the data has shown that using the SVD as a method of post-observation RFI removal is viable. Removing the 50 brightest

modes has reduced the RMS noise in maps by a factor of four. However, care must be taken to exclude data at low u after an SVD based RFI removal, since the sky signal along north-south baselines is likely to be confused for interference due to its low rotation rate.

The consistency of the sky over several nights seems to be quite good. Pairwise differences of nights in December 2007 typically have residual noise of a few milliJanskys. No variable sources were identified over the eight days observed. Further, our pulsar based calibration corrects for any ionospheric variation over the entire primary beam. The major effect which remains uncorrected is the w term, which blurs sources when integrating over long periods. Though we have investigated possibilities for w term projection, including widefield imagining with CASA and an adaptation of the Cosmic Background Imager (CBI) gridder (Myers et al., 2003), w projection carries with it additional difficulties in distorting the power spectrum in ways which are too complicated to account for here.

Our preliminary calculations of the power spectrum used a piecewise-linear fit to model and remove foregrounds at every baseline, from which we calculated an upper limit of $(70\,\mathrm{mK})^2$ at $k = 0.65\,h\,\mathrm{Mpc}^{-1}$, reported in Paciga et al. (2011). However, this was replaced in later analysis with an SVD based foreground subtraction which has fewer assumptions about foregrounds. The SVD also does not carry an implicit $k_\parallel$ scale, allowing the use of a Hermite window function to choose a $k_\parallel$ scale of interest and build up a full 3D power spectrum. Crucially, a simulated EoR signal was injected into the data to quantify the signal loss due to our foreground removal step. With this correction, we reported an upper limit to the HI power spectrum at $z \approx 8.6$ of $(248\,\mathrm{mK})^2$ at $k = 0.5\,h\,\mathrm{Mpc}^{-1}$ (Paciga et al., 2013). While we had devised a simple toy model to put a physical interpretation on our limits, the simulation-corrected limit is not low enough to place strong constraints on the size of the ionization bubbles.

7.2 Prospects at GMRT

The principal limiting factor for EoR observations is still, by far, foregrounds. Throughout this project we have proposed several observing strategies to improve our ability to model and remove them, e.g. drift scanning, mosaicing, and restricted LST ranges, but with more recent data remaining unanalyzed we do not yet have a quantitative idea of how these strategies improve the situation. It is known that the residuals in our foreground subtracted data live primarily at the edge of the primary beam, and can include both point sources (which only contribute to the power at large ℓ) and sidelobe artefacts from bright sources well outside the beam.

While in our most recent analysis we have relied on using SVD modes to remove foregrounds, recently other members of the GMRT-EoR group, led by J. Richard Shaw and Liam Connor, have begun exploring a strategy of observing complete rings of the sky at a single declination and using m-mode decomposition to separate foregrounds from the cosmological signal (Shaw et al., 2013). This analysis is ongoing, and has taken priority over analysing the archival data accumulated on B0823+26, B2217+47, and B2045-16.

We have so far treated our data only as a single redshift bin, spanning $8.1 < z < 9.2$. This is a wide enough range that there may be some evolution of the EoR signal across the band. If the sensitivities necessary for reaching EoR-level signals can be reached, exploring the redshift space dependence will be a natural next step both as a test that the true cosmological signal has been detected and to further probe the reionization history of the universe.

The GMRT still enjoys several benefits over other EoR experiments, including a relatively small field of view and a stable primary beam. Our experience has shown that overcoming the foreground barrier only gets more difficult when these two features are relaxed. However, dedicated EoR experiments such as PAPER, MWA, and, in part, LOFAR, are all active and will contribute to a rich and active field until a measurement of the clean EoR power spectrum can be made.

7.3 Implications for EoR experiments

The most important implication for other EoR experiments is the need to thoroughly model the signal loss that foreground removal strategies may cause. Our SVD method to removing foreground modes can easily reduce the EoR signal by more than 90 per cent if too many modes are removed, but superficially makes it easier to reach the level of the EoR signal (even though the real signal has been thrown away). Any method that removes a large fraction of the signal, as is required to reach EoR level sensitivities, is likely to suffer similarly.

We have corrected for this by injecting a simulated signal and testing how well we can recover the injected power after our foreground removal, which allows us to judge the point at which removing more foreground modes reduces the EoR signal as much as, or more than, the foregrounds. Since this also depends on the way the transfer function of the injected signal is calculated, any EoR result should include a description both of how signal loss was modeled and how the transfer function between input and output power was calculated. Without these corrections it is not possible to interpret any EoR result as a true measurement or limit on the HI power spectrum. Had we not made such a correction in this work, our latest result would have been reported as an upper limit of $(50\,\mathrm{mK})^2$ instead of $(248\,\mathrm{mK})^2$.

Bibliography

Ali, S. S., Bharadwaj, S., & Chengalur, J. N. 2008, MNRAS, 385, 2166

Alvarez, M. A., Pen, U., & Chang, T. 2010, ApJ, 723, L17

Ananthakrishnan, S. 1995, Journal of Astrophysics and Astronomy Supplement, 16, 427

Baek, S., Di Matteo, P., Semelin, B., Combes, F., & Revaz, Y. 2009, A&A, 495, 389

Baek, S., Semelin, B., Di Matteo, P., Revaz, Y., & Combes, F. 2010, A&A, 523, A4

Barkana, R. & Loeb, A. 2001, Phys. Rep., 349, 125

Beardsley, A. P., Hazelton, B. J., Morales, M. F., Arcus, W., Barnes, D., Bernardi, G., Bowman, J. D., Briggs, F. H., Bunton, J. D., Cappallo, R. J., Corey, B. E., Deshpande, A., deSouza, L., Emrich, D., Gaensler, B. M., Goeke, R., Greenhill, L. J., Herne, D., Hewitt, J. N., Johnston-Hollitt, M., Kaplan, D. L., Kasper, J. C., Kincaid, B. B., Koenig, R., Kratzenberg, E., Lonsdale, C. J., Lynch, M. J., McWhirter, S. R., Mitchell, D. A., Morgan, E., Oberoi, D., Ord, S. M., Pathikulangara, J., Prabu, T., Remillard, R. A., Rogers, A. E. E., Roshi, A., Salah, J. E., Sault, R. J., Udaya, S. N., Srivani, K. S., Stevens, J., Subrahmanyan, R., Tingay, S. J., Wayth, R. B., Waterson, M., Webster, R. L., Whitney, A. R., Williams, A., Williams, C. L., & Wyithe, J. S. B. 2013, MNRAS, 429, L5

Bebbington, D. H. O. 1986, MNRAS, 218, 577

Becker, G. D., Rauch, M., & Sargent, W. L. W. 2007, ApJ, 662, 72

Becker, R. H., Fan, X., White, R. L., Strauss, M. A., Narayanan, V. K., Lupton, R. H., Gunn, J. E., Annis, J., Bahcall, N. A., Brinkmann, J., Connolly, A. J., Csabai, I., Czarapata, P. C., Doi, M., Heckman, T. M., Hennessy, G. S., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., McKay, T. A., Munn, J. A., Nash, T., Nichol, R., Pier, J. R., Richards, G. T., Schneider, D. P., Stoughton, C., Szalay, A. S., Thakar, A. R., & York, D. G. 2001, AJ, 122, 2850

Bentum, M., Boonstra, A. J., & Millenaar, R. 2010, in Proceedings of the RFI Mitigation Workshop

Bernardi, G., de Bruyn, A. G., Brentjens, M. A., Ciardi, B., Harker, G., Jelić, V., Koopmans, L. V. E., Labropoulos, P., Offringa, A., Pandey, V. N., Schaye, J., Thomas, R. M., Yatawatta, S., & Zaroubi, S. 2009, A&A, 500, 965

Bernardi, G., Greenhill, L. J., Mitchell, D. A., Ord, S. M., Hazelton, B. J., Gaensler, B. M., de Oliveira-Costa, A., Morales, M. F., Udaya Shankar, R., Subrahmanyan, R., Wayth, R. B., Lenc, E., Williams, C. L., Arcus, W., Balwinder, S. A., Barnes, D. G., Bowman, J. D., Briggs, F. H., Bunton, J. D., Cappallo, R. J., Corey, B. E., Deshpande, A., deSouza, L., Emrich, D., Goeke, R., Herne, D., Hewitt, J. N., Johnston-Hollitt, M., Kaplan, D., Kasper, J. C., Kincaid, B. B., Koenig, R., Kratzenberg, E., Lonsdale, C. J., Lynch, M. J., McWhirter, S. R., Morgan, E., Oberoi, D., Pathikulangara, J., Prabu, T., Remillard, R. A., Rogers, A. E. E., Roshi, A., Salah, J. E., Sault, R. J., Srivani, K. S., Stevens, J., Tingay, S. J., Waterson, M., Webster, R. L., Whitney, A. R., Williams, A., & Wyithe, J. S. B. 2013, preprint (astro-ph/1305.6047)

Bernardi, G., Mitchell, D. A., Ord, S. M., Greenhill, L. J., Pindor, B., Wayth, R. B., & Wyithe, J. S. B. 2011, MNRAS, 413, 411

Bharadwaj, S. & Ali, S. S. 2005, MNRAS, 356, 1519

Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, ApJ, 638, 20

—. 2009, ApJ, 695, 183

Bowman, J. D. & Rogers, A. E. E. 2010, Nature, 468, 796

Bowman, J. D., Rogers, A. E. E., & Hewitt, J. N. 2008, ApJ, 676, 1

Brentjens, M., Koopmans, L. V. E., de Bruyn, A. G., & Zaroubi, S. 2011, in Bulletin of the American Astronomical Society, Vol. 43, American Astronomical Society Meeting Abstracts #217, #107.04

Carilli, C. L., Furlanetto, S., Briggs, F., Jarvis, M., Rawlings, S., & Falcke, H. 2004, New Astron. Rev., 48, 1029

Catinella, B., Haynes, M. P., Giovanelli, R., Gardner, J. P., & Connolly, A. J. 2008, ApJ, 685, L13

Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Nature, 466, 463

Chapman, E., Abdalla, F. B., Harker, G., Jelić, V., Labropoulos, P., Zaroubi, S., Brentjens, M. A., de Bruyn, A. G., & Koopmans, L. V. E. 2012, MNRAS, 423, 2518

Chen, X. & Miralda-Escudé, J. 2004, ApJ, 602, 1

Chengalur, J. N., Gupta, Y., & Dwarakanath, K. S. 2007, Low Frequency Radio Astronomy, 3rd Edition (National Center for Radio Astrophysics)

Choudhury, T. R., Haehnelt, M. G., & Regan, J. 2009, MNRAS, 394, 960

Chuzhoy, L. & Shapiro, P. R. 2006, ApJ, 651, 1

Ciardi, B. & Madau, P. 2003, ApJ, 596, 1

Cohen, A. S., Lane, W. M., Cotton, W. D., Kassim, N. E., Lazio, T. J. W., Perley, R. A., Condon, J. J., & Erickson, W. C. 2007, AJ, 134, 1245

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., & Broderick, J. J. 1998, AJ, 115, 1693

Cooray, A., Li, C., & Melchiorri, A. 2008, Phys. Rev. D, 77, 103506

Cornwell, T. J., Golap, K., & Bhatnagar, S. 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 647

Dalal, N., Pen, U.-L., & Seljak, U. 2010, Journal of Cosmology and Astroparticle Physics, 11, 7

Datta, A., Bowman, J. D., & Carilli, C. L. 2010, ApJ, 724, 526

Datta, K. K., Friedrich, M. M., Mellema, G., Iliev, I. T., & Shapiro, P. R. 2012, MNRAS, 424, 762

de Bruyn, A. G., Katgert, P., Haverkorn, M., & Schnitzeler, D. H. F. M. 2006, Astronomische Nachrichten, 327, 487

de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., Jonas, J., Landecker, T. L., & Reich, P. 2008a, MNRAS, 388, 247

—. 2008b, MNRAS, 388, 247

Delhaize, J., Meyer, M., Staveley-Smith, L., & Boyle, B. 2013, preprint (astro-ph/1305.1968)

Di Matteo, T., Ciardi, B., & Miniati, F. 2004, MNRAS, 355, 1053

Dijkstra, M., Haiman, Z., & Loeb, A. 2004, ApJ, 613, 646

Dillon, J. S., Liu, A., & Tegmark, M. 2012, preprint (astro-ph/1211.2232)

Djorgovski, S. G., Castro, S., Stern, D., & Mahabal, A. A. 2001, ApJ, 560, L5

Douglas, J. N., Bash, F. N., Bozyan, F. A., Torrence, G. W., & Wolfe, C. 1996, AJ, 111, 1945

Dunkley, J., Komatsu, E., Nolta, M. R., Spergel, D. N., Larson, D., Hinshaw, G., Page, L., Bennett, C. L., Gold, B., Jarosik, N., Weiland, J. L., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., & Wright, E. L. 2009, ApJS, 180, 306

Efron, B. 1979, Ann. Statist., 7, 1

Ekers, J. A. 1969, Australian Journal of Physics Astrophysical Supplement, 7, 3

Ewen, H. I. & Purcell, E. M. 1951, Nature, 168, 356

Fan, X., Carilli, C. L., & Keating, B. 2006a, ARA&A, 44, 415

Fan, X., Narayanan, V. K., Strauss, M. A., White, R. L., Becker, R. H., Pentericci, L., & Rix, H. 2002, AJ, 123, 1247

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., & Fukugita, M. 2006b, AJ, 132, 117

Field, G. B. 1959, ApJ, 129, 536

Fomalont, E. B., Goss, W. M., Lyne, A. G., Manchester, R. N., & Justtanont, K. 1992, MNRAS, 258, 497

Friedrich, M. M., Mellema, G., Alvarez, M. A., Shapiro, P. R., & Iliev, I. T. 2011, MNRAS, 413, 1353

Furlanetto, S. R., Lidz, A., Loeb, A., McQuinn, M., Pritchard, J. R., Shapiro, P. R., Alvarez, M. A., Backer, D. C., Bowman, J. D., Burns, J. O., Carilli, C. L., Cen, R., Cooray, A., Gnedin, N., Greenhill, L. J., Haiman, Z., Hewitt, J. N., Hirata, C. M., Lazio, J., Mesinger, A., Madau, P., Morales, M. F., Oh, S. P., Peterson, J. B., Pihlström, Y. M., Tegmark, M., Trac, H., Zahn, O., & Zaldarriaga, M. 2009, Astro2010: The Astronomy and Astrophysics Decadal Survey, Science White Papers, 82

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006a, Phys. Rep., 433, 181

Furlanetto, S. R., Oh, S. P., & Pierpaoli, E. 2006b, Phys. Rev. D, 74, 103502

Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004, ApJ, 613, 1

Gaensler, B. M. & Slane, P. O. 2006, ARA&A, 44, 17

Ghosh, A., Prasad, J., Bharadwaj, S., Ali, S. S., & Chengalur, J. N. 2012, MNRAS, 426, 3295

Gould, D. M. & Lyne, A. G. 1998, MNRAS, 301, 235

Greiner, J., Krühler, T., Fynbo, J. P. U., Rossi, A., Schwarz, R., Klose, S., Savaglio, S., Tanvir, N. R., McBreen, S., Totani, T., Zhang, B. B., Wu, X. F., Watson, D., Barthelmy, S. D., Beardmore, A. P., Ferrero, P., Gehrels, N., Kann, D. A., Kawai, N., Yoldaş, A. K., Mészáros, P., Milvang-Jensen, B., Oates, S. R., Pierini, D., Schady, P., Toma, K., Vreeswijk, P. M., Yoldaş, A., Zhang, B., Afonso, P., Aoki, K., Burrows, D. N., Clemens, C., Filgas, R., Haiman, Z., Hartmann, D. H., Hasinger, G., Hjorth, J., Jehin, E., Levan, A. J., Liang, E. W., Malesani, D., Pyo, T., Schulze, S., Szokoly, G., Terada, K., & Wiersema, K. 2009, ApJ, 693, 1610

Griffen, B. F., Drinkwater, M. J., Iliev, I. T., Thomas, P. A., & Mellema, G. 2013, MNRAS, 431, 3087

Gunn, J. E. & Peterson, B. A. 1965, ApJ, 142, 1633

Guo, Q., Wu, X., Xu, H. G., & Gu, J. H. 2009, ApJ, 693, 1000

Haiman, Z. 2011, Nature, 472, 47

Hankins, T. H. & Rankin, J. M. 2008, preprint (astro-ph/0802.1202)

Harker, G., Zaroubi, S., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Ciardi, B., Jelić, V., Koopmans, L. V. E., Labropoulos, P., Mellema, G., Offringa, A., Pandey, V. N., Pawlik, A. H., Schaye, J., Thomas, R. M., & Yatawatta, S. 2010, MNRAS, 405, 2492

Harker, G., Zaroubi, S., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Ciardi, B., Jelić, V., Koopmans, L. V. E., Labropoulos, P., Mellema, G., Offringa, A., Pandey, V. N., Schaye, J., Thomas, R. M., & Yatawatta, S. 2009, MNRAS, 397, 1138

Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1

Hobbs, G., Lyne, A. G., Kramer, M., Martin, C. E., & Jordan, C. 2004, MNRAS, 353, 1311

Hobson, M. P. & Maisinger, K. 2002, MNRAS, 334, 569

Hogan, C. J. & Rees, M. J. 1979, MNRAS, 188, 791

Hunstead, R. W. 1972, MNRAS, 157, 367

Iliev, I. T., Ciardi, B., Alvarez, M. A., Maselli, A., Ferrara, A., Gnedin, N. Y., Mellema, G., Nakamoto, T., Norman, M. L., Razoumov, A. O., Rijkhorst, E.-J., Ritzerveld, J., Shapiro, P. R., Susa, H., Umemura, M., & Whalen, D. J. 2006a, MNRAS, 371, 1057

Iliev, I. T., Mellema, G., Pen, U., Merz, H., Shapiro, P. R., & Alvarez, M. A. 2006b, MNRAS, 369, 1625

Iliev, I. T., Mellema, G., Pen, U.-L., Bond, J. R., & Shapiro, P. R. 2008, MNRAS, 384, 863

Iliev, I. T., Mellema, G., Shapiro, P. R., Pen, U.-L., Mao, Y., Koda, J., & Ahn, K. 2012, MNRAS, 423, 2222

Jelić, V., Zaroubi, S., Labropoulos, P., Bernardi, G., de Bruyn, A. G., & Koopmans, L. V. E. 2010, MNRAS, 409, 1647

Jelić, V., Zaroubi, S., Labropoulos, P., Thomas, R. M., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Ciardi, B., Harker, G., Koopmans, L. V. E., Pandey, V. N., Schaye, J., & Yatawatta, S. 2008, MNRAS, 389, 1319

Kassim, N. E., Lazio, T. J. W., Ray, P. S., Crane, P. C., Hicks, B. C., Stewart, K. P., Cohen, A. S., & Lane, W. M. 2004, Planet. Space Sci., 52, 1343

Kim, J. & Pen, U. 2009, preprint (astro-ph/0908.1973)

Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2011, ApJS, 192, 18

Kovetz, E. D. & Kamionkowski, M. 2013, Phys. Rev. D, 87, 063516

Lah, P., Chengalur, J. N., Briggs, F. H., Colless, M., de Propris, R., Pracy, M. B., de Blok, W. J. G., Fujita, S. S., Ajiki, M., Shioya, Y., Nagao, T., Murayama, T., Taniguchi, Y., Yagi, M., & Okamura, S. 2007, MNRAS, 376, 1357

Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., & Hernquist, L. 2008, ApJ, 680, 962

Liu, A. & Tegmark, M. 2011, Phys. Rev. D, 83, 103006

—. 2012, MNRAS, 419, 3491

Liu, A., Tegmark, M., Bowman, J., Hewitt, J., & Zaldarriaga, M. 2009a, MNRAS, 398, 401

Liu, A., Tegmark, M., & Zaldarriaga, M. 2009b, MNRAS, 394, 1575

Loeb, A. & Zaldarriaga, M. 2004, Phys. Rev. Lett., 92, 211301

Lonsdale, C. J., Cappallo, R. J., Morales, M. F., Briggs, F. H., Benkevitch, L., Bowman, J. D., Bunton, J. D., Burns, S., Corey, B. E., Desouza, L., Doeleman, S. S., Derome, M., Deshpande, A., Gopala, M. R., Greenhill, L. J., Herne, D. E., Hewitt, J. N., Kamini, P. A., Kasper, J. C., Kincaid, B. B., Kocz, J., Kowald, E., Kratzenberg, E., Kumar, D., Lynch, M. J., Madhavi, S., Matejek, M., Mitchell, D. A., Morgan, E., Oberoi, D., Ord, S., Pathikulangara, J., Prabu, T., Rogers, A., Roshi, A., Salah, J. E., Sault, R. J., Shankar, N. U., Srivani, K. S., Stevens, J., Tingay, S., Vaccarella, A., Waterson, M., Wayth, R. B., Webster, R. L., Whitney, A. R., Williams, A., & Williams, C. 2009, IEEE Proceedings, 97, 1497

Lorimer, D. R., Yates, J. A., Lyne, A. G., & Gould, D. M. 1995, MNRAS, 273, 411

Mack, K. J. & Wesley, D. H. 2008, preprint (astro-ph/0805.1531)

Madau, P., Meiksin, A., & Rees, M. J. 1997, ApJ, 475, 429

Majumdar, S., Bharadwaj, S., & Choudhury, T. R. 2012, MNRAS, 426, 3178

Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993

Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., & Zahn, O. 2008, Phys. Rev. D, 78, 023529

Masui, K. W., McDonald, P., & Pen, U. 2010, Phys. Rev. D, 81, 103527

Masui, K. W., Switzer, E. R., Banavar, N., Bandura, K., Blake, C., Calin, L.-M., Chang, T.-C., Chen, X., Li, Y.-C., Liao, Y.-W., Natarajan, A., Pen, U.-L., Peterson, J. B., Shaw, J. R., & Voytek, T. C. 2013, ApJ, 763, L20

McGreer, I. D., Mesinger, A., & Fan, X. 2011, MNRAS, 415, 3237

McQuinn, M., Lidz, A., Zahn, O., Dutta, S., Hernquist, L., & Zaldarriaga, M. 2007, MNRAS, 377, 1043

McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, ApJ, 653, 815

Mezzacappa, A. 2005, Annual Review of Nuclear and Particle Science, 55, 467

Millenaar, R. P. 2011, SKA Report, `http://www.skatelescope.org/the-location/site-documentation/`

Miralda-Escudé, J., Haehnelt, M., & Rees, M. J. 2000, ApJ, 530, 1

Morales, M. F. 2005, ApJ, 619, 678

Morales, M. F., Bowman, J. D., Cappallo, R., Hewitt, J. N., & Lonsdale, C. J. 2006a, New Astronomy Review, 50, 173

Morales, M. F., Bowman, J. D., & Hewitt, J. N. 2006b, ApJ, 648, 767

Morales, M. F. & Hewitt, J. 2004, ApJ, 615, 7

Morales, M. F. & Wyithe, J. S. B. 2010, ARA&A, 48, 127

Muller, C. A. & Oort, J. H. 1951, Nature, 168, 357

Myers, S. T., Contaldi, C. R., Bond, J. R., Pen, U., Pogosyan, D., Prunet, S., Sievers, J. L., Mason, B. S., Pearson, T. J., Readhead, A. C. S., & Shepherd, M. C. 2003, ApJ, 591, 575

Nityananda, R. 2010, NCRA Technical Report, `http://ncralib1.ncra.tifr.res.in:8080/jspui/handle/2301/484`

Offringa, A. R. 2010, AOFlagger RFI Software, astrophysics Source Code Library

Offringa, A. R., de Bruyn, A. G., Zaroubi, S., van Diepen, G., Martinez-Ruby, O., Labropoulos, P., Brentjens, M. A., Ciardi, B., Daiboo, S., Harker, G., Jelić, V., Kazemi, S., Koopmans, L. V. E., Mellema, G., Pandey, V. N., Pizzo, R. F., Schaye, J., Vedantham, H., Veligatla, V., Wijnholds, S. J., Yatawatta, S., Zarka, P., Alexov, A., Anderson, J., Asgekar, A., Avruch, M., Beck, R., Bell, M., Bell, M. R., Bentum, M., Bernardi, G., Best, P., Birzan, L., Bonafede, A., Breitling, F., Broderick, J. W., Brüggen, M., Butcher, H., Conway, J., de Vos, M., Dettmar, R. J., Eisloeffel, J., Falcke, H., Fender, R., Frieswijk, W., Gerbers, M., Griessmeier, J. M., Gunst, A. W., Hassall, T. E., Heald, G., Hessels, J., Hoeft, M., Horneffer, A., Karastergiou, A., Kondratiev, V., Koopman, Y., Kuniyoshi, M., Kuper, G., Maat, P., Mann, G., McKean, J., Meulman, H., Mevius, M., Mol, J. D., Nijboer, R., Noordam, J., Norden, M., Paas, H., Pandey, M., Pizzo, R., Polatidis, A., Rafferty, D., Rawlings, S., Reich, W., Röttgering, H. J. A., Schoenmakers, A. P., Sluman, J., Smirnov, O., Sobey, C., Stappers, B., Steinmetz, M., Swinbank, J., Tagger, M., Tang, Y., Tasse, C., van Ardenne, A., van Cappellen, W., van Duin, A. P., van Haarlem, M., van Leeuwen, J., van Weeren, R. J., Vermeulen, R., Vocks, C., Wijers, R. A. M. J., Wise, M., & Wucknitz, O. 2013, A&A, 549, A11

Oh, S. P. & Mack, K. J. 2003, MNRAS, 346, 871

Paciga, G., Albert, J. G., Bandura, K., Chang, T.-C., Gupta, Y., Hirata, C., Odegova, J., Pen, U.-L., Peterson, J. B., Roy, J., Shaw, J. R., Sigurdson, K., & Voytek, T. 2013, MNRAS, 433, 639

Paciga, G., Chang, T.-C., Gupta, Y., Nityananda, R., Odegova, J., Pen, U.-L., Peterson, J. B., Roy, J., & Sigurdson, K. 2011, MNRAS, 413, 1174

Pandolfi, S., Ferrara, A., Choudhury, T. R., Melchiorri, A., & Mitra, S. 2011, Phys. Rev. D, 84, 123522

Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012a, ApJ, 753, 81

Parsons, A. R., Backer, D. C., Foster, G. S., Wright, M. C. H., Bradley, R. F., Gugliucci, N. E., Parashare, C. R., Benoit, E. E., Aguirre, J. E., Jacobs, D. C., Carilli, C. L., Herne, D., Lynch, M. J., Manley, J. R., & Werthimer, D. J. 2010, AJ, 139, 1468

Parsons, A. R., Liu, A., Aguirre, J. E., Ali, Z. S., Bradley, R. F., Carilli, C. L., DeBoer, D. R., Dexter, M. R., Gugliucci, N. E., Jacobs, D. C., Klima, P., MacMahon, D. H. E., Manley, J. R., Moore, D. F., Pober, J. C., Stefan, I. I., & Walbrugh, W. P. 2013, preprint (astro-ph/1304.4991)

Parsons, A. R., Pober, J. C., Aguirre, J. E., Carilli, C. L., Jacobs, D. C., & Moore, D. F. 2012b, ApJ, 756, 165

Pawsey, J. L. 1951, Nature, 168, 358

Pearson, T. J. & Kus, A. J. 1978, MNRAS, 182, 273

Pen, U., Chang, T., Hirata, C. M., Peterson, J. B., Roy, J., Gupta, Y., Odegova, J., & Sigurdson, K. 2009, MNRAS, 399, 181

Peterson, J., Pen, U., & Wu, X. 2004, preprint (astro-ph/0404083)

Petrovic, N. & Oh, S. P. 2011, MNRAS, 413, 2103

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., & et al. 2013, preprint (astro-ph/1303.5076)

Platania, P., Burigana, C., Maino, D., Caserini, E., Bersanelli, M., Cappellini, B., & Mennella, A. 2003, A&A, 410, 847

Pritchard, J. R. & Loeb, A. 2010, Phys. Rev. D, 82, 023006

Purcell, E. M. & Field, G. B. 1956, ApJ, 124, 542

Rawlings, S. & Schilizzi, R. 2011, preprint (astro-ph/1105.5953)

Röttgering, H. J. A., Braun, R., Barthel, P. D., van Haarlem, M. P., Miley, G. K., Morganti, R., Snellen, I., Falcke, H., de Bruyn, A. G., Stappers, R. B., Boland, W. H. W. M., Butcher, H. R., de Geus, E. J., Koopmans, L., Fender, R., Kuijpers, J., Schilizzi, R. T., Vogt, C., Wijers, R. A. M. J., Wise, M., Brouw, W. N., Hamaker, J. P., Noordam, J. E., Oosterloo, T., Bahren, L., Brentjens, M. A., Wijnholds, S. J., Bregman, J. D., van Cappellen, W. A., Gunst, A. W., Kant, G. W., Reitsma, J., van der Schaaf, K., & de Vos, C. M. 2006, preprint (astro-ph/0610596)

Roy, J., Gupta, Y., Pen, U., Peterson, J. B., Kudale, S., & Kodilkar, J. 2010, Experimental Astronomy, 28, 25

Salvaterra, R., Haardt, F., & Volonteri, M. 2007, MNRAS, 374, 761

Schroeder, J., Mesinger, A., & Haiman, Z. 2013, MNRAS, 428, 3058

Schwab, F. R. 1984, AJ, 89, 1076

Scott, D. & Rees, M. J. 1990, MNRAS, 247, 510

Shapiro, P. R., Ahn, K., Alvarez, M. A., Iliev, I. T., Martel, H., & Ryu, D. 2006, ApJ, 646, 681

Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2013, preprint (astro-ph/1302.0327)

SKA Expert Panel on Radio Frequency Interference. 2011, SKA Report, `http://www.skatelescope.org/the-location/site-documentation/`

Su, M., Yadav, A. P. S., McQuinn, M., Yoo, J., & Zaldarriaga, M. 2011, preprint (astro-ph/1106.4313)

Sunyaev, R. A. & Zeldovich, I. B. 1975, MNRAS, 171, 375

Swarup, G., Ananthakrishnan, S., Kapahi, V. K., Rao, A. P., Subrahmanya, C. R., & Kulkarni, V. K. 1991, Current Science, 60, 95

Swarup, G. & Venkatasubramani, T. 1991, in Astronomical Society of the Pacific Conference Series, Vol. 17, IAU Colloq. 112: Light Pollution, Radio Interference, and Space Debris, ed. D. L. Crawford, 190

Taylor, G. B., Carilli, C. L., & Perley, R. A., eds. 1999, Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II

Thomas, R. M., Zaroubi, S., Ciardi, B., Pawlik, A. H., Labropoulos, P., Jelić, V., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Harker, G. J. A., Koopmans, L. V. E., Mellema, G., Pandey, V. N., Schaye, J., & Yatawatta, S. 2009, MNRAS, 393, 32

Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2001, Interferometry and Synthesis in Radio Astronomy, 2nd Edition, ed. Thompson, A. R., Moran, J. M., & Swenson, G. W., Jr.

Trac, H. Y. & Gnedin, N. Y. 2011, Advanced Science Letters, 4, 228

Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, ApJ, 757, 101

Tseliakhovich, D. & Hirata, C. 2010, Phys. Rev. D, 82, 083520

van de Hulst, H. C., Muller, C. A., & Oort, J. H. 1954, Bull. Astron. Inst. Netherlands, 12, 117

van de Hulst, H. C., Raimond, E., & van Woerden, H. 1957, Bull. Astron. Inst. Netherlands, 14, 1

van Straten, W. 2004, ApJS, 152, 129

Verheijen, M., van Gorkom, J. H., Szomoru, A., Dwarakanath, K. S., Poggianti, B. M., & Schiminovich, D. 2007, ApJ, 668, L9

Wang, X., Tegmark, M., Santos, M. G., & Knox, L. 2006, ApJ, 650, 529

White, M., Carlstrom, J. E., Dragovan, M., & Holzapfel, W. L. 1999, ApJ, 514, 12

Willis, A. G., Oosterbaan, C. E., & de Ruiter, H. R. 1976, A&AS, 25, 453

Willott, C. J., Delorme, P., Omont, A., Bergeron, J., Delfosse, X., Forveille, T., Albert, L., Reylé, C., Hill, G. J., Gully-Santiago, M., Vinten, P., Crampton, D., Hutchings, J. B., Schade, D., Simard, L., Sawicki, M., Beelen, A., & Cox, P. 2007, AJ, 134, 2435

Willott, C. J., Rawlings, S., & Blundell, K. M. 2001, MNRAS, 324, 1

Wouthuysen, S. A. 1952, AJ, 57, 31

Yatawatta, S., de Bruyn, A. G., Brentjens, M. A., Labropoulos, P., Pandey, V. N., Kazemi, S., Zaroubi, S., Koopmans, L. V. E., Offringa, A. R., Jelic, V., Martinez Rubi, O., Veligatla, V., Wijnholds, S. J., Brouw, W. N., Bernardi, G., Ciardi, B., Daiboo, S., Harker, G., Mellema, G., Schaye, J., Thomas, R., Vedantham, H., Chapman, E., Abdalla, F. B., Alexov, A., Anderson, J., Avruch, I. M., Batejat, F., Bell, M. E., Bell, M. R., Bentum, M., Best, P., Bonafede, A., Bregman, J., Breitling, F., van de Brink, R. H., Broderick, J. W., Bruggen, M., Conway, J., de Gasperin, F., de Geus, E., Duscha, S., Falcke, H., Fallows, R. A., Ferrari, C., Frieswijk, W., Garrett, M. A., Griessmeier, J. M., Gunst, A. W., Hassall, T. E., Hessels, J. W. T., Hoeft, M., Iacobelli, M., Juette, E., Karastergiou, A., Kondratiev, V. I., Kramer, M., Kuniyoshi, M., Kuper, G., van Leeuwen, J., Maat, P., Mann, G., McKean, J. P., Mevius, M., Mol, J. D., Munk, H., Nijboer, R., Noordam, J. E., Norden, M. J., Orru, E., Paas, H., Pandey-Pommier, M., Pizzo, R., Polatidis, A. G., Reich, W., Rottgering, H. J. A., Sluman, J., Smirnov, O., Stappers, B., Tagger, M., Tang, Y., Tasse, C., ter Veen, S., Vermeulen, R., van Weeren, R. J., Wise, M., Wucknitz, O., & Zarka, P. 2013, A&A, 550, A136

Zahn, O., Lidz, A., McQuinn, M., & Dutta, S. a. 2007, ApJ, 654, 12

Zahn, O., Reichardt, C. L., Shaw, L., Lidz, A., Aird, K. A., Benson, B. A., Bleem, L. E., Carlstrom, J. E., Chang, C. L., Cho, H. M., Crawford, T. M., Crites, A. T., de Haan, T., Dobbs, M. A., Doré, O., Dudley, J., George, E. M., Halverson, N. W., Holder, G. P., Holzapfel, W. L., Hoover, S., Hou, Z.,

Hrubes, J. D., Joy, M., Keisler, R., Knox, L., Lee, A. T., Leitch, E. M., Lueker, M., Luong-Van, D., McMahon, J. J., Mehl, J., Meyer, S. S., Millea, M., Mohr, J. J., Montroy, T. E., Natoli, T., Padin, S., Plagge, T., Pryke, C., Ruhl, J. E., Schaffer, K. K., Shirokoff, E., Spieler, H. G., Staniszewski, Z., Stark, A. A., Story, K., van Engelen, A., Vanderlinde, K., Vieira, J. D., & Williamson, R. 2012, ApJ, 756, 65

Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, ApJ, 608, 622

Zaroubi, S., de Bruyn, A. G., Harker, G., Thomas, R. M., Labropolous, P., Jelić, V., Koopmans, L. V. E., Brentjens, M. A., Bernardi, G., Ciardi, B., Daiboo, S., Kazemi, S., Martinez-Rubi, O., Mellema, G., Offringa, A. R., Pandey, V. N., Schaye, J., Veligatla, V., Vedantham, H., & Yatawatta, S. 2012, MNRAS, 425, 2964

Zaroubi, S. & Silk, J. 2005, MNRAS, 360, L64

Zel'Dovich, Y. B. 1970, A&A, 5, 84

Zwaan, M. A., van Dokkum, P. G., & Verheijen, M. A. W. 2001, Science, 293, 1800