

PSA-64 POWER SPECTRUM

ZAKI S. ALI¹, AARON R. PARSONS^{1,2}, JEFF ZHENG, JONATHAN C. POBER⁴, JAMES E. AGUIRRE³, DAVID R. DEBOER²,
 DANIEL C. JACOBS⁸, ADRIAN LIU¹, DAVID F. MOORE³

Draft version August 7, 2014

ABSTRACT

Subject headings:

1. INTRODUCTION

The Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER) is a dedicated experiment to measure the power spectrum of highly redshifted 21 cm emission during the Epoch of Reionization. PAPER is but one experiment that aims to detect this faint signal. Other telescopes that have the same goal are the Giant Meter-wave Radio Telescope (GMRT), the LOW Frequency ARray (LOFAR), and the Murchison Wide-field Array (MWA). PAPER currently consists of 128 dual-polarization antennae in a 100-200MHz band out in the Karoo desert in South Africa.

The current best upper limit on 21 cm signal level is at $(41mK)^2$ at $k = .27hMpc^{-1}$ which was measured by PAPER in a 32 antenna redundant configuration (??). This limit was achieved by using the delay-spectrum technique to remove foregrounds and by using a maximum redundancy array to essentially measure the same Fourier mode repeatedly, boosting sensitivity. In this analysis we employ the same techniques mentioned, as well as introduce an optimal fringe-fringe rate filter to boost sensitivity and make use of improved calibration via the Omnical redundant calibrator package.

The paper is outlined as follows. In section ?? we describe the observations used in this analysis. In ?? we discuss the improvements in this pipeline with respect to the previous analysis of PSA-32 ?. We then move on to the data analysis pipeline in section ?. Section ?? describes the results of our efforts and provides new constraints on EoR. We conclude in ??.

2. OBSERVATIONS

?? Here, we describe the features of the data set used in this analysis. The analysis described in this paper focuses on the 64 antenna array deployed in the Fall of 2012. We used a maximally redundant configuration of the PAPER array (see Figure 1) for this analysis, relying on all of the redundant baselines for the calibration procedure, but only using a subset of the baselines for the power spectrum analysis. The columns are separated by 30 meters and the rows by 5 meters. For the power spectrum analysis we are using the baselines that correspond to the width between two columns (e.g. 49-41) as well as those that correspond to over and up and down one antenna

(e.g. 10-41 and 10-58, respectively). These 154 baselines are instantaneously redundant and therefore they measure the same Fourier modes on the sky. Within a single group of the three types of baselines above, the measurements add coherently, whereas between groups they add in quadrature. Having repeated measurements of the same baseline type greatly increases sensitivity.

The observation of this 64 antenna data set spanned a 135 day period that commenced on 2012 November 8 (JD62456240) and ended 2013 March 23 (JD62456375). Each baseline instantaneously measured the 100-200 MHz band which was divided into 1024 frequency channels of resolution 97.66 kHz and integrated for 10.7 seconds. In this analysis we analyze observations that spanned, in local sidereal time (LST), a range of 1:00 to 10:00 hours. This range corresponds to the "EoR cold patch", in which galactic synchrotron power is minimal (away from the center of the galaxy).

3. SUMMARY OF IMPROVEMENTS FROM PSA32

In comparison to the previous PAPER pipeline (see ?), this analysis took a slightly different approach which included some critical steps to improve our upper limit. In short, the improvements included using a new, refined redundant calibration method (Zheng 2014), increasing the width of the wideband delay filter that removes smooth spectrum foregrounds, weighting the 1st binned data sets, and optimal fringe rate filtering. In section ??, we discuss each of the improvements in more detail.

Figure ?? (TBD) shows the power spectra when each of the steps mentioned above are turned off and for the one where all of them are turned on. We can see the gradual improvement of the power spectra (hopefully).

4. ANALYSIS

Here we describe the analysis pipeline of the data set obtained in 2012/2013 observing season. Data was first run through a preprocessing compression pipeline which reduces the volume of the data by a factor of forty. Afterwards we calibrate the relative phases of our array on the basis of redundancy using a logical approach in delay space, and then get the absolute phase calibration by fitting. We then set the flux scale of our observations by using Pictor A. The logical done above was a rough estimate of the phase calibration and hence we used Omnical to get the calibrations to higher accuracy. This was a major difference in the pipeline from previous iterations on PAPER data. We then remove foregrounds using the delay filtering technique and form power spectra. The details of the above are discussed below and figure ?? shows a block diagram of the analysis pipeline.

¹ Astronomy Dept., U. California, Berkeley, CA

² Radio Astronomy Lab., U. California, Berkeley, CA

³ Dept. of Physics and Astronomy, U. Pennsylvania, Philadelphia, PA

⁴ Physics Dept. U. Washington, Seattle, WA

⁸ School of Earth and Space Exploration, Arizona State U., Tempe, AZ

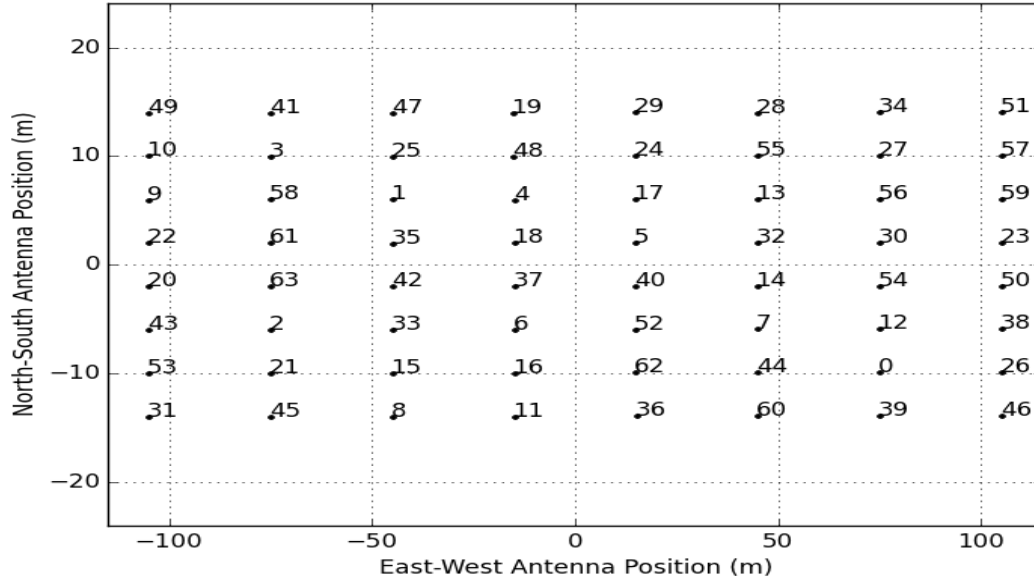


FIG. 1.— Antenna positions for the PAPER 64 observation run.

4.1. Preprocessing Compression

As the number of elements in an interferometer increase, the amount of data to be stored increase by order ΩN^2 , where N is the number antennas. Hence, a doubling of an array, increases the data volume by a factor of 4. For large N arrays, this volume of data could be prohibitive to store on disks and can become a bottleneck in data analysis due it being slow to comb trthrough the data. New methods of data compression and analysis that can either reduce the data volume or provide a streaming analysis pipeline are needed for these large arrays.

PAPER employs a compression method that employs the use of a delay/delay-rate filter described in ?. In short, the algorithm makes use of the fact that for a given maximum baseline (in our case, 300 meters), there is a maximum delay sky emission can enter a baseline, given by

$$\tau_g < \frac{|\mathbf{b}|}{c}. \quad (1)$$

Hence, if we filter out delays that are higher than this maximum delay, we can down sample the frequency axis by the Nyquist rate. This is where reduction in data volume comes in. Similarly, in delay-rate space, there is a maximum delay rate a model point source can have, which is given by,

$$-\omega_{oplus} \frac{|b_y|}{c} \leq \frac{d\tau_g}{dt} \leq \omega_{oplus} \frac{|b_y|}{c}. \quad (2)$$

This equation is valid at an Hour Angle of $h = 0^\circ$ and a declination of $\delta = 0^\circ$, where the maximum delay-rates occur. By filtering out delay-rates that are outside of this range, we are removing sources that have a high rate of change with respect to the rotation of the earth. That is fast transient sources will be removed. We can then decimate in time with respect to the Nyquist rate,

giving us another dimension to decrease data volume. For greater detail, see ?.

Note that the only parameter the compression described above depends on is a baseline orientation and length. For this dataset, we used a baseline oriented along the east-west direction at a latitude of -30° with a length of 300 meter.s Applying the compression to a 10 minute file consisting of 1024 channels over a 100 MHz bandwidth and 60 time samples consisting of 10 second integrations, outputs a file consisting of 203 channels over the same bandwidth and 14 time samples consisting of 42.8 second integrations. This is a factor of 20 reduction in data volume. Originally, this dataset required 40 TB of storage in its raw form. After compression, this data set required 2TB of storage. This reduction in data volume greatly facilitated analysis (how?).

(Show plot of rraw data and compressed data?)

4.2. Calibrations

4.2.1. Overview

Redundant calibration (?, ?) uses redundancy within the array to solve for relative phases and gains of the antennas. To explain redundant calibration, suppose that the baseline between antennas i, j measure a visibility v_{ij} , then we have

$$v_{ij} = g_i g_j^* y_{i-j} + n_{ij}, \quad (3)$$

where g_i, g_j are the complex gains due to antenna i and antenna j , respectively, y_{i-j} is the true visibility measured by perfect antennas i, j for the given baseline type, and n_{ij} is the residual noise from the baseline. If the number of baselines of a given type is much greater than the number of baselines types this problem is overconstrained and g_i, g_j , and y_{i-j} can be solved for. PAPER is in this limit due to the maximally redundant configuration as shown in figure ??.

Redundant calibration comes in two flavors: log calibration and linear calibration. Log calibration, or logcal

for short, takes the logarithm of equation 3 to give a linearized system. Hence, solutions can be solved for by using standar linear algebra techniques. However, this method is biased. On the other hand linear calibration, or lincal for short, is an unbiased method of solving for the complex gain solutions. In this method equation 3 is Taylor expanded about an initial guess for the g_i 's and y_{i-j} 's to give a linearized equation which can be used to solve for the complex gains and sky model.

In this analysis we used a logcal algorithm based in delay space to get a rough calibration of the dataset. This was followed by an absolute calibration to set the overall phase and flux scale using a self calibration. We used model phase centers of Pictor, Fornax A, and Crab Nebula. The absolute amplitude is set by the flux of Pictor A found in ?, whose spectrum is defined by

$$S_\nu = 382 \left(\frac{\nu}{150 \text{ MHz}} \right)^{-.76} Jy. \quad (4)$$

Finally, we used the Omnical calibration package to do another round of redundant calibration to get even more accurate calibration parameters.

4.3. logcal-for lack of a better title

We first perform the same calibration that was done in (?). That is, we use redundancy to do a relative phase⁶ calibration between antennas, which removes the electrical delays from cables in the signal path. Due to redundancy, we can calibrate out all of the per-antennas delays in the signal path relative to two parameters which we call τ_{ns} and τ_{es} . These delays are the relative electrical delays that correspond to baseline delays in the north-south and east-west component for 2 reference baselines (49-10 and 49-41, respectively). The antenna based delay solutions vary as much as a couple nanoseconds day to day when solutions are averaged over hour long timescales withing a day. However, the variations in solutions is worse when only averaging over ten minute time scales. Therefore need for better calibration is required. We use self calibration to derive the two unknow parameters, τ_{ns} and τ_{ew} , by using the Crab Nebula, Fornax A, and Pictor A.

Note that there is no possibility of signal loss (see (?)).

4.4. Gain Calibration

Gain calibration was derived on the basis of redundancy and self calibration. The phase calibrations described above, simultaneously also calibrated for the relative gain variation between antennas. Again we can only calibrate to a fiducial antenna (49) whose gain is defined as unity. We then perform a self calibration to set the flux scale to Pictor A whose spectrum is derived in (?). We use the same methods describes in (?).

Figure ?? shows that dataset beamformed to Pictor A, with log janskies on the y axis and lst on the xaxis for a frequency of $.1 + (120/203) * .1/203$. As can be seen, the day to day variation in the formed beam has a fractional spread of about 10%. This shows the stability of the instrument and the well behaved calibration solutions derived above.

⁶ In actuality, we solve for delays to get around phase wrapping issues. These delays are applied to visibilities as $e^{2\pi i \tau \nu}$

4.5. Omnical

(How did we know that our calibrations were not good enough? Because of the power spectrum? PSA32? We did beamform data to pictorA and say that vs LST, the beamform matched well day to day with a fractional spread of about 10%)

The complex gain solutions found in the previous calibration pipeline were averaged together in time and one solution per frequency was used for the array. This jived with the philosophy that the array was stable in time and frequency. However, upon further review of this data set, it seemed more and more likely that this was not the case anymore. (Is this even true? What specifically? Think Man!)

Due to clues that showed that our data set had time dependent calibration solutions, it was imperative that we do a better job at calibrating our array.

The Omnical redundant calibrator package⁷ (omnical) performs redundant calibration for every time and frequency in a dataset using both logcal and lincal methods as described in ?. It also contains methods on the quality of the solutions by providin a chi-square for the fits to the data.

For this dataset, omnical first performed a logcal (again) to attain a solution per time and frequency. This solution was passed to lincal which iteratively solved for the complex gain solutions. The convergence criteria was when the χ^2 decreased by less than .01%. The χ^2 for the fit used in Omnical is given by

$$\chi^2 = \sum_{ij} |v_{ij} - y_{i-j} g_i^* g_j|^2, \quad (5)$$

which differs from normal nomenclature because we are not inverse varaince weighting. Note that this χ^2 is summing over all baselines and hence giving more weight to higher gains. Note that omnical fits for each of the complex gains and the model visibility, y_{i-j} , for a unique baseline. Using this information, figure ?? shows that the χ^2 is close to 1 for all channels and time (for this day of data). need noise model for this.

Figure ?? shows the gain solutions output by omnical. The amplitude of the gains are roughly order unity through out. These are relative gains between antennas and hence the over flux scale set to Pictor A is still valid. The absolute calibration is still valid.

Since Omnical outputs a model visibility of what a unique baseline should measure, which is derived from the data by removing all of the variation between unique types of baselines and averaging, we are able to use these outputs as our dataset. Infact, this is what is done.

4.6. lstbinning

Need to flesh this out more XXX The Omnical calibration pipeline removes all of the variation between unique baseline types and averages them together, effectively creating a model for what a baseline measures. This is not simulated and is derived from the data. These averaged models are what we use to continue through the analysis pipeline.

The motivation for using these datasets is that they are pre calibrated and have baseline to baseline variation removed from them, including cross talk and other

⁷ <https://github.com/jeffzhen/omnical>

systematics. This is essentially the averaged of what a unique baseline type should measure. This includes RFI events that effect the entire array and sky signal.

We 1stbin these data into 1st bins of 42.8 seconds over the entire observation.

4.7. WB delay filtering

Galactic synchrotron and extragalactic point sources, generally foregrounds, greatly contaminate the EOR signal. They are 5 orders of magnitude (in K) (Pober 2013a) above the expected level of EOR and can hide low order RFI events and cross-talk. Removing foregrounds is a critical step in getting to a detection of EoR. In addition to dominating the EOR signal, foregrounds can corrupt higher-order k modes in the power spectrum measurement from the sidelobes arising from RFI flagging and the finite bandwidth used in the line of sight Fourier transform (corresponding to k_{\parallel}).

PAPER uses the delay filtering technique described in (?) to remove smooth spectrum foregrounds that mask EoR. Taking the fourier transform along the frequency axis for a visibility, which defines the delay transform, maps sky emission in to delay space. Flat spectrum sources in this space are show up as dirac delta functions. This method of localizing sources relies on the fact that for a given baseline, there is a maximum delay points on the sky can come in at. This maximum delay is given light travel time between the two antennas in a baseline. Physically, this corresponds to emission from the horizon. Therefore, all sky emission is localized to this "horizon limit."

Following the technique described in (?), the delay transform of a visibility takes the form

$$\tilde{V} = \int W(\nu)S(\nu)V(\nu)e^{-2\pi i\tau\nu}d\nu \quad (6)$$

$$= \tilde{W}(\tau) * \tilde{S}(\tau) * \tilde{V}(\tau) * \delta(\tau_s - \tau) \quad (7)$$

where $V(\nu)$ is the visibility measured, $W(\nu)$ is a blackman-harris windowing function and $S(\nu)$ is the weighting function that encodes the boolean flags for the data. As mentioned above, a flat spectrum source will show up as a delta function in delay space, but if there is any spectral structure, the shape of that structure will become a convolving kernel in delay space, increasing the footprint of the source in delay space. Since foregrounds are modeled as having a smooth spectrum, the convolving kernel is narrow in delay space, whereas EoR is expected to have a large amount of frequency structure and hence has a wide convolution kernel, scattering signal outside of the horizon limit. By removing smooth spectrum foregrounds by fourier filtering in delay space, we can remove bright foregrounds that keep us from detecting EoR.

On a practical note, the boolean flags in equation 6 are usually not regularly sampled and create structure in delay space due to the uneven sampling and sharp frequency structure. This will cause the normally localized smooth spectrum foregrounds to scatter emission in to otherwise uncorrupted delay modes. As described in ?, we use a CLEAN algorithm that treats the boolean flags as a sampling function and iteratively fits the brightest peak in delay space to some threshold. This does an excellent job in filling in the gaps of the delay transform,

due to the exactly known sampling function, and returns a model of the data without any flags.

In the above CLEAN, we restrict the clean components to fall within 15 ns beyond the horizon. This limit lets CLEAN model foreground emission pushed beyond the horizon (due to spectral structure). This CLEAN model is then subtracted from the delay transform and suppresses smooth spectrum foreground. As seen in figure ??, this filter is responsible for 3 orders of magnitude of foreground removal (in mK) within the horizon. For our 30 meter baselines used in this analysis, the horizon limit is at 100 nanoseconds in delay domain, which at a frequency of 164 MHz, corresponds to $k = \pm 0.057 h Mpc^{-1}$. Talk about signal loss. Using the entire bandwidth increases the degree of foreground separation and is imperative to get the best estimate of foreground isolation. We apply the delay transform defined above to every baseline over the entire 100 Mhz. Figure ?? shows the localization of foregrounds for a 30 meter baseline within the horizon of 100 ns. After the filter is applied we see 3 orders of magnitude in reduction of foreground isolation. This is a factor of 6 in powerspectral!

After the wideband delay filter, we conduct another round of RFI and crosstalk removal which was overshadowed by the foreground signal. For RFI excision we apply a median filter which flags above 3σ . Foregrounds kept us using the traditional method of removing cross talk which consisted of subtracting hour long averages from each visibility. This was due to the fact that some days in the observation had gaps in time owing to some technical difficulties during observations. However, with foregrounds removed we were able to remove 10 minute averages from every visibility within a file (files have a cadence of ten minutes). Normally, hour long averages are needed for foreground contained data to wash out the fringes from the said foregrounds to detect the static phase bias that is crosstalk. For foreground removed data, we do not have this complication and are able to remove the offset by subtracting shorter sums.

4.8. optimal Fringe-rate filter

Before forming power spectra we need to time average visibilities that measure the same k mode on the sky. This is the best way to combine data because we get a \sqrt{N} , where N is the number of samples, gain in sensitivity. This is in opposition to weighting after forming power spectra, where noise beats down as the \sqrt{N} . Rather than a straight averaging in time, we can do better by using a weighted average. Specifically, we want to upweight samples that have higher signal-to-noise.

This is achieved by applying a carefully crafted filter in fringe rate domain, the fourier dual to time. Different patches on the sky correspond to different fringe rates, for a given baseline and frequency. Maximum fringe rates are found along the equatorial plane where zero fringe rates are found around the poles (sources at the poles do not move through the fringes of a baselines). Sources on the other side of the pole, corresponds to negative fringe rates, because they move through the fringe pattern in the opposite direction. By weighing the fringe rates sampled by a baseline by the beam response of the baseline, gives us the optimal fringe rate filter to use for time averaging. See Parsons/Liu 2014 for a detailed discussion.

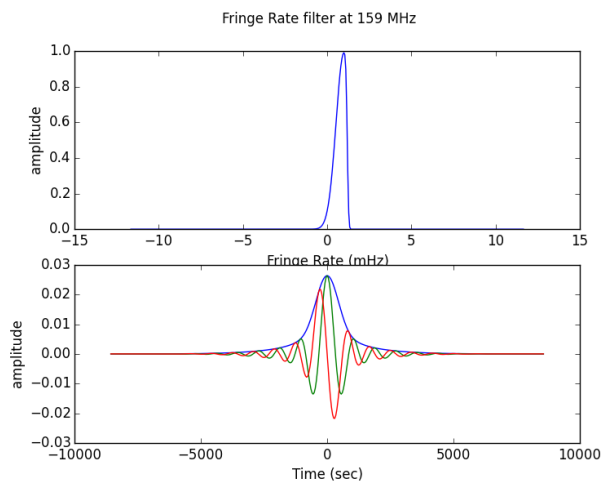


FIG. 2.— slice of a fringe rate filter at a frequency of 159MHz. Top is the filter in fringe rate domain. The bottom consists of the corresponding time domain filter gotten by fourier transforming and windowing with a blackman-harris window to damp the tails.

??

We implement the optimal fringe rate filter by first calculating the fringe rate of every point on the sky and weighting it the beam of a given baseline, summing along constant fringe rate contours. This is then squared to get the effective sensitivity of each fringe rate bin (squaring the power beam). That is we are upweighting fringe rate bins that have greater signal-to-noise. We then fit a gaussian to the optimal filter to have a smooth function, along with tanh function for to have a smooth cut off at the maximum fringe rate for the given baseline and

frequency. Note that this only calculated for a given frequency and scaled to other frequencies, due to the fact that fringe rate scales linearly with frequency via

$$f_{max} = \frac{|\mathbf{b}_{eq}|}{c} \omega_{\oplus} \nu \quad (8)$$

To minimize edge effects in time we fourier transform the optimal fringe rate filter to get the time domain convolution kernel and apply it to each visibility for a given channel and baseline.

The exact implementation of the optimal fringe rate filter, uses a convolution kernel of 400 integrations (should we be??). This has the effect of denoising the data as if it averaged over 4.75 hours with a weighing function that is the beam weighted fringe rate. For each of the baseline types used (30 m baselines of different orientations).

Since the PAPER beam is 45 degrees, and the array is located at a declination of -30° the fringe rates associated with the low signal to noise (down in the beam) correspond to very high and very low/negative fringe rates. Figure ?? shows a cut of the optimal fringe rate at MHz for a 30 m east west baseline. Therefore, the implemented fringe rate filter removes some sky signal, signal associated with fringe rates outside of the ranges shown in Figure ?? . Figure ?? shows that the applied filter removes sky associated with negative fringe rates and very high fringe rates.

5. RESULTS

6. SCIENCE

7. CONCLUSIONS

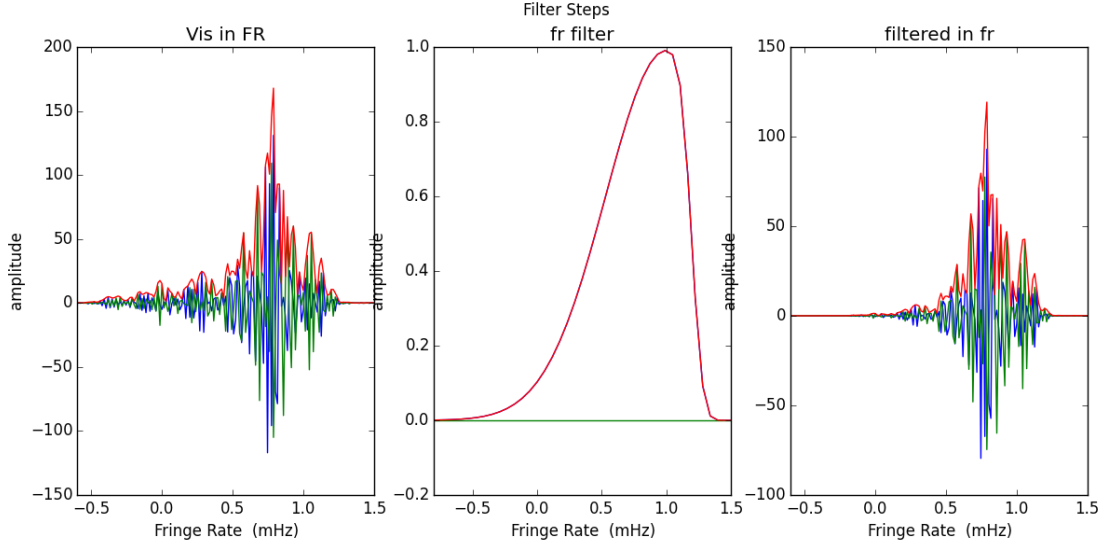


FIG. 3.— slice of a fringe rate filter at a frequency of 159MHz. Shown here is the fringe-rate transform of foreground contained data for a 30 m east-west baseline. Blue and green are the real and imaginary part, respectively and red is the absolute value. Note the maximum and minimum fringerates correspond to the theoretical minimum and maximum for a baseline of this length at 159 MHz. The center panel shows the real (red) and imaginary (green) parts of the fringe rate filter to be applied. Finally, the last panel on the right shows that the fringe rate filtered visibilities. The fringe rate filter is just a weighting applied in fringe rate space and retains foregrounds.

??

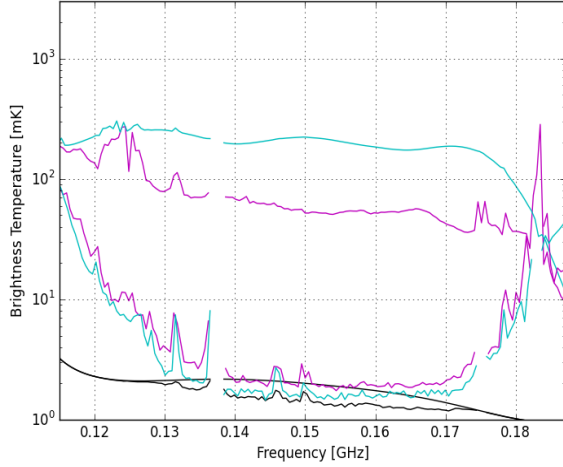


FIG. 4.— Estimates of noise temperature. Magenta is frequency differenced estimate where the cyan is the time differenced estimate. All curves are averaged over all 30m east-west baselines (56) and averaged incoherently in 43s bins of LST from LST 2 to 5 hours with a channel bandwidth of 490 kHz.

??