Multi-redshift limits on the 21cm power spectrum from PAPER

Daniel C. Jacobs[1], Aaron R. Parsons[2,8], James E. Aguirre[3], Zaki Ali[2], Judd Bowman[1], Richard F. Bradley[4,5,6], Chris L. Carilli[7], David R. DeBoer[8], Matthew R. Dexter[8], Nicole E. Gugliucci[5], Pat Klima[5], Adrian Liu[2,11], Dave H. E. MacMahon[8] Jason R. Manley[9], David F. Moore[3], Jonathan C. Pober[2], Irina I. Stefan[10], William P. Walbrugh[9]

ABSTRACT

We present new observations from the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER) telescope which place new lower limits on the HI power spectrum over the redshift range of $7.5 < z < 10.5$, extending previously published single redshift results to cover the full range accessible to the instrument. The epoch of reionization power spectrum is expected to evolve strongly with redshift, and it is this variation with cosmic history that will allow us to begin to place constraints on the physics of reionization. The primary obstacle to this goal are bright foregrounds and the attendant systematics associated with high dynamic range measurements. Here we use foreground removal techniques that take advantage of the large instrumental bandwidth to make a high dynamic range estimate of smooth foreground power. Power spectra at different points across the spectrum reveal the variable efficacy of this subtraction. Noise limited measurements of Δ^2 at k =0.2hMpc^{-1} reach as low as 29mK at redshift 7.5. Most spectra demonstrate a robust foreground removal of foreground signals to the thermal noise limit, leaving systematics on the scale of the noise. We find that in the case of the highest redshift power spectrum, foregrounds begin to dominate over the noise, reflecting the increased difficulty in filtering foregrounds near the edge of the instrumental band.

[1]School of Earth and Space Exploration, Arizona State U., Tempe, AZ

[2]Astronomy Dept., U. California, Berkeley, CA

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

[4]Dept. of Electrical and Computer Engineering, U. Virginia, Charlottesville, VA

[5]National Radio Astronomy Obs., Charlottesville, VA

[6]Dept. of Astronomy, U. Virginia, Charlottesville, VA

[7]National Radio Astronomy Obs., Socorro, NM

[8]Radio Astronomy Lab., U. California, Berkeley, CA

[9]Square Kilometer Array, South Africa Project, Cape Town, South Africa

[10]Cavendish Lab., Cambridge, UK

[11]Berkeley Center for Cosmological Physics, UC Berkeley, Berkeley, CA

Subject headings: reionization

1. Introduction

The Epoch of Reionization, when the first stars ionized the pervasive cosmological Hydrogen in the last global phase change is predicted to be observable in highly redshifted 21 cm radiation. The Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER, XXX)[1] is a low frequency radio interferometer experiment dedicated to opening this window on the universe. Challenges include foregrounds which are brighter by several orders of magnitude and long integration times necessitated by the limited collecting areas of first generation instruments. Direct observation of hydrogen before and during re-ionization is predicted to deliver a wealth of cosmological and astrophysical data, including the nature of the first stellar objects and the timing and rate of galaxy formation, reviews on the physics of reionization as well as theory about the nature of foregrounds maybe be found in Furlanetto et al. (2006); Morales & Wyithe (2010); Pritchard & Loeb (2012).

Telescopes seeking to measure this signal include the Giant Metre-wave Radio Telescope (GMRT; Paciga et al. (2013)), the Low Frequency Array (LOFAR; Yatawatta et al. (2013)) and the Murchison Widefield Array (MWA; Bowman et al. (2013) and Tingay et al. (2013)). PAPER is located in the Karoo desert at the Square Kilometer Array-South Africa cite. PAPER has doubled in size on a yearly basis since 2009, making observations with each array.

Here we report on deep integrations made with a 32 element array in 2011, first described in Parsons et al. (2013), hereafter P14. In that work we described in detail our reduction method which enabled a deep limit on the power spectrum in the presence of bright foregrounds. Using this method we combined 92 nights (1800 hours) of data to place upper limits on the power spectrum of HI in the early universe. However, this analysis was limited to a single redshift (z =7.7). The epoch of reionization signal is expected to evolve strongly with redshift, and it is this variation with cosmic history that will allow us to begin to place constraints on the physics of reionization Pritchard & Loeb (2008); Pober et al. (2014). Therefore, while a detection of the 21 cm signal at even a single frequency would be a tremendous breakthrough, analysis techniques must be developed to capitalize on the wide bandwidths of the current generations of high redshift 21cm telescopes. Using the same data as P14, this paper presents new power spectrum upper limits covering the redshift range $10.5 > z > 7.2$. In section 3 we summarize the observation and reduction methodology, we present the new upper limits in 4, and in 5 we offer conclusions and discussion of future work.

[1] `http:\\eor.berkeley.edu`

2. Observations and reduction

The work here follows the same basic procedure and uses the same underlying data set as P14. Here we provide a quick summary and refer the reader to Parsons et al. (2013) for a more in depth discussion. A general overview of the PAPER system can be found in Parsons et al. (2010), calibration of the primary beam in Pober et al. (2012), and imaging results in Jacobs et al. (2011, 2013a); Stefan et al. (2013). Sensitivity analysis described in Parsons et al. (2012) revealed that for the low gain elements employed by PAPER, a highly redundant "grid" type arrangement offers a significant sensitivity boost. In most interferometers each baseline samples a different mode of the power spectrum; the ideal case for reconstructing images. In a grid, there are many samples of each of these cosmological modes which can be averaged to a high sensitivity before being combined with other different k modes. The PAPER South Africa 32 antenna deployment (PSA32) was arranged in a 4x8 grid with a column spacing of 30m and a row spacing of 4m. In our analysis, as in P14, we include only the three shortest types of spacings where the reionization power is expected to be brightest. This selection includes those between adjacent columns within at least one row of each other, a selection containing 70 ∼30m-long baselines.

Observations spanning the band between 100 to 200-MHz ($13.1 > z > 6.1$) were recorded at a resolution of 48kHz and 10.7s beginning Dec 7, 2011 and ending March 19, 2012 for a total of 92 nights or 1800 hours. WIthin this set we included observations in the LST range 1h - 9h where the sky dominated system temperature is at a minimum. Note that this LST range is slightly shorter than in P14 which extended to LST of 12 hours. These last three hours were found to contribute minimally and be dominated by bright galactic foreground emission and so have been excluded here.

3. Reduction

Here we summarize our data reduction steps, for detailed descriptions see P14.

In several stages throughout the analysis process we take a 2D Fourier transform of the visibility spectra $V(\nu, t)$ into "delay/fringe rate" space where delay is the Fourier dual to frequency and similarly fringe-rate for time. In this space, smooth spectrum sources are physically localized to delays shorter than the light travel time length of the baseline (e.g. sources at the horizon, in the direction of the baseline vector) and fringe rates shorter than the sidereal rate. In this Fourier space, sources are highly localized with deviations from a flat spectrum manifesting as a slight dispersion. The spectrum sampling function, which is uneven due to flagging of interference takes the form of a convolution by a point-spread-function in the same way an imperfect sampling of the uv plane gives rise to the angular psf of an interferometer. To account for this, we use a CLEAN like, iterative, peak-finder and subtraction algorithm which is limited to finding peaks within the physically allowable ranges.

The data analysis pipeline essentially consists of iterative application of the delay/fringe rate transform process, with an ever tightening allowable number of modes, interleaved with stages of averaging (time, frequency, night), before finally computing a power spectrum. This final step takes advantage of the redundant baselines to make an unbiased power spectrum estimate by cross-multiplying identical baselines and then averaging the power spectrum modes. The details of each application of the delay/fringe-rate transform will be laid out the following sections as we provide a brief walk-through of the processing pipeline.

3.1. Initial Averaging

First, the raw data are down-selected to just the 70 30m long baselines described in Section 2. The visibilities are then compressed in the frequency and time directions by removing delay modes above the horizon and highest frequency fringe rates for a 300m baseline. This filtering is done in tandem with a radio frequency interference (RFI) flagging step, using the residuals which have had bright sky-like signals removed to flag 4σ deviations before feeding the flags back into another iteration of the compression step. This mitigates the effects of bright, narrow band, interference being scattered by the delay filter to generate a cleaner averaged data set and results in a time and frequency bin of length 39.6s and width 492.61kHz. This process reduces the data volume by a factor of $\sim$20, or roughly an order of magnitude improvement on traditional time and frequency averaging which in this array would be limited to 100kHz and 10s to avoid averaging away fringes.

3.2. Calibration

We model the gain as a per-antenna amplitude, a phase slope -physically a single time delay- and a single passband for all antennae. Because the array samples correlations redundantly, the relative calibration between antennae is numerically overdetermined and tractable as a linear algebra problem Liu et al. (2010). As described in P14, we compute the ratio between redundant baselines, fit for a gain and phase slope and then algebraically solve for a per-antenna solution. Here we have avoided calibrating each channel independently to preserve as much frequency variation as possible. These solutions vary little over the three month observing period, exhibiting less than 1% rms variation. A single solution derived for for the Dec 7 data set is used for the entire observing run. Time and frequency variation of redundant solutions is explored in detail in Zheng et al (in prep).

Relative calibration is all that is necessary to form a power spectrum, however to set the flux scale, we must form a beam on a bright known point source. By itself, the redundant solution is completely independent of any sky model. The redundant solution contains two free delay parameters and an overall amplitude scale. We fit the two delay parameters to a model of Pictor A, Fornax A, and the Crab Nebula during a time when the sky is dominated by these three sources

while marginalizing over the unknown apparent flux ratio between the three sources[2]. With the delays in place we are now able form a beam on Pictor A and (for each channel) set the overall amplitude to the calibration value of 381Jy found in Jacobs et al. (2013b).

3.3. Foreground Filtering and Night Averaging

Foregrounds are filtered from the calibrated data by removing all bright delay components with light travel times less then the baseline length. Where during the previous compression step a liberal horizon of 300m (1800ns, much longer than the 30 meter baselines under study) was used to calculate the window size, we now choose a window corresponding to the 30m baselines under study. The CLEAN model that is iteratively built is then subtracted from the data leaving residuals which we carry into the next stage. Next, a four hour long running mean is subtracted. This removes excess correlation due to cross-talk in the analog signal chain. The residuals are then flagged once more for RFI before the 92 nights of data are averaged into 36.4 second long local sidereal time (LST) bins. During averaging we found that some LST bins were dominated by very bright data points due to some exceedingly bright outlier samples. As these were very rare we found that removing the 10% brightest from every bin, a small median filter, was quite effective. The source of these outliers is not known, a likely possibility is instability in the analog signal chain stimulated by weather or bright interference, a circumstance that has since been observed in later seasons.

Though the frequency and repeated nightly observations have been averaged to their maximum extent, at 40s integrations the time axis has yet to be fully exploited. Sky-like fringes on a 30m baseline rotate much slower than 40 seconds. Performing a final fringe-rate filter, this time limiting to fringe-rates expected on a 30m baseline (down from 300m in the last iteration) we arrive at a data-set averaged to 900s, the maximum possible while still maintaining coherency. The root mean square of the residual signal (seen in Figure 2) at the end of this process is close to the 3mK level expected given the total integration time.

3.4. Power Spectrum

The output of the forgoing is a single calibrated and foreground filtered sidereal night. The power spectrum is estimated in the delay spectrum of a 10MHz bandwidth range centered on the redshift of interest. To preserve the isolation of any foregrounds which remain, we increase the spectral range by 5MHz on each side and multiply by a Blackman-Harris window thus providing a much higher dynamic range delay spectrum point spread function.

This leaves us with 40 delay samples on each of 70 baselines which are divided into three

[2]Unknown in the sense of a joint uncertainty in source flux and primary beam pattern.

redundant groups. Within these groups we compute the correlation between each mode. The cross correlation of the same delay modes between different redundant baselines provides our estimate of the power spectrum. Cross correlation between different delay modes provides our measure of systematic bias. Sky-like correlations should be identical between all redundant baselines to within the level of the noise. In practice some correlations deviate significantly from the median. These are removed by iteratively dividing the covariance into a model of systematics and a model of sky-like emission and then projecting out large residual modes. This is done by dividing the baselines into different groups such that all cross-multiplications are done without introducing noise bias. For more see Appendix of C of P14.

The residual correlations are all samples of the power spectrum. To estimate the final power spectrum and its uncertainty we compute the mean and variance of many random randomly-selected subsamples, sampling the dimensions: sidereal time, redundant baseline pair, and delay sign[3]

4. Results

4.1. Foreground Filtering and Noise Levels

The root-mean-square of the residual measures how well we've removed foregrounds. In Figure 2 we examine T_{rms} as calculated by differencing between adjacent channels and between redundant baselines. We see that in the last step of coherent integration, the step between 40 second to 789 second integrations, the noise level decreases by the expected factor of $\sqrt{789/40}$. At both stages the noise level found by both methods is roughly consistent with the theoretical level, only deviating significantly at the edges where the effective signal to noise due to the Blackman Harris window drops precipitously.

The theoretical noise level is a straightforward calculation, but given the method employed here, should be given due description. Not every LST bin is observed an equal number of times nor is the noise level constant, as the galaxy moves overhead. the number of points in each bin is mostly determined by the observing schedule, which for PAPER is simply sunrise to sunset, but also by any flagging of data or observing outage (only 92 of the 103 days in the observing period have recorded data). Unfortunately, technical problems limited the retention of the exact cumulative count of points averaged into each LST bin during the averaging process. Here we have chosen to use the first order estimate based on the observing schedule to estimate the number of observations in each LST. Some bins are observed only a few times, while a slim range in the LST range 6 - 9 hours is observed on every night. On average each bin is measured 64 times. Given this accounting of the number of samples $N(lst)$, we estimate the final rms temperature to be

[3]As visibilities are complex, both the positive and negative delays carry information.

$$T_{rms} = \frac{T_g(lst)}{\sqrt{2BtN(lst)}} + T_{rcvr}$$

where B and t are the 496kHz channel width and lst bin length, respectively. We model galactic temperature T_g by summing the de Oliviera-Costa global sky model (de Oliveira-Costa et al. 2008) convolved by the PAPER beam as a function of lst and T_{rcvr}=320K as the system temperature of 550K (found in P14) minus the mean galactic temperature from that time period. The final result is plotted as a dotted line in Figure 2 for before and after fringe rate filtering (t=40s and 788s)

The final fringe-rate step is the last coherent average of the processing routine and represents the last large gain in sensitivity; all following averaging steps are done incoherently on the square of the visibilities where sensitivity is gained at a much slower rate. For this reason the noise curves (bottom three curves of Figure 2) merit closer inspection. At this sensitivity level we see a notable excess of power in the rms difference between baselines. Though the cause of this excess is unknown, it is suggestive of the excess covariance between redundant baselines which we remove in Section 3.4 a hypothesis also supported by the high levels of residual foregrounds in the redshift 10.5 bin where the baseline difference rms is highest.

4.2. Power Spectra

In Figure 3 we show the power spectra at different redshift bins –showing both the spherically averaged power spectrum plotted in P_k units (left) and the final estimate of Δ^2 – and in Figure 4 we see different k mode slices as a function of redshift all plotted with 2σ error bars. The theoretical noise power spectrum is estimated using the method described in Pober et al. (2014) assuming a system temperature of 550K and the observing scheme described in Section 2. Generally, both figures convey a picture of a power spectrum dominated by noise, with foregrounds encroaching from modes closest to the foreground contaminated modes. Foreground dominance appears to grow with increasing redshift, mirroring the rise in baseline-to-baseline rms we found in Figure 2. Recall that the wide-band foreground model we have subtracted is built by weighting the entire band by a Blackman-Harris window function. Foreground removal is much weaker at this frequency, with the model subtracted being centered around 150MHz, contributions from this point in the band are down-weighted by 75%. This filter leads to the notch of missing power near $k_{\|} = 0$ (except at z=7.68 where the foregrounds have been left in for comparison). Only at redshift 10.3 is this notch nearly non-existent. The relative distribution of foreground removal accuracy is also visible in the shape of Δ^2 vs redshift at $k = 0.1$ which is only a factor of two away from the light travel horizon. This bin, which is firmly well into the foreground dominated portion of k space has much the same shape as the T_rms curve in Figure 2, with a minimum near redshift 8.5 and a dramatic rise above redshift 10.

5. Conclusions

With this paper we have extended the redshift coverage of Parsons et al. (2013) to cover the range $10.5 > z > 7.6$. The amount of foreground contamination slowly increases with redshift, paralleling the level of incoherent noise observed. This highlights a possible limitation of the scheme used to subtract the broadband foreground model which trades spectral coverage for isolation between delay modes. Accessing higher redshifts, where there is less spectral information on which to base a foreground model will require some combination of methods for removing foregrounds that make maximal use of the full band without as much down-weighting, as well as a more precise method for removing foregrounds in the narrow band-power spectrum.

With only 32 antenna, this data lacks the sensitivity to exclude all but the most extreme models, however, it does demonstrate the ability of the wideband filtration method to affectively remove foregrounds to the precision needed to integrate a full season to the thermal noise limit. As future observations with PAPER will add sensitivity primarily by increasing to 128 antennae this projects well for our ability to reach design sensitivity. The residual foregrounds analyzed here are all at or near the level of noise and expose the next level of removal required to access a wider redshift range.

REFERENCES

Bowman, J. et al. 2013, Publications of the Astronomical Society of Australia, 30, 31

de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., Jonas, J., Landecker, T. L., & Reich, P. 2008, Monthly Notices of the Royal Astronomical Society, 388, 247, (c) Journal compilation © 2008 RAS

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, Physics Reports, 433, 181, elsevier B.V.

Jacobs, D. et al. 2011, The Astrophysical Journal, 734, L34

Jacobs, D. C. et al. 2013a, eprint arXiv, 1307, 7716, accepted to ApJ, includes two data files

—. 2013b, The Astrophysical Journal, 776, 108

Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., & Hernquist, L. 2008, The Astrophysical Journal, 680, 962

Liu, A., Tegmark, M., Morrison, S., Lutomirski, A., & Zaldarriaga, M. 2010, Monthly Notices of the Royal Astronomical Society, 408, 1029, (c) Journal compilation © 2010 RAS

Morales, M. F. & Wyithe, J. S. B. 2010, Annual review of astronomy and astrophysics, 48, 127, oise

Paciga, G. et al. 2013, Monthly Notices of the Royal Astronomical Society, -1, 1427

Parsons, A. et al. 2010, The Astronomical Journal, 139, 1468

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

Parsons, A. R. et al. 2013, arXiv, submitted to ApJ, 1304.4991v2

Pober, J. et al. 2014, The Astrophysical Journal, 782, 66

Pober, J. C. et al. 2012, The Astronomical Journal, 143, 53

Pritchard, J. & Loeb, A. 2008, Physical Review D, 78, 103511

Pritchard, J. R. & Loeb, A. 2012, Reports on Progress in Physics, 75, 6901

Stefan, I. I. et al. 2013, Monthly Notices of the Royal Astronomical Society, 432, 1285

Tingay, S. et al. 2013, Publications of the Astronomical Society of Australia, 30, 7

Yatawatta, S. et al. 2013, Astronomy & Astrophysics, 550, 136

This preprint was prepared with the AAS LaTeX macros v5.2.

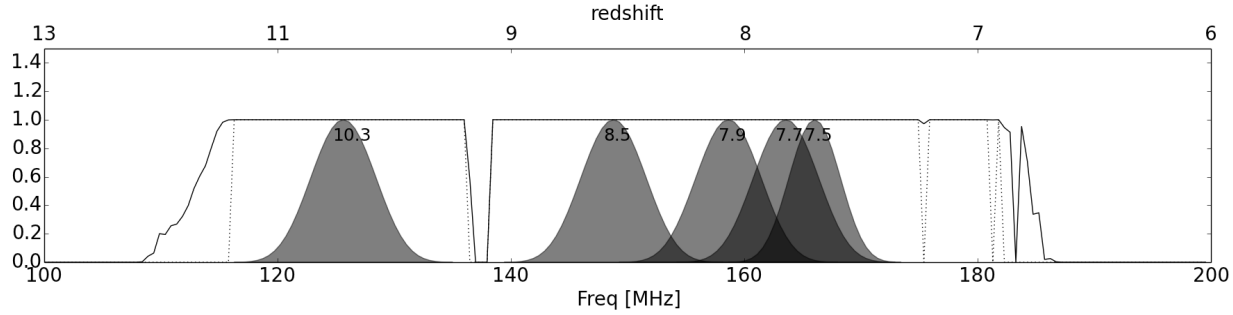

Fig. 1.— The average amount of data remaining after interference flagging over the 3 month period between Dec 2011 and March 2012 (black line) is quite high. Redshift bins (in grey, redshift center indicated with text label) are chosen to include spectral channels with uniform weight, i.e. no missing channels while maximizing coverage over the band. Redshift 7.7 is included for comparison with P14. Channels with no missing data are indicated by the dotted line. Each redshift bin is 20 MHz wide, but weighted by a Blackman-Harris window function which heavily down-weights the outer 10MHz for a Noise Equivalent Bandwidth of 10MHz. The interference is almost exclusively dominated by two features: ORBCOMM satellites at 137MHz and an unidentified intermittent line emitter at 175MHz. The roll off at 115MHz is due the rising noise at low frequencies being incorrectly flagged as interference.

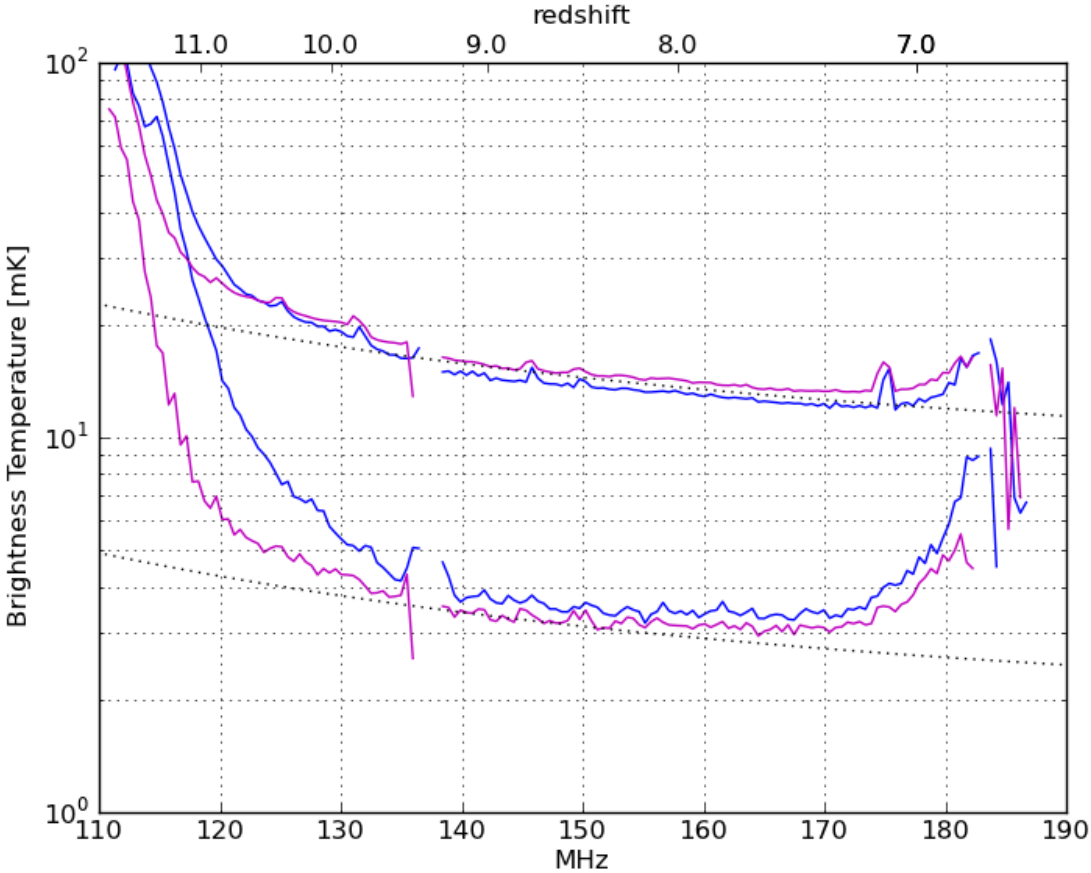

Fig. 2.— Root mean square (rms) noise before and after integrating to 10 minute LST bins (top and bottom sets of curves) indicating the rising significance of foregrounds as we take the last deep integration step. Noise is estimated by differencing adjacent frequencies (magenta) and between redundant baselines (blue), compared with an estimate of the theoretical noise level (dotted). The top three lines show noise after filtering foregrounds and binning into 40s long sidereal bins. At this noise level the frequency and baseline differences are roughly similar, both demonstrate the same small bumps of increased noise due to interference flagging and are consistent with noise over much of the band. The bottom three lines show the noise level after integrating up to the maximum fringe rate of 776s. The step between the top and bottom sets is the last coherent integration with noise largely decreasing by the expected factor $\sqrt{776/40}$ except in the difference between baselines (blue) which demonstrates a clear excess at all frequencies, particularly above redshift 10. With this last large jump in sensitivity we are now seeing the slight dominance of baseline covariance over the spectral derivative rms.

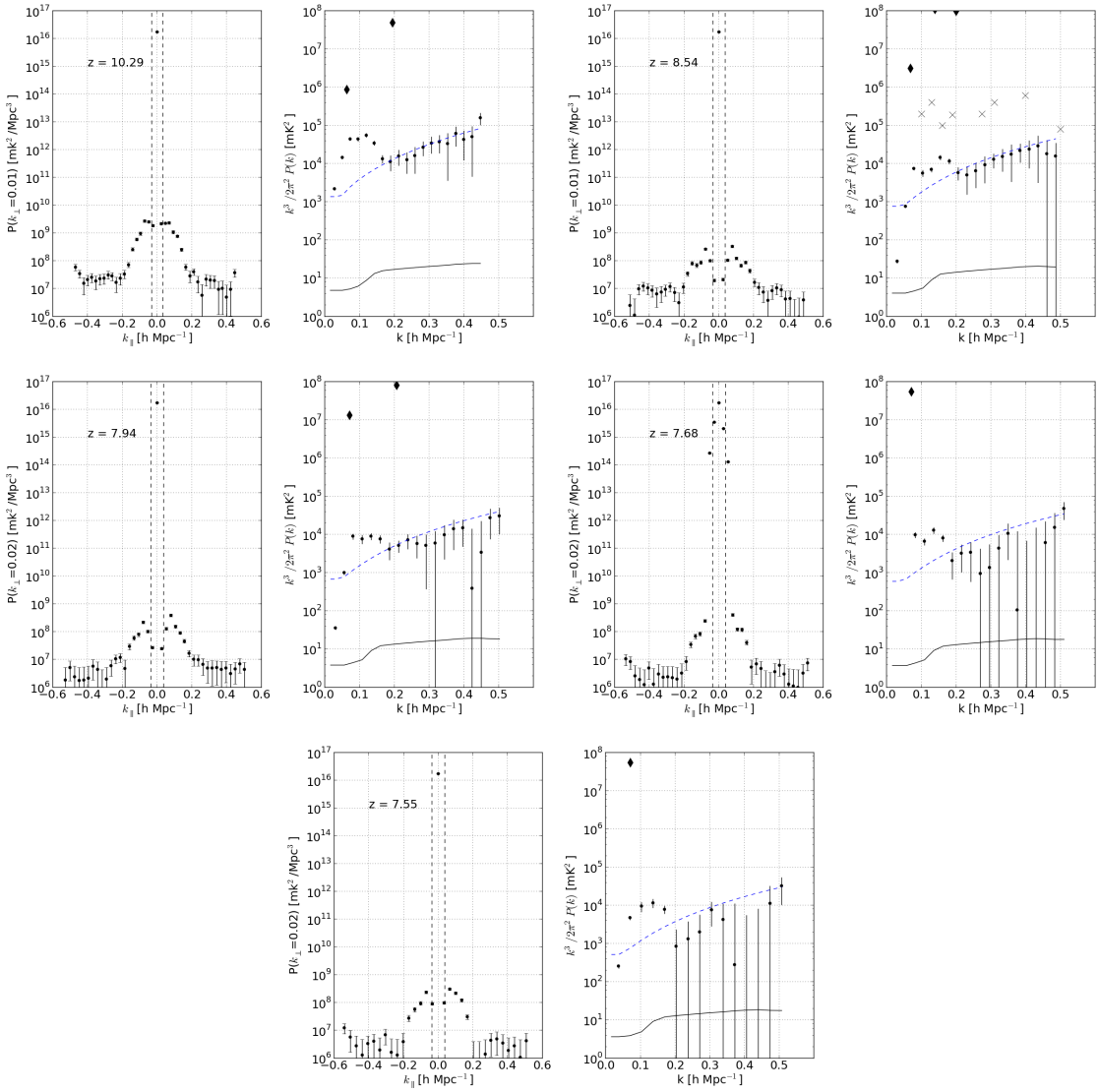

Fig. 3.— Left to right from top: power spectra centered on redshifts 10.3,8.5, 7.9, 7.7 and 7.6 (frequencies: 127, 149.5, 159.5 164.5, and 167 MHz.) All with an effective bandwidth of ±10MHz covering the redshift span Δz ≈1421MHz (B/f^2) which ranges from 0.8 at z =10.3 to 0.45 at z =7.4. The noise curve (blue dashes) is calculated using the method described in Pober et al. (2014) and indicates the 1σ confidence bounds on data points consisting purely of noise; 65% of uncorrelated noise like data points will lie below the curve. Note that due to the weighting of the delay transform, adjacent k bins are 50% correlated. The black line is a fiducial model at 50% ionization (Lidz et al. 2008). The redshift 7.68 bin has been reprocessed for comparison with the P14 result. Note that the LST range processed here is 3 hours less than P14 for a small reduction in sensitivity but large reduction in foreground.

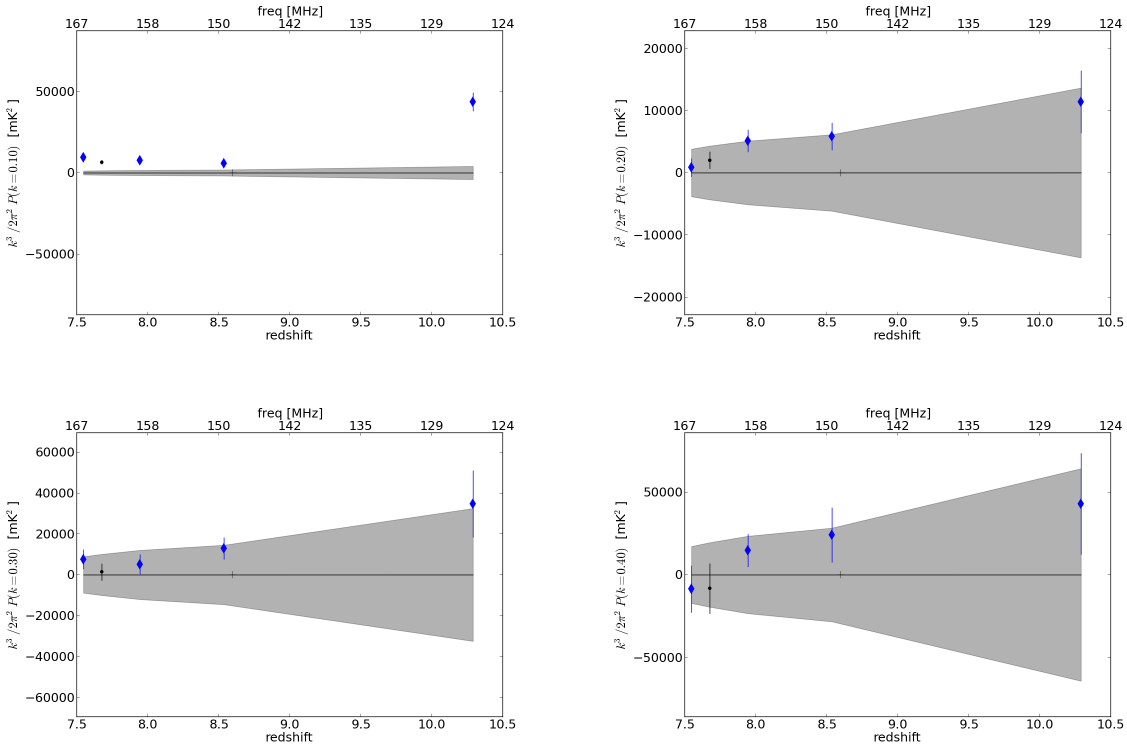

Fig. 4.— Power spectrum amplitude vs redshift at a selection of k modes. Left to right from top, k=0.1,0.2,0.3,0.4 hMpc^{-1}. Paciga 2013 GMRT marked with an 'x', Parsons 2014 PAPER limit marked with thin black, this work marked with thick blue diamonds. The k=0.1 hMpc^{-1} bin (top-left), which samples the delay spectrum at only 2x the maximum horizon delay is foreground dominated with a redshift dependence similar to the T_{rms} residual in Figure 2. The rest are consistent with thermal noise.