target Journal: MNRAS

The Murchison Widefield Array Epoch of Reionization Pipelines

Daniel C. Jacobs[1*], N. Barry[2], A. P. Beardsley[2], G. Bernardi[3,4,5], Judd D. Bowman[1], F. Briggs[6,7], R. J. Cappallo[8], P. Carroll[2], B. E. Corey[8], A. de Oliveira-Costa[9], Joshua S. Dillon[9], D. Emrich[10], B. M. Gaensler[15,7], A. Ewall-Wice[9], L. Feng[9], R. Goeke[9], L. J. Greenhill[5], B. J. Hazelton[2], J. N. Hewitt[9], N. Hurley-Walker[10], M. Johnston-Hollitt[11], D. L. Kaplan[12], J. C. Kasper[13,5], Han-Seek Kim[14,7], P. Kittiwisit[1], E. Kratzenberg[8], E. Lenc[15,7], J. Line[14,7], A. Loeb[5], C. J. Lonsdale[8], M. J. Lynch[10], B. McKinley[14,7], S. R. McWhirter[8], D. A. Mitchell[16,7], M. F. Morales[2], E. Morgan[9], A. R. Neben[9], Nithyanandan Thyagarajan[1], D. Oberoi[17], A. R. Offringa[6,7], S. M. Ord[10,7], Sourabh Paul[18], B. Pindor[14,7], J. C. Pober[2], T. Prabu[18], P. Procopio[14,7], J. Riding[14,7], A. E. E. Rogers[8], A. Roshi[19], N. Udaya Shankar[18], Shiv K. Sethi[18], K. S. Srivani[18], R. Subrahmanyan[18,7], I. S. Sullivan[2], M. Tegmark[9], S. J. Tingay[10,7], C. M. Trott[10,7], M. Waterson[10,6], R. B. Wayth[10,7], R. L. Webster[14,7], A. R. Whitney[8], A. Williams[10], C. L. Williams[9], C. Wu[20], J. S. B. Wyithe[14,7]

ABSTRACT

We present an overview of the Murchison Widefield Array 21 cm Epoch of Reionization analysis methods in which we compare the output of multiple pipelines as applied to a representative selection of data. The focus of this first round of analysis is on building the methodological foundation for detecting the weak statistical signature of cosmological HI at redshifts 6 to 10 in multi-year MWA observations. This paper provides a top level view over the methods and results of multiple, independent, data calibration and reduction pipelines. To assess the accuracy of our methods we split the data analysis steps into the two steps widely considered to present the most challenges, bright foreground removal and power spectrum estimation in the presence of residual foregrounds. Comparing images we see agreement on a significant amount of large scale, galactic, structure, though small differences related to calibration and weighting remain. Features common to all power spectrum results show the continued significance

[1]Arizona State University, School of Earth and Space Exploration, Tempe, AZ 85287, USA

[*]e-mail: daniel.c.jacobs@asu.edu

[2]University of Washington, Department of Physics, Seattle, WA 98195, USA

[3]Square Kilometre Array South Africa (SKA SA), Park Road, Pinelands 7405, South Africa

[4]Department of Physics and Electronics, Rhodes University, Grahamstown 6140, South Africa

[5]Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA

[6]Australian National University, Research School of Astronomy and Astrophysics, Canberra, ACT 2611, Australia

[7]ARC Centre of Excellence for All-sky Astrophysics (CAASTRO)

[8]MIT Haystack Observatory, Westford, MA 01886, USA

[9]MIT Kavli Institute for Astrophysics and Space Research, Cambridge, MA 02139, USA

[10]International Centre for Radio Astronomy Research, Curtin University, Perth, WA 6845, Australia

[11]Victoria University of Wellington, School of Chemical & Physical Sciences, Wellington 6140, New Zealand

[12]University of Wisconsin–Milwaukee, Department of Physics, Milwaukee, WI 53201, USA

[13]University of Michigan, Department of Atmospheric, Oceanic and Space Sciences, Ann Arbor, MI 48109, USA

[14]The University of Melbourne, School of Physics, Parkville, VIC 3010, Australia

[15]The University of Sydney, Sydney Institute for Astronomy, School of Physics, NSW 2006, Australia

[16]CSIRO Astronomy and Space Science (CASS), PO Box 76, Epping, NSW 1710, Australia

[17]National Centre for Radio Astrophysics, Tata Institute for Fundamental Research, Pune 411007, India

[18]Raman Research Institute, Bangalore 560080, India

[19]National Radio Astronomy Observatory, Charlottesville and Greenbank, USA

[20]International Centre for Radio Astronomy Research, University of Western Australia, Crawley, WA 6009, Australia

of wide-field effects, while differences in both imaging and power spectrum convey the importance of calibration algorithms and point towards future work. Using the calibrated and integrated image cubes we apply an inverse covariance technique to make a noise limited estimate of the power spectrum.

1. Introduction

Study of intergalactic Hydrogen in the early universe via the 21 cm line is forecast to provide a wealth of astrophysical and cosmological information. The 21 cm line is both optically thin and spectrally narrow, making possible full tomographic reconstruction. Cosmological Hydrogen is neutral over cosmic time from recombination until reionized by the first batch of UV emitters (stars and accretion disks). Reviews of 21 cm cosmology, astrophysics and observing can be found in Morales & Wyithe (2010); Furlanetto et al. (2006); Pritchard & Loeb (2012); Zaroubi (2013).

Direct detection of HI during the Epoch of Reionization (cosmological redshifts $5 < z < 13$) is currently the goal of several new radio arrays. The LOw Frequency ARray (Yatawatta et al. 2013, LOFAR;), the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. (2014)) and the Murchison Widefield Array (MWA; Tingay et al. (2013); Bowman et al. (2013)) are all currently conducting long observing campaigns.

The analysis of the resulting data presents several challenges. The signal is faint; initial detection is being sought in the power spectrum with thousands of hours (multiple seasons) of integration required. This faint spectral line signal sits atop a continuum foreground four orders of magnitude brighter. At the same time, the instruments are fully correlated phased arrays with wide fields of view that strain the conventional mathematical assumptions of radio astronomy practice. The methods used to arrive at a well calibrated, foreground-free, estimation of the power spectrum are all under development in the sense of the algorithms as well as the implementation.

The path from observation to power spectrum can be roughly divided into two parts: removal of foregrounds and estimation of power spectrum.[1] Recently two sorts of foreground removal have been suggested. Blind filtering, such as the delay/fringe-rate filtering approach (Parsons et al. 2012; Liu et al. 2014a,b), that has been applied to data from the PAPER telescope (Parsons et al. 2014), applies a small amount of knowledge about the instrument to filter modes likely to be contaminated. This method is comparatively robust in the face of uncertainty about the instrument and the sky, at the cost of losing some sensitivity. Meanwhile, full forward modeling and subtraction of sky model such as that implemented for LOFAR (see e.g. Jelić et al. (2008); Yatawatta et al. (2013)), requires a much higher fidelity model of the instrument and the sky (Datta et al. 2010; Vedantham et al. 2012a).

[1]While statistical measures such as Barkana & Loeb (2008) have also been proposed we choose the power spectrum for our initial analysis because the interferometer naturally measures in the Fourier plane.

The MWA analysis approach focuses on direct subtraction of known foregrounds. If successful, it has the benefit of opening the most sensitive power spectrum modes within the "wedge" and substantially improving the ability of early measurements to distinguish between reionization models (Beardsley et al. 2013; Pober et al. 2014). Recent work towards the goal of foreground subtraction includes better algorithmic handling of wide field imaging effects (Tasse et al. 2012; Bhatnagar et al. 2013; Sullivan et al. 2012; Ord et al. 2010b), and continually improving catalogs of sky emission (de Oliveira-Costa et al. 2008; Jacobs et al. 2011; Hurley-Walker et al. 2014; Morgan et al. 2014). Ongoing operation of the next generation low frequency arrays –LOFAR, PAPER and MWA are all in their second or third year of operation– continues to push the refinement instrumental models (e.g. the work of Neben (2015) in mapping the primary beam with satellites) and improve the accuracy of model subtraction. At the same time, more complete surveys of 21 cm reionization foregrounds are currently under way. These include the MWA GLEAM[2] survey and the LOFAR MSSS[3].

Given the challenges of using newly developed methods to reduce data from a novel instrument to make a low sensitivity detection, it is reasonable to consider the question of how one knows one is getting the "right" answer. One option is to generate, as accurately as possible, a detailed simulation of the interferometer output. Full instrument simulation is computationally demanding and difficult to divorce from the analysis methods being tested, which at their core are instrument simulation engines. Development of a completely independent instrumental simulation and is the subject of ongoing work (see e.g.. Thyagarajan et al. 2015), but is not available at scale at this time. The second option is more pragmatic; compare the results of multiple independent pipelines.

Within the MWA collaboration efforts have centered around two, completely independent, paths from raw data to a power spectrum. In this paper we outline each method, leaving the detailed descriptions to other papers. The first pipeline uses Fast Holographic Deconvolution (FHD[4]) for calibration and foreground subtraction, followed by either *ϵppsilon* [5] or *E*MPCOV to estimate the power spectrum. The second pipeline uses an offline version of the MWA Real Time System (RTS) followed by CHIPS[6] to estimate the power spectrum. FHD is described in detail by Sullivan et al. (2012) and the RTS by Ord et al. (2010b). *ϵppsilon* , CHIPS and *E*MPCOV , as applied to the data published here, are described in Hazelton et al 2015, Trott et al 2015, and Dillon et al 2015, respectively.

One benefit from having multiple pipelines is the freedom to focus on different optimization axes. The design of the *ϵppsilon* power spectrum estimator emphasizes speed and relative simplicity,

---

[2]GLEAM: Galactic and Extragalactic All-sky MWA

[3]MSSS: Multi-frequency Snapshot Sky Survey

[4]`github.com/miguelfmorales/FHD`

[5]*ϵppsilon* :Error Propagated Power Spectrum with InterLeaved Observed Noise;

[6]Cosmological HI Power Spectrum

choices motivated by the need to understand the effect of processing decisions such as observation protocol, flagging, and calibration on the **power spectrum**. Using *ϵppsilon* we have discovered and corrected multiple systematic effects visible only in the power spectrum. With the ability to quickly form power spectra on different sets of data *ϵppsilon* has been our primary method for selecting sets of high quality data. Whereas both *ϵppsilon* and CHIPS operate on time-ordered data, *E*MPCOV takes as input single cubes of integrated data (one each for even and odd time samples) which have been selected on the basis statistics like interference and calibration quality and curation with *ϵppsilon* . The *E*MPCOV method uses an empirical estimate of the instrumental covariance to mitigate any remaining residual in the Fourier modes due to instrument mismodeling in the previous steps.

The MWA EoR program has collected more than 1000 hours of data but here we will limit ourselves to a single night (3 hours) an amount which is sufficient to gain insight into the net performance of our calibration algorithms and foreground subtraction models without adding the complexity of a large dataset. Upcoming analyses will focus on going deeper, using analysis techniques built upon the foundation methods described here and in the companion papers.

In **??** we provide a top-level view of our pipeline components, in 2 we summarize our fiducial data set, 3 hours from the 2013 MWA Epoch of Reionization observing program and provide a brief summary of each pipeline component in section 3. In section 4 we examine a comparison of the foreground images and residual power spectra and in section 5 we conclude with a brief presentation of a noise limited power spectrum and a discussion of the similarities and differences and how they impact interpretation of pipeline results.

2. Observing

The MWA is an an interferometric array of phased array tiles operating in the 80-300 MHz radio band. Each tile consists of a 4x4 grid of bow-tie shaped dipoles that are used to form a beam on the sky, steered by an analog delay-line beamformer. The signal is digitized over the entire bandwidth but only 30 MHz are available at any one time. This 30 MHz of bandwidth is broken into 1.28 MHz "coarse" bands by a polyphase filter-bank in the field and sent to the correlator where it is further channelized to 40kHz, cross-correlated and then averaged at 0.5 second intervals. The spectral shape of the coarse polyphase filter is known somewhat imperfectly and is thought to include a small amount of aliasing from adjacent coarse channels, though the exact amount is currently under investigation. This spectral response is corrected to first order during the first post-correlator step, and to second order by the calibration step. More details on the design and operation of the MWA can be found in Lonsdale et al. (2009) and Tingay et al. (2013).

The MWA EoR observing scheme focuses on two 30 MHz tunings, 140-170 MHz and 167-196 MHz (so-called 'low' and 'high') and two minimal foreground regions both -27°declination (near zenith at the MWA's latitude) at RA 0h and 4h (referred to as EoR0 and EoR1). Here we focus on

the 'high' tuning pointing at EoR0, where the high band is chosen for its low sky temperature and EoR0 is chosen for its ease of calibration (lacking bright, resolved sources). During observing, the beam former was set such that the target region repeatedly drifted over the beam. Each drift was about 30 minutes long. This was done for a total of 6 pointings in a night, or about 3 hours. The data included here include the two pointings leading up to the target crossing zenith, the zenith pointing, and then three more pointings after the transit crossing. Data were recorded in 112 second units for a total of 98 snapshots. These snapshots are the basic unit of time on which many operations become independent -eg RFI flagging, FHD calibration and imaging.[7] Each snapshot is flagged for interference using the AOFlagger[8] algorithm and then averaged to 2 seconds and 80kHz. As described in Offringa et al. (2015), the interference environment at the Murchison Observatory is benign and generally requires flagging of about 1% of the data. Though the full set of linear stokes parameters are correlated, and Stokes I images and power spectra are the final product of interest, at this stage of the analysis the instrumental polarizations have been found to be more instructive; only the linear X (east-west) polarization is examined here.

3. Power Spectrum Pipelines

There are many possible paths to a detection of 21 cm reionization but all must in some way remove foregrounds and compute an estimate of the power spectrum[9] . In subtracting a foreground model one must account for: ionospheric distortion, very wide field, primary beam uncertainty, polarization leakage, and calibration accuracy. The MWA pipeline has two independent calibration

[7]Note that this is not true in the RTS case which uses a time interval scaled by the baseline length.

[8]`sourceforge.net/projects/aoflagger`

[9]or some similar statistical measure

Table 1. MWA EoR Observing Parameters

parameter	value	notes
field of view	26°	FWHM, scales as λ
tuning	166-196 MHz	redshift range $7.56 < z < 6.25$
target area	(RA,Dec) 0h00m, -27°00m	
pointing seperation	6.8°	
time and frequency resolution	0.5 s, 40 kHz	
post-flagging resolution	2s, 80 kHz	
time	3 hours on August 23, 2013	a total of 6 pointings, lasting 30m each

and imaging modules which subtract the foregrounds and two power spectrum estimators. All are developed independently, sharing very little code, yet are interconnectable via common data formats to give four possible pipeline paths.These two paths and their interactions are sketched out in Figure 1.

The imaging and foreground subtraction portion of the pipeline can be handled by either of two custom packages. The MWA Real Time System (RTS; Ord et al. (2010b)) was initially designed to make images in real time from the MWA 512. On the de-scoped 128 element array, it has been implemented as an offline system, where it has been adjusted to compensate for the lower filling factor. Fast Holographic Deconvolution (FHD; Sullivan et al. (2012)) is a custom interferometric imaging package developed for wide-field instruments with a focus on accounting for the very wide field of view antenna responses found on phased arrays of dipoles. Both systems were developed in parallel with the construction and commissioning of the MWA to provide a detailed introspection on every aspect of this experimental telescope. Each can calibrate a data set against a model, subtract a model, deconvolve images and use precision models of the instrument informed by the commissioning process including effects such as tile to tile primary beam variation and 0.1dB cable reflections. Foreground inputs include catalogs, images of extended emission and wavelet models of bright, mostly compact, sources. In addition, each has its own unique feature set developed as part of the experimental process.

3.1. Calibration and Imager #1: RTS

The MWA Real Time System (RTS) is a radio interferometry software package specifically written to calibrate and image MWA data (Ord et al. 2010a, Mitchell et al. in prep). The RTS incorporates algorithms intended to address a number of known challenges inherent to processing MWA data, including; wide-field imaging effects, direction-dependant (DD) antenna gains and polarization response, and ionospheric refraction of low-frequency radio waves. Each MWA observation (112s) is processed through a separate instance of the RTS. The RTS is also parallelized over frequency so that each coarse channel (1.28 MHz broken into 40 kHz channels) is processed largely independently of the other coarse channels, with only information about the measured ionospheric offsets communicated between processing nodes. Calibration and model subtraction were based on the TBD catalog.

The RTS calibration strategy is based upon the 'Peeling' technique proposed by Noordam (2004). The brightest (apparent) radio sources in the field of view are sequentially and iteratively processed through a Calibrator Measurement Loop (CML). During each pass through the CML; i) the expected (model) visibilities of known catalogue sources are subtracted from the observed visibilities. For the data processed in this work, $\sim$100 sources are subtracted for each observation. ii) The model visibilites for the targeted source are added back in and phased to the catalog source location. Any ionospheric offset of the source can now be measured by fitting a phase ramp to the phased visibilities. iii) The strongest sources are now used to update the direction-dependant

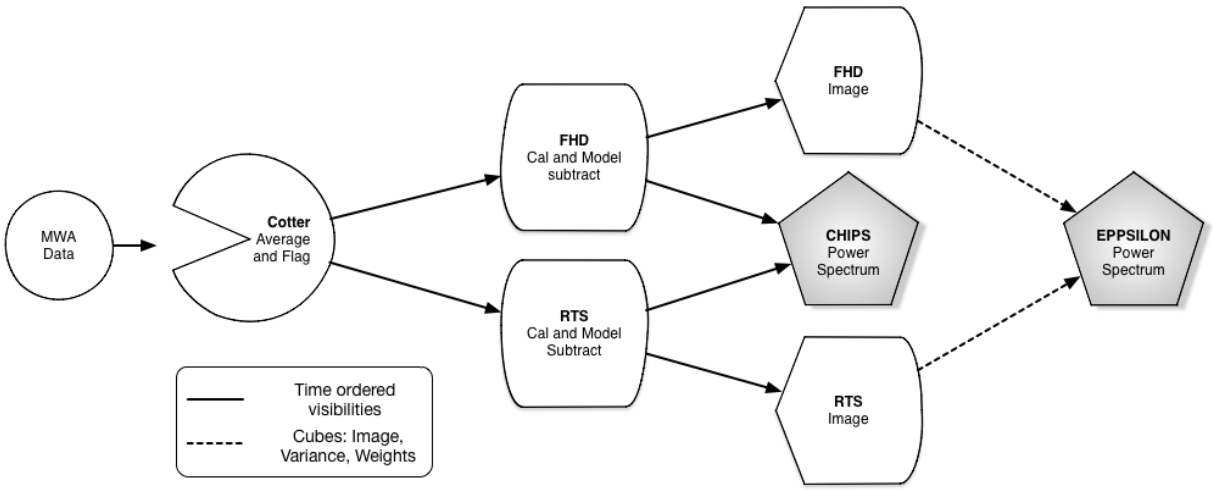

Fig. 1.— Parallel pipelines with cross-connections after foreground subtraction and imaging are compared against each other as a guard against error. Cotter uses AOFlagger to flag RFI and averages by a factor of 8. The averaged snapshots are passed to either FHD or RTS for calibration and imaging, *ϵppsilon* takes the resulting snapshots, averages them in LST and estimates the power spectrum. CHIPS taps into the RTS and FHD to get calibrated and foreground subtracted time-ordered (not yet gridded) visibilities which it then uses to make its own estimate of the power spectrum.

antenna gain terms, while weaker sources are only corrected for ionospheric offsets. For this work, TBD sources are used as full DD calibrators and 100 sources are set as ionospheric calibrators. The CML is repeated until the gain and ionospheric fits converge to stable values. The ∼100 strongest sources are then subtracted from the calibrated visibilities and the residuals are passed to the visibility-based power spectrum described in Section 3.4 and shown in Figure 3. A single bandpass for each tile is found by fitting a 2nd order polynomial to each coarse channel. Calibration and model subtraction parameters are summarized in Table 2. Model subtracted visibilities are passed to the RTS imager and to the CHIPS power spectrum estimator.

The RTS uses a snapshot imaging approach to correct for wide-field and direction-dependant polarization effects. Following calibration, the residual visibilities are first gridded to form instrumental polarization images which are co-planar with the array. These images are then regridded into the HEALPIX frame with wide-field corrections and conversion to Stokes applied through the regridding weights. It is also possible to use the fitted ionospheric calibrator offsets to apply a correction for ionospheric effects across the field during the regridding step, but in this work this correction has not been applied. See Clark.Allen.Arcus (bartTBD) for a more complete description of the RTS imaging algorithms. The resulting image spectral cubes, output on a 112s cadence paired with a spectral image cube of the point-spread function (Fourier dual to the weights in the uvf plane), and the variance (dual to the uvf weights squared). The mean of the image cube is shown in Figure 2, power spectra of with RTS foreground subtraction are shown in the bottom row of Figure 3.

3.2. Imager #2: FHD

Fast Holographic Deconvolution (FHD, Sullivan et al. (2012)) is a calibration and imaging algorithm designed for very wide field of view interferometers with direction- and antenna-dependent beam patterns. FHD has particularly been designed with a focus on producing an accurate measurement of the power spectrum and takes care to export the instrument model to the power spectrum estimation stage for the purposes of error propagation. Like the RTS, FHD uses the beam pattern for gridding visibilities to the u-v plane, and its Hermitian conjugate for de-gridding simulations to form model visibilities. The beam model is composed of the measured antenna response to the electric field for each antenna element and at every fine frequency channel, convolved with the response of the second antenna that forms the visibility. Three data outputs are necessary from gridding in order to calculate the image based power spectrum with accurate error bars: the measured visibilities, gridded with the beam model[10]; the weights, obtained by gridding the beam model; and the variance, obtained by gridding the squared beam model.

The FHD calibration pipeline both measures and removes foregrounds. The calibration model

[10]Note that the resulting image will be tapered by the average primary beam squared

is formed from sources found by deconvolving hundreds of snapshots on the EoR0 field and retaining those common to most of them. This catalog in the MWACS catalog having a primary beam response of 5% or more. Solutions are then computed using the Alternating Direction Implicit technique described in Salvini & Wijnholds (2014). This generates a gain and phase for every channel on every tile, for each 112s snapshot. These solutions are then averaged over all tiles to form a single passband, which corrects for the majority of the spectral dependent effects such as the response of the coarse channel passband. This single passband solution is then divided out of each per tile solution and each residual fit for a 2nd order amplitude polynomial and a first order phase polynomial. This process happens iteratively, with convergence measured by comparing the relative difference between residual visibilities. 10 iterations to converge to a stable residual was the norm. The residual time-ordered visibilites are then passed to CHIPS and to FHD imaging for formation of spectral cubes. FHD splits the samples into even and odd time intervals at a 4s cadence to produce two sets of image and weight cubes which are sent to *ϵppsilon* for power spectrum estimation. Power spectra with FHD foreground subtraction are shown in the top row of Figure 3.

3.3. Power Spectrum #1: *ϵppsilon*

ϵppsilon calculates a power spectrum estimate from image cubes and directly propagates the error bars through the full analysis and is described more fully in Hazelton et al 2015. The design criteria for this method is to make a relatively quick and uncomplicated estimate of the power spectrum to provide a quick turnaround diagnostic. The input to *ϵppsilon* is gridded image cubes for each 112s snapshot, such as are produced by FHD or RTS imaging, in which the data has been split into interleaved time samples (referred to as even and odd cubes) along with matched cubes containing the modeled instrumental weighting and variance. These snapshot healpix cubes are integrated in time keeping pixels with a beam weight of 1% or more, a cut which effectively limits the field of view to $\sim 20^\circ$. The accumulated data, weight and variance cubes are Fourier transformed in two dimensions to take them to uvf space where the spatial covariance matrix is assumed to be diagonal. This is a better assumption if the uv pixel size is well matched to the primary beam size so we restrict the spacing of modes in the spatial DFT to be equal to the inverse of the primary beam field of view. The data (variance) cubes are then divided by the weight cubes (weight cubes squared) to arrive at the best estimates of the sky and variances. Next the sum and difference of the even and odd cubes are computed with variances given by adding the reciprocal of the even and odd variances in quadrature. The difference cube then contains only noise (as long as the time interleaving is fine enough) and the sum cube contains both sky signal and noise.

The next step is to Fourier transform in the frequency direction. Here we choose to use the full 30 MHz spectral window, weighted by a Blackman-Harris window function, which heavily down-weights the outer half of the band to effectively sample 15 MHz; a cosmological redshift range of 0.86. This weighting scheme minimizes the covariance of bright foreground modes between power

Table 2. MWA EoR Calibration and Model subtraction Parameters

parameter	value	free parameters
	RTS	
passband	2nd order poly per coarse channel	48 per tile
gain	amplitude and phase	2 per tile
Direction Dependent	2x2 Jones matrix	4 per DD source
Catalog	TBD	
sources subtracted	TBD	TBD flux cut
	FHD	
passband	fine channel gain spectrum	768 channels for entire array
passband	2nd order poly over full band (1st order for phase)	3 per tile
gain	amplitude and phase	2 per tile
Catalog	MWA Commissioning Survey	
sources subtracted	1000	TBD flux cut

Table 3. Foreground subtraction image differences

Number of sources subtracted FHD/RTS	FHD RMS*	RTS RMS*	diff RMS*
None/None	1.97	1.49	1.08
300/300	1.19	0.59	0.991
1000/300	0.923	0.59	0.817

*The standard deviation of the image in Jy

spectrum modes as described in Thyagarajan et al. (2013); Parsons et al. (2012); Vedantham et al. (2012b), among others. The spectral Fourier transform is dominated by the MWA passband gaps which occur every 1.28 MHz and imparts a harmonic structure to the $k_{\parallel}$ dimension. The Fourier transform of unevenly sampled data is well described by the Lomb & Scargle method which obtains a periodogram in the trigonometric cross products of (sin,cos), resulting in a two-by-two covariance matrix for each Fourier mode containing the cos^2 and sin^2 terms on the diagonal and the $cos \times sin$ cross term in the off-diagonal elements. Lomb (1976); Scargle (1982). Diagonalizing this matrix at each $k_{\parallel}$ mode finds the best estimate of the power spectrum in the presence of missing data. The sky signal power is estimated by the square of the sum cube minus the square of the difference cube, which is mathematically identical to the even/odd cross power if the even and odd variances are identical, while the square of the difference cube provides a realization of the noise power spectrum. Finally the power cubes are averaged averaged, weighting by variance, in $k_x - k_y$ rings to get to a two dimensional $k_{\parallel} - k_{\perp}$ space.

3.4. Power Spectrum #2: CHIPS

The CHIPS power spectrum estimation method computes the maximum likelihood estimate of the 21 cm power spectrum using an optimal estimator formalism and is more completely described in Trott et al 2015. The design criteria for this method were to fully account for instrumental and foreground induced covariance in the estimation of the power spectrum The approach is similar to that used by Liu & Tegmark (2011), but with the key difference of being performed entirely in uv-space, where the data covariance matrix is simpler (block diagonal), and feasible to invert. This approach also allows straightforward estimation of the variances and covariances between sky modes by direct propagation of errors, and requires fewer preparatory analysis steps. CHIPS takes as input calibrated and foreground subtracted time-ordered visibilities. Tapping into the pipeline post-calibration but before imaging CHIPS uses its own internal instrument model to estimate and propagate uncertainty.

The method involves four major steps: (1) Grid and weight time-ordered visibility channel onto a uvw-cube using the primary beam model, (2) compute the least squares spectral (LSS) transform along the frequency dimension to obtain the best estimate of the line-of-sight spatial sky modes (this technique is comparable to that used $\epsilon ppsilon$); (3) compute the maximum-likelihood estimate of the power spectrum, incorporating foregrounds and radiometric noise, averaging k_x and k_y modes into annular modes on the sky, $k_{\perp}$; (4) compute the uncertainties and covariances between power estimates. The first step is the most computationally-intensive, requiring processing of all the measured data. The principle departure point for CHIPS from $\epsilon ppsilon$ is in the much finer resolution of the uv grid. Using an instrument model, CHIPS calculates the covariance between uv samples as a function of frequency. Since the beam and uv sampling function are both highly chromatic, extra precision in this inversion is thought to be highly beneficial. After a line of sight transform similar to that used by $\epsilon ppsilon$, this covariance information is inverted to find the fisher

information, the maximum likelihood power spectrum, and covariances between measurements. The resulting dataset provides an estimate of the power in each $k_\perp, k_\parallel$ mode, shown in the right column of Figure 3.

3.5. Power Spectrum #3: Implicit Covariance

The quadratic estimator method of Liu & Tegmark (2011) treats foreground residuals in maps as a form of correlated noise and simultaneously downweights both noisy and foreground-dominated modes, keeping track of the extra variance they introduce into power spectrum estimates. This technique, accelerated by Dillon et al. (2013), was used in the previous MWA 32T results Dillon et al. (2014). A very similar technique, working on visibilities rather than maps, was used for the recent PAPER 64 results Ali et al. (2015). Dillon et al 2015build on these methods to mitigate errors introduced by imperfect mapmaking and instrument modeling through empirical covariance estimation, assuming all data covariance is sourced by foregrounds after shallow integrations

This final step takes as input FHD calibrated images with foregrounds subtracted as well as possible, split into even and odd time-slices and averaged over many observations using an observation set list curated by examining *ϵppsilon* power spectra. From these cubes, it estimates the frequency-frequency foreground residual covariance in annuli in uvf space, assuming that different uv cells have uncorrelated foreground residuals. This assumption, similar to that made by CHIPS, allows the combined foreground and noise covariance to be inverted directly. The resulting power spectrum, made with the Dillon et al. (2013) fast algorithm and the frequency flagging psuedo-inverse and 2D to 1D binning techniques of Dillon et al. (2014), was used to create the spherically averaged power spectra in Figure 4 which is described in detail in Dillon et al 2015. This 1D power spectrum includes the same data shown in Figures 2 and 3, split into three equal-bandwidth redshift bins. Though well above the predicted cosmological signal level, the measurements are notably consistent with noise across a wide range of k.

4. Results

A heuristic comparison of the images and power spectra reveals several consistent features. A comparison of images at several stages of foreground subtraction is presented in Figure 2. The columns compare between the two pipelines FHD on the left and RTS in the center; the right column shows the difference between the two while the rows are different levels of foreground subtraction. RMS of each image, a measure of the total power, is given in Table 3. The top row is a comparison before any model subtraction. Presented in the raw weighting, without application of any deconvolution, small differences in weighting schemes are perceptible as slightly different point-spread-functions around isolated sources, and broader response around clusters of sources.

Foreground subtraction is highly dependent on the choice of sources included in the model: too

few sources subtracted leaves an excess of power which must be removed as extra free parameters in the covariance step, too many sources leaves open the possibility of mis-subtraction as the number of sources and amount of sky covered increases. In the middle row we have locked both pipelines to the RTS's smaller but more readily diagnostic catalog of 300 sources. This subtraction step decreases the image rms by half and results in a somewhat flatter difference image. In the bottom row we have allowed each system to select from its internal catalog based on nominal operating parameters, for the FHD this means the number of sources increases to 1000 and results in a further decrease in image rms by $\tilde{3}3\%$ and a smaller 20% decrease in the *d*ifference with the RTS image. This is the residual data set which is passed to the power spectrum estimation portions of the pipelines.

The primary feature in the two images shown in Figure 2 is that both consistently present a significant amount of large scale power which, while visually similar in some respects, differs at the 50 to 100% level. The power is generically described as several large "islands" of power (both positive and negative) and which, in the difference image, give the appearance of beginning to dissipate as more sources are subtracted.

The residual, after subtraction of many hundreds of sources, tells a fairly consistent story of a significant amount of large scale power remaining, though the two imagers disagree somewhat on the exact arrangement. One possibility follows from a consideration of the physical layout of the MWA and the approach to calibration used in both imagers. The MWA configuration is a compact core –with tiles separated by < 5 wavelengths in some cases– and a density that falls off radially; a configuration designed to match the expected steep rise of reionization power at large scales. Thus the image weights are highly condensed in the center of the uv plane and make images dominated by large scale structure Beardsley et al. (2013); reconstruction of the large scales is crucial to subtracting modeled foregrounds at small k, where reionization is brightest. The primary question is whether any large-scale structure is "real", rather than some mis-calibration or other artifact that affects the large number of short baselines. Calibration errors in a traditional interferometer having uniformly distributed baselines reveal themselves as side lobes around bright point sources; similar errors on the core-heavy MWA reveal themselves as artifactual large-scale power which is more difficult to distinguish from true emission. In essence, by using only the longer baselines, and a model composed only of point sources, we have performed a calibration on a small amount of data on the periphery of the core and extrapolated the result to the majority of the data at the shortest spacings. It should be no surprise that small differences in calibration and weighting heavily affect the accuracy of the images. These differences are also apparent in the power spectrum.

4.1. Power Spectra

Application of our two independent power spectrum estimators to our two calibration and foreground subtraction pipes gives us a total of four different power spectra (Figure 3). Each power spectrum estimator has been developed to target the output from a "primary" calibration

and foreground subtraction process –the diagonal elements of Figure 3– and have been highly optimized to that up-stream source of data. The off-diagonal power spectra were created using auxiliary links which import the data and the metadata produced by the foreground subtraction step. Since they are less highly optimized, lacking as they do the advantage of a close working relationship, these pathways represent an upper limit on the variance to be expected from small analysis differences and allow us to look for effects common to foreground subtraction.

Properties shared by all are the large amount of power at low $k_{\|}$ roughly at an amplitude of 10^15 mK^2/Mpc^3 and approximately flat in $k_{\perp}$ and the so-called "wedge" shaped linear dependance on baseline length. The wedge is due to the inherently chromatic nature of a wide field instrument scattering smooth spectrum foregrounds; sources entering far from the phase center appear as pixels at higher $k_{\|}$. The solid and dotted lines in the figure indicate the upper boundaries of power from sources at the horizon and at the beam half power point, respectively. With the exception of some instrumental features foreground power is well isolated within this expected boundary.

The two main instrumental systematics are horizontal striping due to missing or poorly calibrated data at the edges of regular coarse passbands and vertical striping due to spectral variation in uvf sampling. The former can be minimized by careful calibration of the passband, the latter by uv rotation synthesis and by accounting for covariance between uvf samples.

The most noticeable difference between foreground removal methods is in the shape of the power spectrum at $k_{\|} = 0$. Where power spectra using data from FHD have a fairly uniform increase in power with decreasing $k_{\perp}$ those using RTS data have a roughly flat spectrum which increases dramatically in the few bins below $k_{\perp} < .002$ or below 20 wavelengths. This corresponds to power on scales of 2.86 °. The number of baselines drops precipitously below 20 wavelengths.XXX

The major difference between the power spectrum methods is in the calculation and minimization of uvf covariance. CHIPS aims to make a more accurate mathematical treatment of the covariance but to do so it must take on more of the instrument modeling. Meanwhile, *ϵppsilon* leaves the modeling to the foreground subtraction step and assumes that, to first order, covariance has been minimized by an additional down-weighting by the primary beam of the instrument; i.e. uv cells are chosen to be sized by the inverse of the primary beam width. Thus one of the largest differences is in the amount of correlation along the vertical, or line-of-sight, direction. Both CHIPs and *ϵppsilon* , when applied to their primary foreground subtraction strategy (FHD for *ϵppsilon* and RTS for CHIPS), have minimal line-of-sight covariance at low $k_{\perp}$ where frequency to frequency variation in uv sampling is small. At high $k_{\perp}$ *ϵppsilon* 's simplistic Fourier Transform reveals the large residual correlation between different $uv\eta$ cells caused by the fact that baseline length changes quickly with frequency. *ϵppsilon* has so far avoided directly accounting for covariance choosing speed over accuracy on long baselines.

XXX TBD add a little discussion of the differences between power spectra when things are getting closer to finalize: focus on 1) things depending on calibration/foreground subtraction, 2) things depending on power spectrum calc. 3) things arising from cor-

relation between fg and ps (probably due to interface stuff like weights or whatever)

Using the speedup provided by *ϵppsilon* , many iterations of "preview" power spectra were used to view the effect of calibration and flagging choices on the power spectrum. For example the wide-field effects, described in detail by Thyagarajan et al. (2015), are clearly visible in power spectra computed by *ϵppsilon* and are seen to be stronger for certain configurations of pointing and sky. Removing these portions of the data eliminated a substantial amount of bleed from the wedge into the window. Using this and other jackknife selections we arrived at a refined data FHD image cube which was then carried into the *E*MPCOV analysis resulting in the power spectra shown in Figure 4. Though well above the predicted signal level the 2σ error bars are mostly consistent with noise. Residual excesses –particularly near low *k*s– are consistent with a fairly aggressive inclusion of points near to the wedge, in this case points up to $0.02Mpc^{-1}$ away from the horizon (the solid black line in Figure 3)were included.

5. Conclusions

Each pipeline is necessarily built on a complex software framework which is only imperfectly described in prose and is susceptible to human error. Comparison of both 2D images and 3D power spectra have allowed us to build confidence in our estimate of the power spectrum and have revealed a number of issues both systematic (related to our understanding of the instrument or foregrounds) and algorithmic (optimizing our use of this knowledge). Lessons learned include:

- in-field calibration vs calibration transfer Comparison of RTS and FHD images continues to reveal the significance of calibration algorithm as well as weighting schemes on the effectiveness of imaging and model subtraction. Differences remaining in the images presented here are likely due to small divergences in calibration method. Current work focuses on finding a robust calibration model upon which the two systems agree and results in better subtraction of the sky model.

- cable reflections One debatable aspect of calibration is the number of free parameters allowed into the spectral dimension. Individual calibration of each channel independently allows the greatest flexibility but has the consequence of possibly adding or subtracting to the spectral line reionization signal. Both calibration pipelines begin by calibrating each channel and then fitting a spectral average. The RTS fits a low order polynomial, piecewise, to each of the 24 1.28MHz sub-band solutions, while FHD fits a similar order polynomial to the entire band's calibration solution. Inspection of power spectra calibrated using the FHD scheme revealed previously unknown spectral features corresponding to reflections on the analog cables at the -20dB level (1.5%). FHD calibration now includes a fit for these reflections and the feature is no longer visible. These features are fully covered by the RTS fit (which uses of order 10 times as many free parameters as FHD).

- full forward modeling for absolute calibration, signal loss One way in which all pipeline results differed from each other is in the overall amplitude of the power spectrum scale. Agreement only occurs when flux scale, weightings, and fourier conventions are all in alignment. Perhaps the most important factor is assessment of signal loss. Unintential or unavoidable down-weighting or subtraction of reionization signal could occur at multiple stages such as bandpass calibration, uvf gridding, or inverse covariance weighting. This loss is best calibrated via forward modeling of a simulated reionization signal and in the process provides verification of the overall power spectrum scale. Such simulations have been used to verify the various steps in the FHD-$\epsilon ppsilon$ pipeline and by stepping through the pipeline at each major operation have been shown to be self-consistent (see Hazelton et al 2015) and suitable for calibration of the other pipelines.

In this overview paper we have provided a top level view of foreground subtraction and power spectrum estimations methods described more completely in companion papers Hazelton et al 2015, Trott et al 2015, and Dillon et al 2015and provided a basis for an apples-to-apples comparison. In this comparison we see that both foreground subtraction methods are able to reliably remove about 50% of the power with a fairly simplistic model but that the reconstruction of the residual large scale power depends heavily on small differences in calibration and imaging algorithm which ultimately limits the accuracy of the reconstruction. In a similar way we use the difference between two independently developed power spectrum pipelines to reveal effects in the power spectrum which seem to be unique to the power spectrum estimation, those common to the calibration and foreground subtraction step, and those which appear to be common to the sky itself and on a believably consistent scale. Though none of the power spectra are identical, the degree of agreement and the success at making nearly noise limited measurements allows us to draw conclusions about the relative quality of different selections of data. Bad data is truly bad and not evidence of an underlying software or algorithmic problem. Using our validated pipeline to make quick estimates of the power spectrum in different selections of data we were able to select a high quality set, with a well understood calibration, for application of the implicit covariance technique and generate a noise limited measurement.

The 1D power spectrum presented here and in Dillon et al 2015is roughly a 500 times deeper then the previous MWA power spectrum Dillon et al. (2014) which was done using roughly the same amount of integration time but only 32 of the present 128 tiles. Future work will focus on refining calibration and weighting schemes to more accurately reconstruct large scale power and building on deeper integrations using data collected in recent observing campaigns.

This work was supported by the U. S. National Science Foundation (NSF) through award AST–1109257. DCJ is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST–1401708. JCP is supported by an NSF Astronomy and Astrophysics Fellowship under award AST-1302774. This work makes use of the Murchison Radio-astronomy Observatory, operated by CSIRO. We acknowledge the Wajarri Yamatji people as the traditional owners of the Ob-

servatory site. Support for the MWA comes from the NSF (awards: AST-0457585, PHY-0835713, CAREER-0847753, and AST-0908884), the Australian Research Council (LIEF grants LE0775621 and LE0882938), the U.S. Air Force Office of Scientific Research (grant FA9550-0510247), and the Centre for All-sky Astrophysics (an Australian Research Council Centre of Excellence funded by grant CE110001020). Support is also provided by the Smithsonian Astrophysical Observatory, the MIT School of Science, the Raman Research Institute, the Australian National University, and the Victoria University of Wellington (via grant MED-E1799 from the New Zealand Ministry of Economic Development and an IBM Shared University Research Grant). The Australian Federal government provides additional support via the Commonwealth Scientific and Industrial Research Organisation (CSIRO), National Collaborative Research Infrastructure Strategy, Education Investment Fund, and the Australia India Strategic Research Fund, and Astronomy Australia Limited, under contract to Curtin University. We acknowledge the iVEC Petabyte Data Store, the Initiative in Innovative Computing and the CUDA Center for Excellence sponsored by NVIDIA at Harvard University, and the International Centre for Radio Astronomy Research (ICRAR), a Joint Venture of Curtin U

REFERENCES

Ali, Z. S. et al. 2015, ArXiv e-prints

Barkana, R. 2009, MNRAS, 397, 1454

Barkana, R. & Loeb, A. 2008, Monthly Notices of the Royal Astronomical Society, 384, 1069

Beardsley, A. et al. 2013, Monthly Notices of the Royal Astronomical Society, 429, L5

Bhatnagar, S., Rau, U., & Golap, K. 2013, ApJ, 770, 91

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

Clark.Allen.Arcus. bartTBD

Datta, A., Bowman, J. D., & Carilli, C. L. 2010, The Astrophysical Journal, 724, 526

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

Dillon, J., Liu, A., & Tegmark, M. 2013, Physical Review D, 87, 43005

Dillon, J. et al. 2014, Physical Review D, 89, 23002

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

Hurley-Walker, N. et al. 2014, PASA, 31, 45

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

Jelić, V. et al. 2008, Monthly Notices of the Royal Astronomical Society, 389, 1319, (c) Journal compilation © 2008 RAS

Liu, A., Parsons, A. R., & Trott, C. M. 2014a, eprint arXiv, 1404.2596

—. 2014b, eprint arXiv, 1404.4372, 19 pages, 7 figures

Liu, A. & Tegmark, M. 2011, Physical Review D, 83, 103006

Lomb, N. R. 1976, Ap&SS, 39, 447

Lonsdale, C. J. et al. 2009, Proceedings of the IEEE, 97, 1497

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

Morgan, J., Hurley-Walker, N., Wayth, R., & MWA. 2014, in American Astronomical Society Meeting Abstracts, Vol. 223, American Astronomical Society Meeting Abstracts # 223, #421.01

Neben, A. R. e. a. 2015, Radio Science

Noordam, J. E. 2004, Ground-based Telescopes. Edited by Oschmann, 5489, 817

Offringa, A. R. et al. 2015, PASA, 32, 8

Ord, S. et al. 2010a, Publications of the Astronomical Society of the Pacific, 122, 1353

Ord, S. M. et al. 2010b, Publications of the Astronomical Society of the Pacific, 122, 1353

Parsons, A. R. et al. 2014, ApJ, 788, 106

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

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

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

Salvini, S. & Wijnholds, S. J. 2014, A&A, 571, A97

Scargle, J. D. 1982, ApJ, 263, 835

Sullivan, I. S. et al. 2012, The Astrophysical Journal, 759, 17

Tasse, C. et al. 2012, Comptes Rendus Physique, 13, 28, académie des sciences

Thyagarajan, N. et al. 2015, ArXiv e-prints

Thyagarajan, N. et al. 2013, The Astrophysical Journal, 776, 6

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

Vedantham, H., Shankar, N. U., & Subrahmanyan, R. 2012a, The Astrophysical Journal, 745, 176

—. 2012b, The Astrophysical Journal, 745, 176

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

Zaroubi, S. 2013, in Astrophysics and Space Science Library, Vol. 396, Astrophysics and Space Science Library, ed. T. Wiklind, B. Mobasher, & V. Bromm, 45

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

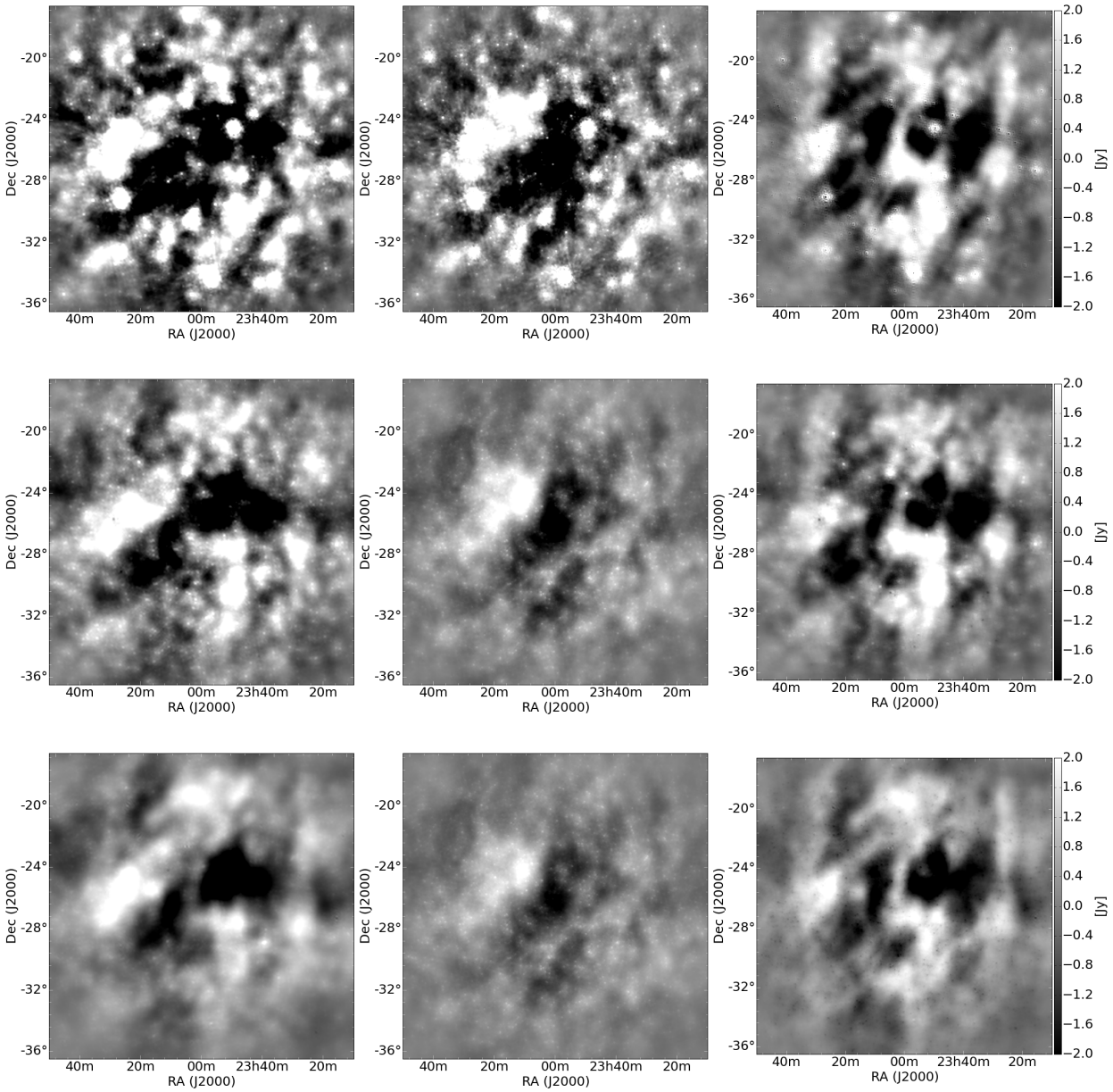

Fig. 2.— A comparison between the image outputs of the FHD (left), RTS (center) and their difference (right) all averaged in the spectral dimension and projected from native healpix to flat sky. The images are dirty, in the sense that catalog sources have been subtracted, but no additional deconvolution has been performed. In the top row, no subtraction has been performed, in the middle row the same catalog of 300 sources have been subtracted, and on the bottom each system is allowed to make its own determination of the appropriate number of sources with FHD increasing the count to $\tilde{1}000$ and RTS staying at 300. Both have been left in the natural weighting used by image-based power spectrum schemes. Most of the *uv* data points sample scales on the degree or larger scales, thus the greatest differences are apparent at large scales.

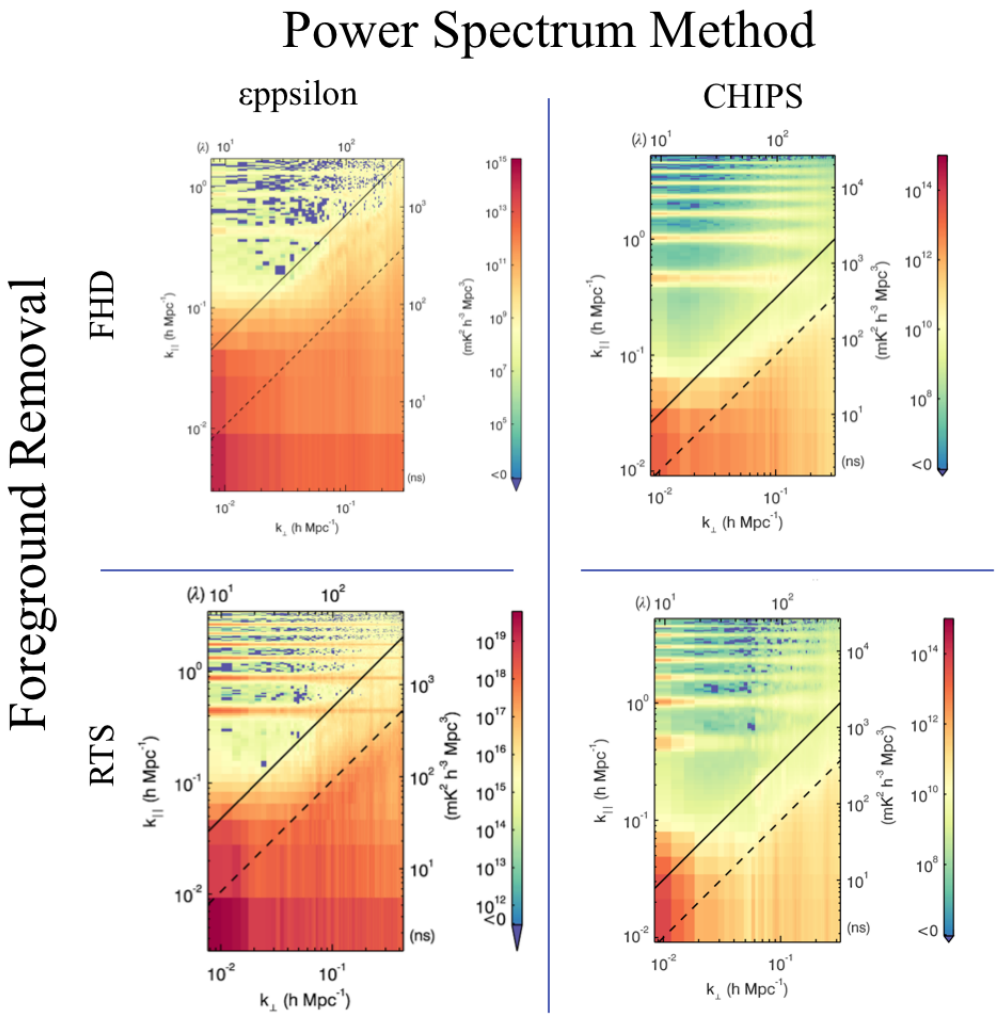

Fig. 3.— MWA power spectra computed using two foreground subtraction methods and two power spectrum estimation methods. In the top row data have been calibrated to a catalog followed by subtraction of a deeper catalog modeled into instrumental space using the Fast Holographic Deconvolution method, in the bottom row calibration and foreground subtraction have been performed with the MWA Real Time System. In the left column, power spectra have been estimated with *εppsilon* , which emphasizes speed and full error propagation, in the right column, CHIPS tries to minimize correlation between k modes.

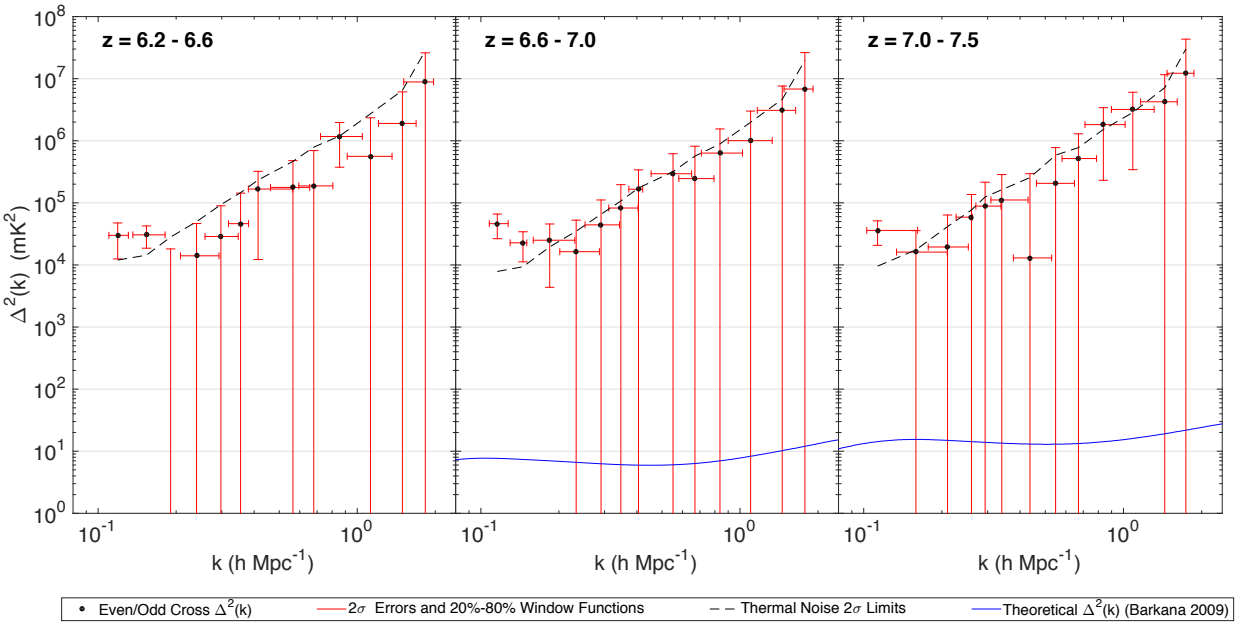

Fig. 4.— These MWA 1D power spectra, of the same data shown in Figure 3 and described in detail by Dillon et al 2015, are computed using: FHD calibration and foreground removal, *epsilon* -based curating and the quadratic estimator method with empirical residual covariance weighting. Also, for comparison an estimate of the noise power spectrum (dashed line) and a fiducial model (blue line) based on that of Barkana (2009), where reionization ends before $z = 6.4$. Though well above our fiducial model level the 2σ error bars are notably consistent with noise rather than foreground leakage. By empirically measuring and projecting out excess covariance, this method has down-weighted remaining foreground-like power regardless of its origin in the foreground or instrument model. Any remaining excesses in this plot are thought to be consistent with a fairly aggressive inclusion of points near to the wedge –in this case points up to $0.02\mathrm{Mpc}^{-1}$ away from the horizon (the solid black line in Figure 3) were included– and known systematics like cable reflections.