



Statistical EOR detection and the Mileura Widefield Array

Miguel F. Morales ^{a,*}, Judd D. Bowman ^a, Roger Cappallo ^b,
Jacqueline N. Hewitt ^a, Colin J. Lonsdale ^b

^a MIT Kavli Institute for Astrophysics and Space Research, 70 Vassar Street, Cambridge, MA 02139, United States

^b MIT Haystack Observatory, Off Route 40 Westford, MA 01886, United States

Available online 31 January 2006

Abstract

Statistical epoch of reionization (EOR) observations are one of the emerging fields of experimental cosmology and could provide unique constraints on the formation of structure and the emergence of the first luminous objects. The primary challenge in observing the 21 cm EOR signature is subtracting the strong astrophysical and instrumental foregrounds. In this paper, we review the three-dimensional statistical EOR analysis, present a multi-stage foreground removal framework, and review the design and early field testing for the Mileura Widefield Array.

© 2005 Published by Elsevier B.V.

Keywords: Cosmology; Theory; Epoch of reionization; 21 cm observations; Foregrounds

Contents

1. Introduction	173
2. Statistical EOR detection	174
3. Foreground subtraction	174
4. Mileura Widefield Array	177
References	178

1. Introduction

Highly redshifted 21 cm neutral hydrogen emission from the epoch of reionization (EOR) has the potential to revolutionize our understanding of galaxy and structure formation and place unique cosmological constraints (Tozzi et al., 2000; Zaldarriaga et al., 2004; Furlanetto et al., 2004a; Bharadwaj and Ali, 2005). Since direct imaging of the EOR will require the collecting area of the square kilometer array, the first generation EOR observatories all

build on the statistical detection techniques developed for the cosmic microwave background anisotropy.

However, there are a number of challenges which must be confronted to reveal the very faint 21 cm EOR signal. Neutral hydrogen emission from the EOR appears at the low radio frequencies as faint fluctuations between 75 and 200 MHz (for redshifts of $18 \rightarrow 6$). This frequency range is commonly used for television, FM radio, and satellite transmission, producing an extremely bright foreground of terrestrial emission. Additionally, turbulence in the Earth's ionosphere refracts low frequency radio waves and must be corrected using techniques similar to wide-field adaptive optics. After the terrestrial sources have been

* Corresponding author.

E-mail address: mmorales@space.mit.edu (M.F. Morales).

removed, there remain a host of astrophysical foregrounds which are ~ 5 orders of magnitude brighter than the EOR signal.

In this paper, we will review the three-dimensional statistical analysis that has been developed for observing the EOR in Section 2, and the foreground mitigation techniques and progress of the US – Australian Mileura Widefield Array in Sections 3 and 4.

2. Statistical EOR detection

Unlike the cosmic microwave background (CMB) emission, which is inherently two-dimensional (sky position), the EOR data is three-dimensional because the redshift of the observed neutral hydrogen emission maps to the line-of-sight distance. This allows us to extend the statistical techniques developed for the CMB to three dimensions, and use power spectrum statistics to study the EOR. These statistical analysis techniques dramatically increase the sensitivity of first generation EOR power spectrum observations, and allow much more detailed studies of the cosmology (Bowman et al., submitted; Furlanetto et al., 2004b; Morales, 2005).

In wide field radio observations, the observed frequency of the redshifted 21 cm line maps to the line-of-sight distance of the source. Ignoring the small effect of peculiar velocities (discussed in depth in Barkana and Loeb, 2005), an approximate image cube can be formed by replacing the observed frequency with the equivalent line-of-sight distance. The resulting image cube is a full three-dimensional tomograph of neutral hydrogen emission in the high redshift universe.

The emission strength from neutral hydrogen during the epoch of reionization depends on the local hydrogen density, the fractional ionization, and the spin temperature of the gas, and gives us a differential specific intensity ΔI at the observed location with respect to the cosmic microwave background radiation. Because the 21 cm line is narrow and the gas is optically thin, we can directly form an image cube which represents the three-dimensional emission of the neutral hydrogen. Furthermore, a three-dimensional Fourier transform can be applied to the image cube to determine the spatial structure of the EOR signal. This gives us four representations of the signal: the observed specific intensity in sky position θ_x, θ_y and frequency Δf (and their inverses u, v, η for the Fourier transform), and cosmological coordinates comoving distance \vec{r} (and inverse \vec{k}). The specific intensity in Fourier space is used for the three-dimensional statistical EOR measurement.

The statistics of the model predictions for the EOR signal fall under the class of spatially *homogeneous* and *isotropic* random fields, since there is no preferred position or direction in space. In the limit that the observed data are from the same epoch, the isotropy of the model implies that the Fourier transform representation of the neutral hydrogen emission will be spherically symmetric since the expectation is the same in all directions. In the absence of

foregrounds the EOR power spectrum could be measured by dividing the individual power measurements in the Fourier space into spherical annuli, and averaging the values within each shell to produce a single power spectrum measurement at the given length scale (Morales and Hewitt, 2004). This reduces the billions of individual power measurements down to of order 10 statistical measurements, and is behind the extraordinary sensitivity of cosmological power spectrum measurements (Bowman et al., submitted). Fig. 1 shows an example of the resulting power spectrum measurement for the Mileura Widefield Array.

The spherical symmetry of the EOR signal is an extraordinarily useful characteristic, and all the statistical characteristics of the signal will share this symmetry (including all moments of the distributions, not just the power spectrum). The spherical symmetry can also be used to differentiate the residual foreground contamination from the EOR signal and is explored in the following section.

3. Foreground subtraction

The foreground subtraction process starts with the image cube and progresses through three stages of foreground subtraction, removing increasingly faint contaminants. The three stages can be described as source removal, spectral fitting, and residual error subtraction.

In the first stage the bright contaminating sources, both astrophysical and man made, are removed. Because the spatial and frequency response of an array is not a delta-function, emission from a bright source will spill over into neighboring pixels and frequency channels. Thus, we need to use the traditional radio astronomy subtraction technique of removing the sources directly from the visibilities to subtract the array sidelobes along with the central emission. The goal of the first foreground removal stage is to subtract the contributions from all sources which can contaminate distant locations in the image cube. The signal strength of the EOR is a few mK, so the astrophysical and RFI sources must be subtracted until the sidelobes are \lesssim mK. This places strong constraints on the spatial and frequency dynamic range of an array, as well as the RFI environment.

Unfortunately, even the faint emission of galaxies below the detection threshold will overwhelm the weak EOR signal (Di Matteo et al., 2002), and we must resort to more powerful subtraction techniques in the second stage to reveal the EOR signal. In every pixel of the image cube, there will be contributions from many faint radio galaxies. The spectrum within one pixel is well approximated by a power-law, and can be fit and removed. Since the EOR signal is bumpy, fitting smooth power law models nicely removes the foreground contribution while leaving most of the cosmological signal (Briggs et al., submitted). There are a number of subtle effects which must be carefully monitored, such as changing pixel size, but this is an effective way of removing the contributions of the faint radio galaxy foreground.

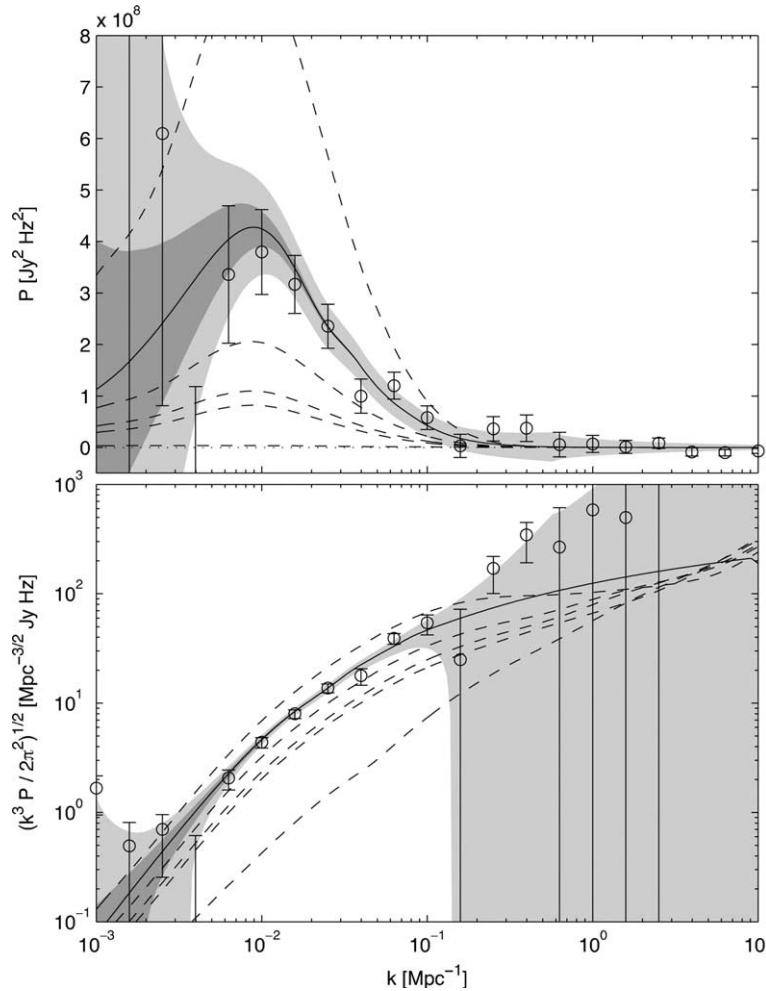


Fig. 1. This figure from Bowman et al. (submitted) shows the sensitivity of a single 360 h power spectrum measurement with the MWA. The signal from a fully neutral intergalactic medium with high spin temperature at a redshift of 8 is shown as the black line, after convolution with the MWA instrumental response. The upper panel is in instrumental coordinates where the measurement noise is Gaussian, while the lower panel shows the same results in the commonly used theoretical units. The dark grey region shows the cosmic variance and the light grey region the thermal noise for five spectral points per decade, as shown by the points and error bars for a single realization. The dashed lines represent models of the power spectrum as reionization progresses from Furlanetto et al. (2004b), with ionized fraction from top to bottom of $x_i = 0.51, 0.0$ (solid), 0.43, 0.38, 0.25, and 0.13. In general, the amplitudes of the model power spectra drop rapidly between $x_i = 0.0$ and 0.13, and then slowly increase with ionization fraction as large bubbles increase the contrast.

While the vast majority of the foreground contamination is removed in the first two analysis stages, residual foreground contamination will remain due to errors in the subtraction process. For example, a misestimate of the spectral slope in the second step will leave residual fore-

ground contamination. While it is impossible to determine what these subtraction errors are individually, we can predict, measure, and remove the *average* effects of this residual foreground contamination from the power spectrum. This subtraction of the average residual errors of the first

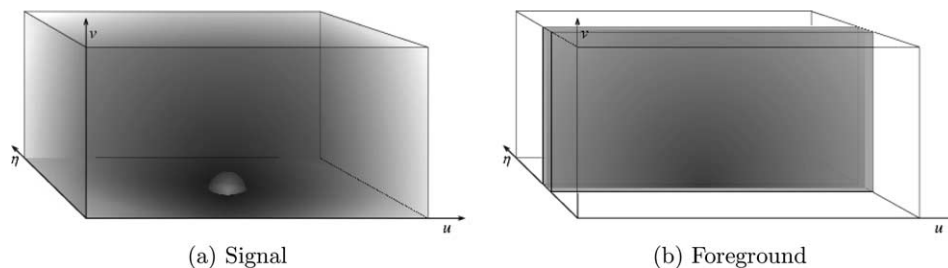


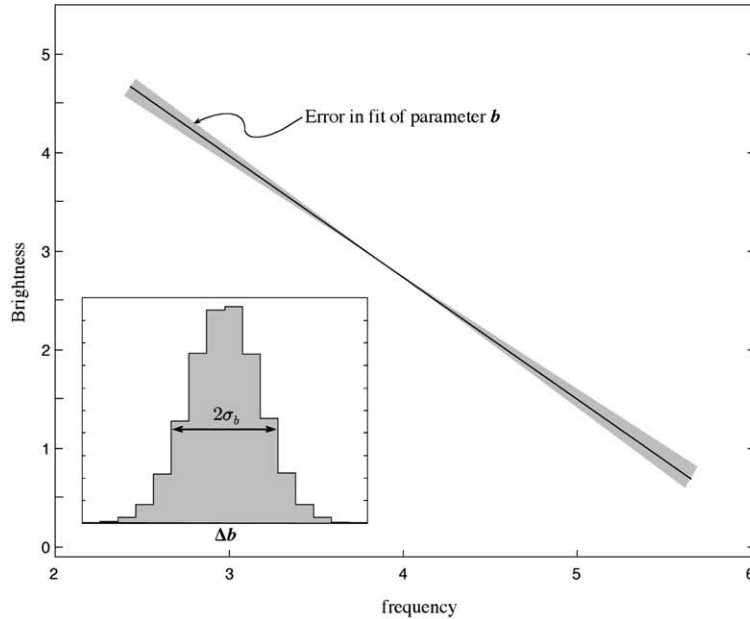
Fig. 2. Panel (a) shows the spherically symmetric power spectrum of the EOR signal (zero is in the center of the lower face), while panel (b) shows the separable-axial power spectrum typical of the residual foregrounds. The power spectrum shapes are known, and the amplitudes can be fit in the parameter estimation stage to separate the residual foreground subtraction errors from the faint EOR signal.

two subtraction stages is made possible by differences between the power spectrum shape of the signal and the residual foregrounds.

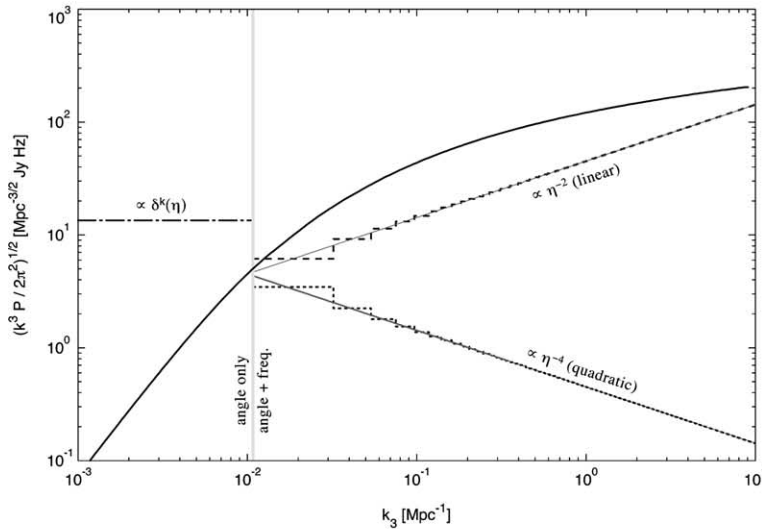
Fundamentally, the spherical symmetry of the EOR signal is based on the combination of narrow line emission and redshift which allows one to form an isotropic image cube, where frequency maps to the line-of-sight distance (see Fig. 2). This mechanism is not present for most of

the foreground contaminants, which either have smooth continuum spectra or are local line sources. These foreground have a *separable-axial* symmetry where the specific intensity $\tilde{I}_F(u, v, \eta)$ for the foreground has the form $A(\eta)B(\sqrt{u^2 + v^2})$ (functions A and B are both complex).

Even after the first two subtraction stages there is residual foreground contamination due to subtraction errors, and these residual foregrounds retain the separable-axial



(a) Spectral Errors



(b) Residual Power Spectrum

Fig. 3. The cartoon in panel (a) shows the true foreground continuum spectrum observed in one pixel as a black line, and the error in fitting the spectral slope b due to thermal noise. The inset shows the expected Gaussian profile of $\Delta b \equiv (b_T - b)$ and the width of the distribution σ_b . Panel (b) then shows the observed EOR signal as a solid black line, and the residual foreground power spectrum due to linear and quadratic fitting errors, which are shown as thick and thin grey lines, respectively. All power spectrum measurements to the left of the vertical line are purely angular (for a bandwidth of 16 MHz) while the measurements to the right of the line use both the angular and frequency information. The dashed lines show the binning effects for the linear and quadratic components of the residual foreground, with the dash-dot line showing the δ^k -function contribution from the offset term. The residual spectral fitting errors are nearly white in the visibility plane, and fall as a high power of η to produce the power spectrum shape shown in Fig. 2 panel (b). The amplitudes of the residual foreground components depend only on the standard deviations of the fitting errors and can be fit in the parameter estimation stage of the analysis.

symmetry. Since we can calculate the shape of these residual foregrounds, we can fit their amplitude in the Fourier cube to identify their contribution. Fig. 2 shows the difference in shape between the power spectrum of the EOR signal and residual foregrounds due to subtraction errors. The detailed shapes of the residual foregrounds can be calculated (Morales and Bowman, submitted), and depend on the standard deviation of the foreground fits. An example of this is shown in Fig. 3.

So in the presence of foregrounds our final stage of the analysis becomes a multi-parameter fit, with each component of the residual foreground and the EOR signal being represented by a corresponding power spectrum template and amplitude. The measurements are then decomposed into template amplitudes to separate the EOR signal from the residual contamination of foreground subtraction errors in the first two stages.

The full analysis of the EOR signal consists of subtracting out the RFI and bright astrophysical sources, fitting the spectra in the image and visibility cubes to remove the effects of faint foregrounds, and then using a parameter fit to separate the cosmological EOR signal from the residual foreground contamination. By fitting both the local parameters in the first two subtraction stages and their standard deviations in the parameter fitting stage, the foreground contamination can be removed to much higher precision.

4. Mileura Widefield Array

Astronomy has a long tradition of going to the ends of the earth for great observing conditions. The Mileura Widefield Array (MWA) is located at the radio quiet Mileura site in the Western Australian Desert, which is one of the great locations for low frequency radio astronomy. The MWA operates in two wavebands, an 800 MHz to 1.6 GHz band for “low” redshift 21 cm work ($z \lesssim 1$) and SKA technical demonstration, and the 80–300 MHz band for EOR observations and other low frequency science drivers.¹ A complete description of the array and all the science drivers can be found at web.haystack.mit.edu/arrays/MWA/index.html.

The low frequency band of the MWA has been optimized for high surface brightness sensitivity and large field of view, and is best described as a precision survey instrument. Each MWA antenna consists of 16 crossed-dipoles in a 4×4 array which are beamformed together using a switchable analog delay line at the antenna. It is this beamforming operation which creates the $15\text{--}45^\circ$ FWHM antenna field of view (depending on frequency), and the phase center can be steered to view any point above an ele-



Fig. 4. This photograph shows an MWA prototype antenna being tested on site in Mileura. The antenna consists of 16 crossed dipoles, and the signals from the dipoles are combined at each antenna by a switchable analog delay line. The resulting beam is $15\text{--}45^\circ$ across, and can be steered by digitally switching the delay lines. Five hundred of these 16 dipole antennas will be placed in a 1.5 km array for the low frequency band of the final MWA observatory.

vation of $\sim 30^\circ$. The frequency response of the antenna has been optimized from 80 to 300 MHz and is sky noise dominated over the full band. A prototype antenna is shown in Fig. 4 undergoing testing at Mileura.

The full MWA array consists of 500 sixteen dipole antennas quasi-randomly distributed within a 1.5 km diameter circle. The signal from each antenna is directly sampled and digital filters are used to select 32 MHz of bandwidth and create four thousand 8 kHz frequency channels (essentially the F part of the FX correlator is performed in the digital receiver). The full bandwidth is then transported to the central processing center. The full antenna field of view is preserved by cross-correlating all 500 antennas, full Stokes, at 32 MHz bandwidth to create 2 billion independent visibility measurements every $1/2$ s in the wide field correlator (WFC). The WFC enables both the wide field of view and allows the antennas to be widely scattered to produce a very high instantaneous spatial dynamic range. The correlator field of view also extends horizon to horizon, allowing the interfering sources in the far side lobes to be imaged and removed as necessary.

Early site characterization and hardware testing is underway, and there are currently three antennas operating in Mileura. Fig. 5 shows unfiltered 1 h spectra from the first antenna. The data are sky noise limited and show the extremely low RFI contamination observed in Mileura.

Ionospheric turbulence refracts radio waves at EOR frequencies, producing wavefront distortions analogous to atmospheric distortion at optical wavelengths. The coherence length of ionospheric fluctuations (or more accurately the Fried length) is typically several km with timescales of a few tens of seconds at EOR frequencies. Since the MWA is smaller than the ionospheric length scale but has a field of view which incorporates many ionospheric patches, the ionospheric distortion is well approximated as a simple

¹ Officially, the low frequency band of the MWA is known as the Mileura Wide-field Array Low Frequency Demonstrator (MWA-LFD). Since only the lower band is applicable for EOR observations it is commonly referred to as simply the MWA without fear of confusion, and we will adopt this convention.

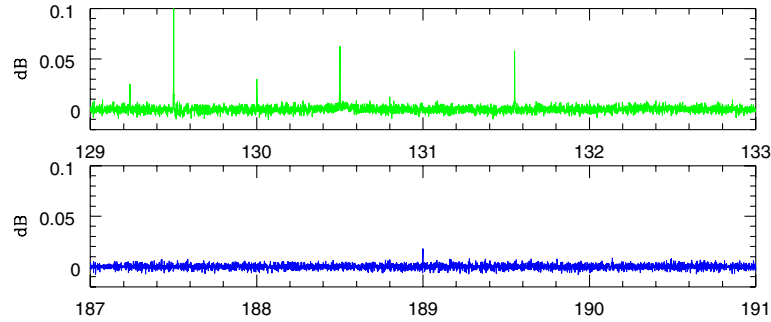


Fig. 5. These spectra from the first MWA test antenna are from a 1-h observation at 131 (top) and 189 MHz (bottom) after bandpass correction (including median filtering to remove a small ripple due to cable reflections), representing EOR observations at redshifts of 9.8 and 6.5, respectively. Note the very small vertical scale.

refractive position shift that varies across the field of view (Londale, *in press*). The ionospheric portion of the calibration can be envisioned as measuring the refractive shift for a large number of calibrator sources across the MWA field of view. Because the MWA has five times the collecting area and 20 times the bandwidth of the VLA 74 MHz system, there is a high density of available calibrator sources. In 10 s the MWA should observe ~ 200 sources within the primary field of view at 10 sigma significance in each 16 kHz band. Since the refractive shift as a function of frequency is known, this gives a two parameter fit to four thousand 10 sigma observations for each of 200 calibrator sources. The observed refractive shifts of the calibrators can then be fit to a low order polynomial to extrapolate the ionospheric calibration across the observed field of view. After correcting for ionospheric effects, deep observations of the calibrator sources are used to determine the gain and polarization calibration for each antenna as a function of sky position.

The MWA data processing path uses the ionospheric and instrumental calibrations to correct the 2 billion visibility measurements from the Wide Field Correlator, and integrates to ~ 10 min observing blocks. These 10 min

blocks are then stored in a 100 TByte database along with details of the observation, such as ionospheric conditions and array performance. For the final analysis the best 10 min observing blocks are selected and integrated to form a single observation a few hundred hours in duration which includes only the best observing conditions. The Mileura Widefield Array has been designed to be a sensitive first generation EOR observatory with excellent dynamic range and very low systematic errors.

References

- Barkana, R., Loeb, A., 2005. *ApJ* 624, L65.
- Bharadwaj, S., Ali, S.K.S., 2005. *MNRAS* 356, 1519.
- Bowman, J.D., Morales, M.F., Hewitt, J. *ApJ* (submitted).
- Briggs, F.H., de Bruyn, A.G., van Haarlem, M. *A&A* (submitted).
- Di Matteo, T., Perna, R., Abel, T., Rees, M.J., 2002. *ApJ* 564, 576.
- Furlanetto, S.R., Zaldarriaga, M., Hernquist, L., 2004a. *ApJ* 613, 1.
- Furlanetto, S.R., Zaldarriaga, M., Hernquist, L., 2004b. *ApJ* 613, 16.
- Londale, C.J. *ASP Conference Series* (in press).
- Morales, M.F., 2005. *ApJ* 619, 678.
- Morales, M.F., Bowman, J.D. *ApJ* (submitted).
- Morales, M.F., Hewitt, J., 2004. *ApJ* 615, 7.
- Tozzi, P., Madau, P., Meiksin, A., Rees, M.J., 2000. *ApJ* 528, 597.
- Zaldarriaga, M., Furlanetto, S.R., Hernquist, L., 2004. *ApJ* 608, 622.