Draft version May 5, 2017
Preprint typeset using LaTeX style AASTeX6 v. 1.0

THE RADIO SKY AT METER WAVELENGTHS: M-MODE ANALYSIS IMAGING WITH THE OWENS VALLEY LONG WAVELENGTH ARRAY

Michael W. Eastwood, Gregg Hallinan

ABSTRACT

I present all-sky maps of the sky from the Owens Valley Long Wavelength Array. These maps are created from the application of m-mode analysis.

1. INTRODUCTION

Studies of the cosmic microwave background (CMB) have given us an unprecedented understanding of the universe during the epoch of recombination (Hinshaw et al. 2013; Planck Collaboration et al. 2014, 2016a). Before arriving at the Earth, the light from the CMB propagates through the intergalactic medium (IGM), intervening galaxies and galaxy clusters, and the Milky Way's interstellar medium (ISM). Careful measurement and removal of contaminating foreground emission is therefore an important component for CMB analysis (Planck Collaboration et al. 2016b). On large angular scales galactic synchrotron, free-free, thermal dust, and anomolous microwave emission contribute. On smaller angular scales point sources such as distant galaxies and quasars contribute (Dickinson 2016).

Propagation effects from the IGM and intervening galaxy clusters can distort the thermal spectrum of the CMB and impart information about the intervening material in those distortions. For instance the measured optical depth to Thomson scattering implies a mean redshift of the Epoch of Reionization (EoR) of ??? TODO: put number and citation. The thermal Sunyaev–Zel'dovich effect does something TODO: thing and citation. The kinetic Synyaev–Zel'dovich effect does another thing TODO: thing and citation.

At redshifts $20 \gtrsim z \gtrsim 7$ neutral hydrogen is observable as a 10 to 100 mK perturbation of the CMB spectrum due to the 21 cm hyperfine structure line of neutral hydrogen (Furlanetto et al. 2006; Pritchard & Loeb 2012). The spatial power spectrum of this perturbation is expected to be dominated by inhomogeneous heating of the IGM at $z \sim 20$ (Fialkov et al. 2014), and by growing ionized bubbles during the EoR at $z \sim 7$ TODO: citation.

TODO: discuss more interesting science to be done

There are numerous ongoing efforts to detect the cosmologically redshifted 21 cm transition including from PAPER/HERA (Ali et al. 2015; DeBoer et al. 2016), LOFAR (Patil et al. 2017), and the MWA (Beardsley et al. 2016).

TODO: discuss global experiments too

These low-frequency 21 cm cosmology experiments are limited by the dynamic range they can achieve against low-redshift sources of emission. The brightness temperature of the galactic synchrotron emission at high galactic latitudes is measured by Rogers & Bowman (2008) as

$$T \sim 300\,\mathrm{K} \times \left(\frac{\nu}{150\,\mathrm{MHz}}\right)^{-2.5}. \tag{1}$$

Therefore experiments conservatively need to achieve at least 4 orders of dynamic range. In theory this kind of dynamic range can be obtained by exploiting the fact that the foreground emission is predominantly synchrotron emission that has a smooth power law spectrum TODO: citation. In practice the interferometer itself imparts frequency structure into the measurements and therefore an experiment's ability to separate the foreground signal from the cosmological signal can be limited by the instrumental spectral smoothness TODO: citation.

Ideally the foreground removal strategy should be informed by the measured spatial and frequency structure of all foreground components. This possibility is limited by the availability of suitable high-fidelity low-frequency sky maps on angular scales ranging from tens of degrees to arcminutes.

The Global Sky Model (GSM) (de Oliveira-Costa et al. 2008; Zheng et al. 2017a) is a data-driven interpolation of various maps between 10 MHz and 100 GHz. However the majority of information contained in the GSM is derived at higher frequencies where the majority of the input maps are from. Below 408 MHz, the interpolation is largely driven by the Haslam 408 MHz map (Haslam et al. 1981, 1982). At lower frequencies, free-free absorption and synchrotron self-absorption becomes increasingly important and hence spectral indices derived at higher frequencies need corrections at lower frequencies.

Guzmán et al. (2011)

TODO: TGSS, MSSS, GLEAM

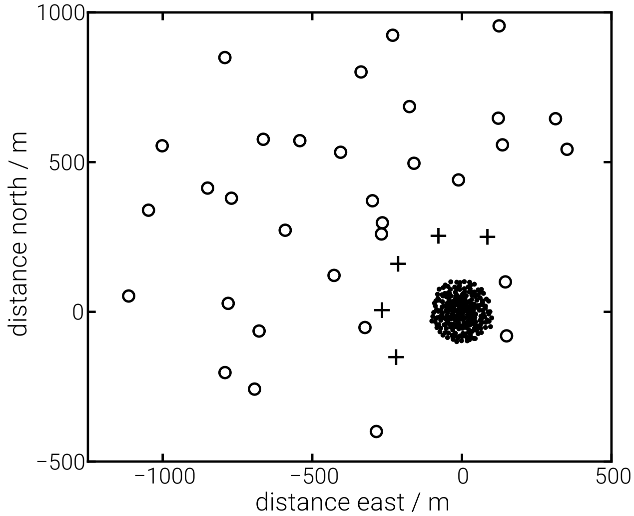

Figure 1. This figure shows the antenna layout for the OVRO LWA. Black dots correspond to antennas within the 200 m diameter core of the array. The 32 open circles are the expansion antennas built in early 2016 in order to increase the longest baseline to 1.5 km. The 5 crosses are antennas equipped with noise-switched front ends.

2. OBSERVATIONS

2.1. *The Owens Valley Long Wavelength Array*

The Owens Valley Long Wavelength Array (OVRO LWA) is a 288-element interferometer located at the Owens Valley Radio Observatory (OVRO) near Big Pine, California. The OVRO LWA is a low-frequency instrument with instantaneous bandwidth covering 27.384 MHz to 84.912 MHz and 24 kHz channelization. Each antenna stand hosts two perpendicular broadband dipoles so that there are 288×2 signal paths in total. These signal paths feed into the 512-input LEDA correlator (Kocz et al. 2015), which allows the OVRO LWA to capture the entire visible hemisphere in a single snapshot image. In the current configuration 32 antennas (64 signal paths) are unused.

The 288 antennas are arranged in a pseudo-random configuration optimized to minimize sidelobes in snapshot imaging. 251 of the antennas are contained within a 200 m diameter core. 32 antennas are placed outside of the core in order to extend the maximum baseline length out to $\sim$1.5 km. The final 5 antennas are equipped with noise-switched front ends for calibrated total power measurements of the global sky brightness. These antennas are used as part of the LEDA experiment TODO: cite Price et al. 2017 to measure the global signal of 21 cm absorption from the cosmic dawn. Figure 1 is a diagram of the antenna configuration.

TODO: cite Gregg's instrument paper

Beginning at 2017-02-17 12:00:00 UTC time, 28 consecutive hours of data was collected. This time was chosen based on the fact that it was raining at OVRO and rain tends to improve the low-frequency RFI environment considerably. During this time the OVRO LWA operated as a zenith-pointing drift scanning interferometer. The correlator dump time was selected to be 13 seconds such that the correlator output evenly divides a sidereal day.

TODO: describe the ionospheric conditions at the time

2.2. *Gain Calibration*

Antenna gain calibration is accomplished using an iterative method independently developed by Mitchell et al. (2008) and Salvini & Wijnholds (2014). The calibration routine is written in the Julia programming language (Bezanson et al. 2017), and is publicly available online[1] under an open source license (GPLv3+).

The antenna complex gains are updated once per day from a 9 hour track of data when Cas A and Cyg A are at high elevations. These two sources are – by an order of magnitude – the brightest radio sources in the northern hemisphere. Therefore the optimal time to solve for the interferometer's gain calibration is when Cas A and Cyg A are at high elevations. A calibration obtained in this manner is good consistent with the following day's calibration to within a 1% error in the flux scale and $1'$ error in the astrometry TODO: double check these numbers. These residual errors are likely driven by differing ionospheric conditions.

The flux scale is tied to the work of Perley & Butler (2016).

Temperature fluctuations within the electronics shelter generate 0.1 dB sawtooth oscillations in the analog gain. These oscillations occur with a variable 15 to 17 minute period. The amplitude of these gain fluctuations is calibrated by smoothing the autocorrelation amplitudes on 45 minute timescales. The ratio of the measured auto-correlation power to the smoothed autocorrelation power defines a per-antenna amplitude correction that is then applied to the cross-correlations.

2.3. *Primary Beam Measurements*

In order to generate wide-field images of the sky, the response of the antenna to the sky must be known. Fortunately drift-scanning interferometers like the OVRO LWA can empirically measure their primary beam under a mild set of symmetry assumptions (Pober et al. 2012). In this work we assume that the primary beam is invariant under north-south and east-west flips, and additionally that the x- and y-dipoles have the same response to the sky after rotating one by 90°. We measure the flux of several point sources TODO: which ones??

[1] https://github.com/mweastwood/TTCal.jl

as they pass through the sky and then fit a beam model composed of Zernike polynomials to those flux measurements. We select the basis functions to have the desired symmetry.

2.4. *Source Removal*

2.4.1. *Cassiopeia A and Cygnus A*

Without removing bright sources from the data, sidelobes from bright sources will dominate the variance in the image. At 74 MHz Cyg A is a 15,000 Jy source (Perley & Butler 2016). A conservative estimate for the confusion limit at 74 MHz with a 10 arcminute beam is 500 mJy (Lane et al. 2012). Therefore we require that Cyg A's sidelobes be at most of -45 dB down from its peak flux to prevent Cyg A's sidelobes from dominating the variance in the image. TODO: Measure Cyg A's sidelobe levels if it is not subtracted

At low frequencies, propagation effects through the ionosphere must be accounted for in order to achieve high dynamic range images. This necessitates the use of direction-dependent calibration and peeling (Mitchell et al. 2008; Smirnov & Tasse 2015). In the dataset used in this paper, scintillation and diffraction events on the timescale of a single integration (13 seconds) are observed. Therefore the direction dependent calibration changes on these timescales and the we must solve for one set of complex gains per source per integration.

The largest angular scale of Cas A is $\sim$8 arcminutes, while the largest angular scale of Cyg A is $\sim$2 arcminutes. With a 10 arcminute resolution, the OVRO LWA marginally resolves both sources. A resolved source model is needed for both sources. We fit a self-consistent resolved source model to eachs ource. This is performed by minimizing the minimizing the variance within an aperture located on each source after peeling. By phasing up a large number of integrations before imaging (at least over 1 hour) it is possible to smear out the contribution of the rest of the sky. We then use NLopt's Sbplx routine (Johnson 2008; Rowan 1990) to vary the parameters in a source model until the variance within the aperture is minimized.

TODO: Show a before and after with the updated source models

TODO: Give a plot with the updated source models

2.4.2. *The Sun*

The Sun can be trivially removed from any map of the sky by constructing the map using only data taken at night. A map of the entire sky can be obtained by using observations spaced 6 months apart. However the OVRO LWA 100 hour dataset consists of 100 consecutive hours of data, and hence roughly half of the data is from the day.

The Sun is a time variable source in radio.

TODO: does the previous routine work on the Sun too? worth a shot

2.4.3. *Sparking Power Lines*

The Owens Valley is an important source of water and power for the city of Los Angeles. Unfortunately this means that high voltage power lines run along the valley to the west of the OVRO LWA. Some of these power line poles have faulty insulators that arc and produce pulsed, broadband RFI. Because these poles exist in the near-field of the array, we have been able to localize some of them by using the curvature of the incoming wavefront. Efforts are currently underway to work with the utility pole owners to have these insulators replaced.

In the meantime it is possible to suppress their contamination in the dataset. The contribution of these RFI sources to the visibilities can be plainly seen by averaging >24 hours of data with the phase center set to zenith. In this way, true sky components are smeared along tracks of constant declination while terrestrial sources (ie. the arcing power lines) are not smeared. For similar reasons, when constructing a map of the sky from >24 hours of data these sparking power lines will produce artifacts in the map that are smeared along tracks of constant declination. However while the amplitude of the RFI will fluctuate with time, its phase will be constant. Therefore if you can use isolate a set of model visibilities for the RFI, it can be removed from each integration by scaling and subtracting these model visibilities from the measured visibilities.

Obtaining a model for the RFI is complicated by the fact that the contaminating sources are at extremely low elevations where the antenna response is maximally inhomogeneous. Therefore it is not enough to know the physical location of the faulty insulator generating the RFI. In addition you must also know the response of each antenna (amplitude and phase) in the appropriate direction. This motivates the use of peeling, which allows the antenna response to be a free parameter.

Therefore model visibilities for the RFI can be obtained by peeling the sources after smearing the visibilities over >24 hours. These model visibilities are then scaled and subtracted from each integration in the data set.

TODO: show an example smeared image before and after peeling

3. IMAGING

The goal of all imaging algorithms is to estimate the brightness of the sky $I_\nu(\hat{r})$ in the direction $\hat{r}$ and frequency ν. A radio interferometer measures the visibilities $V_{ij,\nu}$ between pairs of antennas – here numbered i and j respectively. If the antennas are separated by the baseline $\vec{b}_{ij}$, and $A_\nu(\hat{r})$ describes their response to the

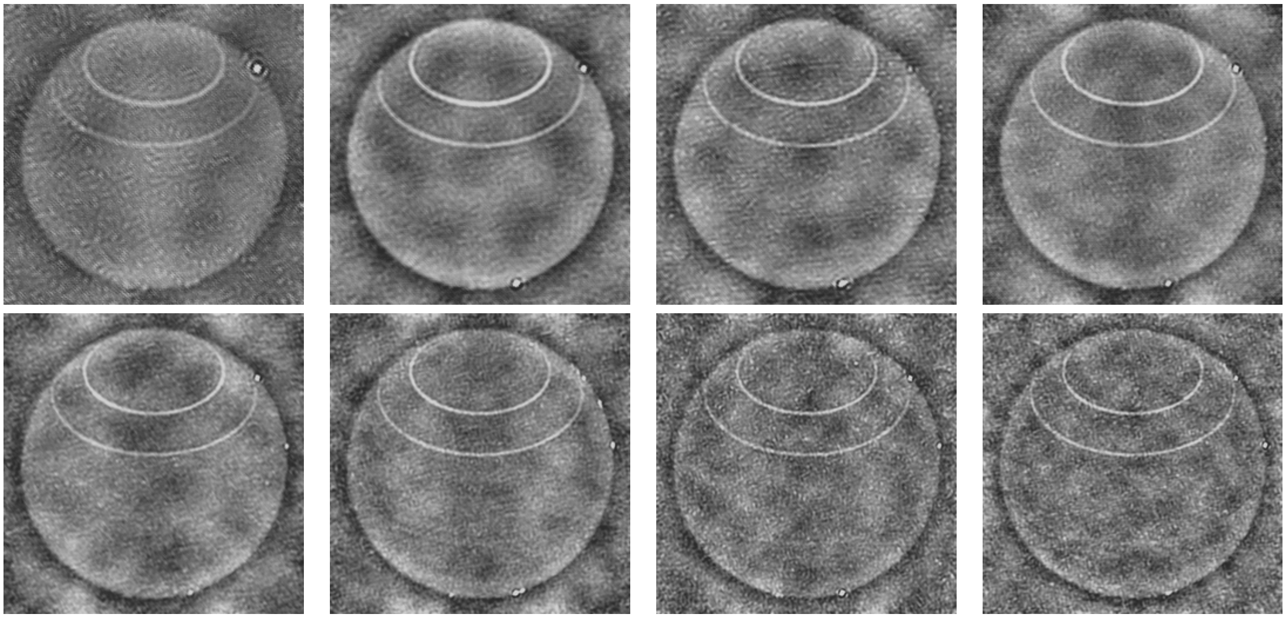

Figure 2. Dirty snapshot images after smearing the visiblities over the entire 100 hour observation window. The two bright source tracks correspond to Cas A (top track) and Cyg A (bottom track) smeared along a ring of constant declination. The sources dotting the horizon are faulty power line insulators.

incident radiation, then

$$V_\nu^{ij} = \int_{\text{sky}} A_\nu(\hat{r}) I_\nu(\hat{r}) \exp\left(2\pi i \hat{r} \cdot \vec{b}_{ij}/\lambda\right) \mathrm{d}\Omega\,. \qquad (2)$$

Imaging the output of a radio interferometer therefore consists of estimating $I_\nu(\hat{r})$ given the available measurements V_ν^{ij}.

Naively one might attempt to solve equation 2 by discretizing, and subsequently solving the resulting matrix equation. If the interferometer is composed of N_{base} baselines, and measures N_{freq} frequency channels over N_{time} integrations then the entire data set consists of $N_{\text{base}} N_{\text{freq}} N_{\text{time}}$ complex numbers. If the sky is discretized into N_{pix} pixels then the relevant matrix has dimensions of $(N_{\text{base}} N_{\text{freq}} N_{\text{time}}) \times (N_{\text{pix}})$. For making single-channel maps with the Owens Valley LWA this becomes a 5 petabyte array (assuming each matrix element is a 64-bit complex floating point number). This matrix equation is therefore prohibitively large, and solving equation 2 by means of discretization is usually impossible TODO: cite Jayce's paper (has he published that yet?) where he does this, but explain why his work is the exception instead of the rule.

Instead it is common to make mild assumptions that simplify equation 2 and ease the computational burden in solving for $I_\nu(\hat{r})$. For example, when all of the baselines $\vec{b}_{ij}$ lie in a plane and the field-of-view is small, equation 2 can be well-approximated by a two-dimensional Fourier transform.

TODO: discuss w-projection

TODO: discuss a-projection

Zheng et al. (2017b)

3.1. *m-Mode Analysis*

On the other hand, transit telescopes can take advantage of a symmetry in equation 2 that greatly reduces the amount of computer time required to image the full-sky with exact incorporation of widefield imaging effects. This technique, called m-mode analysis, also obviates the need for gridding, mosaicing, and multiscale deconvolution. Instead the entire sky is imaged in one coherent synthesis imaging step.

In this context we will define a transit telescope as any interferometer where the response pattern of the individual elements does not change with respect to time. This may be an interferometer like the OVRO LWA where the correlation elements are fixed dipoles, but it may also be an interferometer like the MWA if the steerable beams are held in a fixed position (not necessarily at zenith).

We will briefly summarize m-mode analysis below, but the interested reader should consult Shaw et al. (2014, 2015) for a complete derivation.

For a transit telescope, the visibilities V_ν^{ij} are a periodic function of sidereal time[2]. Therefore it is a very natural operation to compute the Fourier transform of the visibilities with respect to sidereal time $\phi \in (0, 2\pi]$.

$$V_{m,\nu}^{ij} = \int_0^{2\pi} V_\nu^{ij}(\phi) \exp\Big(-im\phi\Big)\, \mathrm{d}\phi \qquad (3)$$

[2] This is not strictly true. Ionospheric fluctuations and non-sidereal sources (such as the sun) will violate this assumption. This paper will, however, demonstrate that the impact on the final maps is mild.

The output of this Fourier transform is the set of m-modes $V^{ij}_{m,\nu}$ where $m = 0, \pm 1, \pm 2, \ldots$ is the Fourier conjugate variable to the sidereal time. The m-mode corresponding to $m = 0$ is a simple average of the visibilities over sidereal time. Similarly $m = 1$ corresponds to the component of the visibilities that varies over half-day timescales. Larger values of m correspond to components that vary on quicker timescales.

TODO: introduce the baseline transfer function because I will want to use it in later sections

It can be shown that there is a discrete linear relationship between the m-modes $V^{ij}_{m,\nu}$ and the spherical harmonic coefficients of the sky brightness $a_{lm,\nu}$.

$$V^{ij}_{m,\nu} = \sum_l B^{ij}_{lm,\nu} a_{lm,\nu} \,, \tag{4}$$

where the transfer coefficients $B^{ij}_{lm,\nu}$ define the interferometer's response to the sky. For example, the transfer coefficients are a function of the baseline and antenna primary beam pattern.

Equation 4 can be recognized as a matrix equation where the transfer matrix $\boldsymbol{B}$ is block-diagonal.

$$\boldsymbol{B} = \begin{pmatrix} m=0 & & & \\ & m=\pm 1 & & \\ & & m=\pm 2 & \\ & & & \ddots \end{pmatrix} \tag{5}$$

The vector $\boldsymbol{v}$ contains the list of m-modes and the vector $\boldsymbol{a}$ contains the list of spherical harmonic coefficients representing the sky brightness. In order to take advantage of the block-diagonal structure in $\boldsymbol{B}$, $\boldsymbol{v}$ and $\boldsymbol{a}$ must be sorted by the value of m.

$$\overbrace{\begin{pmatrix} \vdots \\ m\text{-modes} \\ \vdots \end{pmatrix}}^{\boldsymbol{v}} = \overbrace{\begin{pmatrix} \ddots & & \\ & \text{transfer matrix} & \\ & & \ddots \end{pmatrix}}^{\boldsymbol{B}} \overbrace{\begin{pmatrix} \vdots \\ a_{lm} \\ \vdots \end{pmatrix}}^{\boldsymbol{a}} \tag{6}$$

In practice we now need to pick the set of spherical harmonics we will use to represent the sky. For an interferometer like the OVRO LWA with lots of short baselines, a sensible choice is to use all spherical harmonics with $l \le l_{\max}$ for some $l_{\max}$. The parameter $l_{\max}$ is determined by the maximum baseline length of the interferometer. For an interferometer without short spacings, a minimum value for l might also be used. This $l_{\min}$ parameter should be determined by the minimum baseline length. When creating the maps presented in this paper, we use $l_{\min} = 0$ and $l_{\max} = 1000$.

The interferometer's sensitivity to $l = 0$, however, deserves special consideration. Venumadhav et al. (2016) prove – under fairly general assumptions – that an interferometer is only sensitive to the monopole of the sky brightness if there exists some form of cross-talk or common-mode noise. In fact, the sensitivity of the interferometer is proportional to the amplitude of these effects. For consistency we will include a_{00} while solving Equation 6 for the vector $\boldsymbol{a}$, but we set $a_{00} = 0$ afterwards because we do not have a measurement of cross-talk or common-mode noise.

The size of a typical block in the transfer matrix is $(2N_{\text{base}}N_{\text{freq}}) \times (l_{\max})$. If each element of the matrix is stored as a 64-bit complex floating point number, a single block is 500 MB for the case of single-channel imaging with the OVRO LWA. Compare this number with the 5 PB required for the naive approach. The power of m-mode analysis is the block-diagonal structure of equation 6. By breaking up the problem into N independent blocks, the computational complexity involved in inverting the equation is reduced by a factor N^3. For the case of the OVRO LWA the equation breaks up into $\sim 10^3$ blocks and so we save a factor of $\sim 10^9$ in processing time by using m-mode analysis.

3.2. *m-Mode Analysis Imaging*

Imaging in m-mode analysis essentially amounts to inverting equation 6 to solve for the spherical harmonic coefficients $\boldsymbol{a}$. The linear-least squares solution, which minimizes $\|\boldsymbol{v} - \boldsymbol{B}\boldsymbol{a}\|^2$, is given by

$$\hat{\boldsymbol{a}}_{\text{LLS}} = (\boldsymbol{B}^*\boldsymbol{B})^{-1}\boldsymbol{B}^*\boldsymbol{v} \,, \tag{7}$$

where * indicates the conjugate-transpose. However, usually one will find that $\boldsymbol{B}$ is not full-rank and hence $\boldsymbol{B}^*\boldsymbol{B}$ is not an invertible matrix. For example, an interferometer located in the northern hemisphere will never see a region of the southern sky centered on the southern celestial pole. The m-modes contained in the vector $\boldsymbol{v}$ must contain no information about the sky around the southern celestial pole, and therefore the act of multiplying by $\boldsymbol{B}$ must destroy some information about the sky. The consequence of this fact is that $\boldsymbol{B}^*\boldsymbol{B}$ must have at least one zero eigenvalue, which means it is not invertible.

Another way of looking at the problem is that because the interferometer is not sensitive to part of the southern hemisphere, there are infinitely many possible solutions to equation 6 that will fit the measured data equally well. So we need to regularize the problem and apply an additional constraint that prefers a unique solution. For example, you may prefer that in the absence of any information the sky should be set to zero or you may prefer that the sky should be set to some prior expectation.

3.2.1. *Tikhonov Regularization*

The process of Tikhonov regularization minimizes $\|\boldsymbol{v} - \boldsymbol{B}\boldsymbol{a}\|^2 + \epsilon\|\boldsymbol{a}\|^2$ for some arbitrary value of $\epsilon > 0$ chosen by the observer. The solution that minimizes this expression is given by

$$\hat{\boldsymbol{a}}_{\text{Tikhonov}} = (\boldsymbol{B}^*\boldsymbol{B} + \epsilon\boldsymbol{I})^{-1}\boldsymbol{B}^*\boldsymbol{v}\,. \tag{8}$$

Tikhonov regularization adds a small value ϵ to the diagonal of $\boldsymbol{B}^*\boldsymbol{B}$, fixing the matrix's singularity. We can see this by using the singular value decomposition (SVD) of the matrix $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$. Equation 8 becomes

$$\hat{\boldsymbol{a}}_{\text{Tikhonov}} = \boldsymbol{V}(\boldsymbol{\Sigma}^2 + \epsilon\boldsymbol{I})^{-1}\boldsymbol{\Sigma}\boldsymbol{U}^*\boldsymbol{v}\,.$$

The diagonal elements of $\boldsymbol{\Sigma}$ are the singular values of $\boldsymbol{B}$ and so the contribution of each singular component to the Tikhonov-regularized solution is scaled by $\sigma/(\sigma^2 + \epsilon)$, where σ is the singular value for the given singular component. Tikhonov regularization therefore acts to suppress any component for which $\sigma \lesssim \sqrt{\epsilon}$. If $\sigma = 0$, the component is set to zero.

In practice the measurement $\boldsymbol{v}$ is corrupted by noise with covariance $\boldsymbol{N}$. For illustrative purposes we now assume that $\boldsymbol{N} = n\boldsymbol{I}$ for some scalar n. In this case the covariance of the Tikhonov-regularized spherical harmonic coefficients $\boldsymbol{C} = \langle \hat{\boldsymbol{a}}^*_{\text{Tikhonov}}\hat{\boldsymbol{a}}_{\text{Tikhonov}}\rangle - \langle\hat{\boldsymbol{a}}_{\text{Tikhonov}}\rangle\,\langle\hat{\boldsymbol{a}}^*_{\text{Tikhonov}}\rangle$ is

$$\boldsymbol{C} = n\boldsymbol{V}\boldsymbol{\Sigma}^2(\boldsymbol{\Sigma}^2 + \epsilon\boldsymbol{I})^{-2}\boldsymbol{V}^*\,. \tag{9}$$

Each singular component is scaled by a factor of $\sigma^2/(\sigma^2+\epsilon)^2$. In the absence of Tikhonov regularization ($\epsilon = 0$) singular components with the smallest singular values – the ones that the interferometer is the least sensitive to – actually come to dominate the covariance of the measured spherical harmonic coefficients. Tikhonov regularization improves this situation by downweighting these components.

While Tikhonov regularization will force unmeasured modes to zero, if a prior map of the sky already exists, it will be preferrable to instead minimize $\|\boldsymbol{v} - \boldsymbol{B}\boldsymbol{a}\|^2 + \epsilon\|\boldsymbol{a} - \boldsymbol{a}_{\text{prior}}\|^2$.

3.2.2. *L-Curves*

Tikhonov regularization requires the observer to pick the value of ϵ. If ϵ is too large then too much importance is placed on minimizing the norm of the solution and the least-squares residuals will suffer. However if ϵ is too small then the problem will be poorly regularized and the resulting sky map may not represent the true sky. Picking the value of ϵ therefore requires understanding the trade-off between the two norms.

This trade-off can be analyzed quantitatively by trialing several values of ϵ and computing $\|\boldsymbol{v}-\boldsymbol{B}\boldsymbol{a}\|^2$ and $\|\boldsymbol{a}\|^2$ for each trial. An example is shown in Figure TODO: create a figure. The shape of this curve has a characteristic L-shape, and as a result this type of plot is called an L-curve TODO: reference. The ideal value of ϵ lies near the turning point of the plot. At this point a small decrease in ϵ will lead to an undesired rapid increase in $\|\boldsymbol{a}\|^2$, and a small increase in ϵ will lead to an undesired rapid increase in $\|\boldsymbol{v} - \boldsymbol{B}\boldsymbol{a}\|^2$.

In practice, the L-curve should be used as a guide to estimate a reasonable value of ϵ. However better results can often be obtained by tuning the value of ϵ. For instance increasing the value of ϵ can improve the noise properties of the map by down-weighting noisy modes. Decreasing the value of ϵ can improve the resolution in the map by up-weighting the contribution of longer baselines, which are likely fewer in number. In this respect choosing the value of ϵ is analagous to picking the weighting scheme in traditional imaging where robust weighting schemes can be tuned to similar effect.

3.2.3. *The Moore-Penrose Pseudoinverse*

The Moore-Penrose pseudoinverse (denoted in this paper with a superscript †), is commonly applied to find the minimum-norm linear-least squares solution to a set of linear equations. This can be used in place of Tikhonov regularization as

$$\hat{\boldsymbol{a}}_{\text{Moore-Penrose}} = \boldsymbol{B}^\dagger\boldsymbol{v}\,. \tag{10}$$

Much like Tikhonov regularization, the Moore-Penrose pseudoinverse sets components with small singular values (below some user-defined threshold) to zero. Components with large singular values (above the user-defined threshold) are included in the calculation at their full amplitude with no down-weighting of modes near the threshold. The essential difference between using the Moore-Penrose pseudoinverse and Tikhonov regularization is that the pseudoinverse defines a hard transition from on to off. Modes are either set to zero or included in the map at their full amplitude. On the other hand Tikhonov regularization smoothly interpolates between these behaviors. Because of this, Tikhonov regularization tends to produce better results in practical applications.

3.3. *m-Mode Analysis with the OVRO LWA*

3.4. *CLEAN*

While m-mode analysis imaging automatically deconvolves large angular scales, point sources do not get deconvolved by the imaging process. This is a limitation of the set of basis functions we have chosen to represent the sky – the spherical harmonics. Point sources carry power to large values of l, but the regularization will eventually begin to downweight spherical harmonics with l larger than what the interferometer is sensitive to. This consequence of this is that point sources will appear in the maps with some characteristic point spread function. This point spread function may be computed with

$$\boldsymbol{a}_{\mathrm{PSF}}(\theta,\phi) = (\boldsymbol{B}^*\boldsymbol{B} + \epsilon\boldsymbol{I})^{-1}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{a}_{\mathrm{PS}}(\theta,\phi)\,, \tag{11}$$

where $\boldsymbol{a}_{\mathrm{PSF}}(\theta,\phi)$ is the vector of spherical harmonic coefficients representing the point spread function at the spherical coordinates (θ,ϕ), and $\boldsymbol{a}_{\mathrm{PS}}(\theta,\phi)$ is the vector of spherical harmonic coefficients for a point source at (θ,ϕ) given by

$$a_{lm,\mathrm{PS}}(\theta,\phi) = Y^*_{lm}(\theta,\phi)\,. \tag{12}$$

TODO: include a few images of the computed PSF for my maps

TODO: describe how we incorporate this into major and minor iterations

4. ERRORS

4.1. *Terrestrial Interference*

When writing down equation 2, it is implicitly assumed that the correlated voltage fluctuations measured between pairs of antennas are generated by astronomical sources of radio emission. In practice, this implicit assumption can be violated. For instance a low-frequency interferometer located in the vicinity of an arcing power line will see an additional contribution from the radio-frequency interference (RFI) generated by the arcing process. Similarly common-mode pickup along the analog signal path of the interferometer may generate an additional spurious contribution to the measured visibilities. While the amplitude and phase of these contaminating signals may fluctuate with time, they do not sweep across the sky at the sidereal rate characteristic of astronomical sources.

The first step in equation 8 is to compute $\boldsymbol{B}^*\boldsymbol{v}$. In this step we compute the projection of the measurement $\boldsymbol{v}$ onto the space spanned by the columns of $\boldsymbol{B}$. Each column of $\boldsymbol{B}$ describes the interferometer's response to a corresponding spherical harmonic coefficient of the sky brightness distribution. Therefore the act of computing $\boldsymbol{B}^*\boldsymbol{v}$ is to project the measured m-modes onto the space of m-modes which could be generated by astronomical sources. The degree to which a source of terrestrial interferer will contaminate a map generated using m-mode analysis imaging is determined by its amplitude after projection.

For instance, a bright interfering source might contribute $\boldsymbol{v}_{\mathrm{terrestrial}}$ to the measured m-modes. However, if $\boldsymbol{v}_{\mathrm{terrestrial}}$ is actually perpendicular to all of the columns of $\boldsymbol{B}$, there will be no contamination in the map because $\boldsymbol{B}^*\boldsymbol{v}_{\mathrm{terrestrial}} = 0$. In practice this is unlikely. In general the contamination is proportional to the overall amplitude of the interference ($\|\boldsymbol{v}_{\mathrm{terrestrial}}\|$) and the degree to which the interference mimics an astronomical signal ($\|\boldsymbol{B}^*\boldsymbol{v}_{\mathrm{terrestrial}}\|/\|\boldsymbol{v}_{\mathrm{terrestrial}}\|$).

These astronomical sources are anchored to the celestial sphere and therefore rotate across the sky over the course of a sidereal day. Terrestrial source of interference are not anchored to the celestial sphere.

This characteristic motion of astronomical sources is a fundamental component to m-mode analysis imaging

radio emission is astronomical. Astronomical sources are anchored to the celestial sphere and therefore rotate across the sky over the course of a sidereal day. However, in a practice

m-mode analysis imaging implicitly assumes that the correlation between any two pairs of antennas

However it is instructive to analyze the contaminating effect the RFI will have on m-mode analysis maps of the sky. The correlation observed on a given baseline and frequency channel for a single source of RFI is

$$V(t) = I(t)B^{\text{near-field}} \tag{13}$$

However we need a strategy to mitigate the RFI before imaging. A constant source attached to the horizon has a fringe rate of zero. The only possible astronomical origins for a source with zero fringe rate are:

1. a point source located at the north or south celestial poles, and
2. a ring of constant declination.

Therefore while snapshot images will show the RFI source on the horizon, extended integrations will show the RFI as rings of constant declination. We can take advantage of this discrepancy to separate astronomical emission from terrestrial RFI.

4.2. *Ionospheric Scintillation*

One of the key assumptions made by m-mode analysis is that the sky is completely static. We assume that the only time-dependent behavior is the rotation of the Earth, which slowly rotates the sky through the fringe patterns of the interferometer. At low frequencies the ionosphere violates this assumption. In particular, ionospheric scintillation will cause even static sources to exhibit significant variability.

The correlation observed on a given baseline for a single point source is

$$V_\nu(t_{\mathrm{sidereal}}) = I_\nu B_\nu(t_{\mathrm{sidereal}}), \tag{14}$$

where I_ν is the flux of the source at the frequency ν, and B_ν is the baseline transfer function. The transfer function is a function of the direction to the source, which is in turn a function of the sidereal time t_{sidereal}. If the source is varying, from intrinsic variability or due to scintillation, than the source flux is also a function of the time coordinate t such that

$$V_\nu(t_{\mathrm{sidereal}}) = I_\nu(t) B_\nu(t_{\mathrm{sidereal}}), \tag{15}$$

where $t_{\text{sidereal}} = (t \mod 23.9345\,\text{hours})$.

Now assume we have observed with our interferometer for a single sidereal day. In order to compute the m-modes we must Fourier transform with respect to sidereal time. In a real measurement this is a discrete Fourier transform of the observed correlation with respect to time (where the sum over time is restricted to a sidereal day). TODO: fix normalization

$$V_{\nu,m} = \sum_t V_\nu(t) e^{-imt} \tag{16}$$

Define $V^{\text{static}}_{\nu,m}$ to be the observed m-modes if the source was actually static ($I_\nu(t) \equiv I_{\nu,0}$). Then as a consequence of the Fourier convolution theorem

$$V_{\nu,m} = \sum_{m'} V^{\text{static}}_{m'} I_{\nu,m-m'}. \tag{17}$$

This will tend to scatter power between m-modes. Simulate images of point sources with this

4.3. *Ionospheric Refraction*

4.4. *Beam Errors*

5. RESULTS

5.1. *Sky Maps*

#	ν / MHz	$\Delta\nu$ / MHz
1	36.528	0.024
2	41.760	0.024
3	46.992	0.024
4	52.224	0.024
5	57.456	0.024
6	62.688	0.024
7	67.920	0.024
8	73.152	0.024

Table 1. Summary of the produced maps

5.2. *Error Maps*

TODO: Cite Abhilash, Eastwood 2016

6. DISCUSSION

REFERENCES

Ali, Z. S., Parsons, A. R., Zheng, H., et al. 2015, ApJ, 809, 61

Beardsley, A. P., Hazelton, B. J., Sullivan, I. S., et al. 2016, ApJ, 833, 102

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. 2017, SIAM Review, 59, 65

de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008, MNRAS, 388, 247

DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2016, ArXiv e-prints, arXiv:1606.07473

Dickinson, C. 2016, ArXiv e-prints, arXiv:1606.03606

Fialkov, A., Barkana, R., Pinhas, A., & Visbal, E. 2014, MNRAS, 437, L36

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181

Guzmán, A. E., May, J., Alvarez, H., & Maeda, K. 2011, A&A, 525, A138

Haslam, C. G. T., Klein, U., Salter, C. J., et al. 1981, A&A, 100, 209

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

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19

Johnson, S. G. 2008, The NLopt nonlinear-optimization package, http://ab-initio.mit.edu/nlopt

Kocz, J., Greenhill, L. J., Barsdell, B. R., et al. 2015, Journal of Astronomical Instrumentation, 4, 1550003

Lane, W. M., Cotton, W. D., Helmboldt, J. F., & Kassim, N. E. 2012, Radio Science, 47, RS0K04

Mitchell, D. A., Greenhill, L. J., Wayth, R. B., et al. 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 707

Patil, A. H., Yatawatta, S., Koopmans, L. V. E., et al. 2017, ApJ, 838, 65

Perley, R. A., & Butler, B. J. 2016, ArXiv e-prints, arXiv:1609.05940

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16

—. 2016a, A&A, 594, A13

—. 2016b, A&A, 594, A25

Pober, J. C., Parsons, A. R., Jacobs, D. C., et al. 2012, AJ, 143, 53

Pritchard, J. R., & Loeb, A. 2012, Reports on Progress in Physics, 75, 086901

Rogers, A. E. E., & Bowman, J. D. 2008, AJ, 136, 641

Rowan, T. 1990, PhD thesis, Department of Computer Sciences, University of Texas at Austin

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

Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2014, ApJ, 781, 57

Shaw, J. R., Sigurdson, K., Sitwell, M., Stebbins, A., & Pen, U.-L. 2015, PhRvD, 91, 083514

Smirnov, O. M., & Tasse, C. 2015, MNRAS, 449, 2668

Venumadhav, T., Chang, T.-C., Doré, O., & Hirata, C. M. 2016, ApJ, 826, 116

Zheng, H., Tegmark, M., Dillon, J. S., et al. 2017a, MNRAS, 464, 3486

—. 2017b, MNRAS, 465, 2901

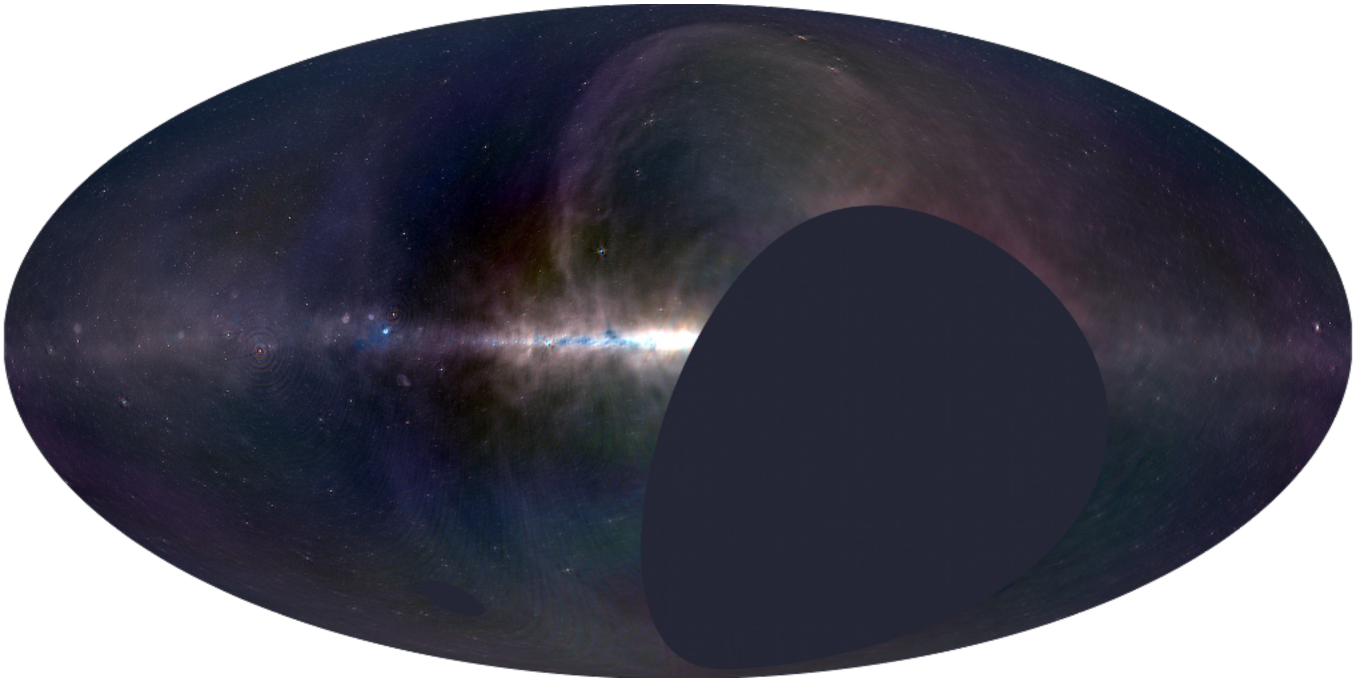

Figure 3. three color map

maps/rgb-map.png maps/rgb-map.png

maps/rgb-map.png maps/rgb-map.png

maps/rgb-map.png maps/rgb-map.png

maps/rgb-map.png maps/rgb-map.png

Figure 4. single channel maps

maps/rgb-map.png maps/rgb-map.png

maps/rgb-map.png maps/rgb-map.png

maps/rgb-map.png maps/rgb-map.png

maps/rgb-map.png maps/rgb-map.png

Figure 5. jackknifed error maps