Fast Power Spectrum Estimation from Real 21cm Data

Josh Dillon, MIT

with
Adrian Liu (Berkeley), Chris Williams (MIT),
Max Tegmark (MIT), Jackie Hewitt (MIT)

and
Ed Morgan, Al Levine, Miguel Morales, Steven Tingay, Gianni Bernardi, Judd Bowman, Frank Briggs,
David Emrich, Daniel Mitchell, Divya Oberoi, Thiagaraj Prabu, Randall Wayth, Rachel Webster
21cm Cosmology

- Probes the unknown astrophysics of reionization and the first stars and galaxies
- 21cm tomography can eventually become the most precise test of cosmological models

Marcelo Alvarez, Ralf Kaehler, Tom Abel
The first detection will be statistical.

Barkana (2009), Morales & Wyithe (2010)
In Practice...

Cosmological Signal
...the cosmological signal is very dim.
And the contaminants are very bright.

- Instrumental Noise
- Bright Point Sources
- Galactic Synchrotron
- Unresolved Point Sources

Cosmological Signal
In an ideal world, there’s an optimal estimator...

\[ \hat{p}^\alpha \equiv M^{\alpha \beta} x^T C^{-1} Q^\alpha C^{-1} x \]

**Invertible Normalization Matrix**

**Inverse Covariance Weighting**

**Quadratic Power Spectrum Estimator**

**Data**

**Fourier Transform and Bin**

*Liu & Tegmark (2011)*
...with well-understood error properties.

\[ \text{Cov}(\hat{p}) = MFM^\top \]

Contains all the errors and error covariances

Fisher Information calculated from the covariance models:

\[ F^{\alpha\beta} = \frac{1}{2} \text{tr} \left[ C^{-1} Q^\alpha C^{-1} Q^\beta \right] \]

Liu & Tegmark (2011)
Sounds hard. Why bother?
The Signal
The Signal

Just like the cosmological signal, there’s power on many scales.
Blue noise represents noise for high angular frequency noise.
The Foregrounds

Low (spectral) frequency foregrounds dominate over signal and noise.
Naïve Filtering

Does an OK job, but throws out lots of information.
Inverse Variance Weighting

Preserves as much information as possible.
Yet despite all that,

It’s hard to measure $P(k)$.
It’s hard to measure $P(k)$.

- Large data volume

$$F^{\alpha \beta} = \frac{1}{2} \text{tr} \left[ C^{-1} Q^\alpha C^{-1} Q^\beta \right]$$
It’s hard to measure $P(k)$.

- Large data volume
- Uncertain foreground parameters
It’s hard to measure $P(k)$.

- Large data volume
- Uncertain foreground parameters
- Incomplete $uv$ coverage
It’s hard to measure \( P(k) \).

- Large data volume
- Uncertain foreground parameters
- Incomplete \( uv \) coverage
- Radio frequency interference
It’s hard to measure $P(k)$

- Large data volume
- Uncertain foreground parameters
- Incomplete $uv$ coverage
- Radio frequency interference
- Foreground leakage into the EoR window
It’s hard to measure $P(k)$.

- Large data volume
- Uncertain foreground parameters
- Incomplete $uv$ coverage
- Radio frequency interference
- Foreground leakage into the EoR window
- Optimal binning to spherical $P(k)$
It's hard to measure $P(k)$.

- Large data volume
- Uncertain foreground parameters
- Incomplete $uv$ coverage
- Radio frequency interference
- Foreground leakage into the EoR window
- Optimal binning to spherical $P(k)$

For the rest, see Dillon, Liu, et. al. (2013)
This Data

- Taken in March 2010 with the MWA-32T prototype array
- Approximately 3-5 hours of observation per frequency band
- (For more, see Williams et al. (2012) or Chris Williams’s MIT thesis.)
Fast Power Spectrum Estimation

1. Generate lots of random data cubes from the model covariance, exploiting symmetries

All in $O(N\log N)^*$
Fast Power Spectrum Estimation

1. Generate lots of random data cubes from the model covariance, exploiting symmetries

2. Calculate our quadratic estimator

\[ q^\alpha \equiv x^\top C^{-1} Q^\alpha C^{-1} x \]

All in \( O(N\log N) \)*
Fast Power Spectrum Estimation

1. Generate lots of random data cubes from the model covariance, exploiting symmetries

2. Calculate our quadratic estimator

\[ q^\alpha \equiv x^\top C^{-1} Q^\alpha C^{-1} x \]

- Approximate the Q matrix with a padded FFT

All in \( O(N\log N) \)*
Fast Power Spectrum Estimation

1. Generate lots of random data cubes from the model covariance, exploiting symmetries

2. Calculate our quadratic estimator

\[ q^\alpha \equiv x^T C^{-1} Q^\alpha C^{-1} x \]

- Approximate the Q matrix with a padded FFT
- Use the conjugate gradient method to calculate \( C^{-1} x \)

All in \( O(N \log N) \)*
Fast Power Spectrum Estimation

1. Generate lots of random data cubes from the corrupted model covariance, exploiting symmetries.

2. Calculate our quadratic estimator

\[
\mathbf{x}^\top \mathbf{C}^{-1} \mathbf{Q}^{-1} \mathbf{x} \approx \tilde{N}_{lm} = \frac{\lambda^2 T^2_{\text{sky}} \delta_{ij} \psi_0 (k_x \Delta x/2) \psi_0 (k_y \Delta y/2) \delta_{lm}}{A^2_{\text{ant}} (\Omega_{\text{pix}})^2 n_x n_y \Delta \nu} \frac{t_l}{t_i}. \quad (B10)
\]

\[
\mathbf{C}_v = \mathbf{F}^\top \text{diag}(\mathbf{F}c) \mathbf{F}_v. \quad (A1)
\]

\[
\ln \left( \frac{\eta_i}{\eta_j} \right) = \ln \left( \frac{\nu_i}{\nu_j} \right) \approx \frac{\nu_0 + \Delta \nu_i}{\nu_0 + \Delta \nu_j} - 1 = \frac{1}{\nu_0} (\Delta \nu_i - \Delta \nu_j). \quad (53)
\]

\[
N_{ij} = \frac{\lambda^2 T^2_{\text{sky}}}{A^2_{\text{ant}} (\Omega_{\text{pix}})^2} \sum_{x, y} \sum_{m} \sum_{l} e^{i k_x x_j} e^{i k_y y_j} e^{-i k_x x_i} e^{-i k_y y_i} j_0^2 (k_x \Delta x/2) j_0^2 (k_y \Delta y/2) \delta_{lm} \frac{t_l}{t_i (\lambda_i)}. \quad (B9)
\]

\[
N_{ij} = \frac{\delta_{ij}}{\Delta \nu} \frac{\lambda^2 T^2_{\text{sky}} \psi_0^2}{A^2_{\text{ant}} \Omega_{\text{pix}} n_x n_y} \int j_0^2 (k_x \Delta x/2) j_0^2 (k_y \Delta y/2) \times \frac{1}{i (k_{\perp}, \lambda_i)} \frac{dk_x dk_y}{(2\pi)^2} \quad (B7)
\]

All in \(O(N\log N)^*\)
Fast Power Spectrum Estimation

All in \(O(N \log N)\)*

Approximate the \(Q\) matrix with a padded FFT

\[ P_N(U + N)P = 1 + N \]
\[ = 1 + \tilde{N} \]
\[ = 1 + \tilde{\Upsilon} \]

(C2)

Since applying \(P_N\) only requires multiplying by the inverse square root of a diagonal matrix and Fourier transforming in two dimensions, the complexity of applying \(P_N\) to a vector is less than \(O(N \log N)\).

2. Constructing a Preconditioner for \(U\)

The matrix \(U\) (Equation 55) can be written as the tensor product of three Toeplitz matrices, one for each dimension, bookended by two diagonal matrices, \(D_U\). Furthermore, since \(D_U\) depends only on frequency (as we saw in Section III D.2), its effect can be folded into \(U_c\) such that

\[
D_U [U_c \otimes U_c \otimes U_c] D_U = U_c \otimes U_c \otimes U_c. 
\]

(C3)

It is generally the case that the number of columns is comparable with \(N\). This assumption fairly compact arrangement be optimal for 2D resolution on the table, if we describe the data sources. For the only, we can there

\[
\text{matrix that looks like } I + \tilde{\Upsilon} \text{ we can make it look like } I.
\]

So can take \(I + \tilde{\Upsilon} + \tilde{\Gamma}\), where \(\Gamma = P_N P_P P_N\), and turn it into \(I + \tilde{\Upsilon} + \tilde{\Gamma}\). Looking at \(\Gamma\),

\[
\Gamma = N^{-1/2} \hat{F} \lambda_c \hat{v} \hat{v}^T N^{-1/2} = \lambda_c \hat{N}^{-1/2} \hat{v} \hat{v}^T \hat{N}^{-1/2} \otimes \hat{v} \hat{v}^T,
\]

where \(\lambda_c\) is the singular value that we are considering and where \(\hat{v} = F \hat{v}\).

Again, we will look at a preconditioner of the form:

\[
P_T = I - \beta \Pi
\]

where

\[
\Pi = \left( \hat{N}^{-1/2} \hat{v} \hat{v}^T \hat{N}^{-1/2} \right) \otimes \hat{v} \hat{v}^T.
\]

(C21)

This time, the \(\hat{N}^{-1/2}\) matrices do pass through the eigenvectors to cancel another one out. We now exploit the spectral similarity of foregrounds and the fact that \(\hat{v} \hat{v}^T = \hat{v} \hat{v}^T \otimes \hat{v} \hat{v}^T = 1\) to obtain

\[
P_T D_U P_T^T = \tilde{\Upsilon} + \lambda_c \hat{v} \hat{v}^T (\beta^2 - 2 \beta \tilde{\Gamma}).
\]

(C22)

This is very useful because it means that if we pick \(\beta\) properly, we can get one of the two changes that cancel the \(\tilde{\Gamma}\) terms we expect when we calculate the full effect of \(P_T\) and \(P_N\) on \(U_c + \hat{U}\). Noting that the real eigenvalue of \(\tilde{\Gamma}\) is \(\lambda_c = \lambda_c \hat{N}^{-1/2} \hat{v} \hat{v}^T \hat{N}^{-1/2}\), we define \(\lambda_c = \lambda_c \hat{v} \hat{v}^T \hat{N}^{-1/2} \hat{v} \hat{v}^T\). Multiplying our preconditioner by our matrices, we find that the equality of the single eigenvalues yields another quadratic equation for \(\beta\):

\[
1 + \lambda_c = 1 - 2 \beta + \beta^2 + (\beta^2 - 2 \beta + 1) \tilde{\Gamma} + \lambda_c \frac{\lambda_c}{\tilde{\Gamma}} (\beta^2 - 2 \beta).
\]

(C23)

Solving the quadratic equation, we get

\[ P_U = I - \sum_k \left( 1 - \sqrt{1 + \lambda_k} \right) \tilde{v}_k \tilde{v}_k^T \otimes \tilde{v}_k \tilde{v}_k^T \]

(C12)

where the pair of \(\epsilon^\pm\) matrices pick out a particular \(\epsilon^\pm\) cell. If we want to generalize to more eigenvectors of \(U_c\), we simply need to keep subtracting off sums of matrices on the right-hand side of Equation (C12):

\[ P_U = I - \sum_k \left( 1 - \sqrt{1 + \lambda_k} \right) \tilde{v}_k \tilde{v}_k^T \otimes \tilde{v}_k \tilde{v}_k^T, \]

(C13)

This works because every set of vectors corresponding to a value of \(k\) is orthogonal to every other set. Each term in the above sum acts on a different subspace of \(C\), independent of all the other terms in the sum. If the relevant matrices are orthonormal, it

\[
P_U = I - \sum_k \left( 1 - \sqrt{1 + \lambda_k} \right) \tilde{v}_k \tilde{v}_k^T \otimes \tilde{v}_k \tilde{v}_k^T,
\]

(C24)

Finally, generalizing to multiple eigenvalues and taking advantage of the orthornormality of the eigenvectors, we have

\[
P_T = I - \sum_k \left( 1 - \sqrt{1 + \lambda_k} \right) \tilde{v}_k \tilde{v}_k^T \otimes \tilde{v}_k \tilde{v}_k^T.
\]

(C25)

The result of this somewhat complicated preconditioner is a reduction of the condition number of the matrix to be inverted by many orders of magnitude (see Figure 5).

Lastly, we include Fourier transforms at the front and the back of the preconditioner, so that the result when multiplied by a real vector, returns a real vector. Therefore, the total preconditioner we use for \(C\) is

\[
F^T P_T D^T P_T F (R + U + N + G) P^T P_T^T P^T F.
\]

(C26)

which can be interpreted as a set of matrices describing spectral coherence, each localized to one point source, and all of which are spatially uncorrelated. And likewise, we can write down \(G\) as:

\[
G = \sum_{\lambda \neq 0} \left( \lambda_c \tilde{v}_k \tilde{v}_k^T \otimes \tilde{v}_k \tilde{v}_k^T \right).
\]

(C15)

We now make two key approximations for the purposes of preconditioning. First, we assume that all the \(\lambda_c\) eigenvalues are the same, so \(\tilde{v}_k \approx \tilde{v}_k\) for all \(k\), all of which are also taken to be the same as the eigenvectors that appear in the preconditioner for \(U\) in Equation (C13). Second, as in Section C.2, we are only interested in acting upon the largest eigenvalues of \(R\) and \(G\). To this end, we will ultimately only consider the largest values of \(\lambda_c \tilde{v}_k \tilde{v}_k^T \otimes \tilde{v}_k \tilde{v}_k^T\) which will vastly reduce the computational complexity of the preconditioner.

Our strategy for overcoming the difficulty of the different bases is to simply add the two perpendicular parts of the matrices and then decompose the sum into its eigenvalues and eigenvectors. We therefore define

\[
\Gamma = R + G
\]

(C16)

(choosing the symbol \(\Gamma\) because it looks like \(R\) and \(G\)). Given the above approximations, we can reexpress \(\Gamma\) as follows:

\[
\Gamma \approx \sum_k \left( \Gamma_{\lambda_c \tilde{v}_k \tilde{v}_k^T} \otimes \tilde{v}_k \tilde{v}_k^T \right)
\]

(C17)

where we have defined each \(\Gamma_{\lambda_c \tilde{v}_k \tilde{v}_k^T}\) as:

\[
\Gamma_{\lambda_c \tilde{v}_k \tilde{v}_k^T} = \sum_{\lambda \neq 0} \left( \lambda_c \tilde{v}_k \tilde{v}_k^T \otimes \tilde{v}_k \tilde{v}_k^T \right).
\]

(C18)

Due to the high spectral coherence of the foregrounds, only a few values of \(k\) need to be included to precondition for \(\Gamma\). Considering the limit on angular box size imposed by the flat sky approximation and the limit on angular resolution imposed by the array size, this should require at most a few eigenvalue decompositions of matrices no bigger than about \(10^4\) entries on a side. Moreover, those eigenvalue decompositions need only be computed once and then only partially stored for future use. In practice, this is not a rate limiting step, as we see in Section III E.2.

We now write down the eigenvalue decomposition of \(\Gamma\):

\[
\Gamma = \sum_k \left( \sum_{\lambda \neq 0} \lambda \tilde{v}_k \tilde{v}_k^T \right) \tilde{v}_k \tilde{v}_k^T.
\]

(C19)

Before we attack the general case, we assume that only one value of \(\lambda_{\epsilon^\pm}\) is worth preconditioning—we generalize to the full \(P_T\) later. We now know that if we have a
1. Generate lots of random data cubes from the model covariance, exploiting symmetries

2. Calculate our quadratic estimator

\[ q^{\alpha} \equiv x^T C^{-1} Q^{\alpha} C^{-1} x \]

- Approximate the Q matrix with a padded FFT
- Use the conjugate gradient method to calculate \( C^{-1} x \)
- Develop covariance matrices that can be multiplied by quickly
- Develop a preconditioner for the CGM

All in \( O(N \log N) \)*
Fast Power Spectrum Estimation

1. Generate lots of random data cubes from the model covariance, exploiting symmetries

2. Calculate our quadratic estimator

3. Monte Carlo many quadratic estimators to get error bars and window functions.

Use $\text{Cov}(q) = F$

To avoid $F^{\alpha \beta} = \frac{1}{2} \text{tr} \left[ C^{-1} Q^\alpha C^{-1} Q^\beta \right]$

All in $O(N \log N)^*$
It works as fast as advertised.

Dillon, Liu, & Tegmark (2013)
We want to preserve the EoR Window, but...

What is the EoR Window?
What is the EoR Window?
What is the EoR Window?

Frequency Resolution

$k_\parallel$ (cMpc$^{-1}$)  $k_\perp$ (cMpc$^{-1}$)
What is the EoR Window?

[Diagram showing a grid with axes labeled $k_{||}$ (cMpc$^{-1}$) and $k_\perp$ (cMpc$^{-1}$), with the color intensity indicating frequency resolution and angular extent.]
What is the EoR Window?

Frequency Resolution

Angular Extent

No Baselines
What is the EoR Window?

- Angular Extent
- Frequency Resolution
- Foregrounds
- No Baselines
What is the EoR Window?

- **Angular Extent**
- **Frequency Resolution**
- **Foregrounds**
- **No Baselines**

The diagram illustrates the relationship between angular extent and frequency resolution, highlighting the "Wedge" region where foregrounds and no baselines are present.
What is the EoR Window?

- Frequency Resolution
- Angular Extent
- Foregrounds
- No Baselines

"Wedge"
The wedge is the imprint of the chromaticcility of the synthesized beam.

\[ k_\parallel = \begin{bmatrix} \sin \theta_{\text{field}} & \frac{D_M(z)E(z)}{D_H(1 + z)} \end{bmatrix} k_\perp \]
The wedge is the imprint of the chromaticity of the synthesized beam.

\[ k_{||} = \left[ \sin \theta_{\text{field}} \frac{D_M(z)E(z)}{D_H(1 + z)} \right] k_{\perp} \]
Recall...

\[ \hat{p}^\alpha \equiv M^{\alpha\beta} x^\top C^{-1} Q^\alpha C^{-1} x \]

and

\[ \text{Cov}(\hat{p}) = M F M^\top \]

- Smallest errors, but errors are correlated
- Hard to cut out foregrounds
- Decorrelated errors.
- Each band power represents a mutually exclusive yet collectively exhaustive piece of information.
A good estimator preserves the EoR Window.

\[ M \sim I \quad M \sim F^{-1/2} \]
A good estimator preserves the EoR Window.

\[ M \sim I \quad M \sim F^{-1/2} \]
Results:
The wedge evolves with frequency in just the way we expected.
Results:

We set power spectrum limits across a wide range of scales and redshifts.
Results:

We set power spectrum limits across a wide range of scales and redshifts.

Best Limit So Far:

\[ \Delta(k) \leq 0.26 \, \text{K} \]

at 95% confidence
What’s next?

• The power spectrum estimation methods are fast, rigorous, and generally applicable.

• The same tool that estimates power spectra can also forecast sensitivity for future experiments.

• Will become increasingly important when, with larger volumes of data and more sensitivity (e.g. with HERA or The Omniscope) we start to narrow in on a detection.