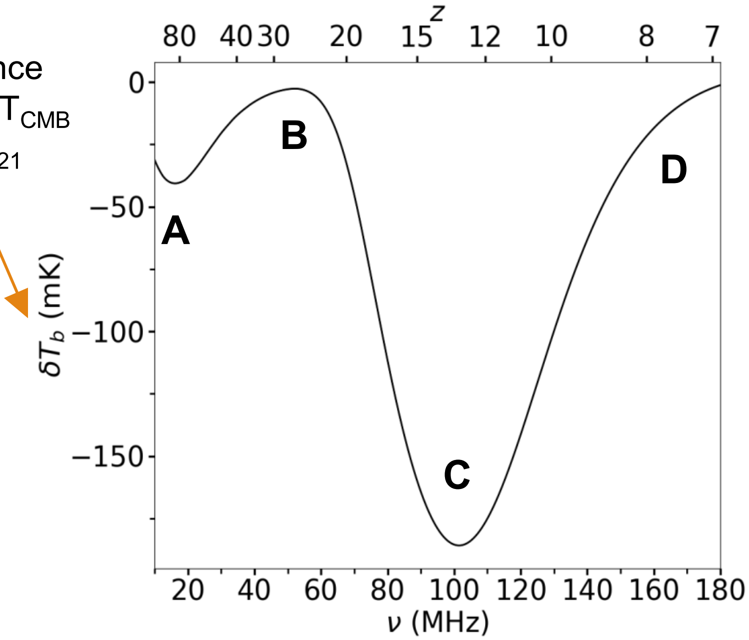
Separating the global 21-cm signal from strong foregrounds and instrument systematics using an SVD/MCMC pipeline

KEITH TAUSCHER*, DAVID RAPETTI,
JACK O. BURNS, ERIC R. SWITZER

Why pursue the global 21-cm signal?

- Interaction of excitation temperature, T_S, of HI's 21-cm transition with radiation fields produces signal which opens up the first billion years after recombination to new inquiry

Difference between T_{CMB} and T_{21}



A: Collisions between H atoms (as measured by the kinetic temperature of the gas) couple less to T_S as Universe expands: $T_S \rightarrow T_{CMB}$

B: First stars ignite, inducing a coupling between their Lyman-α radiation and T_S

C: Heating sources, including the first black holes, push cold gas temperature toward T_{CMB}

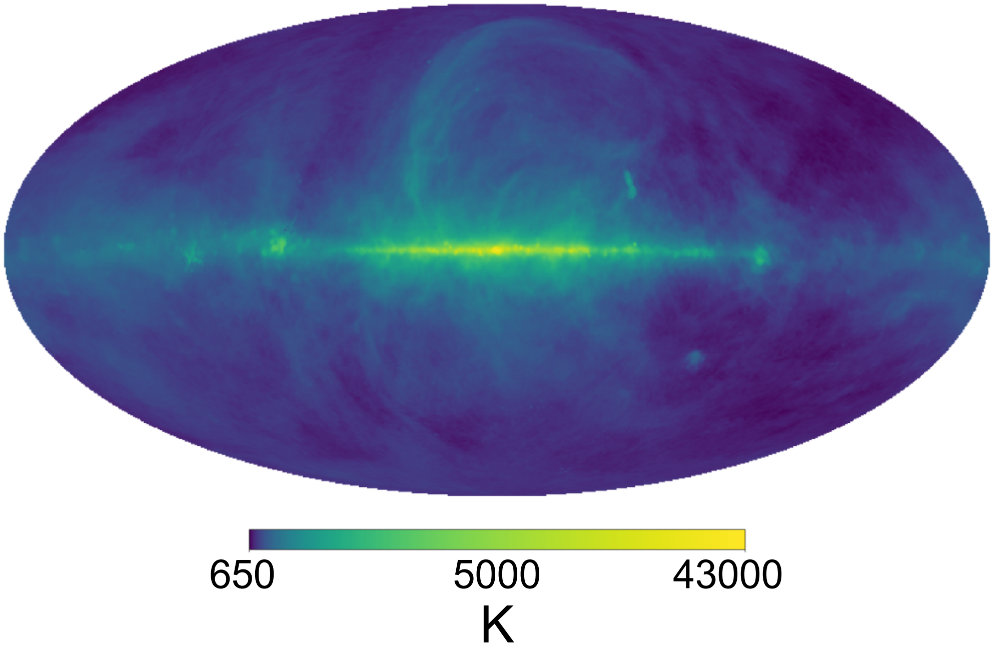
D: Reionization drives signal to zero since only neutral hydrogen emits 21-cm radiation

Signals generated with Jordan Mirocha's Accelerated Reionization Era Simulations (ares) code available at https://bitbucket.org/mirochaj/ares

Why hasn’t the 21-cm signal been measured yet?

- Strong galactic foregrounds:
  - Combine with beam chromaticity to produce complicated spectral structure
  - Mix with radiometer systematics through uncertainties in calibration parameters
- Foreground model must be observation-dependent to take advantage of data’s structure and generate significant results

Galaxy map from Haslam et al. (1982) scaled to 80 MHz



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

Our solution: induced structure and training sets

- To differentiate the signal from foreground emission and other systematic effects, one can induce structure in the data, e.g. through:
  - Introduction of channels which show systematic effects but no signal (e.g. Stokes parameters)
  - Modulation of systematic effects through some experimental mechanism (e.g. rotation)

- To take advantage of this structure, for each effect, use Singular Value Decomposition (SVD) on simulated training sets to create an ordered set of custom basis functions with which to fit each effect
  - SVD equivalent to Principal Component Analysis (PCA) and EigenValue Decomposition (EVD)

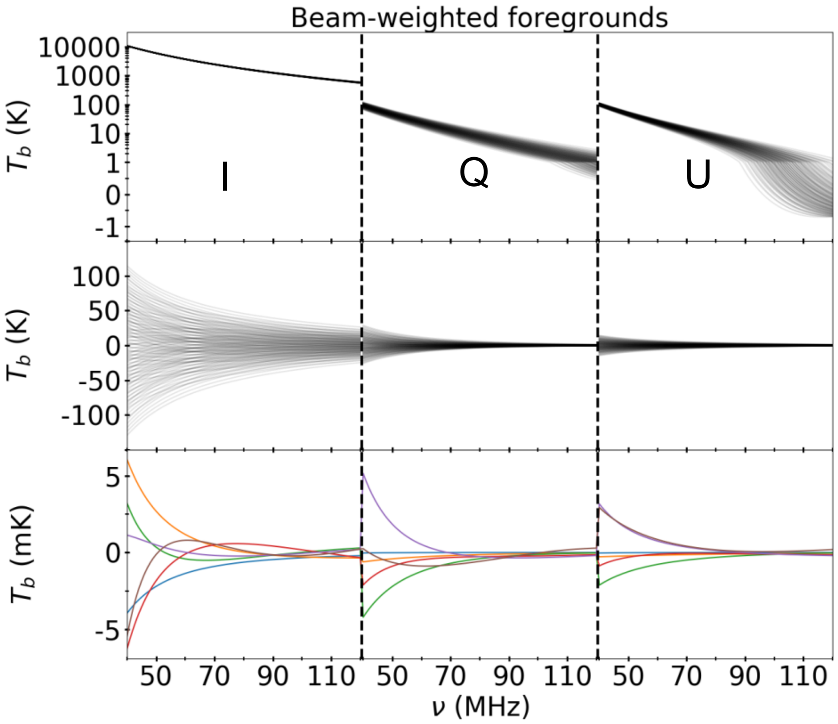
Training set

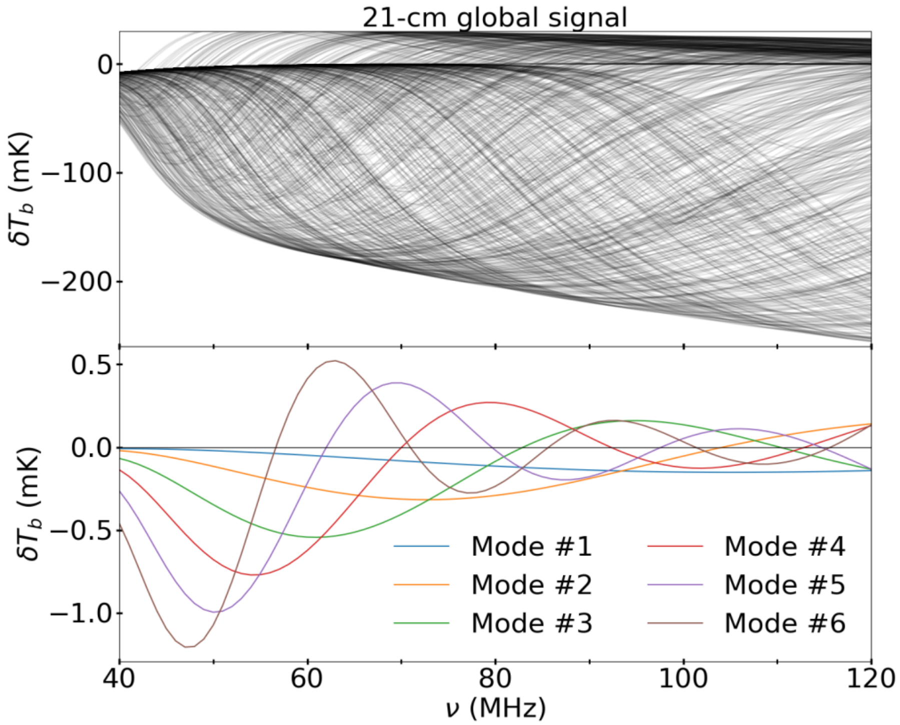
$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{T}$$

Basis vectors
(orthonormal)

Mode importances
(diagonal)

Sample training sets and SVD basis functions





From paper accepted to ApJ: arxiv:1711.03173

From eigenmodes to signal constraints

- Once modes are chosen, the posterior distribution of the SVD coefficients parameters must be computed using a likelihood function through:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \mathcal{L}(\boldsymbol{y}|\boldsymbol{x})$$

- If the likelihood is Gaussian and the model is linear, the posterior is also Gaussian and there are analytical expressions for its mean and covariance

- If the likelihood is not Gaussian or the model is nonlinear, some other technique for exploring this distribution must be used
  - Since physical signal models are often slow to compute, by comparison, the SVD models have the potential to speed up methods such as MCMC sampling

Model selection

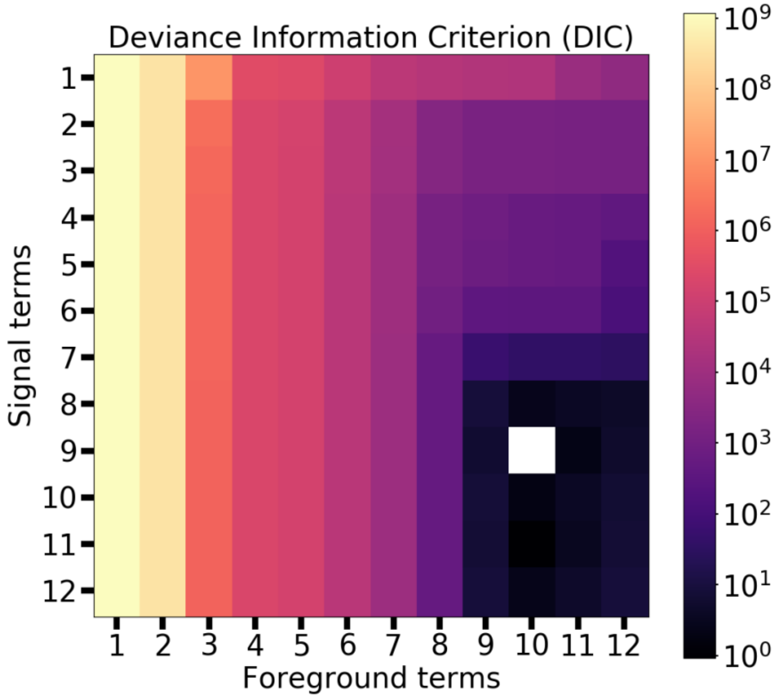
- One must choose a number of modes to use for each set of basis vectors

- Minimizing the Deviance Information Criterion (DIC) is our most consistent way of yielding unbiased fits

$$\mathrm{DIC} = -2\ln\mathcal{L}_{\max} + 2p$$

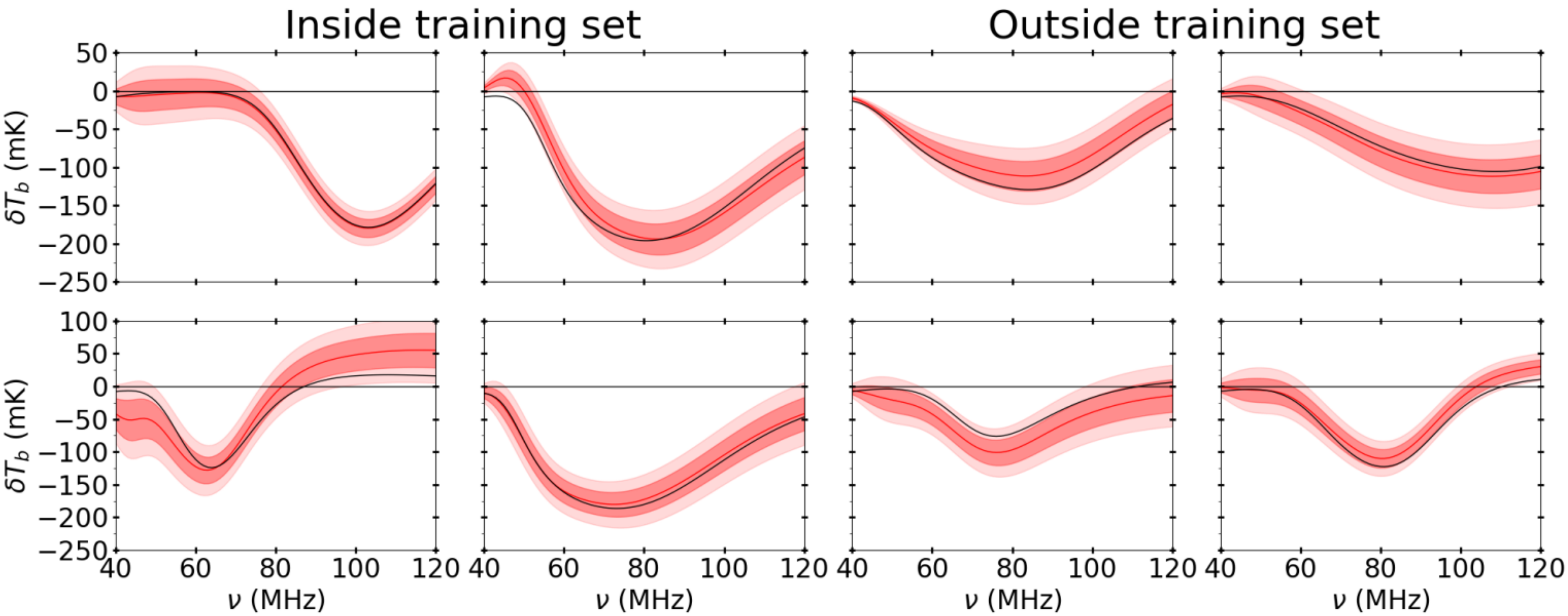
$\mathcal{L}_{\max}$: Maximum likelihood

p: Total number of parameters



From paper accepted to ApJ: arxiv:1711.03173

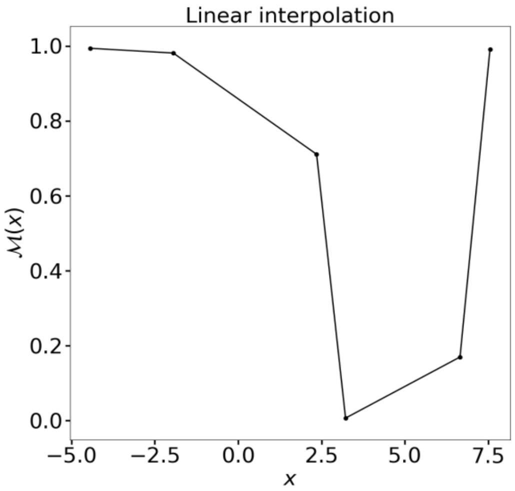
68% and 95% confidence signal fits

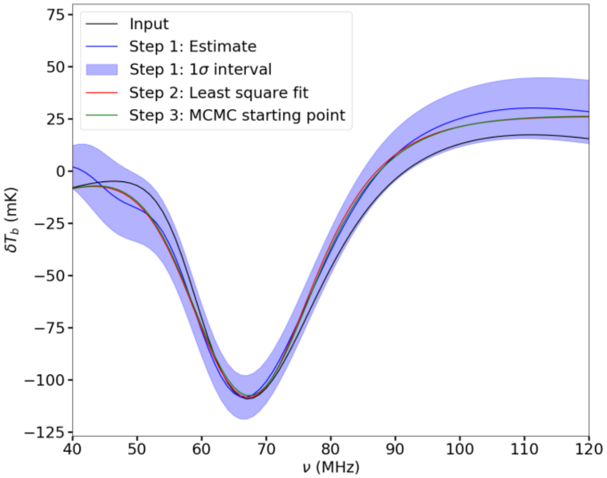


From paper accepted to ApJ: arxiv:1711.03173

Moving into physical parameter space

- While the process described yields fits to the signal in frequency space, more is needed to transform that fit into physical parameter space
- Our solution is to generalize linear interpolation to arbitrary input and output dimension and use it to perform a least square fit by interpolating between training set curves





Next steps and conclusions

- With estimate of physical parameters governing signal, we will initialize a Markov Chain Monte Carlo (MCMC) sampler to map out the probability distribution in this space.
  - Without such a starting point, the uncertainties in 21-cm signal parameter space are so large that an MCMC sampler may wander endlessly searching for the maximum likelihood region
  - This yields joint confidence intervals on all of the parameters as well as on any quantities derived from them.

- The pylinex code transforms the problem of extracting a signal from data into one of providing accurate and sufficiently dense training sets for each data component (signal+systematic effects)

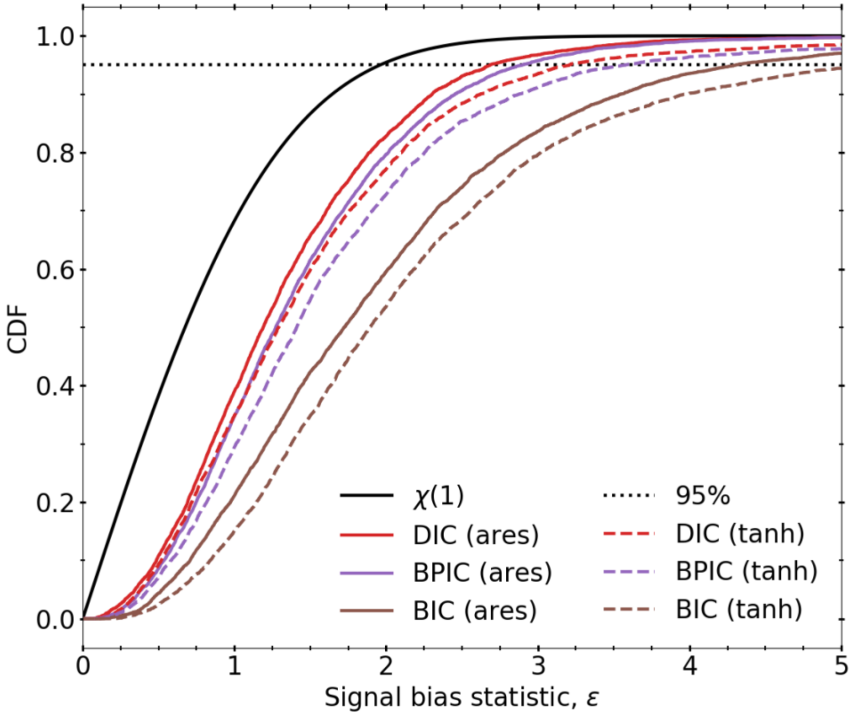
- Using pylinex to compute the global 21-cm signal from simulated sky-averaged data for all 4 Stokes parameters at various rotation angles, we extract it within 30 mK to 95% confidence 95% of the time

Backup slides

Signal bias statistic

- The signal bias statistic is a measure of the root mean square error weighted bias of the signal fit

$$\varepsilon_{21\text{-cm}} = \sqrt{\frac{\boldsymbol{\delta}_{21\text{-cm}}^T \boldsymbol{C}^{-1} \boldsymbol{\delta}_{21\text{-cm}}}{N_\nu}}$$

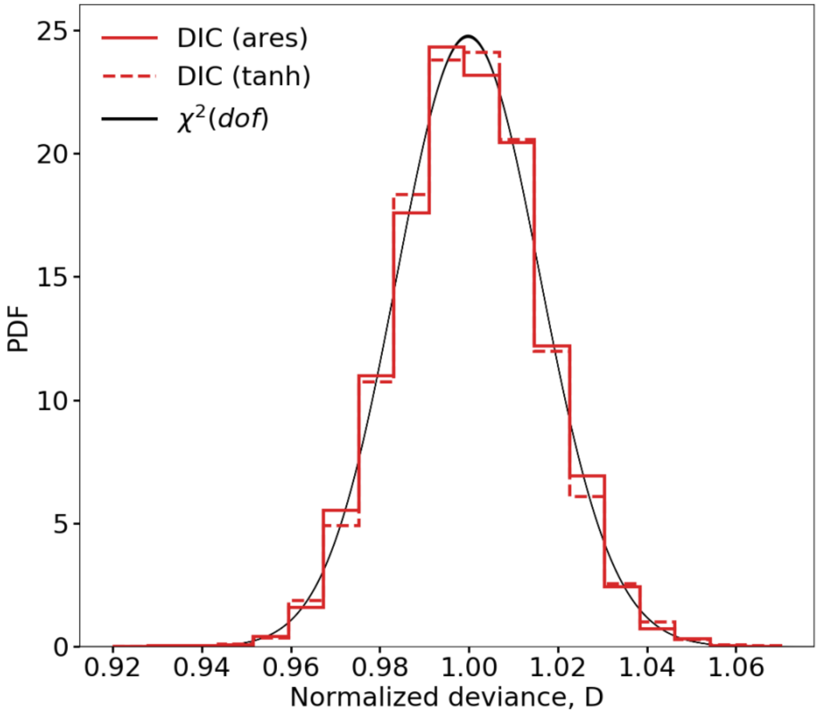


From paper accepted to ApJ: arxiv:1711.03173

Normalized Deviance

- The deviance normalized by the degrees of freedom contains information about how well the training sets fit the data

$$D = \frac{\boldsymbol{\delta}^T \boldsymbol{C}^{-1} \boldsymbol{\delta}}{N_{\rm dof}}$$



From paper accepted to ApJ: arxiv:1711.03173