University of Colorado Boulder

Moving away from analytical a priori foreground models in signal extraction

Keith Tauscher*, David Rapetti, Jack Burns, & Neil Bassett

*Keith.Tauscher@colorado.edu

Main problems of signal shape extraction

1. Measurement effects: simple spectral shapes of individual foreground sources are distorted by antenna beams
   a. For a deterministic (pre-fitting) correction of the data for beam-averaging to succeed, the beam and foreground must be known to the -50 dB level or better

2. Must be statistically limited: high confidence extraction requires every $\gtrsim$1 mK effect be modeled
   a. Specific model⇔effect relationships are preferred over generic models
   b. Generic models must be shown to accurately describe each effect they purport to model

Traditional approach to measurement (EDGES)

- EDGES attempts to calibrate out beam chromaticity from single spectra and then fit a polynomial or polynomial-like model simultaneously with a chosen signal model.

$$BCF(\nu) = \frac{\int_0^{\pi}\int_0^{2\pi} B(\nu,\theta,\phi)\; T_{\mathrm{FG}}(\nu,\theta,\phi)\;\sin\theta\; d\phi\; d\theta}{\int_0^{\pi}\int_0^{2\pi} B(\nu_n,\theta,\phi)\; T_{\mathrm{FG}}(\nu,\theta,\phi)\;\sin\theta\; d\phi\; d\theta}$$

Beam chromaticity factor

- The beam corrections rely on correct sky model temperatures and simulations of the antenna beam.

Model of beam weighted foreground

$$M_{\mathrm{BWFG}}(\nu) = \sum_{k=0}^{N-1} a_k\; \nu^{-2.5+k}$$

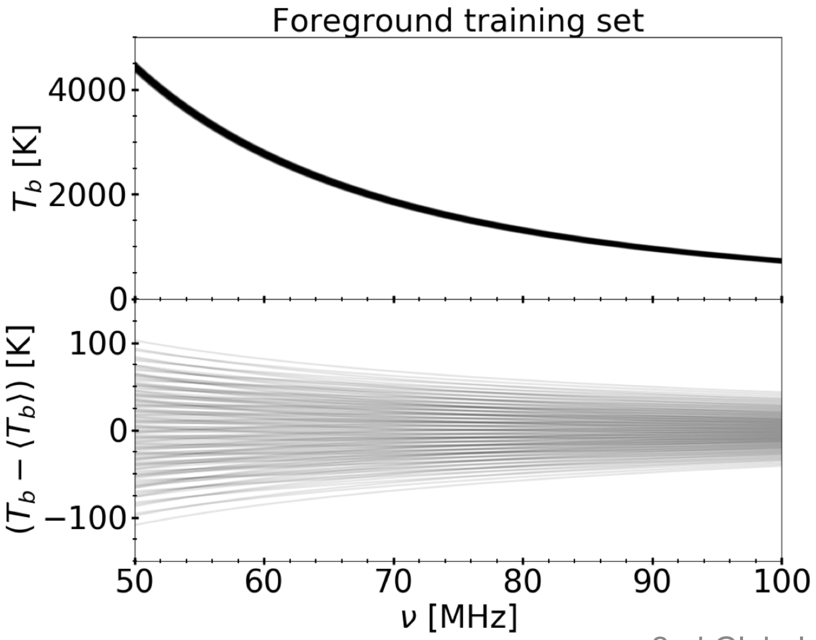
Approach of our new method, SVD/MCMC

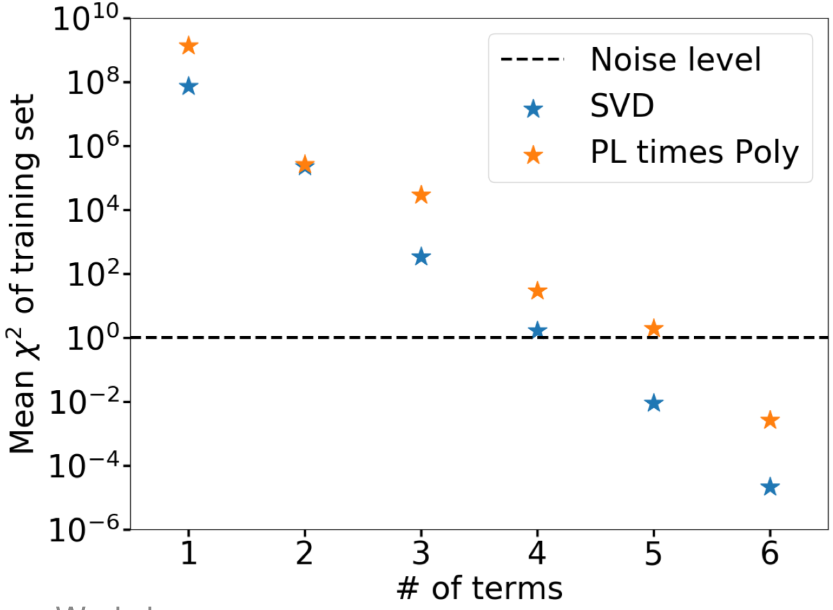
- Use a large number of spectra at once.
- **Create models specifically suited for a given dataset by simulating training sets for its spectra instead of relying on an a priori model.**
  - Training sets can be arbitrarily complex, removing necessity for perfectly smooth foregrounds.
  - Instead of trying to remove beam effects, the beam effects are included in the model itself.
- Generalize so that any simulable effect can be included in such a way that only simulations need to change, without requiring change to any free parameters.
- Develop well documented, widely applicable, easy-to-use, open-source code implementing these processes so anyone can use them: https://bitbucket.org/ktausch/pylinex

SVD/ MCMC = Big Data + Simulations

Using SVD to generate optimal basis vectors

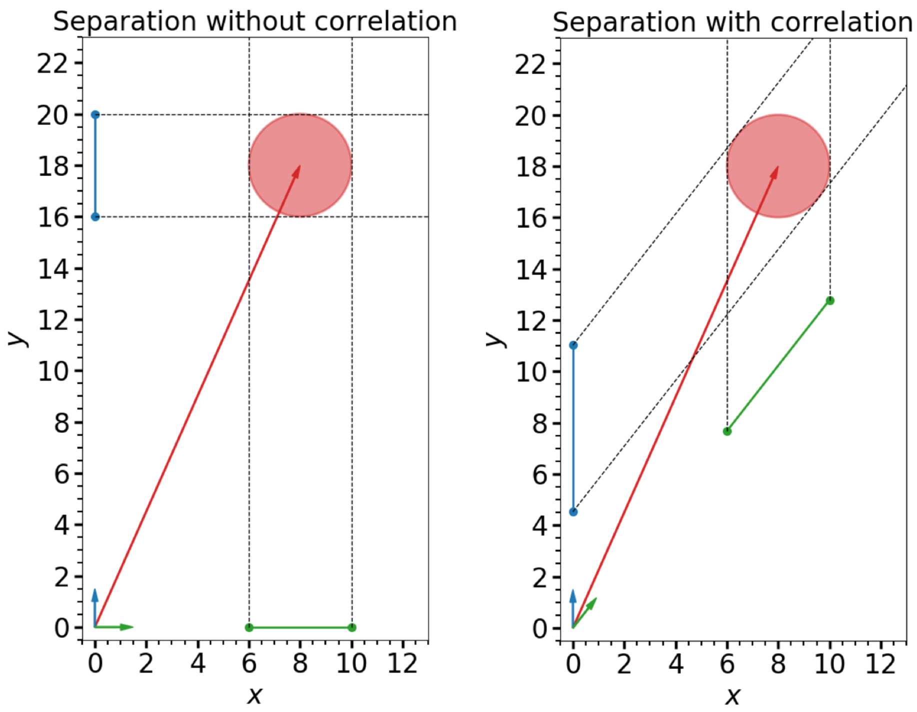
- Singular Value Decomposition (SVD) is a factorization of a training set that provides the optimal basis vectors with which to fit that training set.





Simultaneously fitting foreground and signal

- Once we have SVD models for both foreground and signal, we fit them simultaneously to the data.
- The uncertainties in this separation of the two components depend on how similar their models are.
  - **If the foreground and signal training sets are different enough and the experiment is designed well enough, then the signal can be constrained rigorously.**



Drift-scan measurements

- Using drift-scan measurements is an aspect of experimental design that can lower the overlap/similarity between foreground and signal.

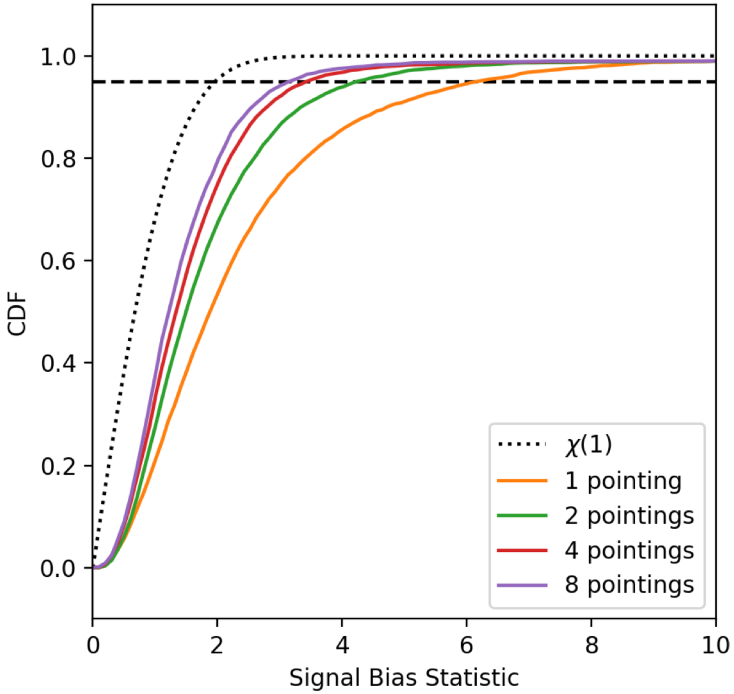
- Time introduces valuable structure into sky-averaged data because the antenna beam points at different points in the sky at different times.

- Different spectra have the same signal but different, **yet correlated**, foregrounds.

Effect of multiple pointings

- Pointing the antenna at multiple independent directions is a discrete form of drift-scan.

- The forecasted errors on signal confidence intervals lead to less bias when simulating data with multiple antenna pointings instead of one.



Other benefits of SVD/MCMC generalization

- Since our pipeline does not require a foreground model to be given beforehand (only training sets), **data aspects with no obvious extension in the polynomial-like approach can be utilized**.
- One of these effects is polarization, which is measured in terms of the Stokes parameters I, Q, U, and V.

Single antenna experiments (e.g. EDGES, SARAS)

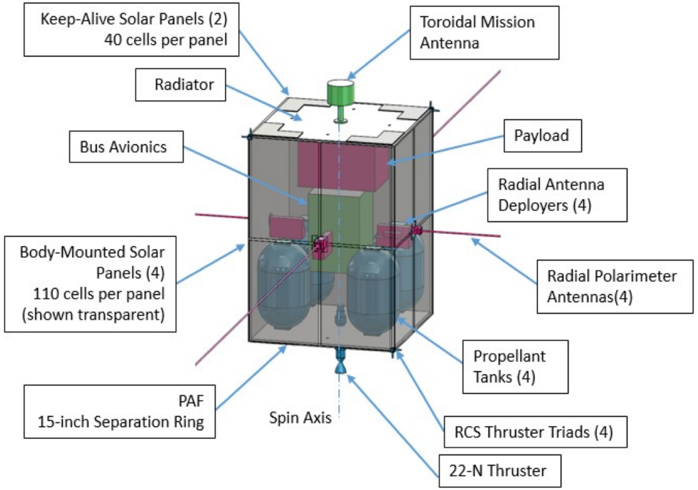
$$I = \langle |E_X|^2 + |E_Y|^2 \rangle$$

$$Q = \langle |E_X|^2 - |E_Y|^2 \rangle$$

$$U = 2\ \mathrm{Re}(E_X^* E_Y)$$

$$V = 2\ \mathrm{Im}(E_X^* E_Y)$$

Dual antenna experiments (e.g. DAPPER, CTP)



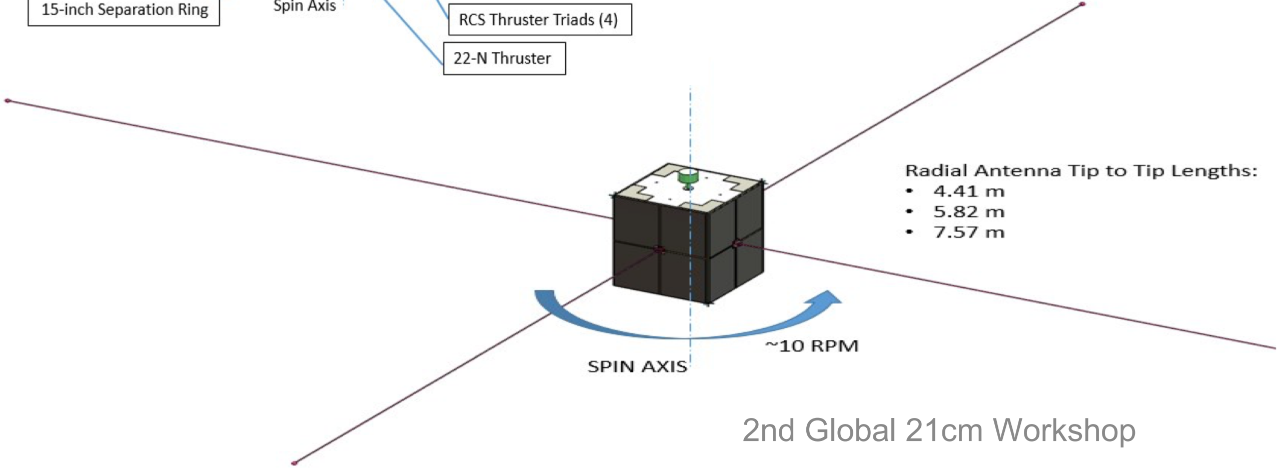
80 cm

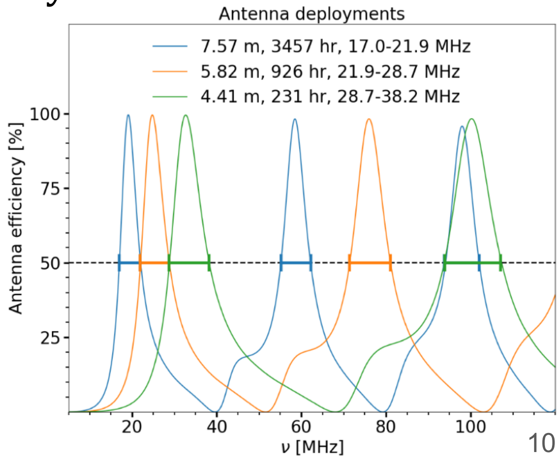
Spacecraft

- Deep Space Explorer bus by Bradford Space Industries.
- High impulse, high ΔV.

Antennas

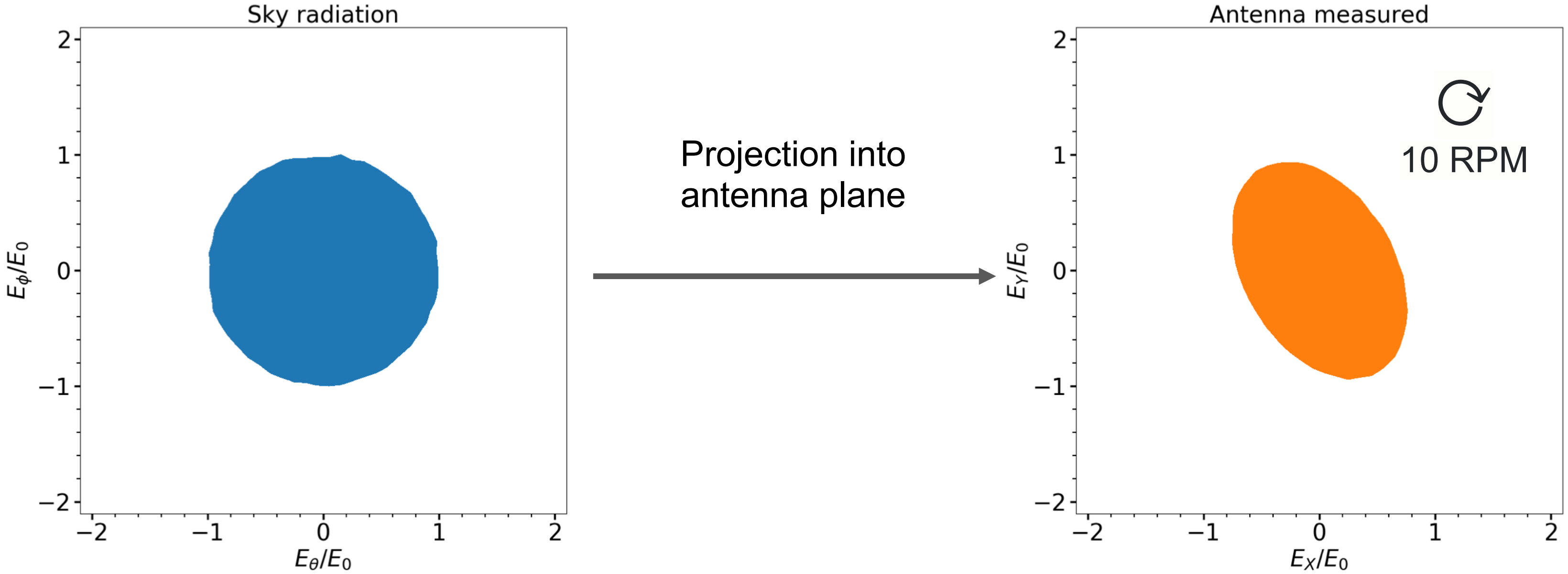
- Deployable, spinning, wire boom antennas arranged in 2 orthogonal, co-linear pairs.
- 3 length deployments to “tune” instrument.





Dynamical induced polarization

- Projection onto the instrument's antennas induces a polarization signal measured by dual antenna instruments, which can help constrain foreground.



SVD takes advantage of structure

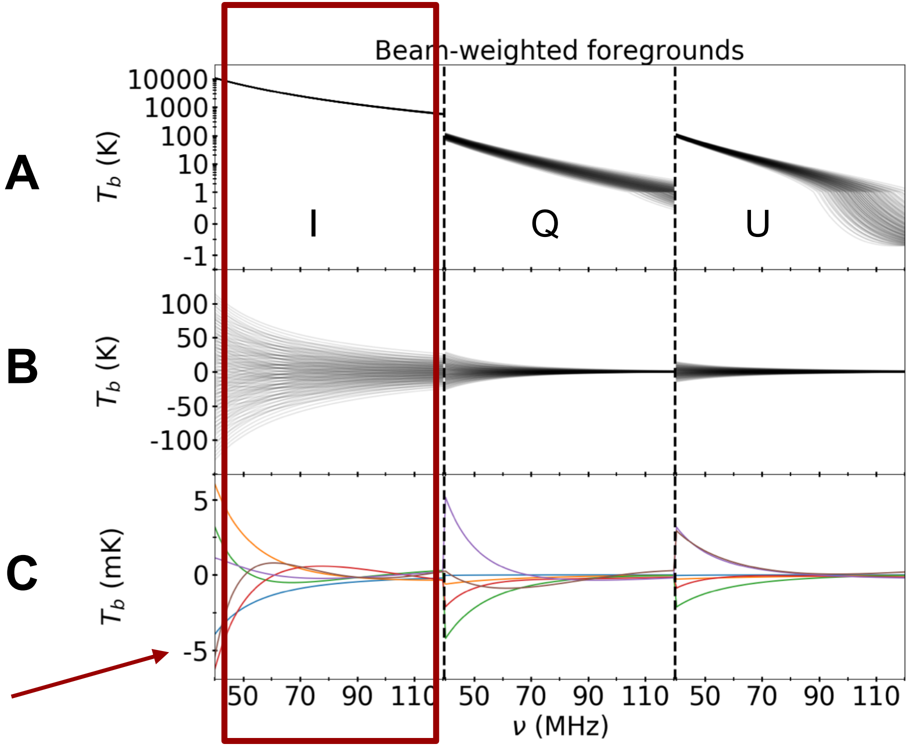
- SVD naturally provides a model which accounts for connections between Stokes parameters

A: Training set of Stokes parameters accounting for induced polarization.

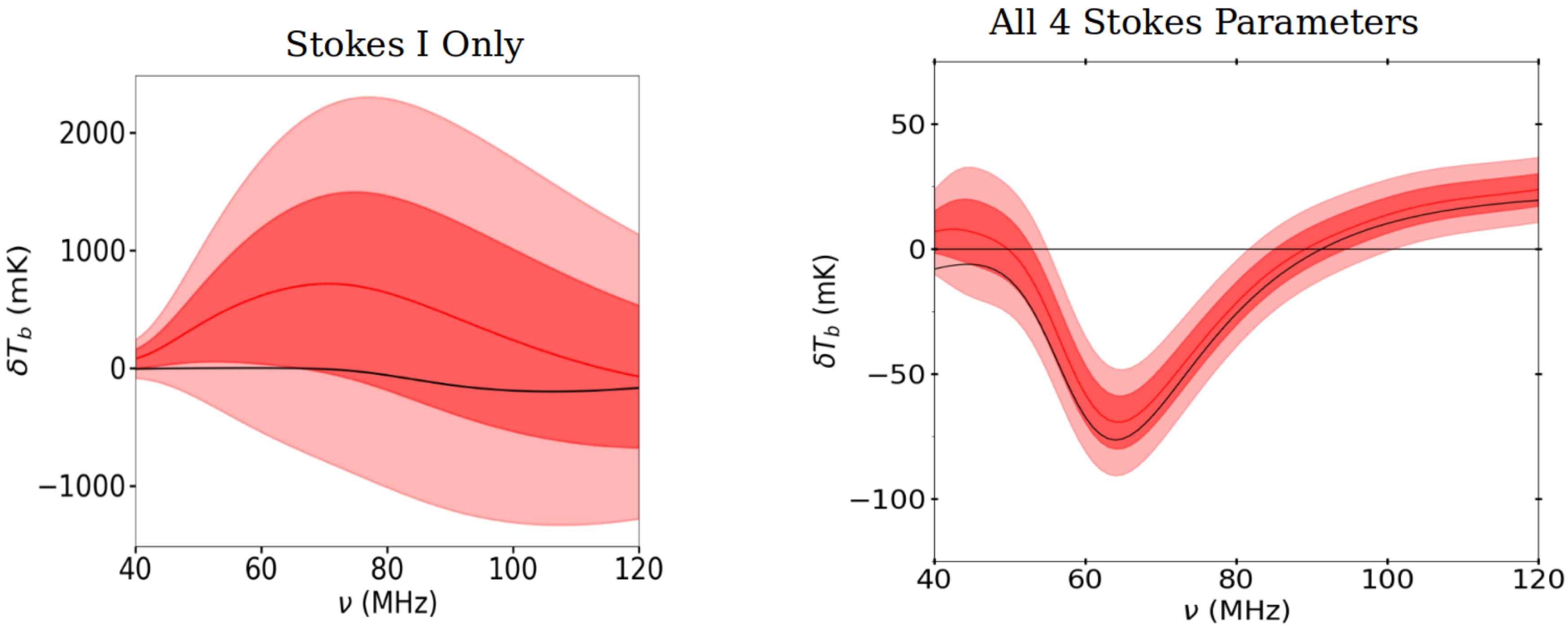
B: Training set with average subtracted

C: Optimal basis vectors to fit training set, provided SVD

Signal exists only here!



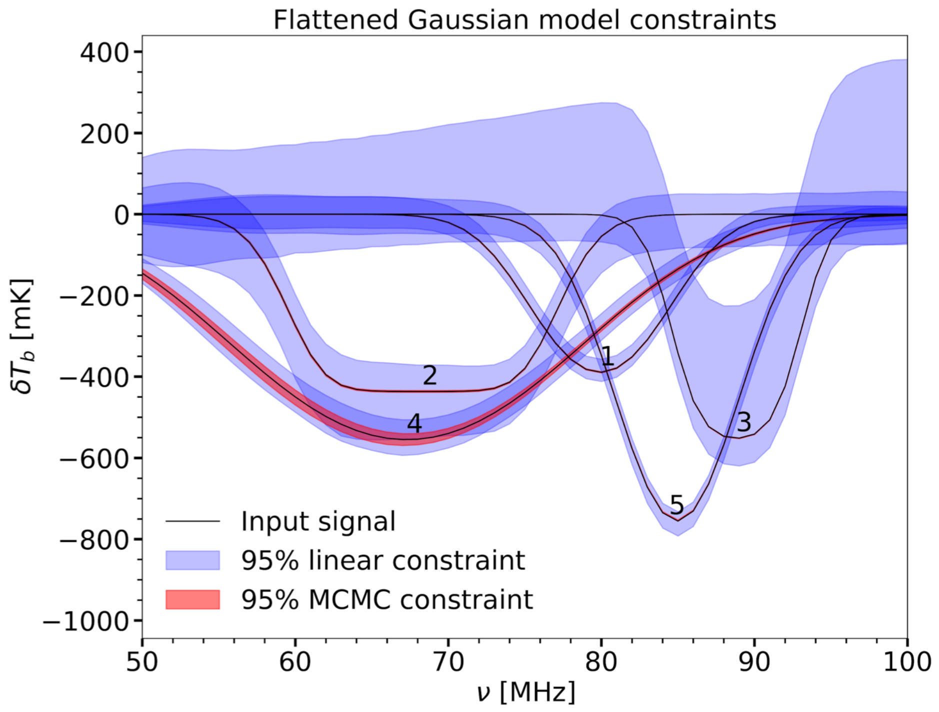
Effects of induced polarization data on constraints



Note the large scale difference between the plots.

Results with signal model from EDGES paper

- When using a training set of flattened Gaussian models defined as in EDGES Nature paper (Bowman et al. 2018), we obtain the confidence intervals on the right.
- These simulations include both Stokes parameter measurements and multiple antenna pointings.



A new goodness-of-fit statistic for 21-cm cosmology

- Traditional chi-squared statistic: $$\chi^2 = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{y_k - \mathcal{M}_k}{\sigma_k}\right)^2$$

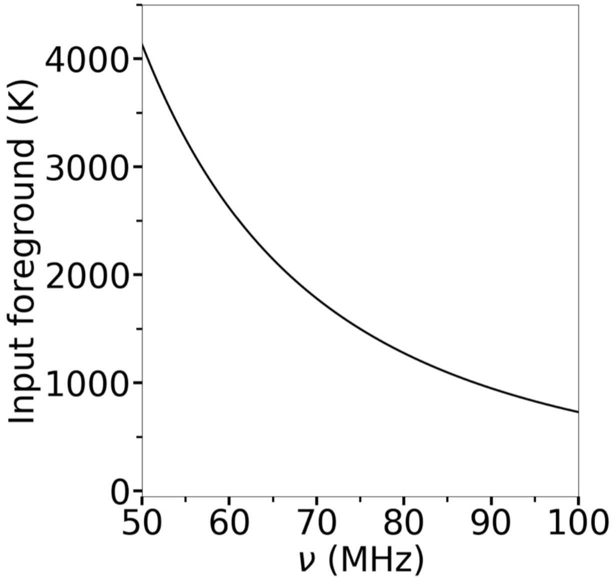
Data   Error   Model

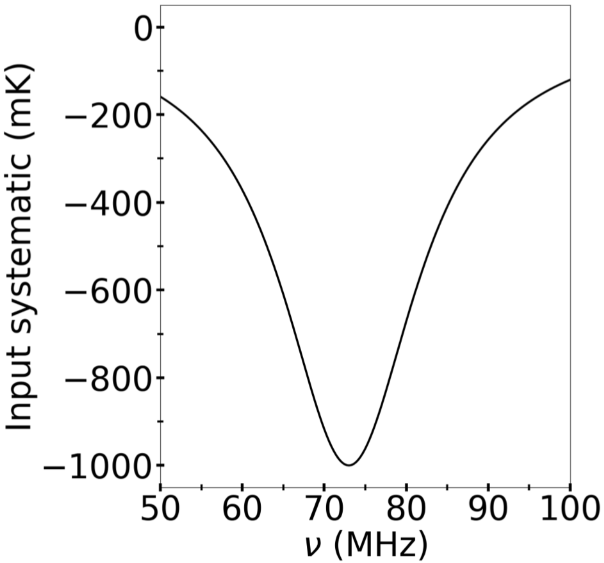
- Use as much data information as possible ➞ average/bin as little as possible
  - Averaging/binning less creates larger noise, which weakens the chi-squared statistic
- New psi-squared statistic designed to look at the squared channel-to-channel correlations of residuals instead of their absolute values (see code at https://bitbucket.org/ktausch/psipy)

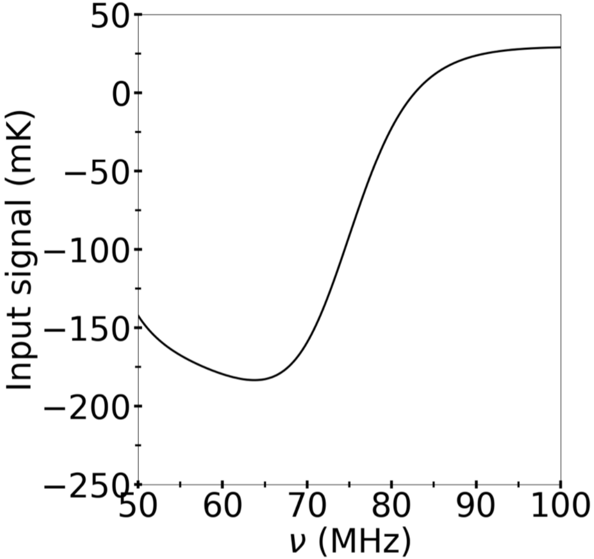
$$\psi^2 = \sum_{k=1}^{N-1}\left(\frac{N-k}{N-1}\right){\rho_k}^2 \quad \text{where} \quad \rho_k = \frac{1}{N-k}\sum_{q=1}^{N-k}\left(\frac{y_q - \mathcal{M}_q}{\sigma_q}\right)\left(\frac{y_{q+k} - \mathcal{M}_{q+k}}{\sigma_{q+k}}\right)$$

Utility of psi-squared statistic

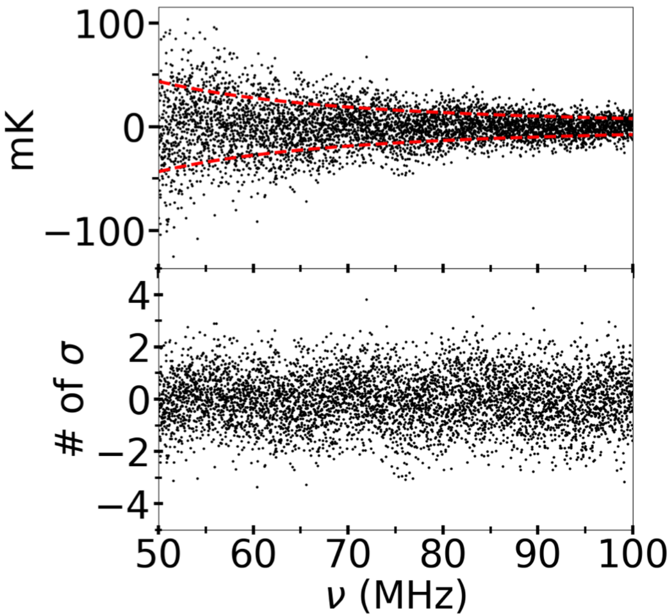
We performed fits on simulated datasets containing 1) a foreground, 2) a ground plane resonance, and 3) a 21-cm signal.







Effect on statistics of not fitting all data components



F.G.+Sys. Fit

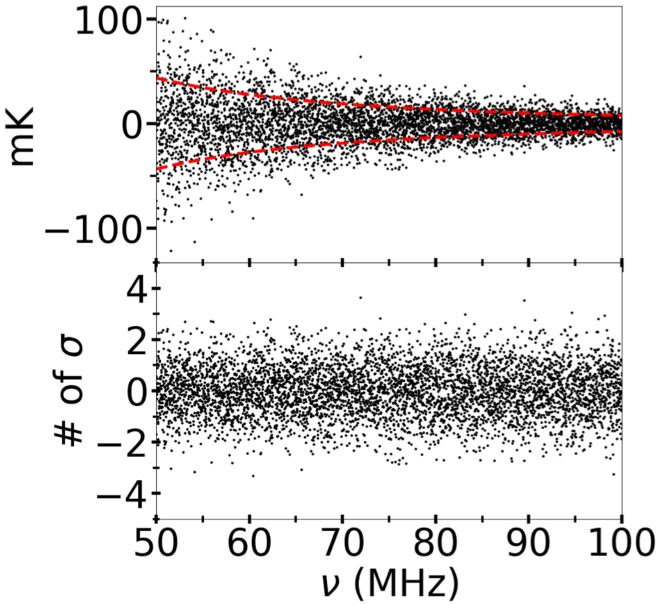
$$\chi^2 = 1.002 = \mu_{\chi^2} + 0.12\sigma_{\chi^2}$$

$$\psi^2 = 1.261 = \mu_{\psi^2} + 4.94\sigma_{\psi^2}$$

F.G.+Sys.+Signal Fit

$$\chi^2 = 0.982 = \mu_{\chi^2} - 0.88\sigma_{\chi^2}$$

$$\psi^2 = 0.961 = \mu_{\psi^2} - 0.73\sigma_{\psi^2}$$



Summary and conclusions

- We have been developing a training set-based pipeline for extracting the 21-cm global signal from sky-averaged spectral data that **allows any effect that can be simulated to be included in the analysis**.
- In our simulations, we include **multiple antenna pointing directions** (precursor to drift-scan measurements) and **Stokes parameter measurements** because they decrease the overlap between foreground and signal, lowering uncertainties.
- In the past year, we have for the first time completed the pipeline to extract physical parameters from our 21-cm signal constraints in frequency space (see David R.'s talk)
- The **new psi-squared statistic** should allow for more unaveraged/unbinned data to be analyzed at once **without losing goodness-of-fit discerning power**.
    - It is designed to detect low-level, wide-band features that typify 21-cm signal experiments' residuals.
    - When used for this purpose, psi-squared is more sensitive than chi-squared.

References

"An absorption profile centred at 78 megahertz in the sky-averaged spectrum". Bowman, J.D., Rogers, A.E.E., Monsalve, R.A., Mozdzen, T.J., Mahesh, N., 2018, *Nature*, 555, 67-70.

"Results from EDGES High-band. I. Constraints on Phenomenological Models for the Global 21 cm Signal". Monsalve, R.A., Rogers, A.E.E., Bowman, J.D., Mozdzen, T.J., 2017, *ApJ* 847 64.

"Spectral index of the diffuse radio background between 50 and 100 MHz". Mozdzen, T.J., Mahesh, N., Monsalve, R.A., Rogers, A.E.E., Bowman, J.D., 2019, *MNRAS* 483 4411.

"Global 21cm signal extraction from foreground and instrumental effects I: Pattern recognition framework for separation using training sets". Tauscher, K., Rapetti, D., Burns, J.O., Switzer, E., 2018, *ApJ* 853 187.

"A new goodness-of-fit statistic and its application to 21-cm cosmology". Tauscher, K., Rapetti, D., Burns, J.O., 2018, *JCAP* 1812, 015.

Backup slides

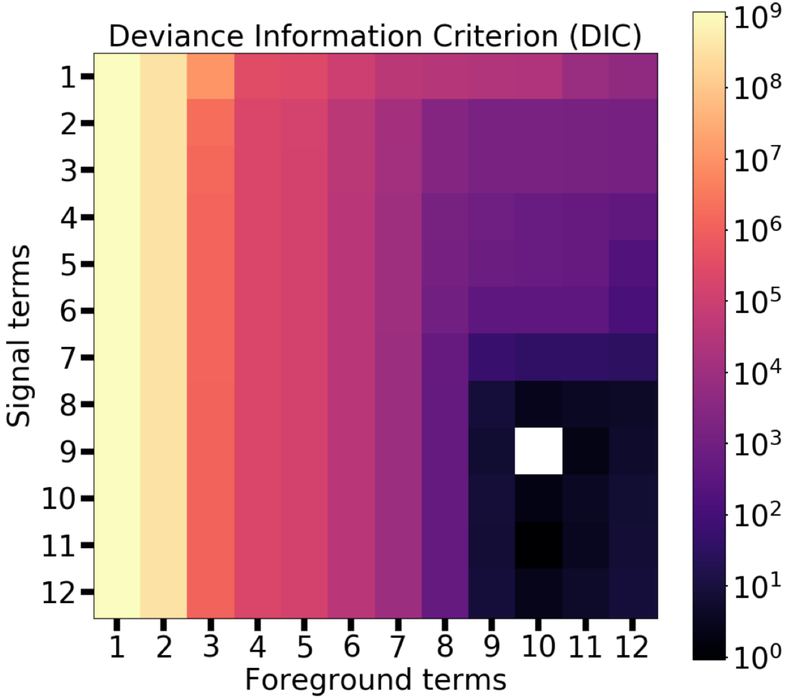
Model selection

- 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

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

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

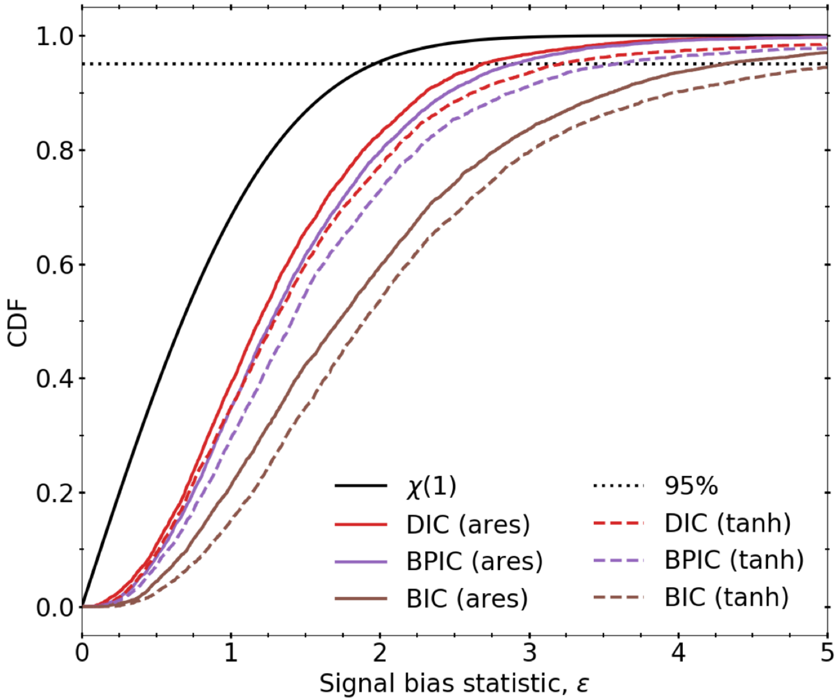
p: Total number of parameters



Signal bias statistic

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

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



Note on intrinsic polarization

- Polarization intrinsic to the foregrounds enters into data for all 21-cm global signal experiments, whether they are single or dual-antenna experiments, as pointed out by Spinelli et al. (2019).*
- When measuring with two antennas, the polarization signal rotates as the instrument rotates, instead of polarized sources going into and out of view.
- Induced polarization is expected to have a much larger effect (~10% of foreground) than intrinsic polarization (~1% of foreground) on polarized Stokes parameters.

*Spinelli, M., Bernardi, G., & Santos, M.G., *MNRAS* 489 4007 (2019)