Searching for the one in a million: extracting the cosmic signal from Dark Ages Radio Explorer data

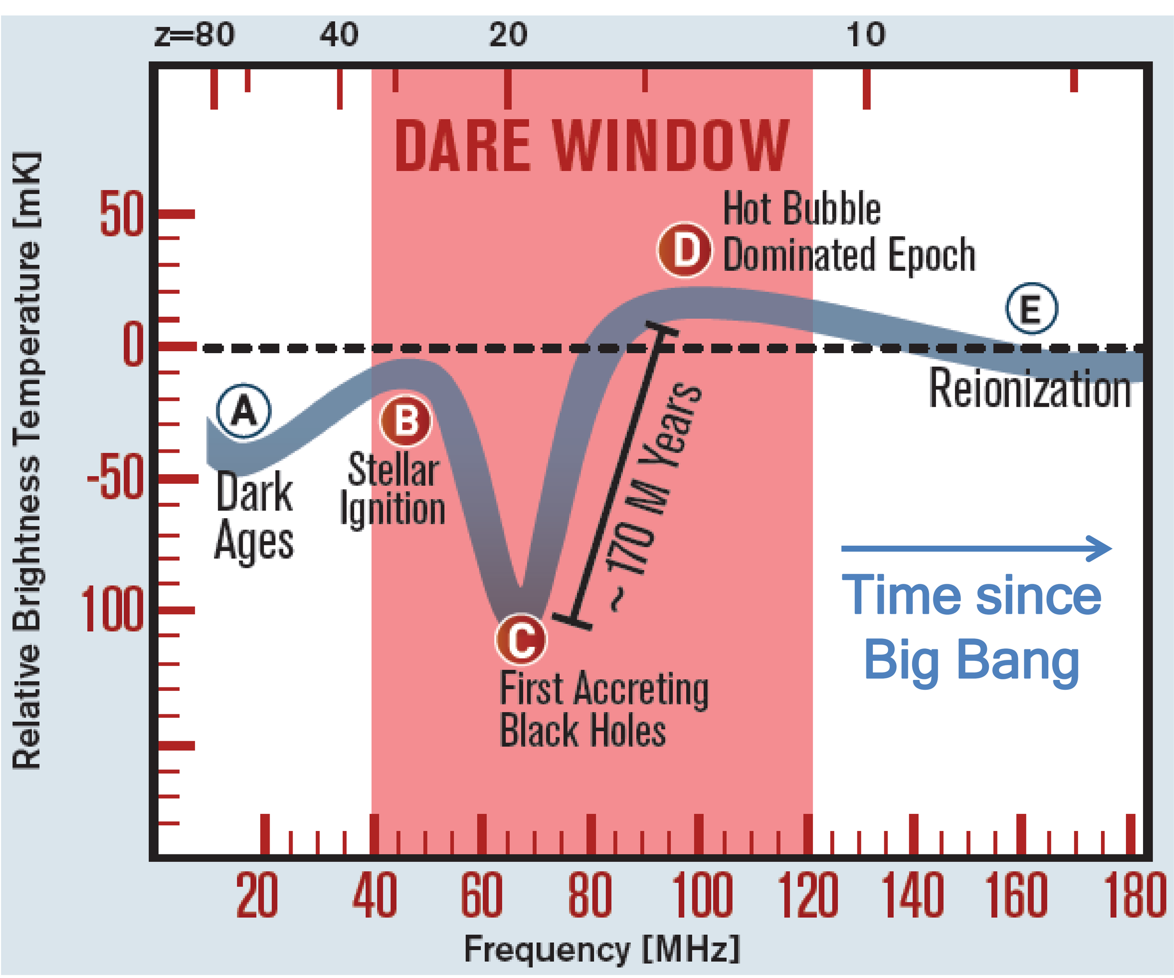
Geraint Harker,[1,3]* Jonathan Pritchard[2,3] and Jack Burns[1,3]

[1]Center for Astrophysics and Space Astronomy, University of Colorado Boulder; [2]Harvard-Smithsonian Center for Astrophysics; [3]NLSI Lunar University Network for Astrophysics Research (LUNAR)

Summary

The Dark Ages Radio Explorer (DARE) will study the cosmic dawn from lunar orbit by detecting the 21-cm (1420 MHz) hyperfine line from neutral hydrogen in the intergalactic medium in the early Universe, redshifted so that it reaches us in the 40-120 MHz radio band. Many other sources of radiation at these frequencies act as foregrounds to the cosmic signal, the strongest being man-made radio-frequency interference, which DARE avoids by collecting data in the shielded zone over the lunar farside. *Astrophysical foregrounds are still around a million times more intense than the redshifted 21-cm signal*, however. To achieve the dynamic range needed to remove them, *we have developed a sensitive foreground removal algorithm based on the Markov Chain Monte Carlo technique*, which makes use of the properties of the signal and foregrounds, and simultaneously fits the properties of the science instrument. All DARE's science objectives can be achieved with a three-year mission, and interesting constraints on the 21-cm signal can be found within one year.

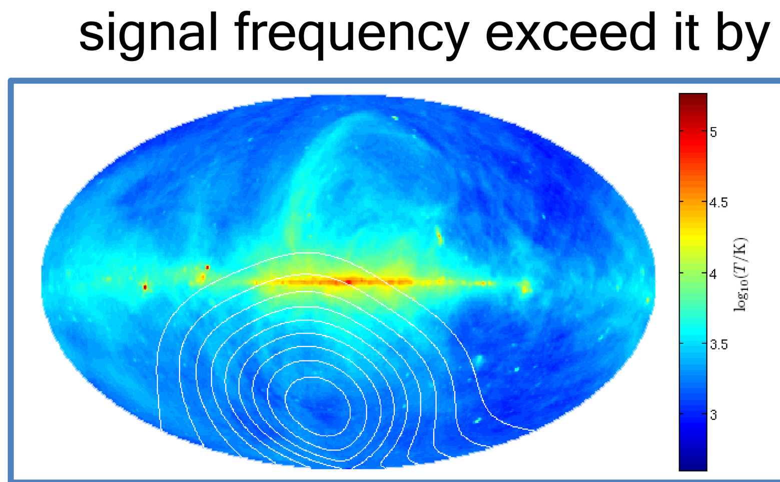
21-cm science from the Cosmic Dawn



Frequency and temperature of turning points B to D in redshifted 21-cm spectrum → information about the properties of the first stars and black holes in the Universe, through their effects on hydrogen which forms most of the intergalactic medium. The first sources drive the gas into absorption, then into emission. Finally, they erase the signal by ionizing the hydrogen.

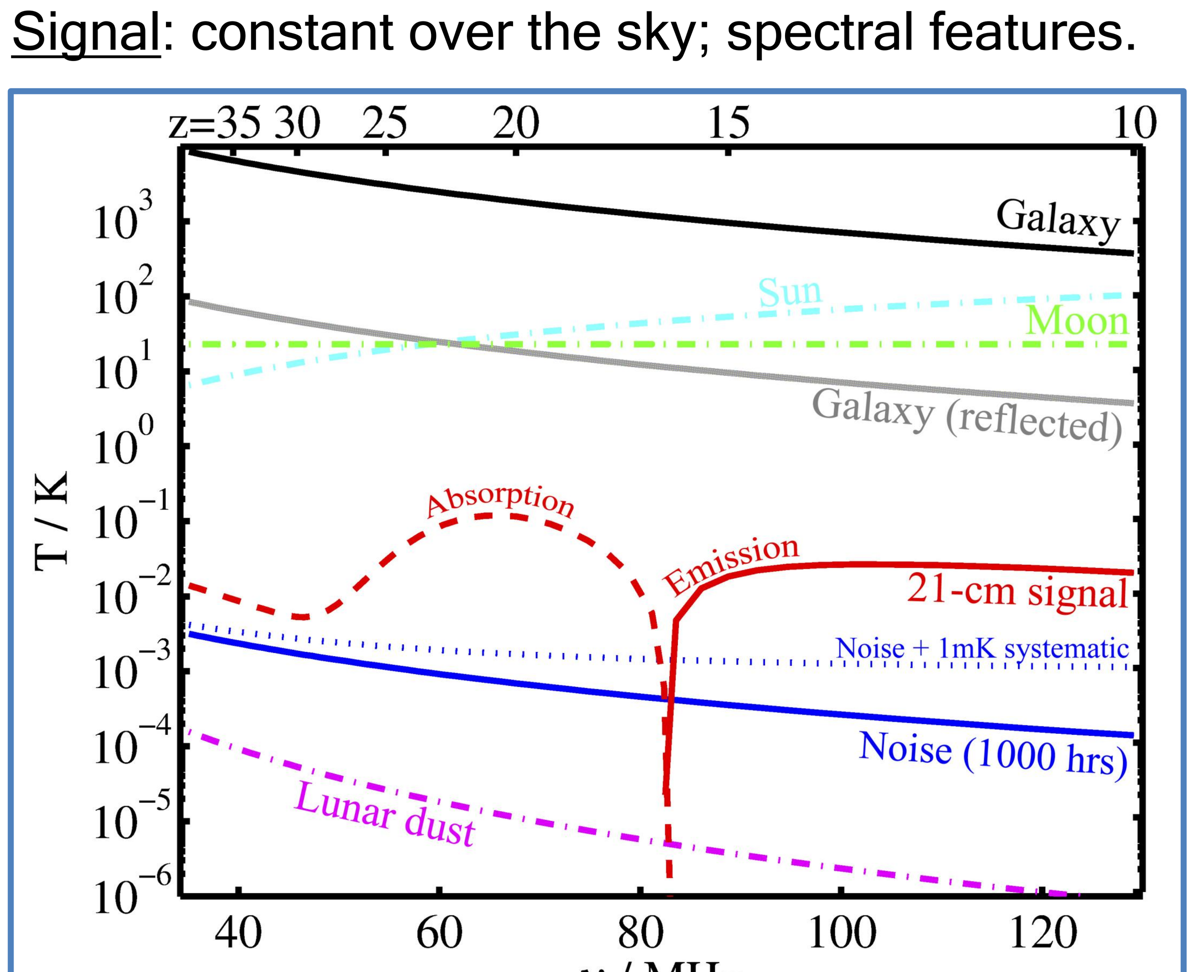
Foregrounds

- Radio power written as $P=k_B T_b \Delta\nu$ (k_B Boltzmann constant, $\Delta\nu$ frequency bandwidth, T_b 'brightness temperature').
- Astrophysical sources at the redshifted 21-cm signal frequency exceed it by ≈10^4-10^6.



Brightness temperature of the sky at 70 MHz. The spatial structure mainly comes from Galactic synchrotron emission. White contours show the power pattern of a dipole antenna.

Useful contrasting properties for separation:
Foregrounds: smooth in frequency; vary spatially.
Signal: constant over the sky; spectral features.



Various foregrounds at DARE frequencies, compared to the signal and to a noise spectrum at a representative level.

Model includes the Galaxy, extragalactic sources, Sun, Moon, and reflections of other foregrounds by the Moon. Impact of lunar dust particles on DARE → negligible noise spectrum.

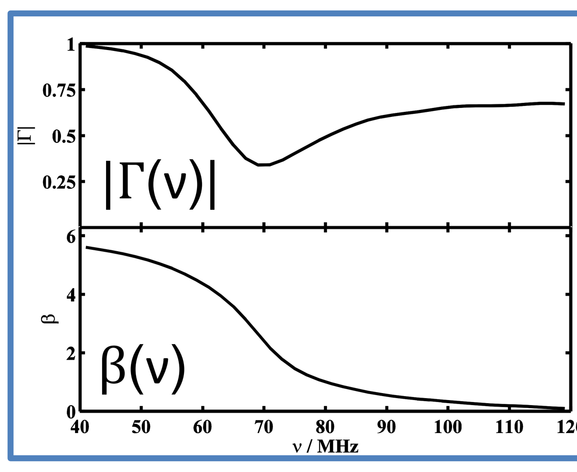
*geraint.harker@colorado.edu

Instrument response and data model

DARE takes independent spectra from 8 different directions in the sky. The measured antenna temperature is given by

$$T_{ant} = [1-|\Gamma|^2]T_{sky} + [1+2|\Gamma|\varepsilon\cos\beta + |\Gamma|^2]T_{rcv}$$

where T_{rcv} is the receiver temperature and ε, $\beta(\nu)$ and $|\Gamma(\nu)|$ depend on the properties of the instrument. We take $\varepsilon = 0.1$.



Frequency response of the DARE instrument, computed from an electromagnetic model of the antenna system. This is designed to be smooth as a function of frequency, so as not to compromise the smoothness of the foregrounds. We model each of these as the ten lowest frequency coefficients of its discrete cosine transform.

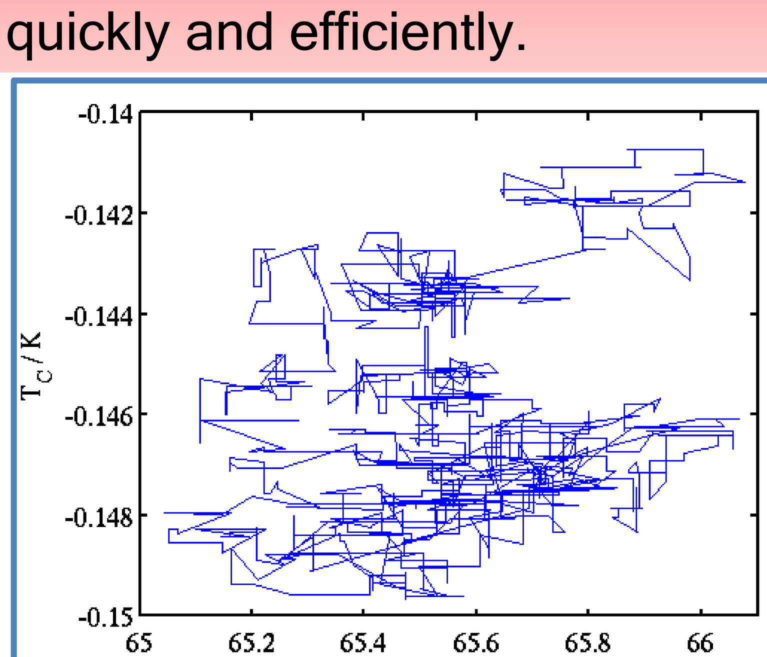
To generate a synthetic DARE data set:

- Find parametrized models for the signal (6 parameters), foregrounds (45), instrument (22)
- Use these to make spectra for the 8 regions
- Add noise according to radiometer equation

Then, recover signal parameters from noisy data by searching 73-dimensional parameter space.

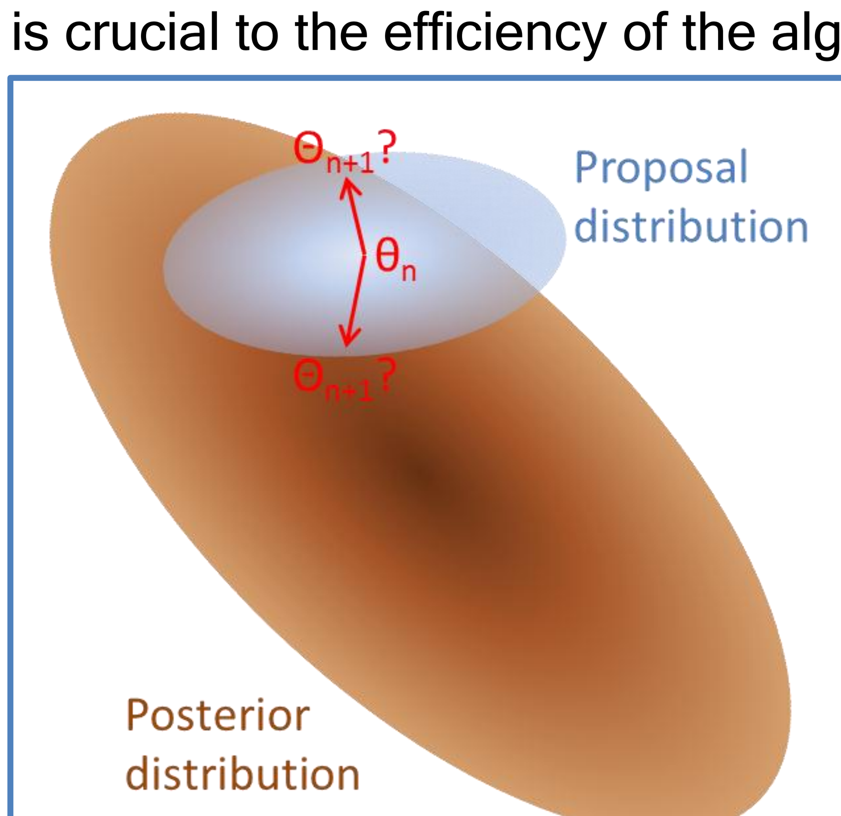
Markov Chain Monte Carlo (MCMC) method

We want to know the probability distribution of our model parameters given the data (the posterior probability distribution). MCMC is a way of drawing samples from the posterior distribution by taking a random walk through parameter space. It can explore high-dimension spaces relatively quickly and efficiently.



The path taken by an MCMC chain through a two-dimensional slice of parameter space. Each link in the chain is a sample from the posterior parameter distribution. The chain spends most of its time in regions of high probability, but is able to make excursions in order to fully explore the posterior.

At each step, a proposal for the position of the next link in the chain is drawn from a proposal distribution, then accepted or rejected based on the value of the posterior probability computed at that point. The form of the proposal distribution is crucial to the efficiency of the algorithm.

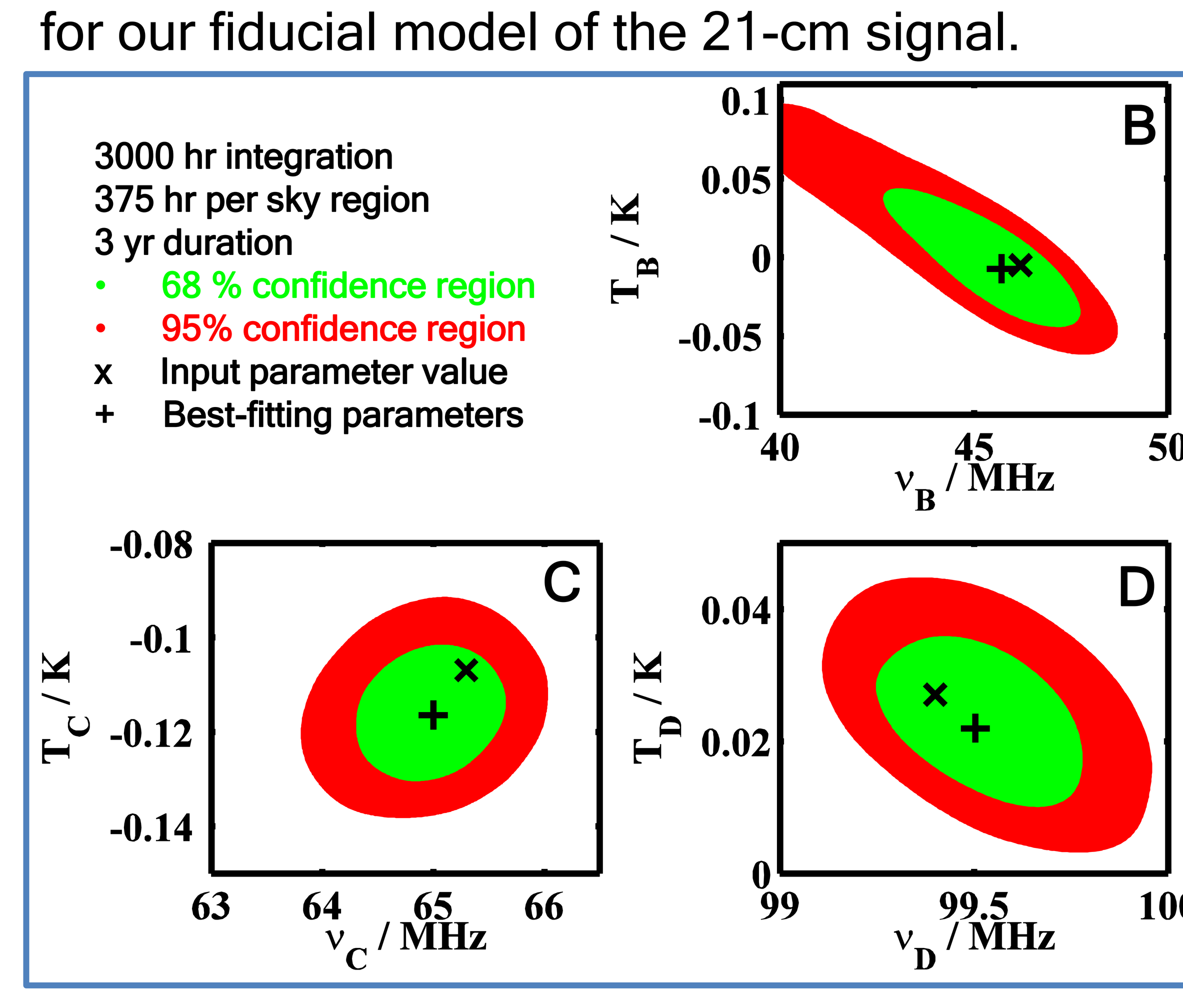


- Θ_n is the position of the chain at the n^{th} step.
- The proposal distribution should be well matched in size and shape to the posterior distribution, so we adapt the proposal distribution as the chain grows using knowledge of the existing samples.

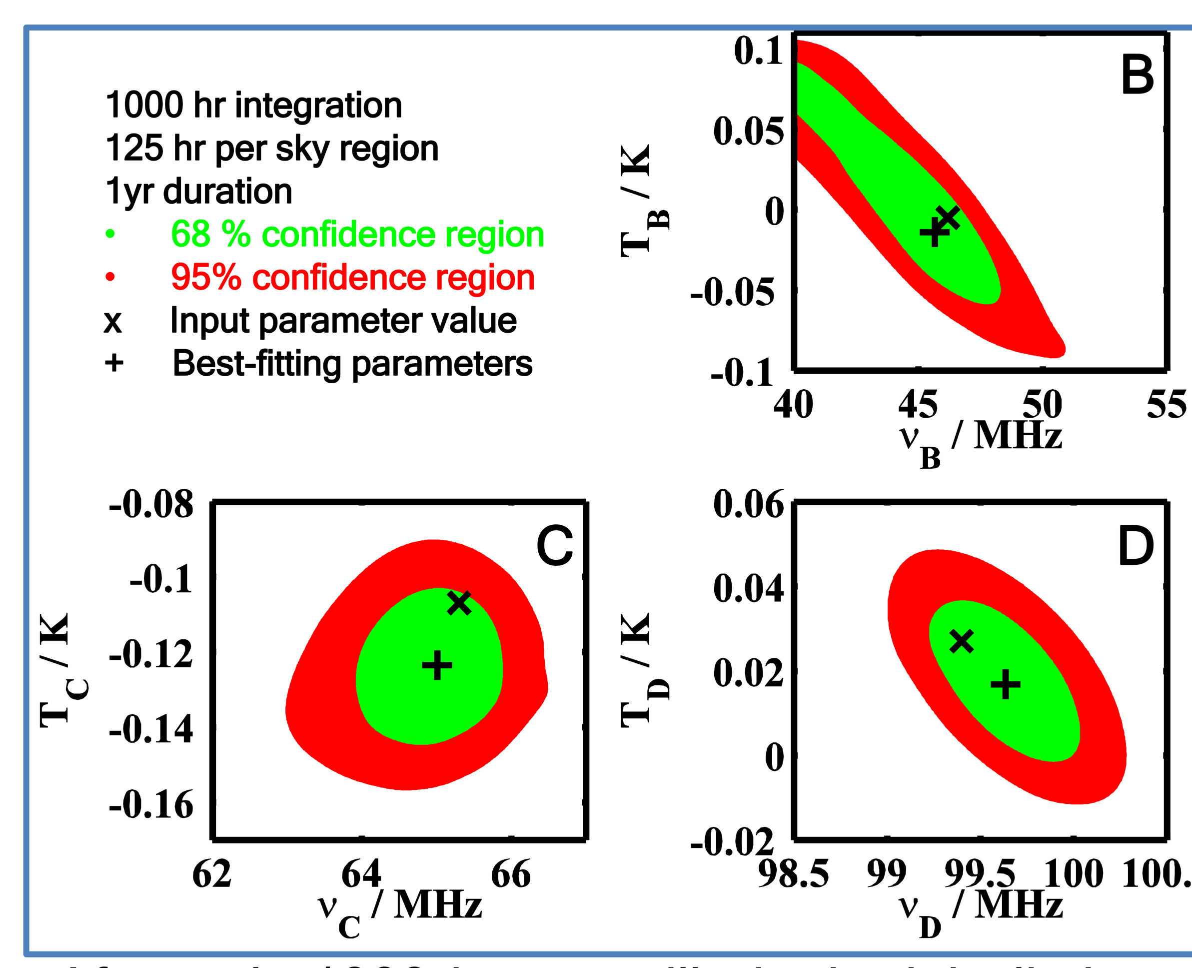
- If the proposal distribution is too narrow in some direction, the chain can only take tiny steps and parameter space is explored too slowly.
- If the proposal distribution is too broad, the chain tries to take huge steps outside the probable region, so many proposals are rejected.

Results

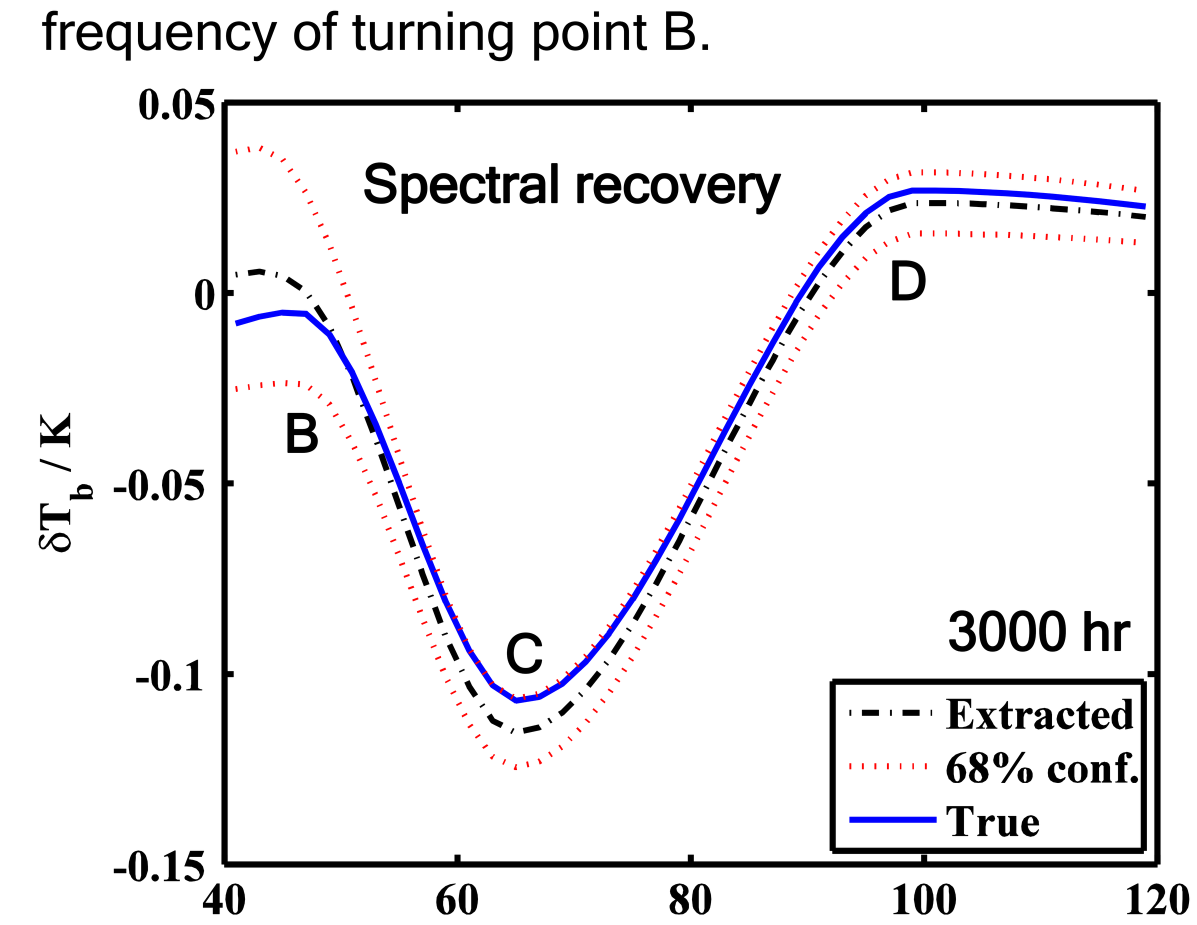
We show posterior probability contours for the frequency and temperature of turning points B, C and D computed from synthetic DARE data sets for our fiducial model of the 21-cm signal.



After 3000 hr, all three turning points are well constrained: to within 0.5 MHz for turning point D, 1 MHz for C and 5 MHz for B.



After only 1000 hr, we still obtain tight limits on turning points C and D, and an upper limit on the frequency of turning point B.



The shape of the cosmological signal is well determined. *This will allow us to rule out a large range of currently plausible models for the UV and X-ray properties of early sources.* The method is also very generally applicable to similar experiments at higher frequencies.

University of Colorado Boulder