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

Keith Tauscher*[(1)], Jack O. Burns[(1)], David Rapetti[(1,2)], and Eric R. Switzer[(3)]
(1) Center for Astrophysics and Space Astronomy, Department of Astrophysical and Planetary Sciences, Boulder, CO 80309
(2) NASA Ames Research Center, Mountain View, CA 94035
(3) NASA Goddard Space Flight Center, Greenbelt, MD 20771

The sky-averaged (global) highly redshifted 21-cm line from neutral hydrogen is a key probe of aspects of the Universe which we do not yet understand well. For instance, since the signal arises from a hyperfine transition of neutral hydrogen, its strength is proportional to the neutral fraction, making it a probe of reionization. Its brightness temperature at a given redshift also yields information about the thermal state of the InterGalactic Medium (IGM) at the corresponding time. Despite the great scientific value of the global signal, it has not yet been detected due to the fact that it must be separated from very strong beam-corrupted Galactic foregrounds (4-5 orders of magnitude larger than the signal) and instrumental calibration errors. In many previous analysis approaches, foregrounds and calibration errors, referred to collectively as "systematics" below, were modeled using polynomials, requiring them to be known to high precision. In contrast, here we present a data pipeline for global signal experiments which simulates "training sets" and then performs Singular Value Decomposition (SVD) on them to produce systematics basis vectors specifically tailored to each individual experimental observation. In this way, instead of requiring the systematics to be known precisely, our pipeline requires the modes of variation of these systematics to be known precisely, a task better suited for simulations and measurements to address. Similarly, simulations of many different signals generated through varying parameters of the *ares* code yield modes of variation of the global 21-cm signal. Using the SVD modes for both the systematics and the signal, the pipeline uses a linear weighted least squares fit to find an initial estimate of the global 21-cm signal with errors. Then, using this first estimate as guidance, the pipeline performs multiple Markov Chain Monte Carlo (MCMC) fits which allow us to make robust inferences (means, confidence intervals, etc.) on physical parameters governing the behavior of the signal.