Data analysis and foreground removal algorithms for 21-cm cosmology experiments

Geraint J. A. Harker
Center for Astrophysics and Space Astronomy, University of Colorado Boulder

Cosmology experiments using the highly redshifted 21-cm line of hydrogen are taking a range of observational and instrumental approaches. The simplest, at first sight, are experiments aimed at measuring the sky-averaged ('global') signal using a single, well-characterized and precisely calibrated antenna, either on the ground or in space. Multiple-antenna and interferometric global signal experiments aim to trade off the increased complexity against improved resolution, and a better handle on e.g. the behaviour of the ionosphere and the instrumental beam. Studying the 21-cm fluctuations requires interferometry but still allows for many different approaches, e.g. using drift scans versus tracking individual fields, using redundant baselines versus aiming for minimal redundancy, or trading data volume against uv coverage and field of view.

Despite this diversity, 21-cm experiments share a lot of the same challenges in extracting cosmological information from the data, arising from the fact that the 21-cm signal is buried under astrophysical foregrounds which are orders of magnitude more intense. This is what places such stringent requirements on instrumental calibration and the dynamic range of the observations. Removing foregrounds inevitably removes some of the signal, and so it is important to understand the properties of the foregrounds, and how their removal affects the cosmological information we can glean from the signal. Since 21-cm experiments will often also constitute our deepest observations of the foregrounds, we must be able to rigorously validate our foreground models and infer 21-cm signal model parameters from the same data. A number of methods have been developed to do this, ranging from 'blind' signal separation in which (as far as possible) the data dictate the form of the foregrounds, to completely parametric techniques in which explicit models are required for all the components of the signal. We will review some of the advantages and disadvantages of the various methods and highlight the performance of some of them in removing the foregrounds from realistic, synthetic data.