3rd URSI AT-AP-RASC, Gran Canaria, 29 May – 3 June 2022

Processing HERA Data with Machine Learning and Baseline Dependent Averaging

Paul La Plante*[(1)], Bryna J. Hazelton[(2)(3)], Joshua S. Dillon[(1)], Lister Chen[(1)], and The HERA Collaboration
(1) Department of Astronomy, University of California, Berkeley, CA
(2) Department of Physics, University of Washington, Seattle, WA
(3) eScience Institute, University of Washington, Seattle, WA

1 Extended Abstract

The Hydrogen Epoch of Reionization Array (HERA) is a 21 cm radio interferometer endeavoring to measure Cosmic Dawn and the Epoch of Reionization. When fully constructed, it will consist of 350 dual-polarization antennas recording 6,144 frequency channels over 187.5 MHz of bandwidth. To avoid decoherence for the longest baselines, the fundamental time spacing of the correlator is designed to be 2 seconds. However, recording the full cross-correlation visibility matrix for all baseline pairs every 2 seconds is prohibitively expensive given the storage constraints on site, and is a much higher time cadence than required to avoid decoherence of short baselines, which make up the majority of HERA's data.

To reduce the amount of disk space required to save HERA data, we have implemented a baseline-dependent averaging (BDA, [1]) scheme which uses different integration times for different baselines inside of the HERA correlator. Essentially, longer baselines have a shorter integration time to reduce the decoherence that occurs, and vice versa. As part of the BDA scheme implemented in HERA, we have also included fringe-stopping of each individual antenna. This allows for recording visibilities that have been phased to the "center" of a common time window for all baselines.

In this talk, we present verification of the BDA scheme developed for use in the HERA correlaetor, as well as recent advancements in the HERA data processing pipeline that have been used to accommodate BDA data. Among these changes are: (1) developing methods of data handling that present a uniform time grid to calibration methods, and (2) the creation of machine learning (ML)-based methods that are able to handle the native time resolution of the correlator. These ML methods are used both for identifying radio frequency interference (RFI) events in data, as well as flagging other problems with antennas. We also present updates on the HERA Real Time Processing (RTP) system [2, 3], which have been made to support the implementation of these methods. These improvements include a more flexible way of using graphics processing units (GPUs) for various compute-intensive steps, as well as improved methods for transporting data using the Librarian data management system.

References

[1] S. J. Wijnholds, A. G. Willis, and S. Salvini, "Baseline-dependent averaging in radio interferometry," *MNRAS*, **476**, May 2018, pp. 2029-2039, doi: 10.1093/mnras/sty360.

[2] P. La Plante, P. K. G. Williams, and J. S. Dillon, "Developing a Real Time Processing System for HERA", *Radio Science Letters*, **2**, 2020, doi:10.46620/20-0041.

[3] P. La Plante, P. K. G. Williams, M. Kolopanis, *et al.*, "A Real Time Processing System for Big Data in Astronomy: Applications to HERA", *Astronomy and Computing*, **36**, July 2021,doi:10.1016/j.ascom.2021.100489.