

Correlator Ramp and Specification Tradeoffs

HERA Team

May 15, 2018

Executive Summary

HERA is operated in “seasons” that are denoted as:

Table 1: HERA Operating Seasons

H0C	season which ended in March 2017	used the first 19 antennas (with their slightly wonky feeds).
H1C	season which ended in April 2018	start with 32, end with 71 with old feeds/architecture
H2C	season to begin October 2018	start with 59, end with 189 with new feeds/architecture
H3C	season to begin August 2019	350 elements with new feeds/architecture

H0C and H1C used the recycled PAPER equipment and beginning H2C the new hardware and configuration will be installed. To accommodate these updates H2C will be broken into two phases. The ramp for the build-out is summarized in Table 2 and the antennas at the start of the phase are shown in Fig. 1.

Table 2: Build-out Summary. Total bandwidth = 187.5 MHz within 50-250 MHz

	Date	# ants	Channels	Int	Total data¹	Comments
H2C.A	1 Oct 18	59-120	1536/122.1kHz	10 s	0.11 PB	Fringe-stopping hooks in, but not implemented. Even-odd implemented. Walshing implemented.
H2C.B	1 Jan 19	120-189	3072/60.0kHz	10 s	0.93/0.18 PB	Fringe-stopping and baseline-dependent averaging also implemented. Goal to present data as same as H2C.A for DS users.
H3C	1 Aug 19	350	6144/30.5kHz	2 s	94/5.7 PB	Fringe-stopping, baseline-dependent averaging, even-odd, Walshing implemented

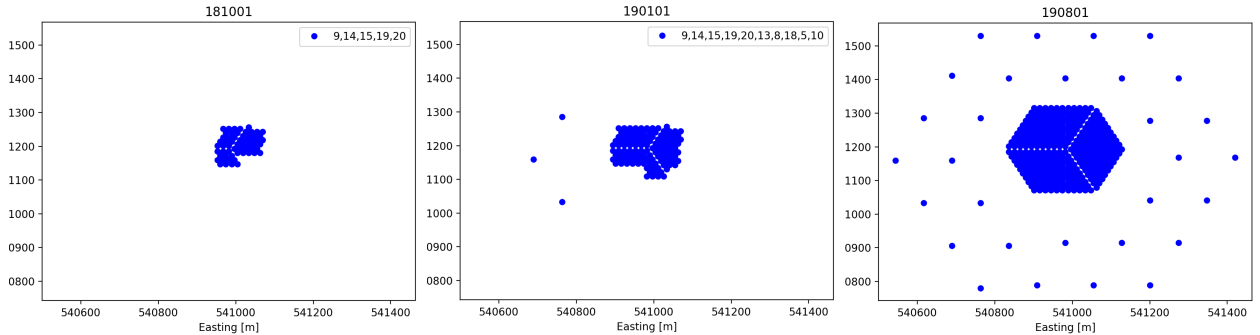


Figure 1: Planned element rollout. Left shows H2C.A comprising 59 elements by 1 Oct 2018. Right shows H2C.B comprising 120 elements by 1 Jan 2019. The legend lists the node numbers.

¹For H2C.B and H3C, the 2 values shown are the raw “naïve” data rate, and including baseline-dependent averaging with 5% maximum decorrelation for a source at 20° off-zenith. H3C data volume assumes 6 months of observing.

Introduction

The specifications for the HERA correlator and real time processing (RTP) have important implications for science with HERA, particularly for image-based calibration, foreground subtraction and power spectra. These specifications also drive the data rate of HERA, and therefore the time and costs associated with processing and storing the data.

The effect on imaging and related activities is driven by the amount of decorrelation of the visibilities caused by integrating in time and frequency. Decorrelation causes *baseline-dependent* changes in the instrument response (effective primary beam) that are challenging to model and will limit the accuracy of image-based calibration and foreground subtraction, which rely on high-quality visibility simulations.

This memo explores trade-offs along four primary axes:

1. **Number of Frequency Channels:** Since the bandwidth is fixed (50-250 MHz), the number of channels is directly related to the channel width. Larger numbers of channels (narrower channel widths) lead to lower decorrelation and higher data rates. The number of channels in the current correlator is 2048 (122kHz wide channels).
2. **Integration Time:** The final length of the integrations written by the correlator (after fringe-stopping if that is being done, see next item). Longer integration times lead to more decorrelation and lower data rates. The integration time of the current correlator is 10s.
3. **Fringe Stopping:** Fringe stopping removes the first-order decorrelation effect from integrating in time by phasing each visibility (at a very high time cadence, e.g. 100ms) to a fixed point on the celestial sphere before the final time integration step. Fringe stopping significantly decreases the decorrelation for a fixed final integration time. Fringe stopping has no effect on the data rate, but it requires more sophisticated firmware on the correlator (to handle the phasing). The current correlator does not do fringe stopping.
4. **Baseline Dependent Averaging:** Baseline dependent averaging refers to a practice of averaging baselines for longer times depending on their fringe-rate, with integration times selected to be integer factors of the correlator integration time that keep the visibility below some specified level of decorrelation. This means that the decorrelation on short baselines will be *increased* relative to doing no baseline dependent averaging, but it leads to a substantial decrease in the final data rate for HERA, which is dominated by short baselines. Baseline dependent averaging would be implemented in RTP, not the correlator, so it would not decrease the correlator data rate but it would decrease the data rate after RTP. Software adjustments would be required to handle baseline dependent averaging in down-stream code, but the software team is optimistic that most of the handling for it could be done in `pyuvdata` in a way that reduced the impact on most users. The current RTP does not do baseline dependent averaging.

The plots and tables shown in this memo were produced using code and notebooks located in the ProjectFiles repository². Interested parties are encouraged to run their own scenarios.

² https://github.com/HERA-Team/ProjectFiles/tree/master/spec_calcs

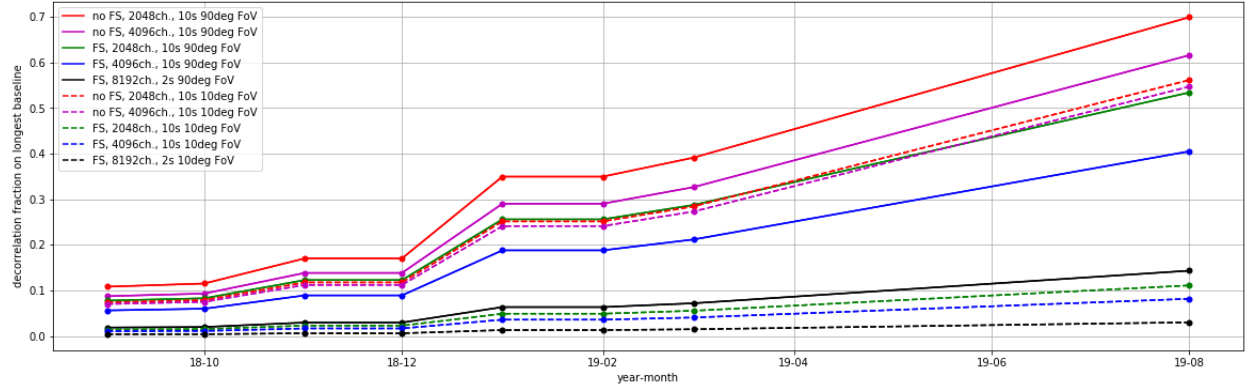


Figure 2: Decorrelations on the longest baseline over the course of the buildout for various correlator options. The red line shows the current correlator specifications (2048 channels, 10s integration, no fringe stopping), the blue line shows the imaging-driven specifications (4096 channels, 10s integrations with fringe stopping) for h2c and the magenta and green lines give two intermediate options. Finally, the black line shows an aggressive imaging-driven specification for the full array (8192 channels, 2s integrations with fringe stopping). Solid lines show the decorrelation at the horizon while the dashed lines show the decorrelation 10 degrees away from the zenith. This plot is generated in the `h2c_stats.ipynb` notebook².

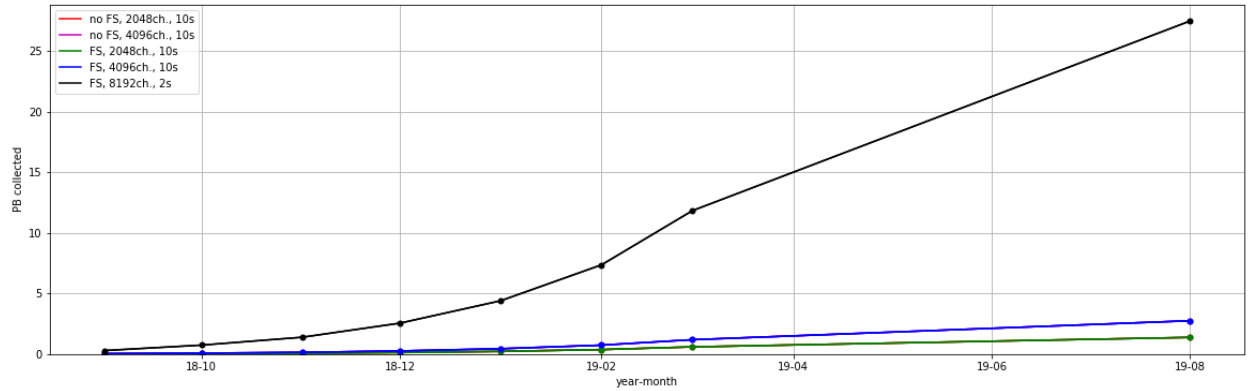


Figure 3: Correlator data volumes on the longest baseline over the course of the buildout for various correlator options. The line colors are the same as in 2, but the green line is on top of the red line and the blue line is on top of the magenta line because the data rate depends only on the number of frequency channels and the final integration time (not on fringe stopping). The data volumes after RTP can be decreased substantially by baseline-dependent averaging. This plot is generated in the `h2c_stats.ipynb` notebook².

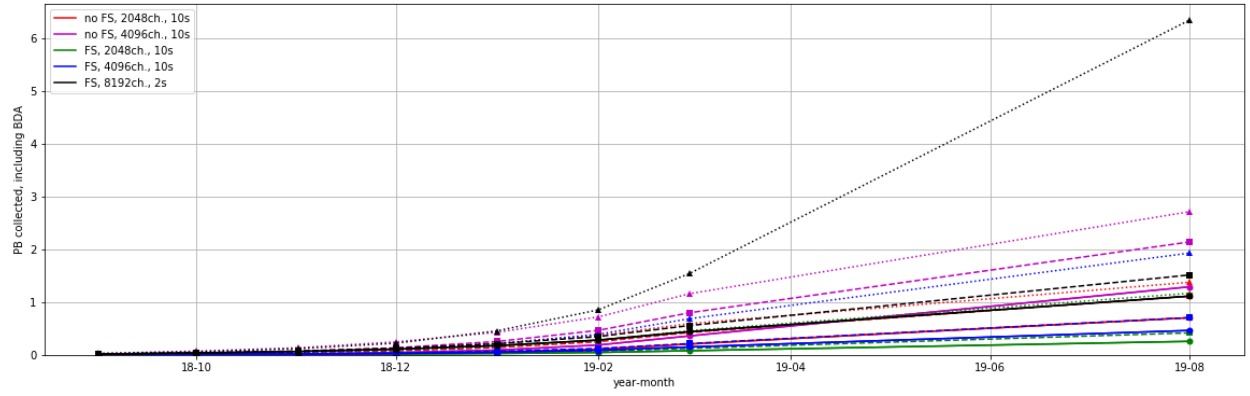


Figure 4: Final correlator data volumes including the effects of baseline-dependent averaging. Line colors are the same as in the previous figures (corresponding to different correlator specifications). Solid, dashed, and dotted lines correspond to 10%, 5%, and 1% maximal decorrelation for a source 20° off-zenith. The baseline-dependent averaging calculation can be seen in `bda_calc.py` script, and the plot is generated in the `h2c_stats.ipynb` notebook².