URSI GASS 2020, Rome, Italy, 29 August - 5 September 2020

Full-scale SKA1-Low Science Data Processing Workflow

M. Dolensky[(1)], R. Dodson* [(1)], R. Tobar[(1)], A. Wicenec[(1)], T. An[(2)] and R. Wang[(3)]
(1) ICRAR, University of Western Australia, Australia; e-mail: markus.dolensky@icrar.org; richard.dodson@icrar.org; rtobar@icrar.org; andreas.wicenec@icrar.org
(2) Shanghai Astronomical Observatory, China; e-mail: antao@shao.ac.cn
(3) Oak Ridge National Laboratory ORNL, USA; e-mail: wangr1@ornl.gov

Extended Abstract

The world's currently most powerful, monolithic supercomputer SUMMIT at Oak Ridge National Laboratory (ORNL) simulated an observation of the early Universe [1]. Simulating the workload of a data reduction pipeline was an excellent opportunity to test architectural elements of the planned Square Kilometre Array (SKA) radio telescope. The tested software components included the novel execution framework DALiuGE [2] developed at ICRAR and the Adaptable Input/Output System (ADIOS) provided by ORNL.

The experiment simulated a typical 6 hour astronomical observation with the planned layout of more than 130000 dipole antenna that will comprise the low frequency radio interferometer SKA1-Low in the Western Australian outback. A major challenge of the science data processing workflow was to generate the data volume of the future telescope. The team included collaborators of the Shanghai Astronomical Observatory (SHAO) who drew on earlier experience gained on the previously most powerful supercomputer Tianhe-2. This time 27360 Nvidia V100 GPUs were employed in parallel. Each GPU ran an instance of OSKAR2, a simulator developed at the Oxford e-Research Centre. A total of 2.6 Petabytes of synthetic measurements were generated within three hours of walltime. The execution framework engine DALiuGE orchestrated the processing tasks on 4561 compute nodes. After filtering and averaging operations the resulting data product was 113 Terabyte in size. It was further processed by the *cimager* task of Yandasoft provided by CSIRO/CASS, which synthesized a series of image planes (Figure 1). The experiment demonstrated that DALiuGE is capable of deploying and controlling workflows at SKA1-scale, both, in streaming and batch processing mode and that it scales to the largest existing supercomputer. The ADIOS library achieved a peak write speed of 130 GB/s into a single data product in MeasurementSet format.

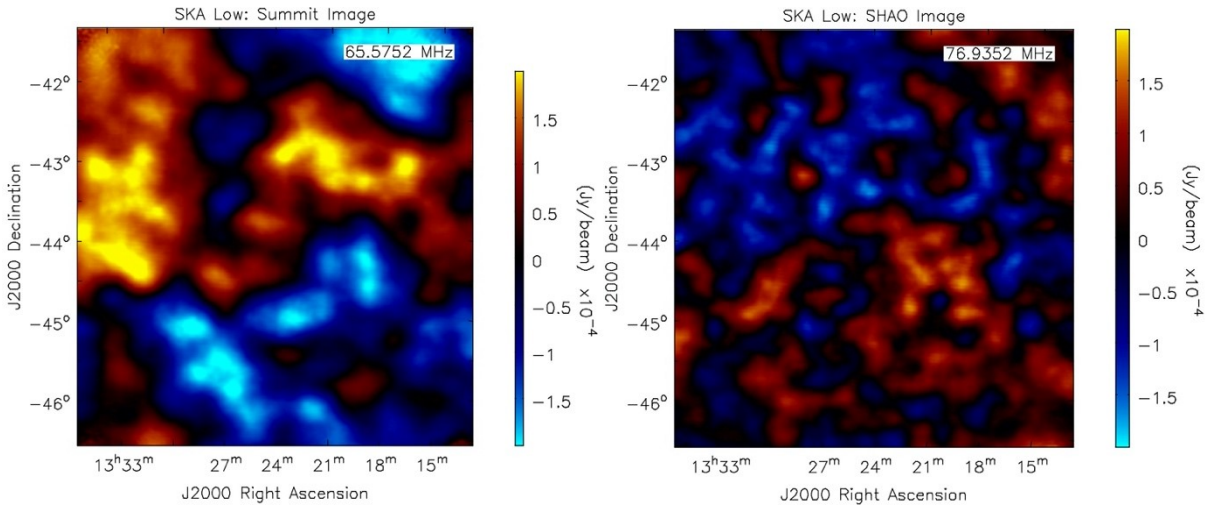

Figure 1. The recovered simulated radio signal generated from the 21 cm simulation at 66 and 77 MHz.

References

[1] R. Wang, A. Wicenec, T. An, SKA shakes hands with Summit, *Science Bulletin*, 2020, arXiv:1912.12591.

[2] C. Wu et al., "DALiuGE: A Graph Execution Framework for Harnessing the Astronomical Data Deluge.", *Astronomy and Computing*, 20, pp.1-15., Feb. 2017, arXiv:1702.067617.