Astronomical Data Analysis Software and Systems XXVII
ASP Conference Series, Vol. 522
Pascal Ballester, Jorge Ibsen, Mauricio Solar, and Keith Shortridge, eds.

High-performance Pipeline Processing for ASKAP

Matthew T. Whiting

CSIRO Astronomy & Space Science, Sydney, NSW, Australia;
`Matthew.Whiting@csiro.au`

Abstract. The Australian Square Kilometre Array Pathfinder (ASKAP) is an innovative wide-field, high-data-rate radio-synthesis telescope, that requires high-performance processing pipelines for calibration and imaging. These pipelines have been demonstrated with commissioning and Early Science observations, and are regularly used by science team members.

1. High-performance processing for ASKAP

The Australian Square Kilometre Array Pathfinder (ASKAP, DeBoer et al. 2009) is a radio-synthesis telescope located at the Murchison Radio-astronomy Observatory (MRO) in Western Australia, featuring thirty-six 12m dishes equipped with Phased-Array Feed receivers. These provide a wide, 30 square degree instantaneous field-of-view by forming up to 36 separate dual-polarisation beams at once. This results in a high data rate: 70 TB of correlated visibilities in an 8 hour observation, requiring high-performance calibration & imaging pipelines.

The ASKAPsoft package has been developed specifically for meeting this challenging problem. Particular design decisions in the imaging algorithms, and the framework under which they run, have been made to meet processing requirements on memory and run-time, and scientific requirements such as dynamic range. The processing for ASKAP is described in Guzman et al. (2019), and the ASKAPsoft Science Data Processor codebase is available through the CSIRO Data Access Portal[1].

The ASKAPsoft software will run in pipelines on the Cray XC30 '*Galaxy*' supercomputer at the Pawsey Supercomputing Centre in Perth, Western Australia. Since ASKAP will be primarily a survey telescope, these pipelines must run in near-real-time, commencing once an observation is completed and completing in time to process the next. All calibration and imaging will be done in these pipelines, followed by the formation of astronomical catalogues, and the resulting data products (calibrated visibilities, images, cubes, and catalogues) will be stored in the CSIRO ASKAP Science Data Archive (see Huynh et al. (2019)).

[1] ASKAPsoft SDP 0.19.5: `http://doi.org/10.4225/08/5a0a447510141`

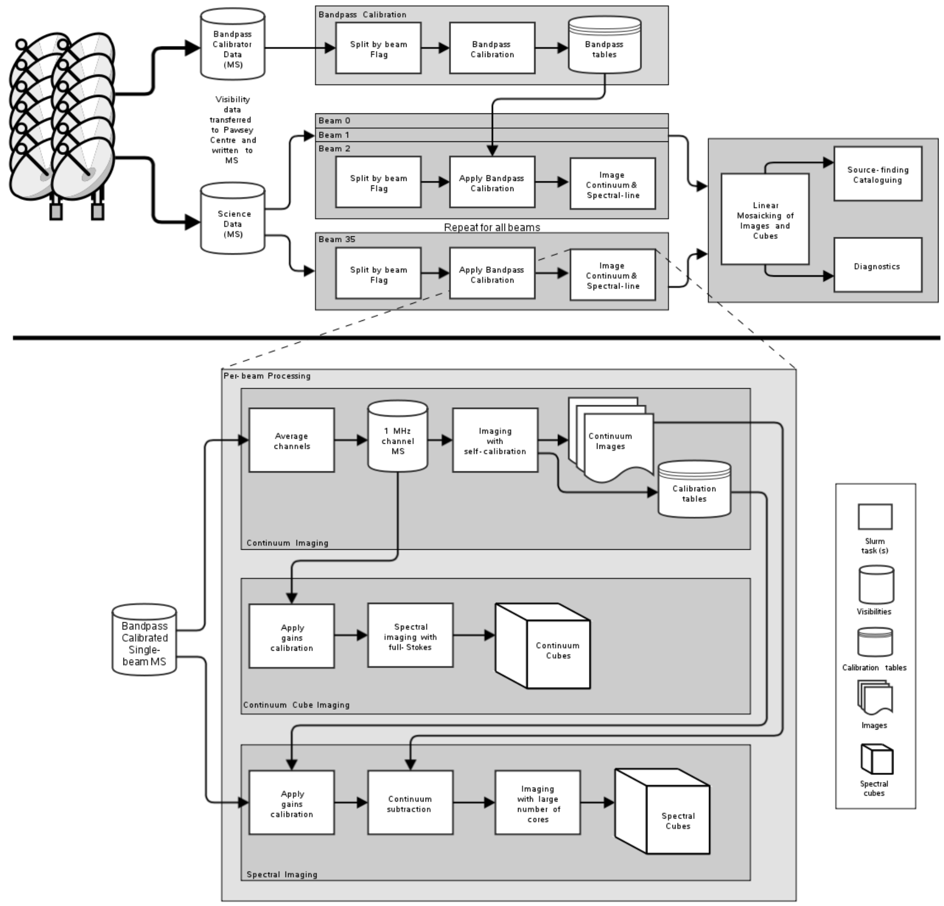

Figure 1. (top) A cartoon of the processing pipeline used for ASKAP Early Science, showing the steps required for end-to-end processing. The telescope writes a measurement set (MS) for offline processing at the Pawsey Centre. The bandpass is calibrated against an observation of the radio galaxy PKS B1934−638, the standard southern hemisphere flux density calibrator at these radio frequencies, then each of the 36 beams is processed separately, before being mosaicked together. Catalogues are then formed, and diagnostic & validation scripts are run. (bottom) For each beam, several imaging pipelines can be run: continuum; continuum-cube; and/or spectral-line pipelines, to produce image data products. The continuum pipeline includes self-calibration, to further refine the gains which are then applied to the visibility datasets.

2. ASKAP Early Science Pipelines

ASKAP is currently conducting its Early Science program with an array of 12 PAF-equipped antennas. This is a series of small surveys that allow ASKAP's science teams to address key scientific questions in the lead-up to the full-scale Survey Science Projects with the full 36-antenna array, and at the same time contribute to the development of ASKAP operations and processing.

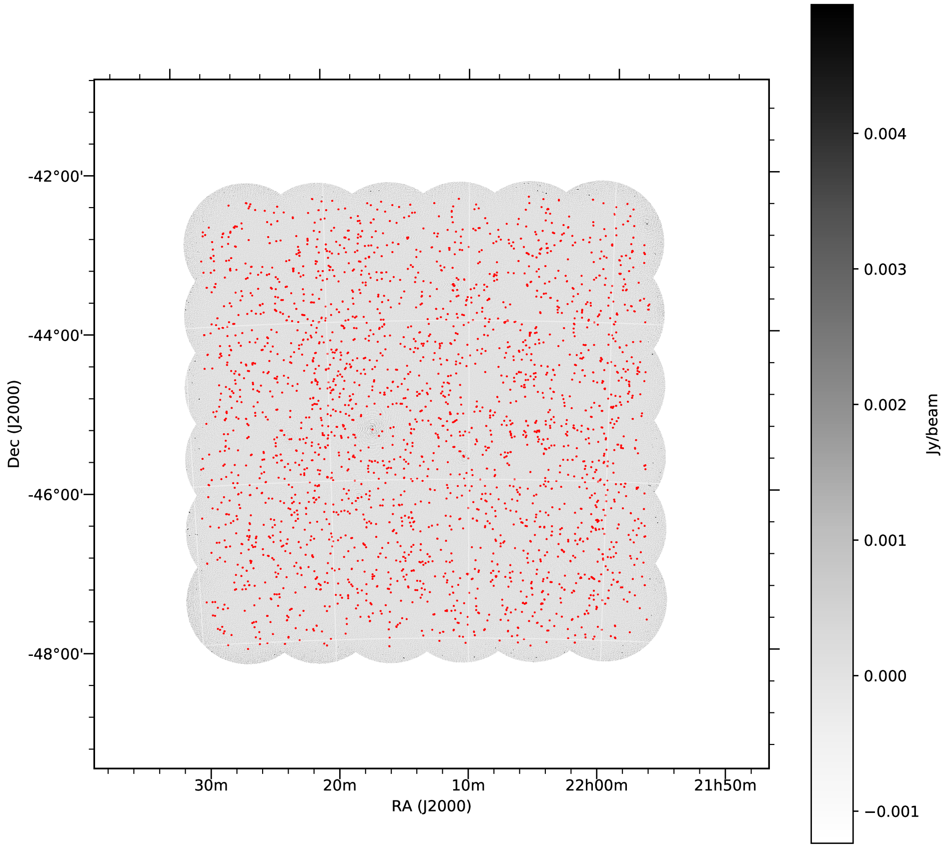

Figure 2. An example 36-beam continuum image, as processed by the ASKAPsoft pipeline. This is a single 12 hour observation of the field surrounding NGC 7232, taken at a frequency of 1400 MHz with a 48 MHz bandwidth: a total of 2643 components above 0.52 mJy/beam are detected in an area of 31.4 square degrees, and are shown by the red circles. This plot is produced automatically by the pipeline as part of the diagnostics process.

The ASKAP pipelines that have been developed for Early Science processing provide a simplified front-end to a processing workflow utilising the ASKAPsoft tasks. They launch a set of jobs to run on the *Galaxy* Cray XC30 as a pre-defined workflow, with a limited set of parameters exposed to the user. This lowers the bar to entry for science team members, making it easier for the science teams to understand the data and start doing science.

A schematic of these pipelines is shown in Fig. 1. The pipeline has a large amount of configurability, so that parts of it can be switched on and off as required, and the user has control of more than 400 configuration parameters to ensure the processing meets requirements.

There are three key components to running the ASKAPsoft tasks:

- **Shell scripts:** The bash scripts control the creation of the jobs, and ensure parameters are consistent across the entire pipeline run.

- **SLURM jobscripts:** Created by the bash scripts, these are scheduled on the supercomputer queue with appropriate dependencies.
- **Parameter sets:** Individual files for each task that parameterise the task. Created at run-time by the SLURM jobscript.

These pipelines have been used to publish datasets in the CASDA archive – the final step of ingesting into the archive is included as part of the pipeline workflow. An example of such data is shown in Fig. 2. This is a continuum image of the field surrounding NGC 7232, overlaid with the continuum component catalogue generated by the source-finding application SELAVY (Whiting & Humphreys 2012). Such images are produced by the pipeline as diagnostic images, and provided to the end-user through CASDA. Datasets from this and related observations are available now through CASDA[2]

3. Towards full ASKAP and SKA

The current pipeline-processing workflow is based on offline processing, where the processing is disconnected from the observation. As larger numbers of telescopes become incorporated into the array, the processing will need to move to an `online` mode of operation, where observations are processed immediately upon completion of the data ingest.

The pipelines will then need to evolve to better adapt to the available resources, and handle the additional complexity of real-time pipelines operating simultaneously with near-real-time processing. The ASKAP telescope and its pipelines will thus again play an important pathfinding role as we move towards the even more complex SKA Science Data Processor.

Acknowledgments. We acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site. This work owes much to the ASKAP computing team and to the ASKAP team who built this wonderful instrument. The ASKAP Commissioning & Early Science (ACES) team, and the wider ASKAP science teams, are to be thanked for their on-going work in commissioning the telescope and pipelines (without which the pipelines would not be in the shape they are). This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

References

DeBoer, D., et al. 2009, Proceedings of the IEEE, 97, 1507

Guzman, J. C., et al. 2019, in ADASS XXVII, edited by P. Ballester, J. Ibsen, M. Solar, & K. Shortridge (San Francisco: ASP), vol. 522 of ASP Conf. Ser., 31

Huynh, M., et al. 2019, in ADASS XXVII, edited by P. Ballester, J. Ibsen, M. Solar, & K. Shortridge (San Francisco: ASP), vol. 522 of ASP Conf. Ser., 263

Whiting, M., & Humphreys, B. 2012, Publications of the Astronomical Society of Australia, 29, 371

---

[2]The CASDA search page is at `https://data.csiro.au/dap/public/casda/casdaSearch.zul`, and datasets for the early science spectral-line project can be found at `http://hdl.handle.net/102.100.100/52548?index=1` (for images and visibilities) and `http://hdl.handle.net/102.100.100/52547?index=1` (for catalogues).