ASKAPsoft pipeline gets ready for the pilot surveys

Optimising the pipelines to support near real time processing of ASKAP pilot survey data

Mark Wieringa, Eric Bastholm, Paulus Lahur, Daniel Mitchell, Stephen Ord, Wasim Raja, Max Voronkov, Matthew Whiting (ASKAP SDP team)

(credit CSIRO)

CSIRO ASTRONOMY AND SPACE SCIENCE
www.csiro.au

The rapid growth of the ASKAP data volume of a typical 10h observation has required optimisations of code, pipeline and data layout to keep data processing times reasonable. We show what we've done so far and what is planned for the near future.

ASKAP current status (Oct 2019)

The ASKAP construction project has officially finished, with all 36 antennas operational with 36 simultaneous beams and 288 MHz bandwidth in the 700-1800 MHz range. Spectral resolution is 1 MHz in continuum mode (288 channels) and 18.5 kHz in line mode (15552 channels). Successful operation at full capability (all beams, antennas and bandwidth) from data capture to science ready data in the archive has been demonstrated for several fields.

The ASKAP-X project (extended modes) has just started and will extend the bandwidth to 336 MHz and commission correlator zoom modes and other new capabilities like commensal coherent FRB detection, non-contiguous bands, trading bandwidth vs beams, tied array, performance improvements and the processing cluster upgrade.

ASKAP is currently conducting the Pilot Surveys, subsets of the 8 large survey projects scheduled to run over the next few years. The aim is to verify that telescope and data processing performance is sufficient to reach the survey goals and identify changes in survey strategy, operations and processing pipelines needed before the full surveys are started.

Data Processing Model

Science data processing is done using the ASKAPsoft package on the Galaxy cluster (CRAY XC30) at the Pawsey Centre in Perth. Galaxy has 472 CPU nodes with 20 cores and 64 GB memory each and about 7 PB of storage. Specifications for its replacement have been drawn up and we expect an upgraded machine in about a year.

The original ASKAP processing model (ASKAP-SW-0020) envisaged a one-size-fits-all pipeline that would read the data once and produce calibrated image cubes in the archive. During commissioning it became clear a more traditional and flexible approach was needed where the visibilities are stored. We now have a pipeline with MPI-enabled tasks as building blocks.

The pipeline tasks do flagging, splitting and averaging, bandpass calibration, continuum imaging and self-calibration, continuum subtraction, spectral cube imaging and mosaicking as well as source finding and quality control. See Fig.1 for the data flow. The new approach requires a larger disk buffer to allow for intermediate data, quality checks and reprocessing where needed.

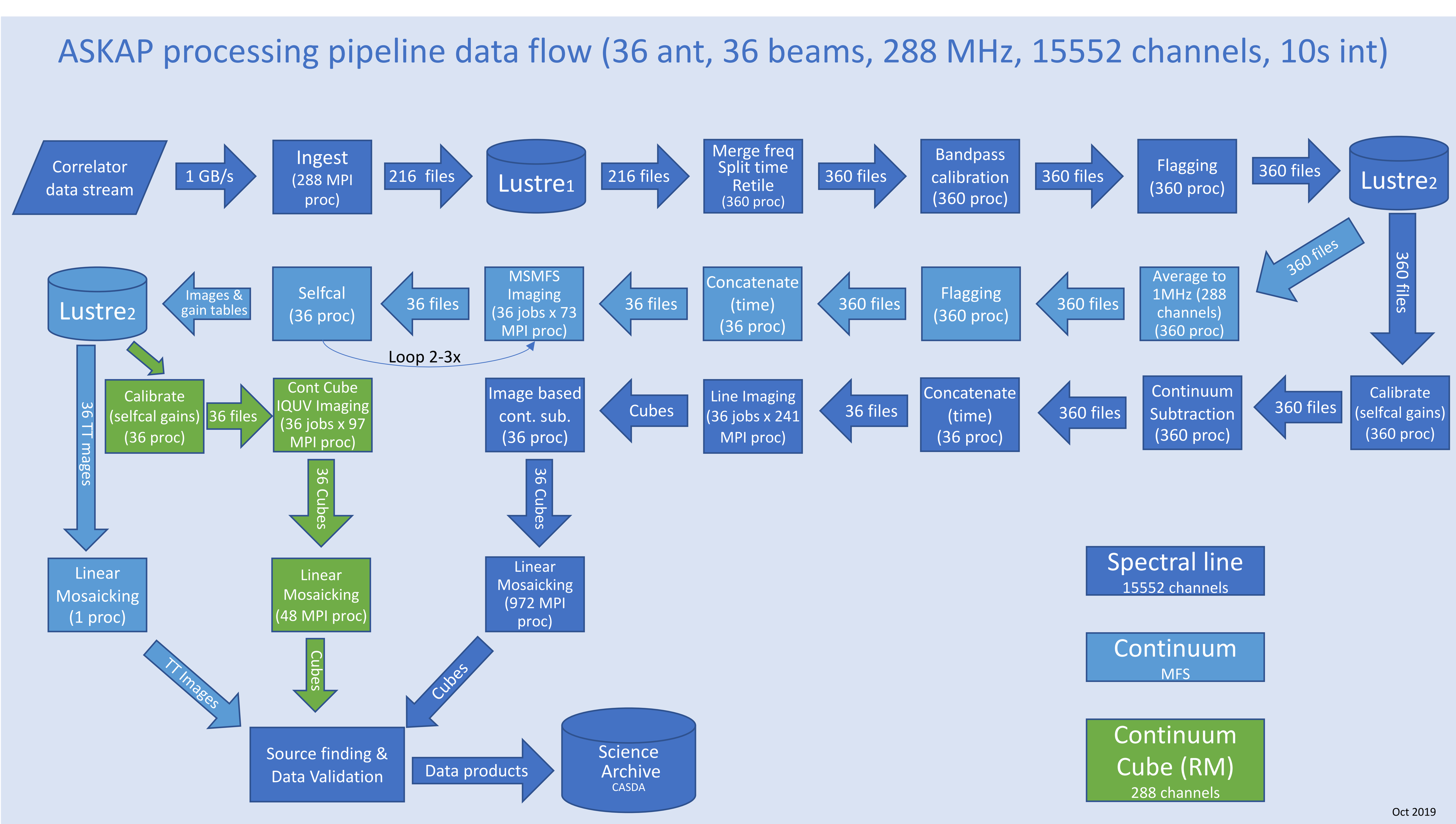

Figure 1: ASKAP processing pipeline dataflow. The Continuum self calibration generates the calibration and continuum model needed for the Spectral line processing. The continuum cubes allow study of sources with complex spectra and polarisation (Rotation Measure Synthesis).

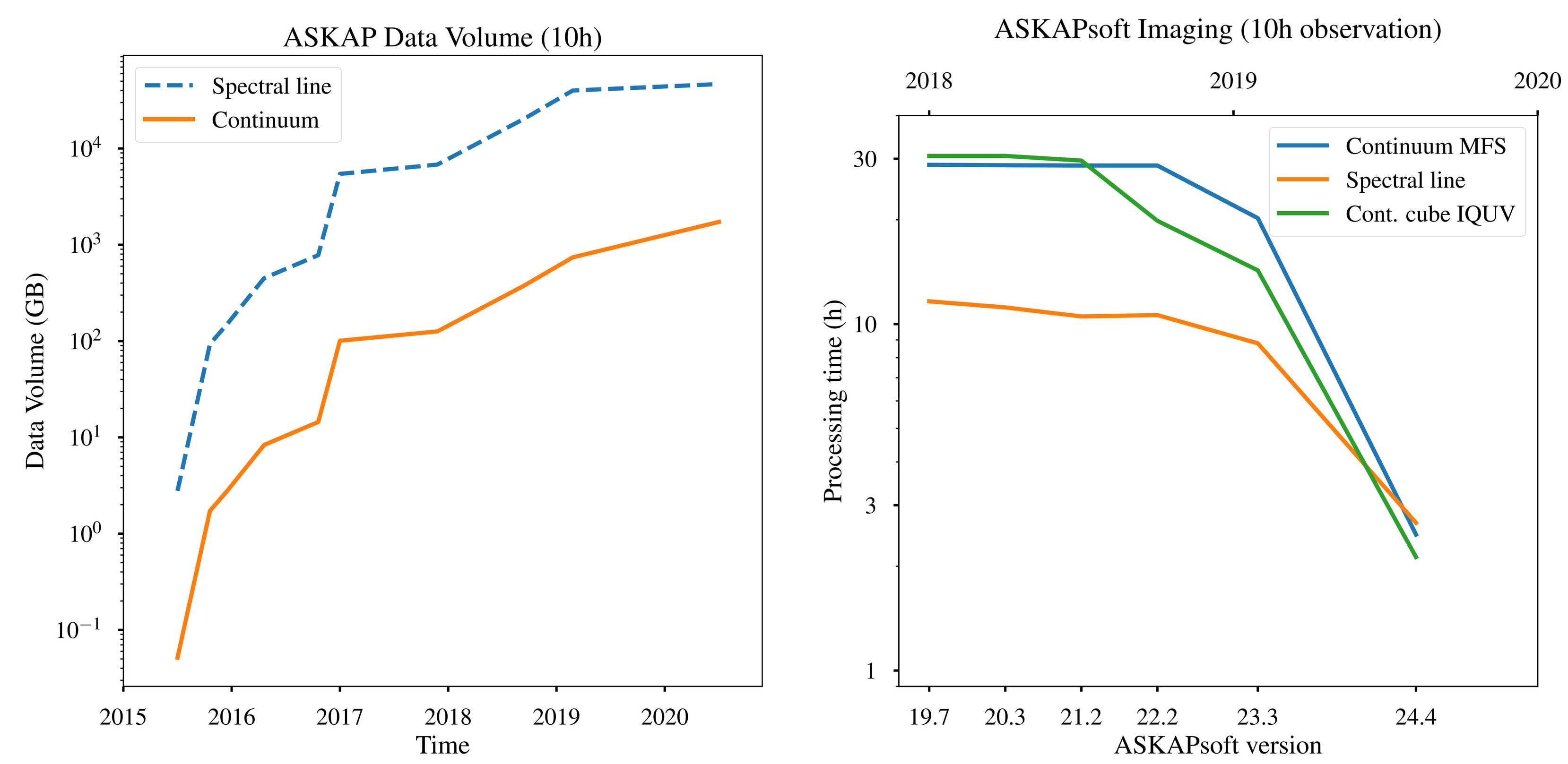

Figure 2: (left) Growth of ASKAP data volume for a 10h observation in spectral line and continuum mode as more bandwidth and more antennas come online (right) Decrease in processing time for imaging a 36 antenna, 288 MHz, 10h observation, using resources scaled to process all 36 beams on the Galaxy cluster at once.

Optimising the pipeline

Aim: keep the processing time from ingest to science ready archive data similar to the observing times for long spectral line observations (avoid backlog)

Use embarrassingly parallel nature of the data:

- Process 36 beams in parallel until linear mosaicking stage
- Split data in time (1h) for flagging, calibration and cont. subtraction

Use MPI to distribute work:

- Imager can flexibly distribute and iterate over work units (Taylor terms, channels) to fit job to available resources
- Distribute 36 imaging jobs over 468 (36 x 13) nodes with up to 20 workers/node and up to 64 channels per worker
- Minimise cross worker dependencies – e.g., use parallel writes to fits cubes.

Optimise DATA layout in MeasurementSet to match imaging access: retile from [many channels, fewer rows] required for ingest process to keep up with correlator to [single channel, many rows] optimal for imaging.

Parallelise tasks: calibration, mosaicking

Optimise imaging algorithms and parameters:

- Peak finding, beam patch size, gain, scales, w-snapshot/w-projection etc.

Code optimisations derived from profiling:

- standard serial techniques (loop order, move object creation out of loop etc),
- avoiding duplicate calculations (e.g., caching convolution kernels)

Outlook

Improvements in processing speeds tend to lead to higher expectations of the science products: better deconvolution (lower sidelobe levels, more scales), increased resolution, full polarisation processing etc. Over the next year a moderate increase in the data rate is expected, the bandwidth will increase by 48 MHz and the integration time will halve. This means the quest for better performance will not end.

Main areas for further improvement:

- Reduce memory use for Continuum Imaging (to fit more processes on a node)
- Optimise mosaicking (speed, memory); parallelise flagging, cont. subtraction
- Gain/Bandpass Calibration (use more efficient solvers, better constraints)
- Use OpenACC/OpenMP to parallelise Clean minor cycle (demonstrated)
- Multi-Scale MFS imaging: optimise MFS gridding, MPI grid merge

FOR FURTHER INFORMATION
Mark Wieringa
e mark.wieringa@csiro.au
w https://www.atnf.csiro.au/projects/askap/index.html

REFERENCES
ASKAPSoft documentation:
https://www.atnf.csiro.au/computing/software/askapsoft/sdp/docs/current/index.html
ASKAPSW-0020: https://www.atnf.csiro.au/projects/askap/ASKAP-SW-0020.pdf

ACKNOWLEDGEMENTS
We acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site.
This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.