Science & Technology Facilities Council

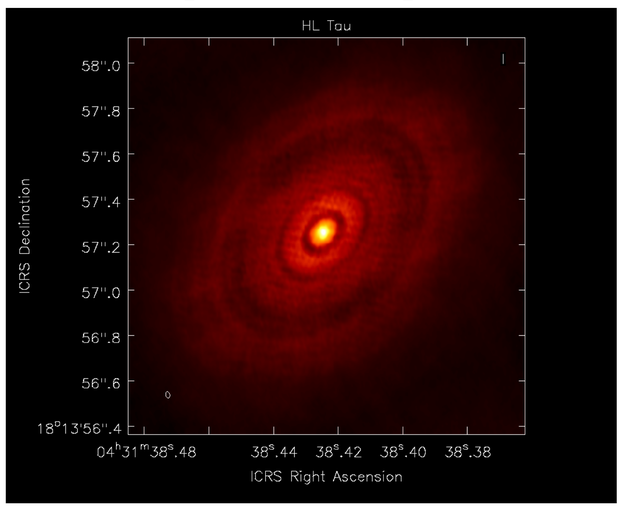
www.stfc.ac.uk

The ALMA Science Pipeline

V. C. Geers, S. Williams (UK ATC), D. Muders, F. Wyrowski (MPIfR), G. Kosugi,
T. Nakazato, K. Sugimoto (NAOJ), L. Davis, C. A. Hales, B. Kent, J. Kern (NRAO)

Overview

The Atacama Large Millimeter Array (ALMA) is a radio interferometer composed of 66 antennas located on the Chajnantor plateau at 5 km altitude in northern Chile. With baselines of up to 16 km, ALMA provides unprecedented spatial resolution and sensitivity in the mm/submm wavelength range to study star and galaxy formation. The ALMA project intends to reach a broad astronomical community. An automatic reduction pipeline is required to deliver science quality data products to radio expert & non-expert astronomers.



HL Tau Science Verification Data calibrated and imaged by the Cycle 4 Pipeline

Current Status

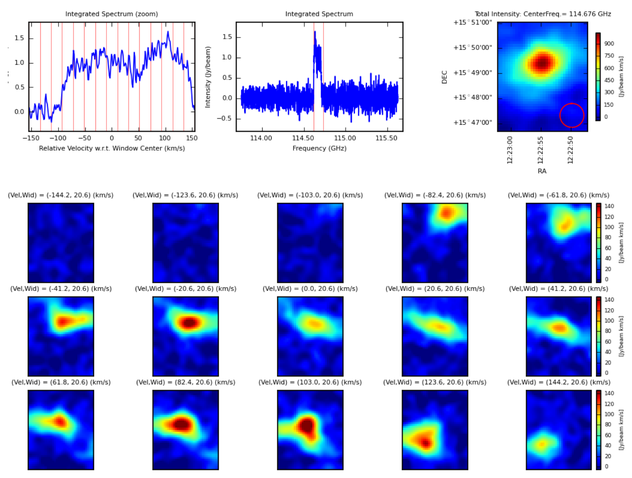
As of Cycle 4, Oct 2016, ALMA Pipeline (in CASA 4.7) provides for interferometry and single-dish:

- calibration & flagging
- images of the calibrator sources
- continuum / spectral line images of science targets

Recent imaging improvements:

- continuum frequency range finding
- dynamic range based clean thresholds

Plans for Cycle 5/6 improvements include auto-boxing in clean-mask selection, fine-tuning clean threshold heuristic, improved QA scoring, improved flagging, and linear polarimetry.



Single-dish channel maps

Pipeline Team

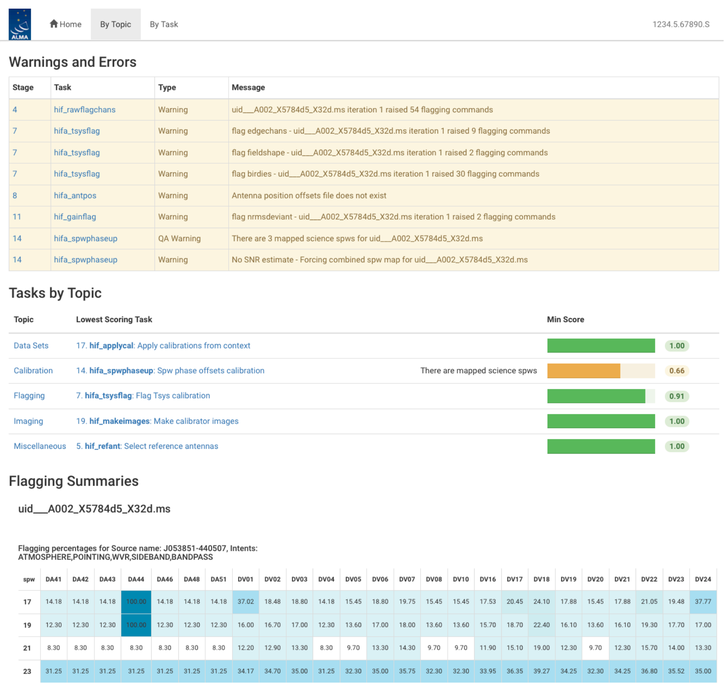
The ALMA pipeline team consists of a distributed group of developers from all ALMA partners, a working group to define requirements and goals, and a large group of testers at the ALMA Regional Centers (ARCs) and the Joint ALMA Observatory.

Pipeline Design

The ALMA Pipeline[1,2,3] is part of the CASA (Common Astronomy Software Applications) data reduction package[4], which consists of a set of C++ tools with an iPython interface, presenting a set of data reduction tasks. The pipeline tasks are designed to be modular, each can be run via task interface or automated python script. CASA includes post-processing tools for data analysis.

Heuristics

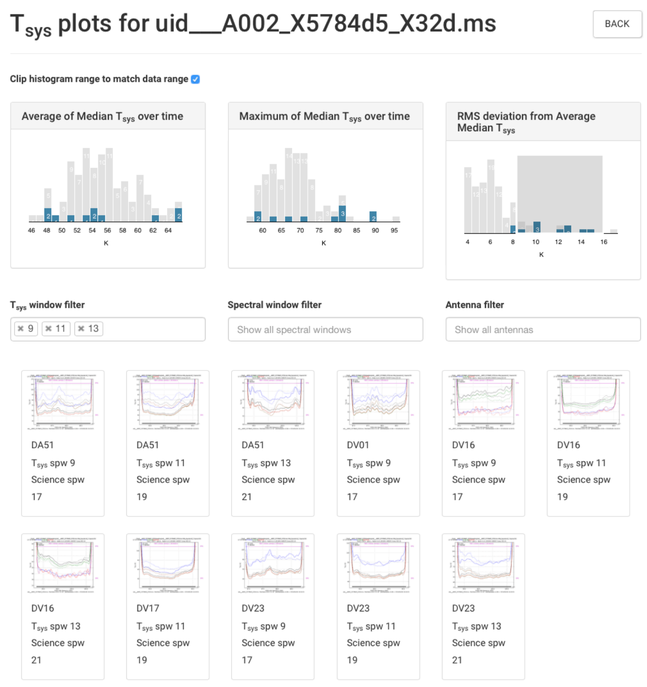
The pipeline heuristics capture the expert knowledge needed to process radio interferometry and single dish data and to perform the zero spacing combination. The algorithms and recipes are being iterated through the Pipeline Working Group.



By-Topic summary of warnings, scores & flagging

Web Log

The pipeline processing results are presented as a hierarchical set of web pages facilitating navigation from high-level summaries down to detail pages illustrating the data structure and contents, and as well as the calibration and imaging steps.



Histogram filtered plots

Quality Scores

The data processing and products are evaluated with Quality Assessment scores ranging from 0 (awful) to 1 (perfect), based on metrics such as RMS, S/N ratios, etc. Each reduction step, each topic (calibration, flagging, imaging) and the overall run are tagged with a score. Low scores point to problem areas to be inspected.

Book Keeping

To handle the complex observing mode setups, the pipeline keeps track of the metadata and the intermediate processing results, through special context, data domain, and results objects. A calibration library keeps track of all calibrations generated throughout the pipeline run and how to apply them to the data.

Pipeline Use

The CASA Pipeline is designed to take advantage of a range of processing capabilities, from high end workstation to small clusters, with various (parallelized) pipeline tasks able to benefit from high system memory, high bandwidth file I/O, and multiple cores. This provides the flexibility to support different use cases and processing environments:

At Joint ALMA Observatory

At the JAO, 83% of ALMA observing projects are processed with the pipeline, significantly reducing the manual workload on the science QA verification team.

At home institutes

PIs receive a data package produced by JAO with assistance, verification, and delivery[5,6] by their local ARC. The package contains the processing results ready for analysis, and a script to re-run the pipeline with desired modifications.

On Amazon Web Services

CASA and the ALMA Pipeline can be run on Amazon Web Services, visit our web portal at:
https://casa.nrao.edu/casa_aws_introduction.shtml
for info on:

- how to get started with pre-made Amazon Machine Images for specific CASA releases
- AWS instances & storage & monitoring
- recommended AWS configurations w.r.t. balancing RAM/CPU needs against costs

References

[1] http://adsabs.harvard.edu/abs/2012ASPC..461..185D
[2] http://adsabs.harvard.edu/abs/2014ASPC..485..383M
[3] http://adsabs.harvard.edu/abs/2015ASPC..499..355S
[4] http://adsabs.harvard.edu/abs/2007ASPC..376..127M
[5] http://adsabs.harvard.edu/abs/2014SPIE.9152E..0JP
[6] http://adsabs.harvard.edu/abs/2014SPIE.9149E..0ZS

Visit casa.nrao.edu

for more information on ALMA Pipeline and CASA.

Thank you to Todd Hunter (NRAO) for feedback on this poster.

ALMA Atacama Large Millimeter / submillimeter Array

Max-Planck-Institut für Radioastronomie

Science & Technology Facilities Council UK Astronomy Technology Centre

国立天文台 NAOJ National Astronomical Observatory of Japan

NRAO