

Imaging in CASA



Urvashi Rau

**National Radio
Astronomy Observatory**

CASA Users Committee

3rd October 2019



Topics of current focus

The past year :

ALMA pipeline support for Cycle 7 :

- Ephemeris source imaging
- Numerous numerical edge-effects with cube partitioning
- Automasking wrap-up
- Debugging mosaic PB issues

VCLASS Epoch 1.1 and Epoch 1.2 imaging :

- Time-dependent pointing corrections (homogeneous and heterogenous)
- Numerous wideband mosaic imaging details (PB-in-PSF, position-dependent PSF, conjbeams)

Current Topics :

- Focus on parallelization and resource management
 - Numerical consistency, memory use, unifying operational structures
- Algorithms and features
 - Address the root causes of various mosaicing numerical/algorithm symptoms
 - Joint Single dish and interferometric imaging
- Functional and stakeholder verification tests , numerical characterization

Parallelization and Resource Management

Cube parallelization

- Make numerics identical to serial cube mode, no 'too many files', etc..
- Single mechanism for cube partitioning (consolidate chunking and parallelization)

Prepare for self-calibrations in the pipelines

- Clean up interfaces w.r.to parallelization, savemodel and restarts
- Experiment with a parallel write approach, otherwise, default to serial writes.

Memory use and compute time

- Mem reduction for VLASS : Separate memory use of the major and minor cycle
- Compute speedup for MTMFS (via simple algebraic reduction of a math step...)

Other

- Bug fixes for several " List of MS vs concatenated MS " issues/differences w/wo parallelization.
- Ephemeris imaging with parallelization (required for ALMA Cycle 8)
- Clean up the PySynthesisImager scripting interface to better allow for atomic operations
 - Useful for scripting, algorithm prototyping, evaluating HTC operation solutions
 - e.g. ALMA pipeline makes the PSF ~10 times (we should remove PB calcs from makePSF)

Algorithms – New features, R&D

Separating Algorithm R&D from production CASA :

- Classify topics as experimental versus ready for production
- Schedule either within CASA or hand over to ARDG until ready for production
- Numerical characterization and analysis prior to inclusion in production tclean.
- (Existing features and oddities are being handled on a case-by-case basis)

Topics of recent focus :

- **Spectral cube weighting** per channel (CASA)
- **VLASS pointing corrections** for Epoch 1.1 (casa 5.7) and 1.2 (casa 5.6) wideband mosaics
 - ARDG : Homogenous & heterogenous time-dependent pointing corrections for ‘awproject’
 - CASA : Write two levels of verification tests. Evaluate same experimental features for ‘mosaic’
- **Joint Single-Dish and Interferometer imaging** for Spectral Cubes and MTMFS
 - ARDG : WidebandSDINT : Prototyped, published, tested at SD+INT workshop in Leiden.
 - CASA : Wrap above scripts into a CASA task (scheduled for casa 5.7)
- **Wideband Mosaics** with position-dependent wideband PSFs
 - CASA/ARDG : Simulations and tests to demonstrate/isolate the root algorithmic issues
Investigation of solutions is ongoing.
- Ongoing : Full Mueller imaging for ALMA and VLA + CFCache library for awproject (ARDG)

Verification Tests (more info in Sandra Castro's talk)

Functional Verification Tests

Ongoing : Finding and filling gaps in our imaging test suite (casa 5.6, and continuing)

- Added ~30 tests for spectral cube and wideband mosaics with three gridders
- Added a few tests for weighting schemes regularly used by ALMA cubes
- test_tclean.py now tests about 240 different use cases.

Stakeholder Verification Tests

Purpose : Creatively test algorithmic options exercised by CASA's dominant stakeholders

- Define tclean calls on real data, with agreed-upon measurables and metrics/tolerances.
- To be used to detect and quantify numerical effects of code changes

- Example :

ALMA : Measure flux accuracy, PB levels, noise, PSF beam ratio,
for a single pointing, mosaic and ephemeris source.

Ensure consistency across different runtime modes (parallelization, chunking)

VLASS : Use one row-of-pointings to test pointing corrections, wideband pbcor,
Spectral index for a single pointing versus joint mosaic, etc.

Status : Initial versions of tests and infrastructure are in progress (to be used from 5.7 onwards)

Numerical Characterization – Short Memos (as part of sci-development)

Purpose : Rigorous analyses of algorithmic features, prior to user-testing

Move away from ad-hoc experimental changes for specific use cases.

Examples :

- **Re-restoration by an external restoring beam had a scaling error**
 - Impact of the bug was analyzed analytically and with simulations, and guidelines provided to users
- **Improving the use of scalebias for MS-Clean and MT-MFS (with MS)**
 - Math, with imaging examples of the numerical effects, including understanding edge-cases.
- **Weighting schemes for spectral cubes**
 - Test and evaluate PSF beam shape changes with different weighting schemes and binning
- **Algebraic reduction of a compute step within MTMFS (wideband imaging)**
 - Demo runtime speedups, identify situations of maximal benefit, confirm unchanged numerics
- **Wideband Mosaics (and position-dependent PSFs) [work is ongoing]**
 - A summary of the math along with an explanation of the problem (and symptoms)
- **Rectangular Mosaics [work is ongoing]**
 - Review/revise gridding of rectangular images ('mosaic' and 'awproject'), characterize numerics.
- **Memory separation of major and minor cycle [work is ongoing]**
 - Tests to assess usefulness/impact (all gridders/deconvolvers) and numerical equivalence.

Future Infrastructure work (and relation to CNGI/ngCASA)

Timeline :

- No more major refactors or rewrites.
- We have until the end of 2021 to work on all semi-major infrastructure changes.
- From 2022 until CNGI and ngCASA are ready for use (~few yrs) we will go into 'support mode'.

Some topics to deal with in the next year+ :

- Move the simulator visibility prediction, setjy, and ft to use the current imager framework
- Consolidate parallelization and memory management schemes for cube and continuum
 - Channel chunking vs cube partitioning vs Multi-MS for the frequency axis
 - List-of-MS vs concatMS vs reftable partitioning for the row axis
 - Find some workable solution to (or workaround for) parallel writes to an MS
- Define clearer functional units/steps to enable experimentation and diverse usage patterns
 - Move towards more modular operations (consistent with casa6 and CNGI/ngCASA)
- New algorithms as needed by ALMA and VLASS (e.g. polarization imaging)

[Note : Similar topics are being evaluated for the rest of CASA as well]