ASTRON

Netherlands Institute for Radio Astronomy

LOFAR Imaging Busy Week XXI

Toward a direction dependent pipeline

Summary of activities

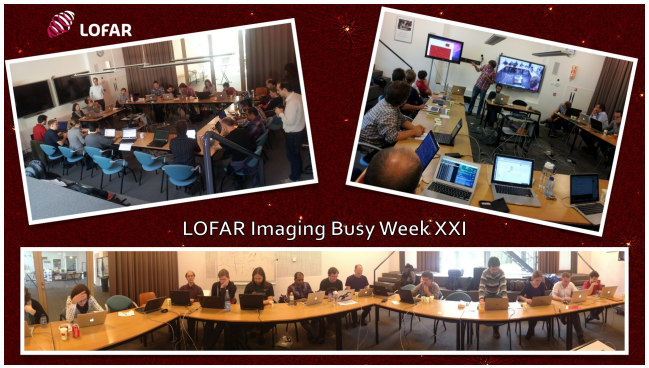
George Heald
(on behalf of the Busy Week crew)
LSM, 17/09/2014

ASTRON is part of the Netherlands Organisation for Scientific Research (NWO)

LOFAR

Busy week XXI: overview

- 1-5 September 2014 at ASTRON
  - Monday: overview of CITT progress, input from science teams
  - Tuesday-Friday: Group work, progress/summary discussions
- Participants: Pepe Sabater, Martin Hardcastle, Tim Shimwell, Leah Morabito, Wendy Williams, Björn Adebahr, Cyril Tasse, Andreas Horneffer, Stefan Fröhlich, Rocco Coppejans, Francesco de Gasperin, Soobash, Marisa Brienza, Pandey, Aleksandar Shulevski, Reinout van Weeren, Alexander Drabent, Maaijke Mevius, George Heald, Nicolas Vilchez, Bas van der Tol, Tammo Jan Dijkema, David Rafferty, Carmen Toribio, Emanuela Orrù



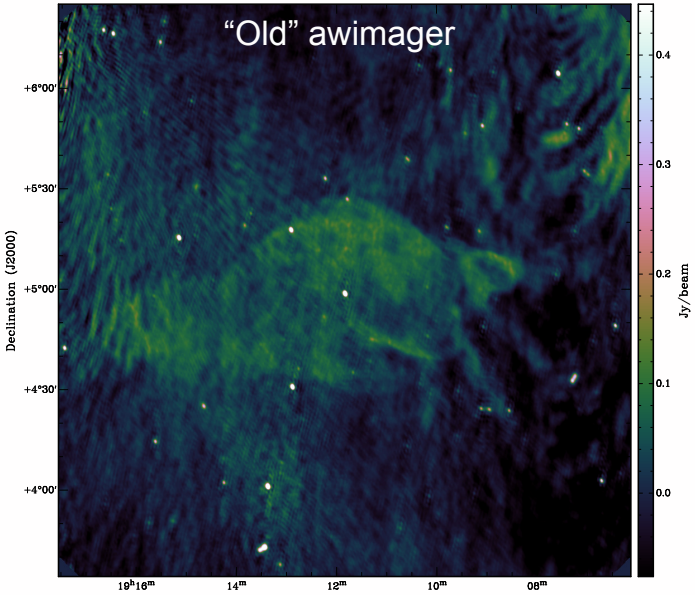
Topics

- awimager
- NDPPP
- Calibration strategy
- EoR tools
- Selfcal
- LoSoTo + LSMtool
- Beam
- Phase screen
- Smart demixing

… stay tuned for two bonus rounds!

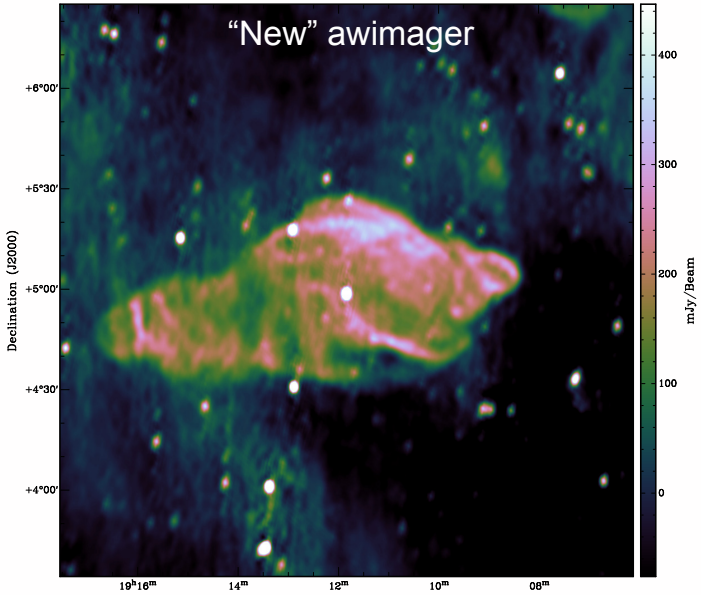
(1/9) awimager

- Multiscale clean using new awimager (Jess Broderick et al)



(1/9) awimager

- Multiscale clean using new awimager (Jess Broderick et al)



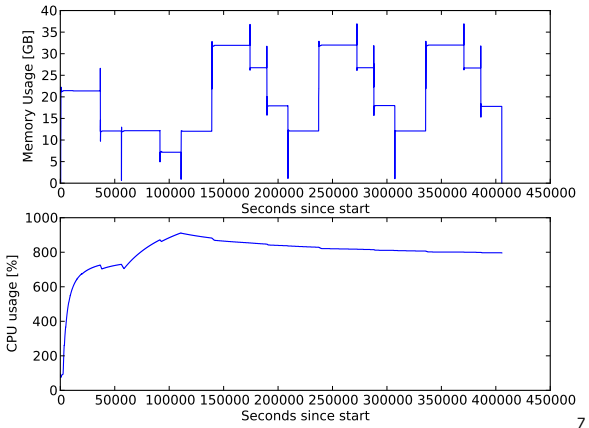
(1/9) awimager

- Test python beam interface (Maaijke Mevius et al)

…see slides from Maaijke after this presentation

(1/9) awimager

- Check memory usage (Björn Adebahr)
  - Used psutil python package to track resource allocation
  - tested nterms=1,2
  - some ‘warmup’ time (several minutes) observed
  - only uses all cores when nterms=2 (memory allocation error?)
  - smaller memory footprint than before, but still high
  - memory usage scales with image size and data size
  - slower than casa (30-50x)

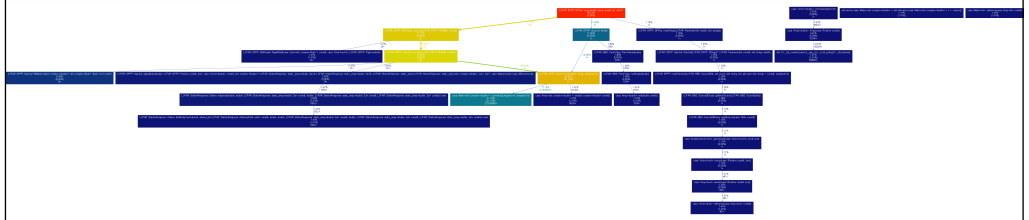


(1/9) awimager

- Update: Stefan Fröhlich (Jülich) utilizing GNU profiler to assess runtime in awimager internals (toward identifying bottlenecks)

(2/9) NDPPP & gaincal

- GNU profiler run on NDPPP shows expected behavior (good!) and a much clearer structure



- Initial assessment shows that a fairly high time is spent in the casa data structures (awimager more than NDPPP), so memory operations seem to be a chokepoint. Further work needed though.

(2/9) NDPPP & gaincal

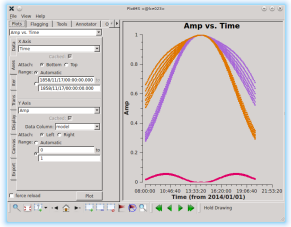
- Test solint >1, commonscalarphase, model data

… they work!

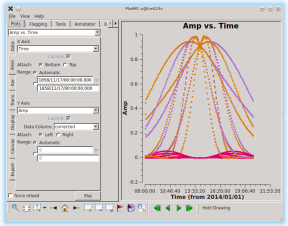
- test beam + predict (Andreas Horneffer et al)

NDPPP Beam-Issues

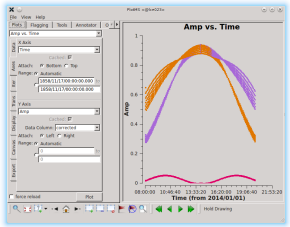
BBS



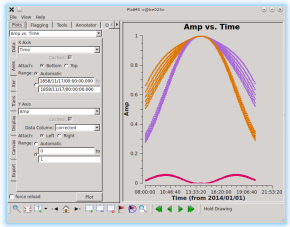
NDPPP – start



NDPPP – after fixing frequencies

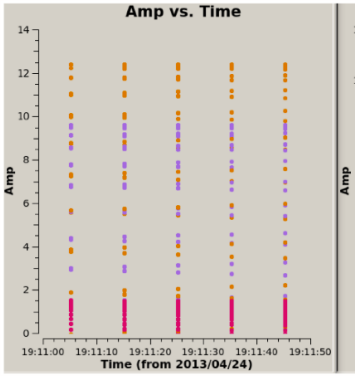


NDPPP – final (fixed reference frame)

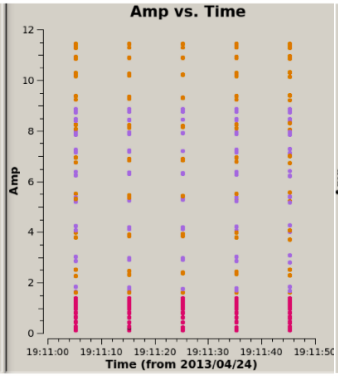


(2/9) NDPPP & gaincal

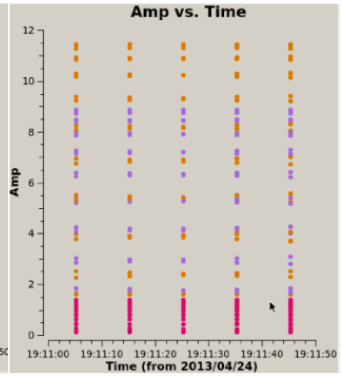
- onebeamperpatch option (Reinout van Weeren et al)



onebeamperpatch=True



Two patches (reference)



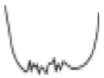
onebeamperpatch=False

(2/9) NDPPP & gaincal

LOFAR ASTRON

- test convergence (Martin Hardcastle et al)

- When minimum in landscape is not well defined, iteration stalls. This is detected and iteration stops.

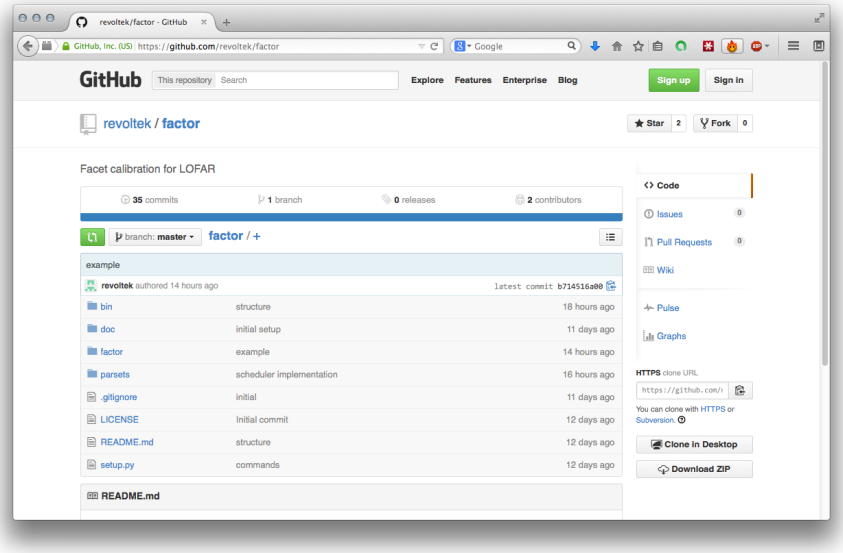


- Result: when it works it works well, with very weak dependence on number of iterations, or tolerance parameter. However in some cases stall detection is triggered too fast? Under investigation.

(3/9) Calibration strategy

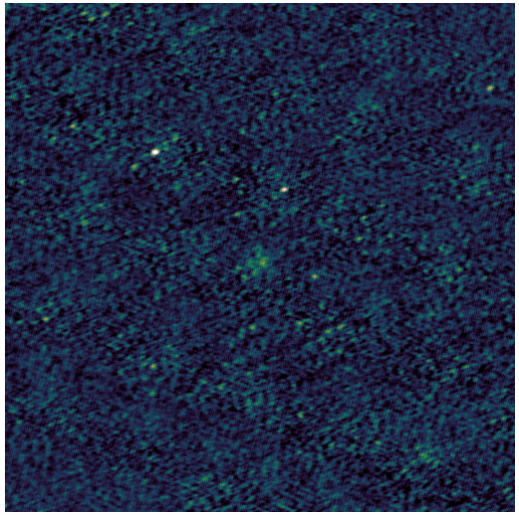
LOFAR ASTRON

- New project in development: Factor (de Gasperin et al)



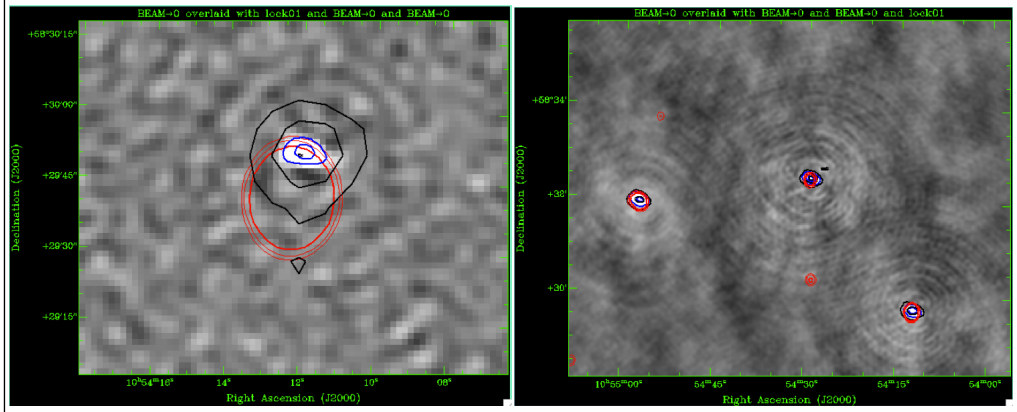
(4/9) EoR tools

- excon: good & fast mechanism to identify timeslots with terrible ionospheric behavior
- This test: Toothbrush field (van Weeren et al)
- 20” pixels, 1 min frames, outer uv cut 10 klambda



(5/9) Selfcal tests

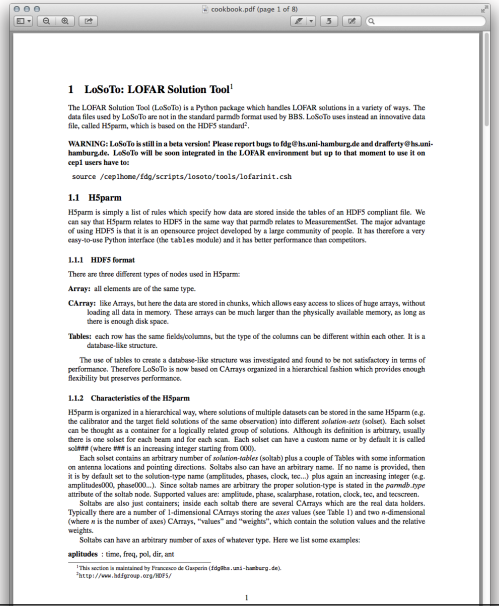
- Source positions shift (Nicolas, Elizabeth et al)
- Solution: start at higher resolution



(6/9) LoSoTo & LSMtool

- LoSoTo: outlier rejection module implemented, see docs

(Francesco's github repository, to be merged into LOFAR cookbook at next release)



cookbook.pdf (page 1 of 8)

1 LoSoTo: LOFAR Solution Tool[1]

The LOFAR Solution Tool (LoSoTo) is a Python package which handles LOFAR solutions in a variety of ways. The data files used by LoSoTo are not in the standard parmdb format used by BBS. LoSoTo uses instead an innovative data file, called H5parm, which is based on the HDF5 standard[2].

WARNING: LoSoTo is still in a beta version! Please report bugs to fdg@hs.uni-hamburg.de and dradferty@hs.uni-hamburg.de. LoSoTo will be soon integrated in the LOFAR environment but up to that moment to use it on cep1 users have to:

```
source /cep1home/fdg/scripts/losoto/tools/lofarinit.csh
```

1.1 H5parm

H5parm is simply a list of rules which specify how data are stored inside the tables of an HDF5 compliant file. We can say that H5parm relates to HDF5 in the same way that parmdb relates to MeasurementSet. The major advantage of using HDF5 is that it is an opensource project developed by a large community of people. It has therefore a very easy-to-use Python interface (the `tables` module) and it has better performance than competitors.

1.1.1 HDF5 format

There are three different types of nodes used in H5parm:

Array: all elements are of the same type.

CArray: like Arrays, but here the data are stored in chunks, which allows easy access to slices of huge arrays, without loading all data in memory. These arrays can be much larger than the physically available memory, as long as there is enough disk space.

Tables: each row has the same fields/columns, but the type of the columns can be different within each other. It is a database-like structure.

The use of tables to create a database-like structure was investigated and found to be not satisfactory in terms of performance. Therefore LoSoTo is now based on CArrays organized in a hierarchical fashion which provides enough flexibility but preserves performance.

1.1.2 Characteristics of the H5parm

H5parm is organized in a hierarchical way, where solutions of multiple datasets can be stored in the same H5parm (e.g. the calibrator and the target field solutions of the same observation) into different *solution-sets* (solset). Each solset can be thought as a container for a logically related group of solutions. Although its definition is arbitrary, usually there is one solset for each beam and for each scan. Each solset can have a custom name or by default it is called sol###### (where ### is an increasing integer starting from 000).

Each solset contains an arbitrary number of *solution-tables* (soltab) plus a couple of Tables with some information on antenna locations and pointing directions. Soltabs also can have an arbitrary name. If no name is provided, then it is by default set to the solution-type name (amplitudes, phases, clock, tec...) plus again an increasing integer (e.g. amplitudes000, phase000...). Since soltab names are arbitrary the proper solution-type is stated in the *parmdb_type* attribute of the soltab node. Supported values are: amplitude, phase, scalarphase, rotation, clock, tec, and tecscreen.

Soltabs are also just containers; inside each soltab there are several CArrays which are the real data holders. Typically there are a number of 1-dimensional CArrays storing the *axes* values (see Table 1) and two *n*-dimensional (where *n* is the number of axes) CArrays, "values" and "weights", which contain the solution values and the relative weights.

Soltabs can have an arbitrary number of axes of whatever type. Here we list some examples:

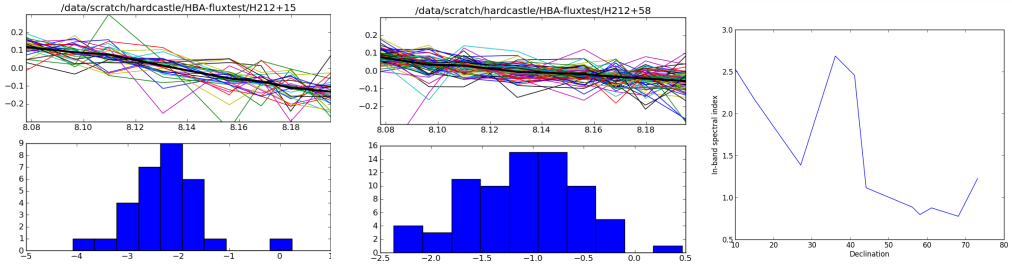
aplitudes : time, freq, pol, dir, ant

[1] This section is maintained by Francesco de Gasperin (fdg@hs.uni-hamburg.de).
[2] http://www.hdfgroup.org/HDF5/

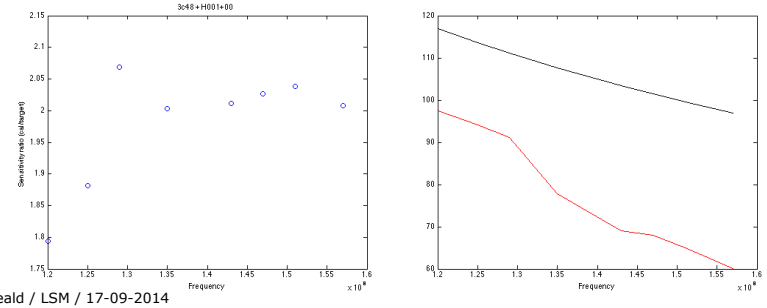
1

(7/9) Beam

- Grating lobe issue noted by Ger, we checked for it in MSSS data:



- Obtained code from Stefan Wijnholds, now assessing impact



(8/9) Phase screen

LOFAR ASTRON

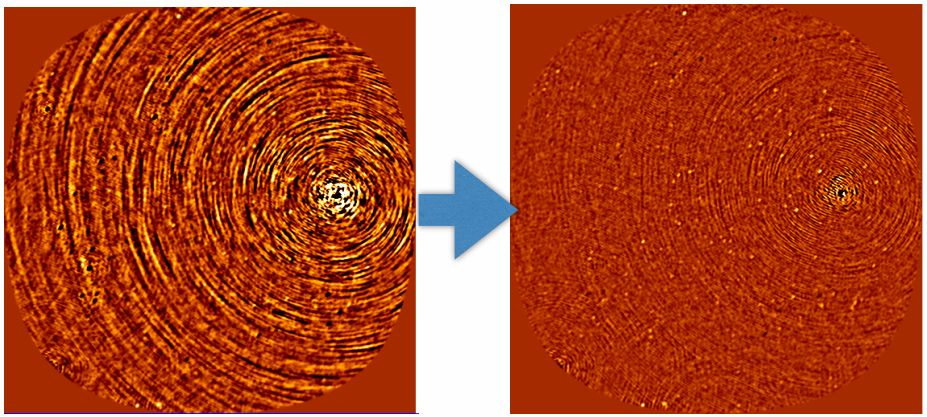
- HBA phasescreen applied to Toothbrush data (David Rafferty, Maaijke Mevius)
- Results still in progress
- Early optimism came from improper phase application
- Most recent results suggest minimal improvement from screen (?)
  - Ongoing…

(9/9) Smart demix

- Check on smart demix performance and do final verification for inclusion in normal observations (Manu & Carmen)
- Sources were not being subtracted (due to different amplitude scale between BG and Cobalt)
  - now shown that subtraction occurs when scale is adjusted (demix.defaultgain)
  - conclusion: it works, but last steps of investigation underway to determine whether new parameters are optimised
- Feedback / assistance from users welcome!
  (science support is overloaded…)

Bonus #1: killMS

- killMS (Cyril Tasse): direction dependent calibration scheme
- subtraction of calibrated skymodel; simultaneous solve in all directions
- Test 1: 1.25h, 10sb (36 ch), 106 sources (in 10 directions), HBA (1700 BL)
  Runtime: 20 minutes



Bonus #2: LOFAR software stack

- Official releases of LOFAR imaging software are being deployed on LTA connected systems:
  - SURFsara Grid cluster
  - NIKHEF Grid cluster
  - RUG grid cluster
  - RUG HPC cluster (Milipede)
  - Jülich HPC cluster (JUROPA)
- Build script is available for the LOFAR community
  - Builds dependencies from source
  - Depending on local installation building certain dependencies can be disabled
  - Uses read-only access to LOFAR SVN
- Resources for support are limited
  - Contact details on wiki page
  - Personal workstation installations supported on best effort basis only
  - Installations on (shared) cluster environments more likely to be considered
- http://www.lofar.org/wiki/doku.php?id=public:software_stack_installation

Full summary on wiki!