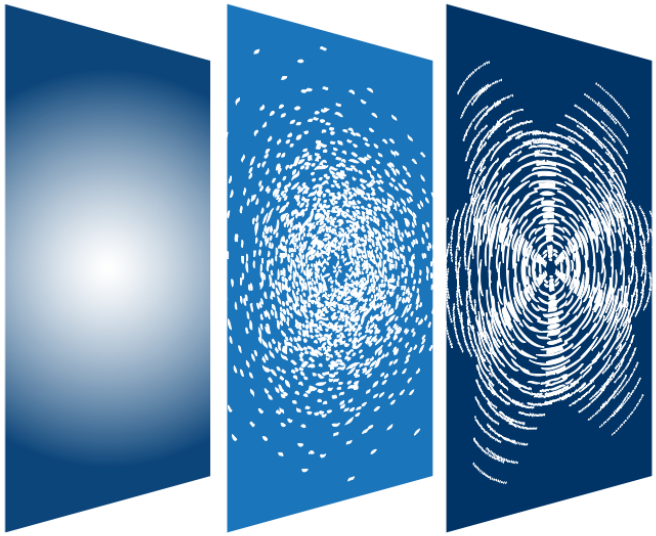


CASA

Common Astronomy Software Applications

Performance

Current Status and Future Plans

Jeff Kern
NRAO

Hardware this Week

- The computer on the table is not the system you have been running against.
- We have been using the NM Cluster [13 nodes] this week.
  - 16 Core Intel Xeon 2.53 GHz
  - 64 GB RAM
- The IO subsystem is Lustre
  - 667 TB
  - 40 Gbit/s Infiniband NIC
- Cluster available for external use:
  - https://info.nrao.edu/computing/guide/astronomySupport/clusterSchedNM

Memory

- CASA assumes that the visibility data is too large to hold in memory.
  - True for modern ALMA and VLA data sets
  - Implies that imaging is the dominant stage in terms of memory usage

- CASA attempts to hold only the necessary information consistent with efficient operation in memory.
  - For basic imaging (no A- or W-terms, single scale, single frequency term) the peak memory usage is approximately 24 bytes / pixel
    - 1 Double Precision Complex Grid
    - 1 Single Precision Complex Grid
  - For spectral line this is per output channel

Image Size	Memory Required
3k x 3k	206 MB
6k x 6k	864 MB
10k x 10k	2.4 GB

Data Access (Disk I/O)

- Although some stages of processing access only a fraction of the data, many operations require accessing the full visibility data
  - For many operations the I/O access either in terms of absolute data rate or number of transactions is the limiting factor in CASA performance.
- CASA is working on improvements which decrease the data access requirements, and ensuring that reads are as efficient as possible.

	Peak Data Rate	OS Realized Data Rate	Time for 500 GB file
SATA Disk	115 MB/s	60 MB/s	2h 22 m
Raid	200-500 MB/s	375 MB/s	22 m
Lustre (10 GB)	1.2 GB/s	900 MB/s	9.5 m

One The Fly Calibration

- Approach: Decrease the amount of I/O to improve overall performance
- Remove need to call *applycal* just to see if your calibration "worked"
- Specify calibration instructions for multiple calibration tables in a file, to be processed on-the-fly in a single execution
- Example calibration file:

```
caltable='ngc5921.bcal' calwt=True tinterp='nearest'
caltable='ngc5921.fluxscale' calwt=True tinterp='nearest' fldmap='nearest'
# comment allowed in callib file
caltable='ngc5921.gcal' calwt=True field='0' tinterp='nearest' fldmap=[0]
caltable='ngc5921.gcal' calwt=True field='1,2' tinterp='linear' fldmap=[0,1,1,3]
```

- Further documentation on the calibration library syntax and limitations can be found in Appendix G of the CASA Cookbook:
  - http://casa.nrao.edu/docs/UserMan/casa_cookbook018.html

OTF Calibration in PlotMS:

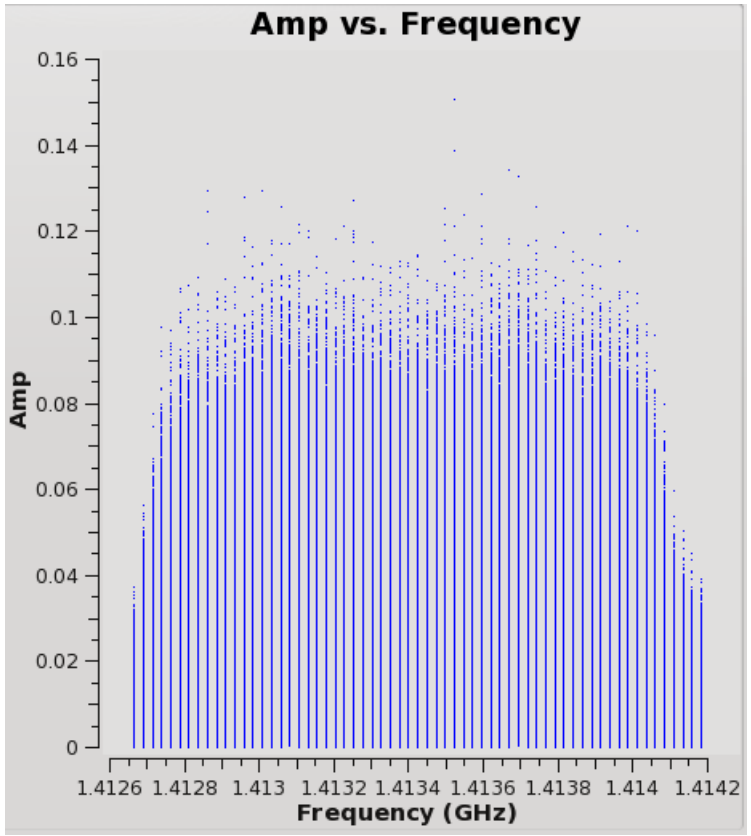
View calibrated data without first running applycal

- Specify this calibration library file in plotms with the ***callib*** parameter
- View corrected data by setting the ***datacolumn*** to 'corrected'
- Some cal tables (BPOLY, GSPLINE) are not yet supported
- Example plotms command:

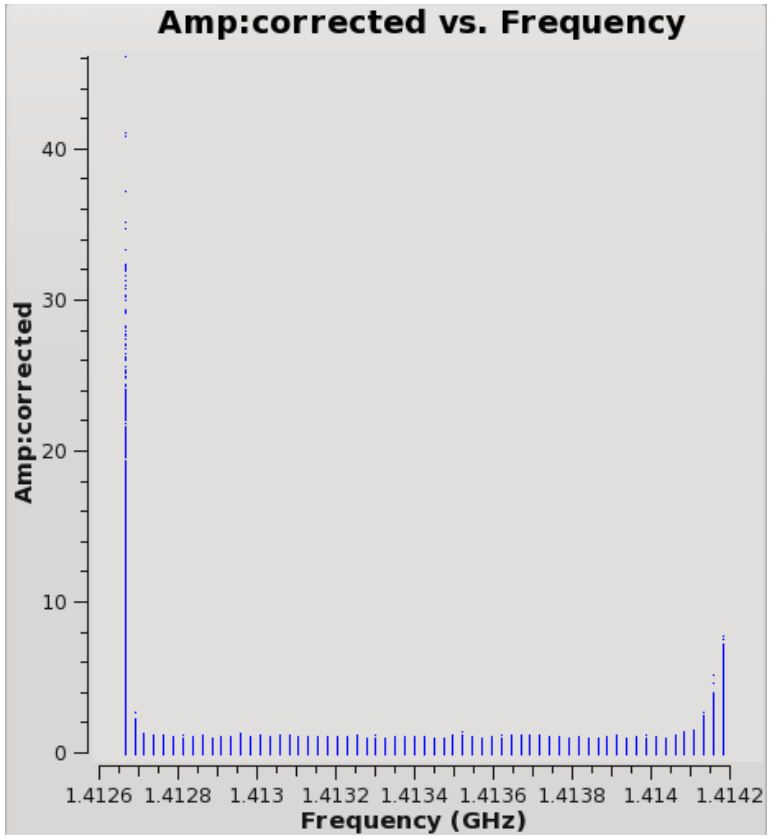
```
plotms(vis='ngc5921.ms', xaxis='frequency', yaxis='amp',
ydatacolumn='corrected', field='N5921_2, antenna='*&*', callib='callib.txt')
```

OTF Calibration in Plot MS

Observed Amplitudes



Corrected Amplitudes



Threaded Gridder

- tClean is a new experimental clean implementation.
  - Will eventually replace clean and the primary implementation

- Several gridders make use of a threaded implementation
  - Thread: Single process space, multiple cores, sharing memory
  - Mosaic, GridFT (basic) use 4 threads by default
  - W-Projection will use as many as possible by default
  - Controlled through *OMP_NUM_THREADS* variable

Multi-Process Implementation

- CASA now has a multi-process implementation based on MPI
  - http://casa.nrao.edu/docs/UserMan/casa_cookbook011.html
- Filling through calibration are implemented through the Multi-MS and are completely data parallel.
  - Available for general use, but still in "early adoption" phase.
- tClean has a "parallel" parameter which if set to True and running in an MPI environment will use multiple cores.
  - Demonstrated up to 106 cores so far.
  - Resource usage requires care
  - Continuum has been fairly well tested, cube is being tested currently.

Questions

This VLA mosaic produced using AW-Projection and CASA parallelization framework.

Major cycle decreased nearly 50x
(10 days to 5 hours).

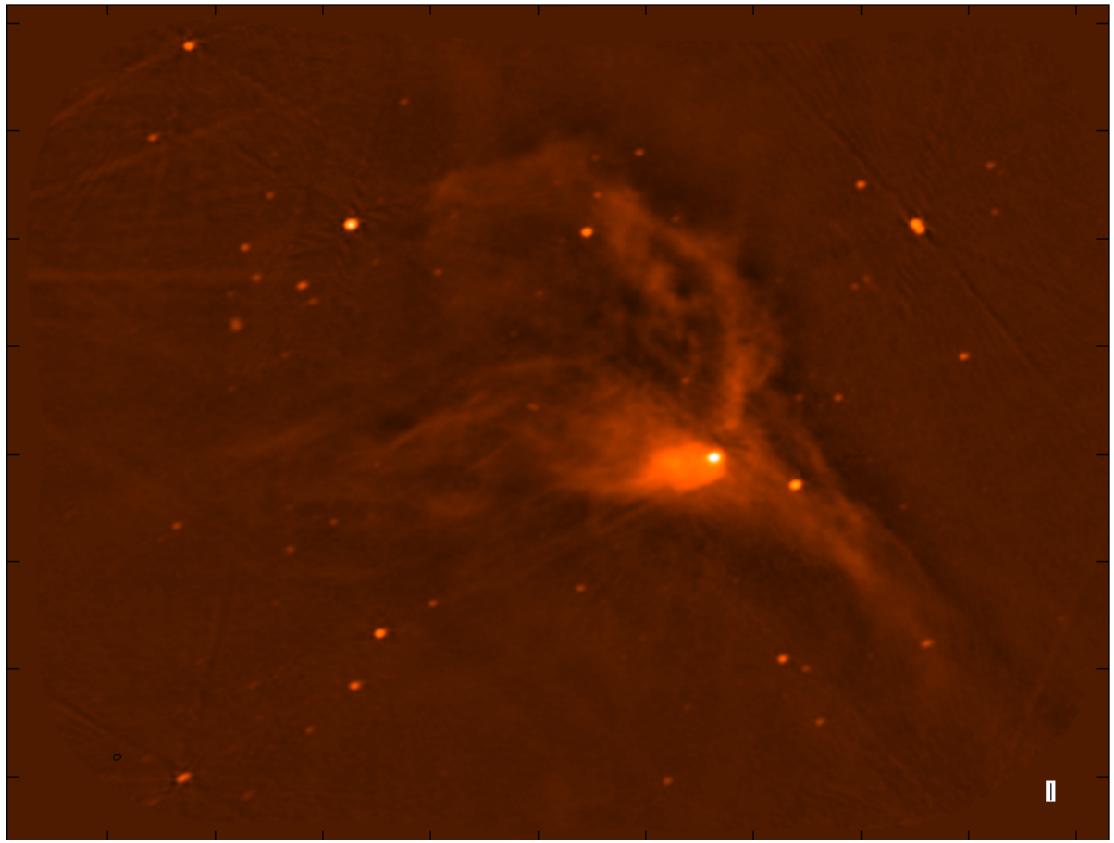


Image Courtesy Bhatnagar, Rao, Rupen, Roshi, Golap (NRAO), Green(Cambridge,UK)