ALMA Development Program

Jeff Kern

CASA Team Lead

Atacama Large Millimeter/submillimeter Array
Expanded Very Large Array
Robert C. Byrd Green Bank Telescope
Very Long Baseline Array

Opportunities for Software Development

- VLA upgrades post construction:
  - Hardware (Feeds, correlator electronics,)
  - Online System
    - Improved observing modes
    - Efficiencies, additional meta-data.
  - Postprocessing System
    - New Algorithms

It is reasonable to expect that we will find similar opportunities for improvement in ALMA.

Scope of Talk

- Planned and possible improvements to online system
- Planned developments within CASA
- Suggestions for offline capabilitied.

Very Long Baseline Interferometery

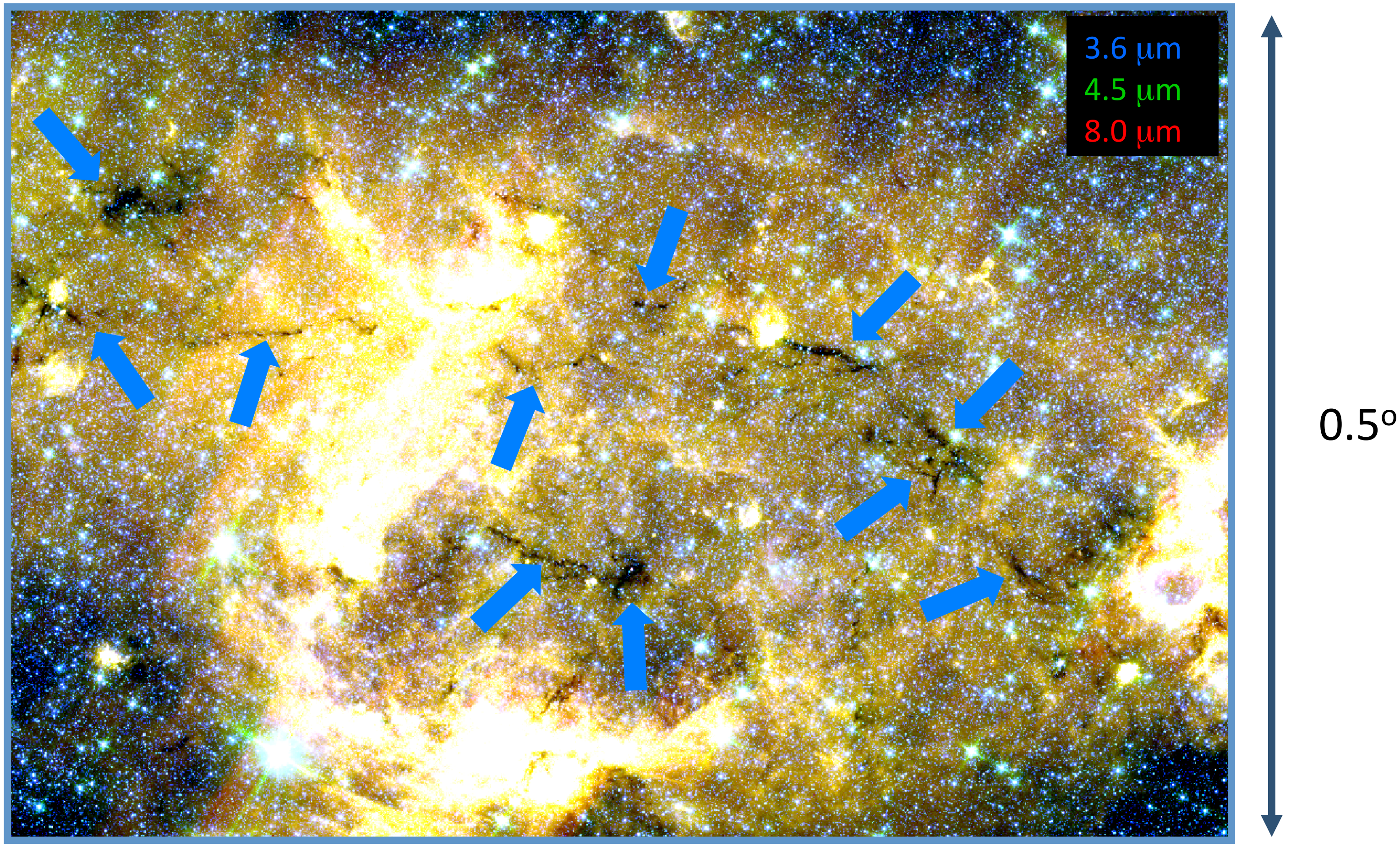
- The Event Horizon Telescope has been funded
  - Called out in Astro 2010 for development
  - Project to phase ALMA for beamformed science.
  - Currently developing implementation plan to submit to ALMA board.
- In addition to the hardware portions there is a significant software component.
  - Phasing Algorithms
  - Translation of .vex files to online system
  - Support of correlator protocol for application of phasing commands
  - Various bits of hardware control

Large Area Surveys

- Large area surveys are a key component of the ALMA mission.
  - distribution of gas in galaxy clusters
  - how gas cycles in and out of individual galaxies
  - formation of molecular clouds form and stars form

Large Area Surveys

Infrared Dark Clouds:
A Galactic web of star formation

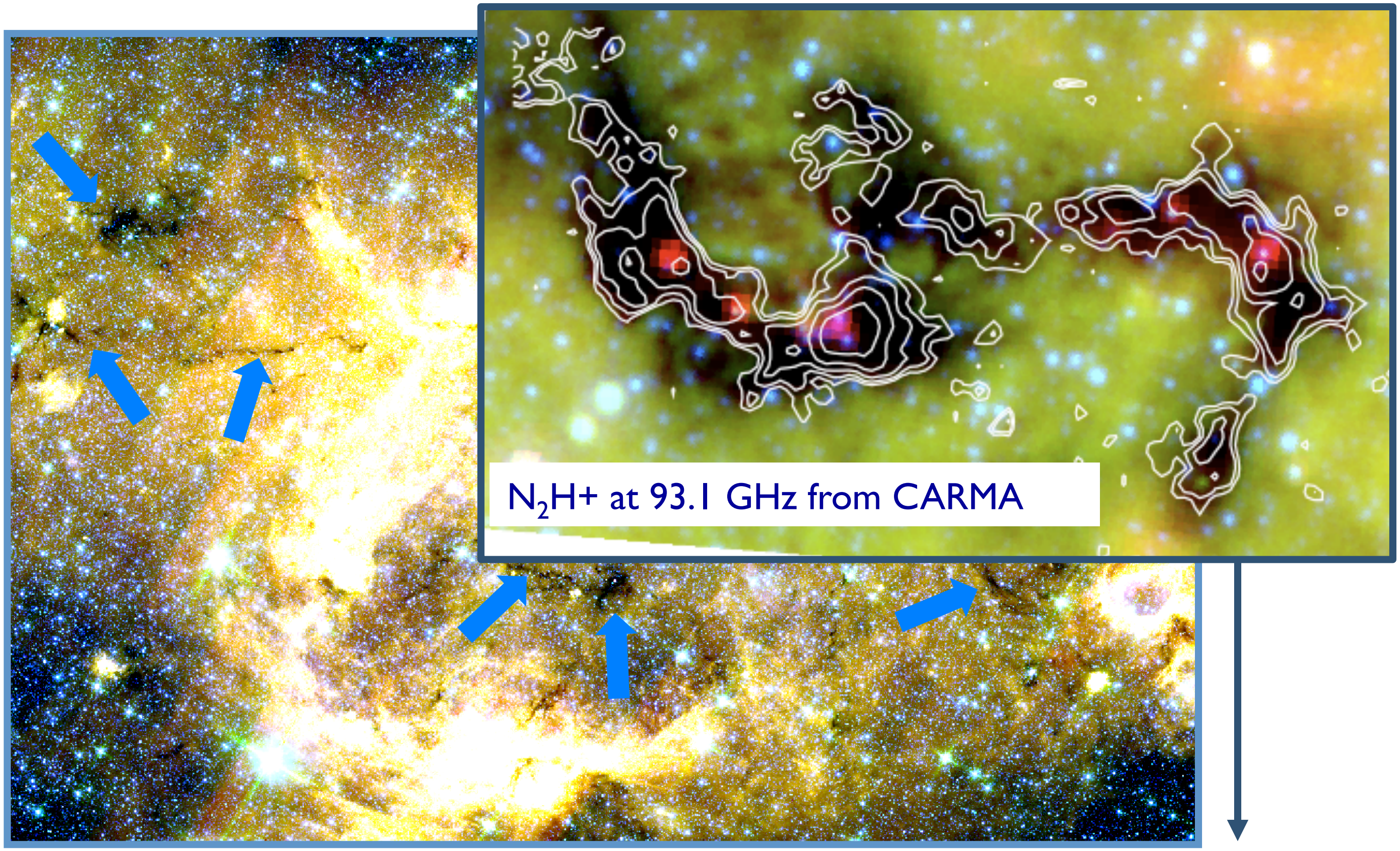


ALMA

NRAO

Large Area Surveys

Infrared Dark Clouds:
A Galactic web of star formation



Large Area Surveys

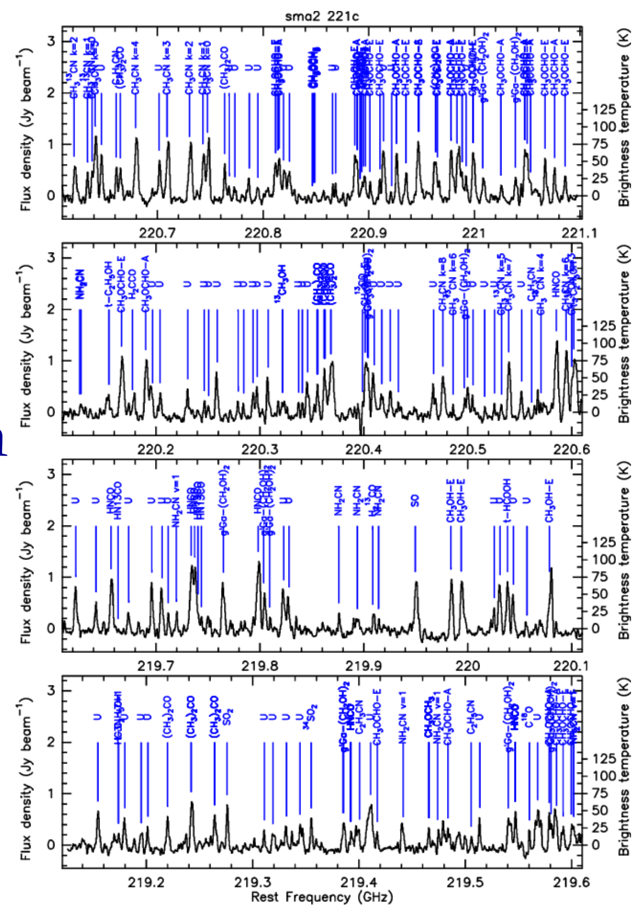
- Time required to map a region depends on:
  - the field of view
  - mapping speed
- Increasing either of these will increase the effectiveness of ALMA in generating large area suvey
  - Use Focal Plane Arrays (FPAs) to increase field of view
  - Mapping speed determined by minimum dump durtion for on-the-fly (OTF) mapping.
    - Increase required throughput from correlator to archive

Large Area Surveys (Example)

- OTF map of 1 square degree in CO(3-2) in 50 hours would:
  - require twice the current maximum data rate (64 MB/s).
  - produce raw data ~20 TB
  - produce a ~40 TB measurement set
- Resulting image would have 100 Mpixels per channel!

Spectral Line Surveys

- Spectral lines are critical tools:
  - probing physical conditions
  - understanding astrochemistry
  - discerning the chemical building blocks of life
- Use of the wide bandwidths plus high spectral and angular resolution is limited by the rate that data can be accessed from the correlator.
  - Restricted by network hardware, ability to process and archive.
    - Processing 1 GB/s will require ~1000 CPUs in CBE.



2 GHz of bandwidth toward a massive star forming region (Brogan et al. in prep.)

ALMA

NRAO

Expanded Bandwidth

- Eventually Moore's Law will enable us to naturally relax the data rate constraints on the ALMA telescope.

We could apply development resources to accelerate the pace of this advancement.

Data Rate Improvements

- ALMA (and EVLA) are fundamentally limited by the amount of data that we can transport and afford to store.
- ALMA– Data rate of 1 GB/s would immediately improve mapping speed, spectral line surveys, and transient phenomena.
- Physical Transport Solution:
  - Hardware Upgrade: 1 GB/s to10 Gb/s (Swithes, NICs)
  - Improvements to operational computers and software
  - Data transmission speed to the local "spool" archive
- Software / Analysis Solution:
  - What types of pre-archival data analysis could be done to allow higher effective data rates from the instrument without increasing the volume of data to store:

Transient Phenomena

- In Astro2010 identified as key discovery space
- ALMA's 16-ms dump time is a good match for transient science.
  - Limited by current maximum data rate.
  - Small field of view could be corrected by FPAs.
- Data limitations could be mitigated by an online event detection scheme.
  - Also decreases impact of data archiving.

Online System: Observing Modes

- The ALMA construction project will deliver a substantial set of observing modes:
  - Standard Interferometery (both the BL Array and the ACA)
    - "Continuum" and Spectral Line Modes
  - Pointed Mosaics
  - Total Power Observing modes
    - OTF Mapping, nutating subreflectors, …
  - OTF Interferometery

ALMA has spectacular instantaneous UV coverage, wide bandwidth, low slew times, and multi-array support.

Are there novel ways to utilize these resources to open new windows on the millimeter and submillimeter sky?

Common Astronomy Software Applications (CASA)

Support the efficient calibration, editing, imaging, and analysis of NRAO's newest interferometers: EVLA and ALMA.

Provide a framework for continued research in interferometric astronomical data analysis

Provide, as a service to the community, a flexible reduction package capable of supporting the wide variety of interferometers in use today

CASA

- CASA is a suite of applications
  - C++ to do most of the hard work
  - Python "tool" interface for flexible access to the compiled layers
  - Python "task" interface for simpler access to routine tasks
- Viewer for presenting and working with image and cube data
- plotMS allows plotting of visibilities on various axis.

CASA Development

- CASA is currently managed and funded as a joint effort of ALMA and the EVLA.
- Development priorities are set by a Science Steering committee with representation from both projects, as well as an at large member.

NRAO and now CASA have primarily focused on turning raw visibilities into calibrated images and cubes.

CASA Processing

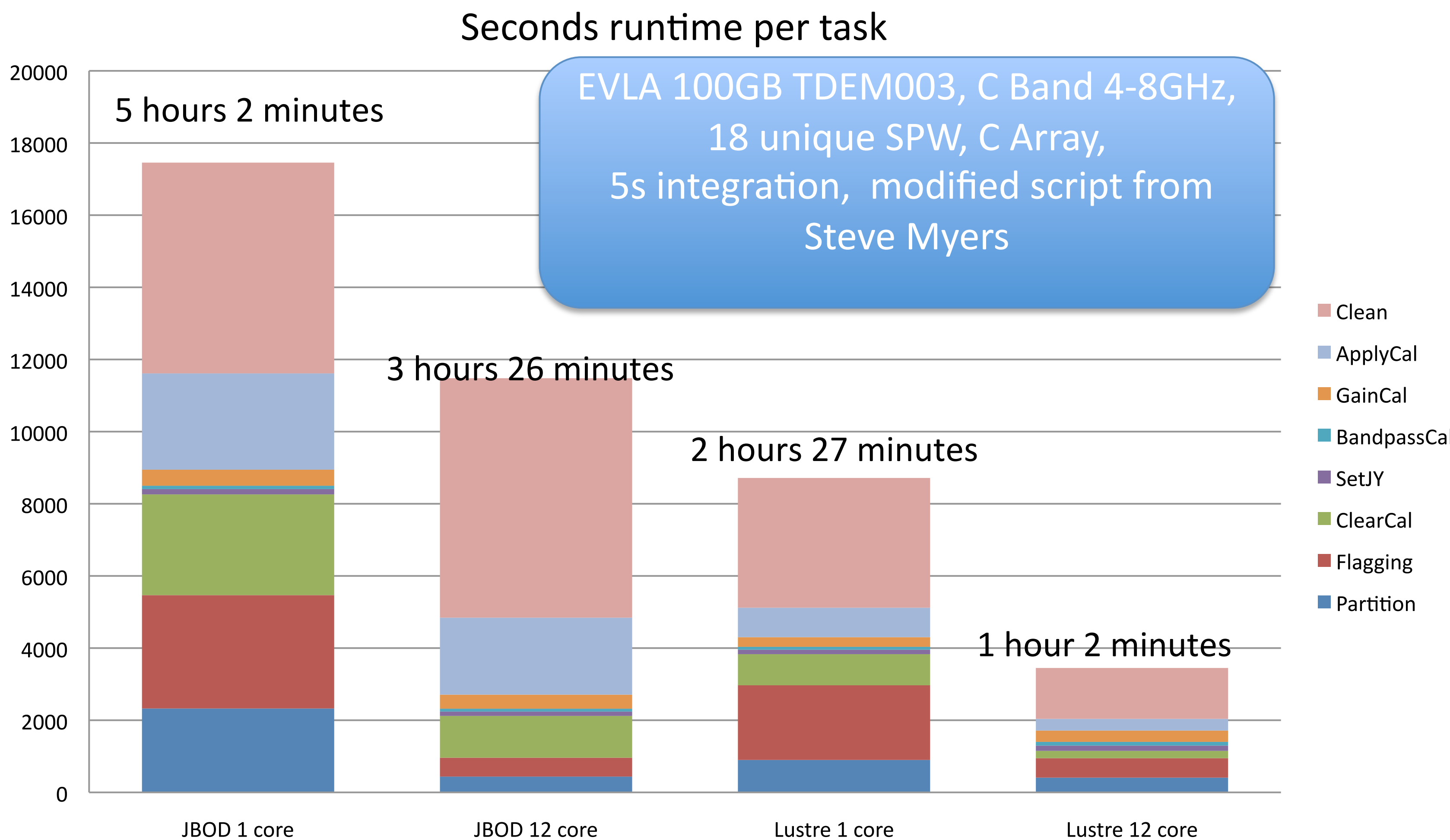
- Processing data can be split into two portions:
  - Editing and Calibration
    - Almost entirely input output dominated
  - Imaging
    - Can be I/O dominated, but for interesting cases is primarily compute limited.

Example: 500 GB Measurement Set on a SATA disk takes ~2.4 hours just to read the data (~12 min on Lustre filesystem)

CASA Parallelization

- Data analysis is an "Embarrassingly Parallel" problem.
- Implementation of parallel CASA is ongoing
  - Continuum and Cube cleaning
    - Simple cases (MFS not yet implemented)
  - Application of calibration
  - Flagging

CASA Parallelization Performance



Graphics Processing Units

- GPU application in gridding stage can only help if the I/O requirements can be met.
- May be possible to reduce memory footprint of convolution based algorithms using a GPU.

- Currently working with MeerKat on testing of GPU based algorithms.
  - Other groups are also investigating (LOFAR)

Algorithm Development

- Producing thermal noise limited images from ALMA requires new techniques and algorithms.
- Pointing self cal (in development by ARDG) is likely to be required for sources which fill the primary beam.
- A-Projection will likely be required for mosaics
- OTF Interferometery is still in early days
  - Data rate limitations will have a large effect here.

To some extent this effort will be driven by the data, but it is likely that to optimize ALMA science new imaging and calibration algorithms will need to be designed and implemented.

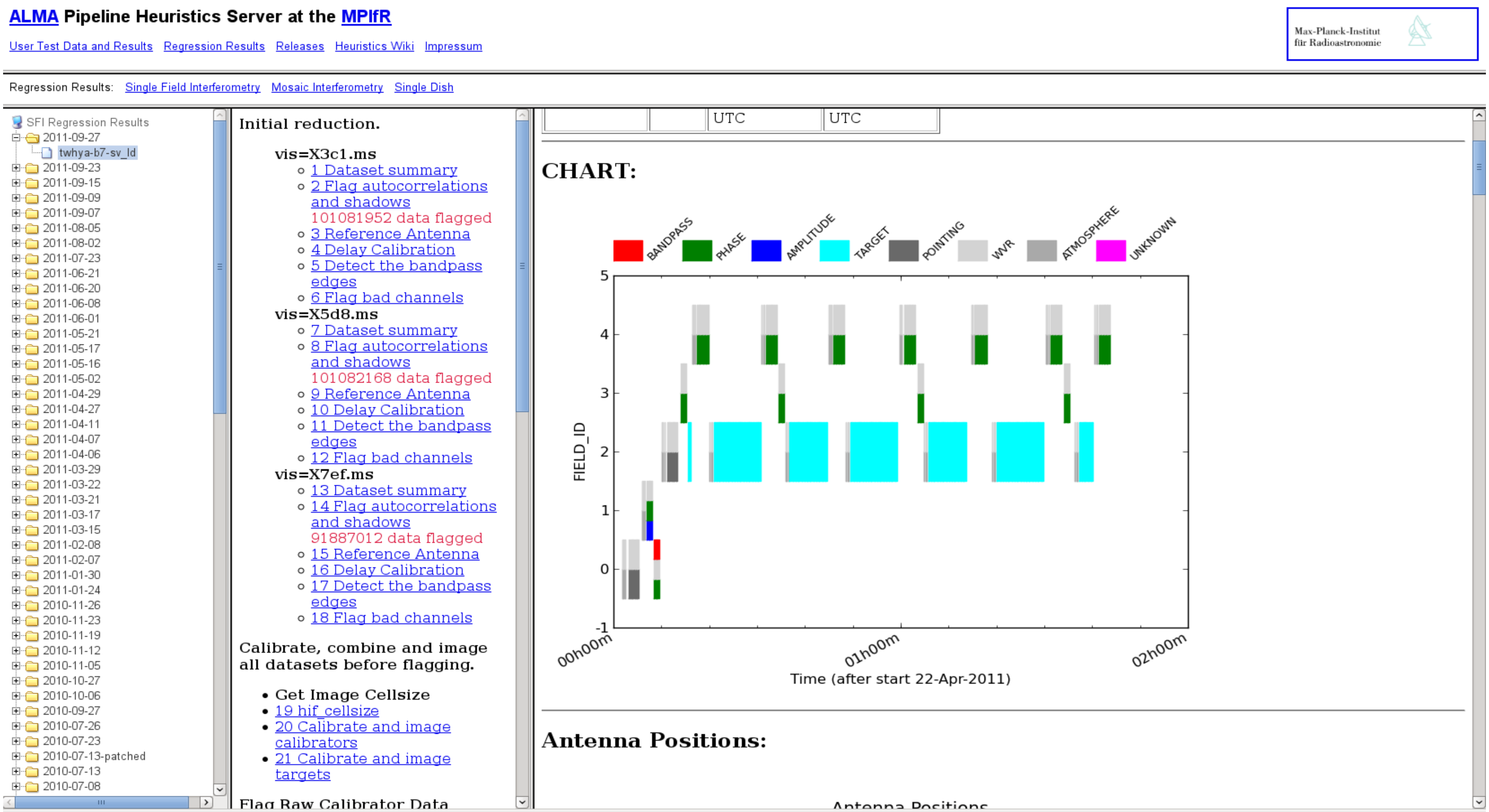
Auto Flagging

- Probably less of an issue for ALMA, but significant issue for EVLA.
- Implemented Cropping algorithm in Time - Frequency plane.
- Application of online flags generated by ALMA system will probably be sufficient.
  - Perhaps a need for flagging known "birdies"

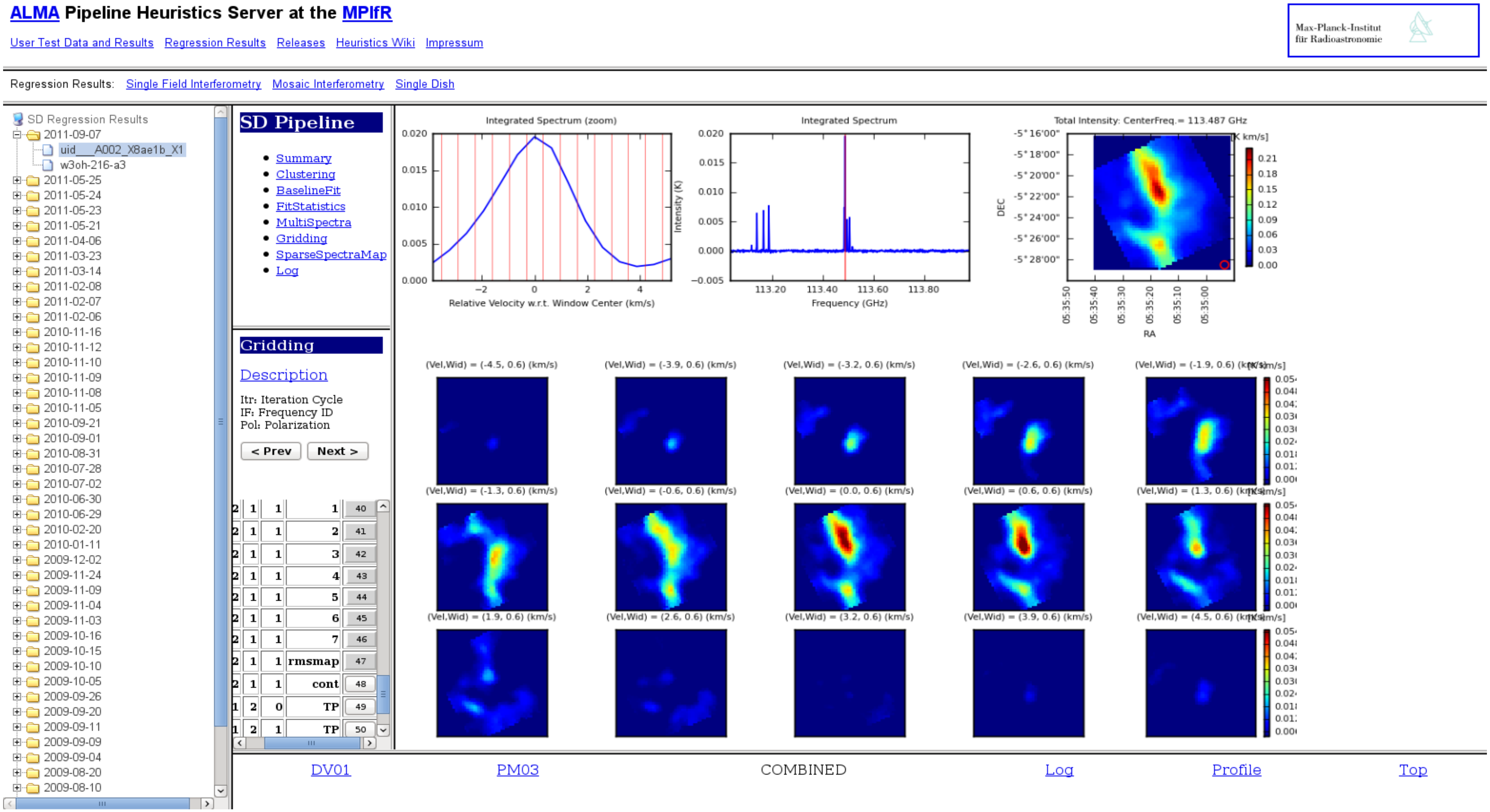
Pipeline Analysis

- ALMA Construction will deliver a standard pipeline.
  - Produced calibrated images to the user
  - Reprocessing at NAASC will be supported.
- Design is intended to be extensible, allowing design and testing of new approaches.
  - Expect considerable contribution from the community in development of "best practices."

Pipeline Results



Pipeline Results



Pipeline Results

ALMA Pipeline Heuristics Server at the MPIfR

User Test Data and Results | Regression Results | Releases | Heuristics Wiki | Impressum

Max-Planck-Institut für Radioastronomie

Regression Results: Single Field Interferometry | Mosaic Interferometry | Single Dish

SFI Regression Results
- 2011-09-27
  - twhya-b7-sv_ld
- 2011-09-23
- 2011-09-15
- 2011-09-09
- 2011-09-07
- 2011-08-05
- 2011-08-02
- 2011-07-23
- 2011-06-21
- 2011-06-20
- 2011-06-08
- 2011-06-01
- 2011-05-21
- 2011-05-17
- 2011-05-16
- 2011-05-02
- 2011-04-29
- 2011-04-27
- 2011-04-11
- 2011-04-07
- 2011-04-06
- 2011-03-29
- 2011-03-22
- 2011-03-21
- 2011-03-17
- 2011-03-15
- 2011-02-08
- 2011-02-07
- 2011-01-30
- 2011-01-24
- 2010-11-26
- 2010-11-23
- 2010-11-19
- 2010-11-12
- 2010-11-05
- 2010-10-27
- 2010-10-06
- 2010-09-27
- 2010-07-26
- 2010-07-23
- 2010-07-13-patched
- 2010-07-13
- 2010-07-08

- 23 Flag noisy phases
  1068 rows flagged
- 24 Flag baseline timestamps with outlying amplitudes in calibrator data

vis=X5d8.ms
- 25 Flag noisy visibilities
  108 rows flagged
- 26 Flag noisy phases
  360 rows flagged
- 27 Flag baseline timestamps with outlying amplitudes in calibrator data

vis=X7ef.ms
- 28 Flag noisy visibilities
- 29 Flag noisy phases
  108 rows flagged
- 30 Flag baseline timestamps with outlying amplitudes in calibrator data

Compute final results

vis=X3c1.ms
- 31 Individual calibrator cleaned integrated maps
- 32 Calibrate and image individual targets

vis=X5d8.ms
- 33 Individual calibrator cleaned integrated maps
- 34 Calibrate and image individual targets

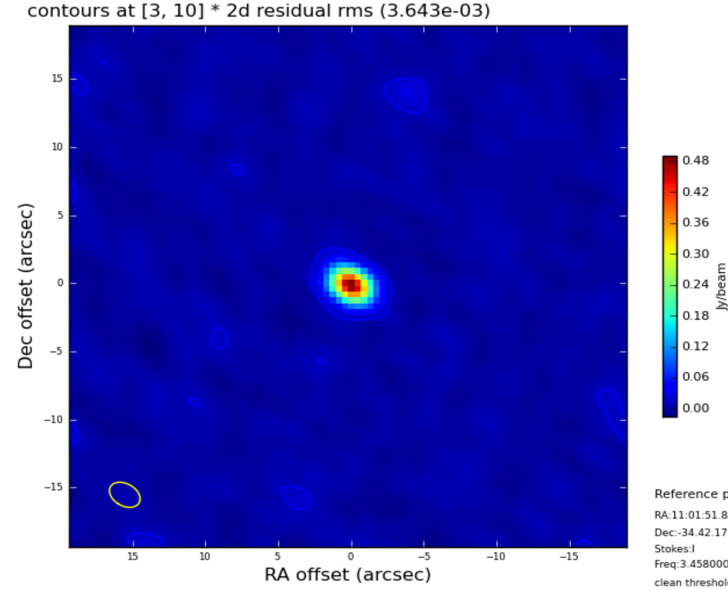
vis=X7ef.ms
- 35 Individual calibrator cleaned integrated maps
- 36 Calibrate and image individual targets

Calibrate, combine and image all datasets.

Image Maps (Dec Offset vs RA Offset)

Back | Clear All Filters

Field:TW Hya (TARGET) SpW:21 (a) Stokes:I - Clean image
contours at [3, 10] * 2d residual rms (3.643e-03)



Reference position:
RA:11:01:51.84498068
Dec:-34.42.17.16088989
Stokes:I
Freq:3.45800000e+11 Hz
clean threshold: 7.74e-03

Image 5 / 8 TW Hya: Dec Offset vs Ra Offset for Map Clean image, Polarisation I, Channels All, Spectral Window 21

VAO and CASA

- VAO is working on desktop integration in the coming year.
  - Beginning dialog with VAO in late October to define scope and requirements.

GUI Interface

- CASA has a requirement to produce a GUI interface to the package.
- We have agreed to work with Dr. Emmanuel Pietriga, INRIA, France a Human Computer Interface expert on design and implementation.

Interfacing with CASA

- CASA is extendable by outside users.
- Currently the ability to reuse this code is limited
  - Python can be shared and reused via the "buildmytask" executable in CASA.
- Sharing compiled code is currently more difficult
  - Working on a "plugin" infrastructure to support executables and improve python support
  - dbus based interface to the Viewer allows reuse of the CASA viewer as a display tool.

Visualization

- "Reduced" data sets continue to grow in size.
  - The 1 TB cube is not far off.
- None of the viewers currently in widespread use by the radio community support data sets of this size.