CASA Status

ASAC Meeting, March 9, 2010 – Mitaka, Japan

Crystal Brogan (subsystem scientist)

NRAO/North American ALMA Science Center

Atacama Large Millimeter/submillimeter Array

Expanded Very Large Array

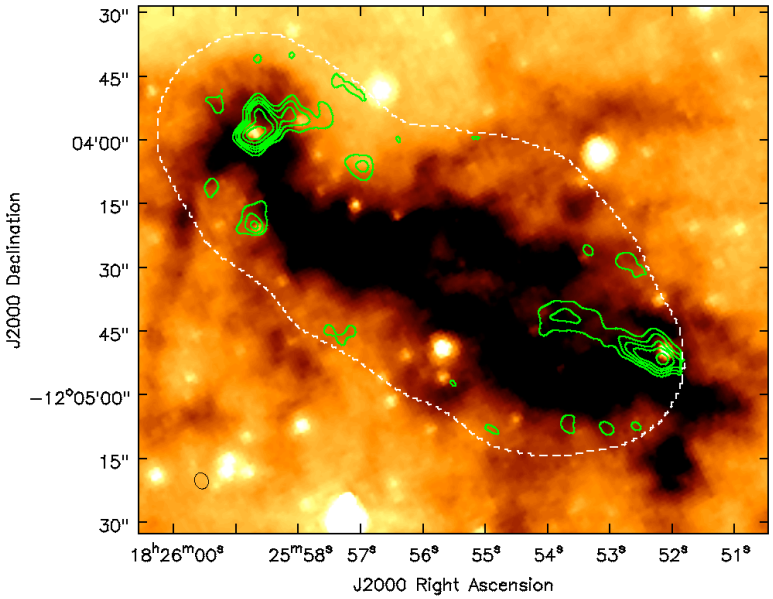
Robert C. Byrd Green Bank Telescope

Very Long Baseline Array

What is CASA?

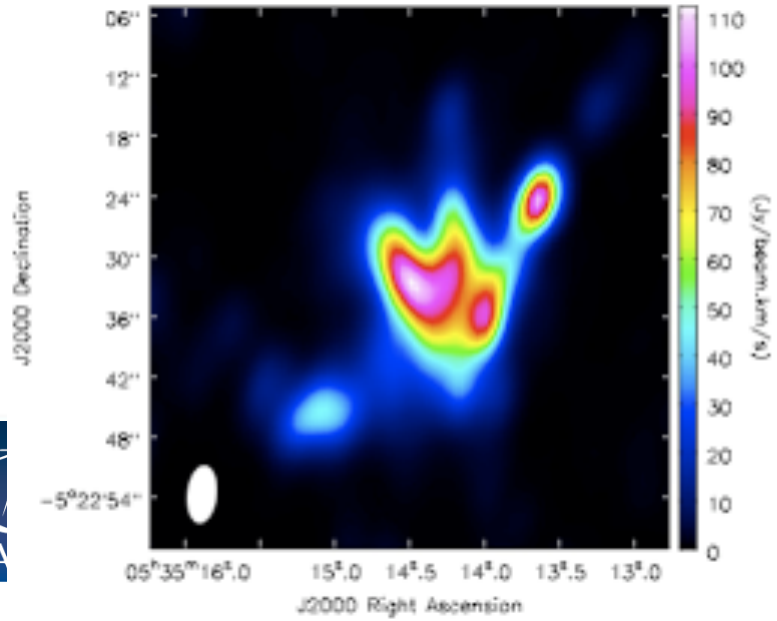
- CASA is the post-processing package for ALMA and EVLA
- It is a suite of applications for the reduction and analysis of radio-astronomical data (derived from the former AIPS++ package)
- The algorithms are written in C++; interface in python/ipython/Qt
- It is fully scriptable, with in-line help and scientist-written documentation (notably the user manual/cookbook)
- Telescope data (visibility and single-dish) are stored in a MeasurementSet (MS); a filler converts ALMA/raw data to the MS
- It contains functionality for manipulating/plotting/… core infrastructure data types (e.g., Images, Tables, Measures, …)
- Extensive interferometric calibration and imaging capabilities implemented via the Hamaker, Bregman, Sault formalism (Measurement Equation)
- It contains image analysis and other mathematical functionality

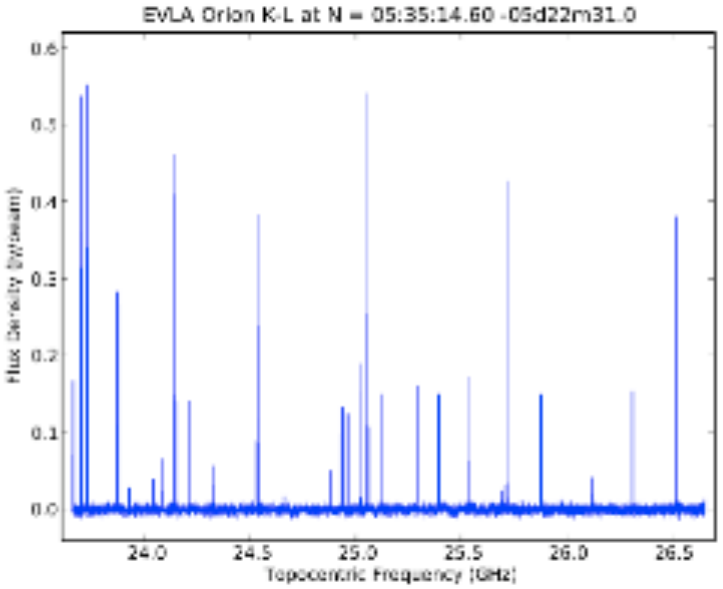
Production of Scientific Images in CASA



SMA ^{12}CO (2-1) integrated intensity (green countours) superposed on a GLIMPSE 8 μm image of the Infrared Dark Cloud (IRDC) G19.3+0.07. Six-pointing SMA mosaic.

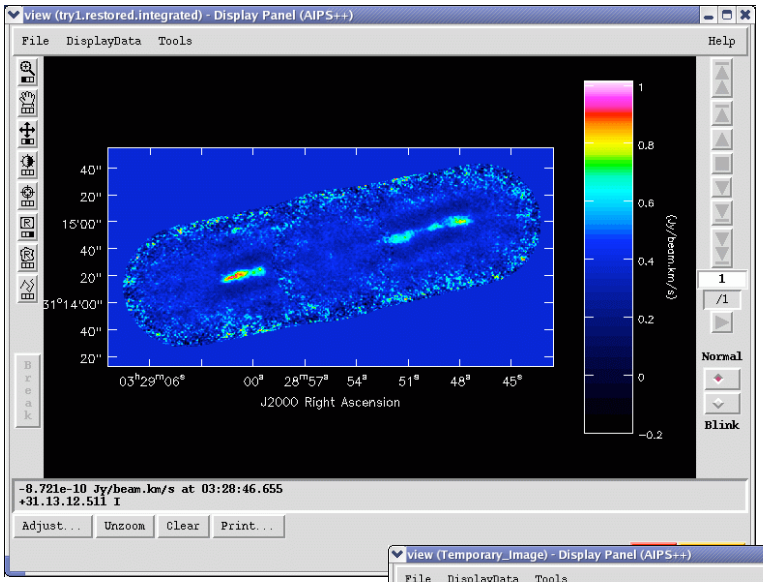
EVLA demo science: Orion Hot Core integrated intensity and spectrum: 24k channels!



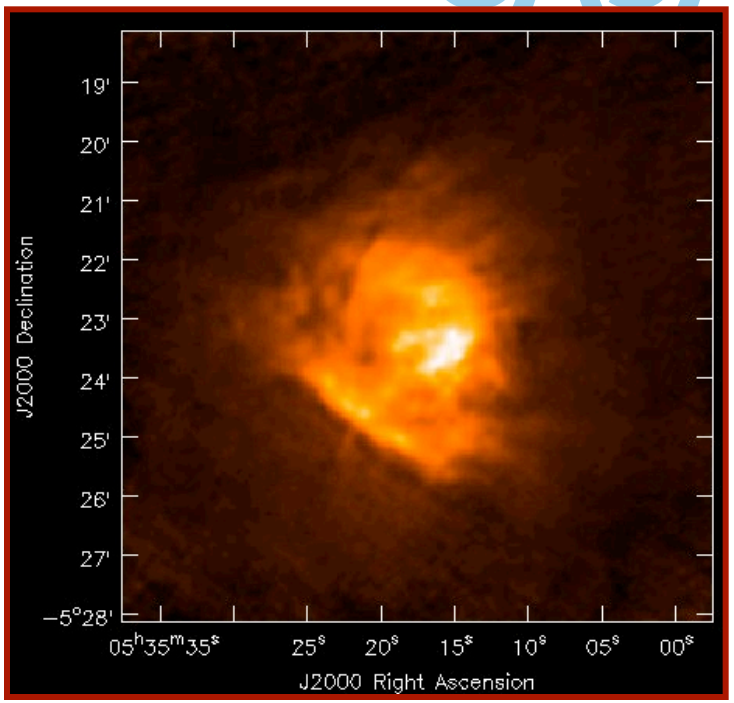


Single Dish/Interferometer Combination

CASA

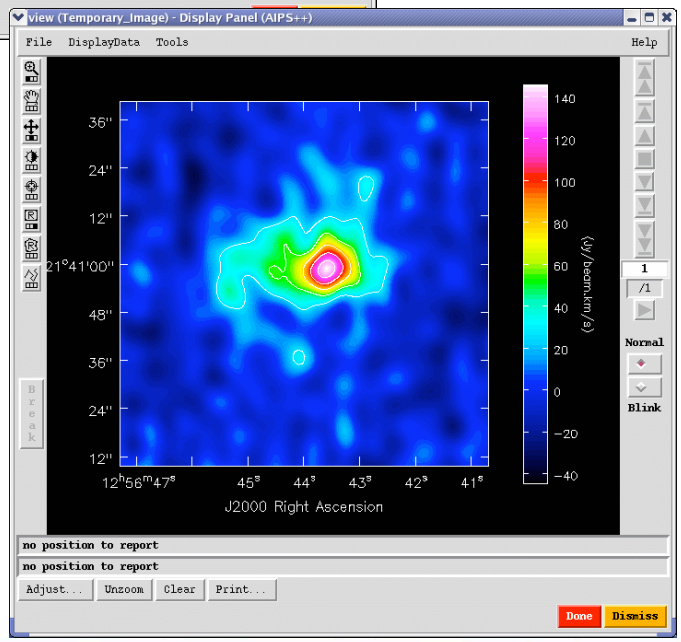


VLA SiO(1-0) mosaic, CLEAN deconvolution



Orion nebula continuum mosaic: VLA mosaic + GBT single dish on-the-fly image. Combined mosaic image produced using joint deconvolution, multi-scale CLEAN

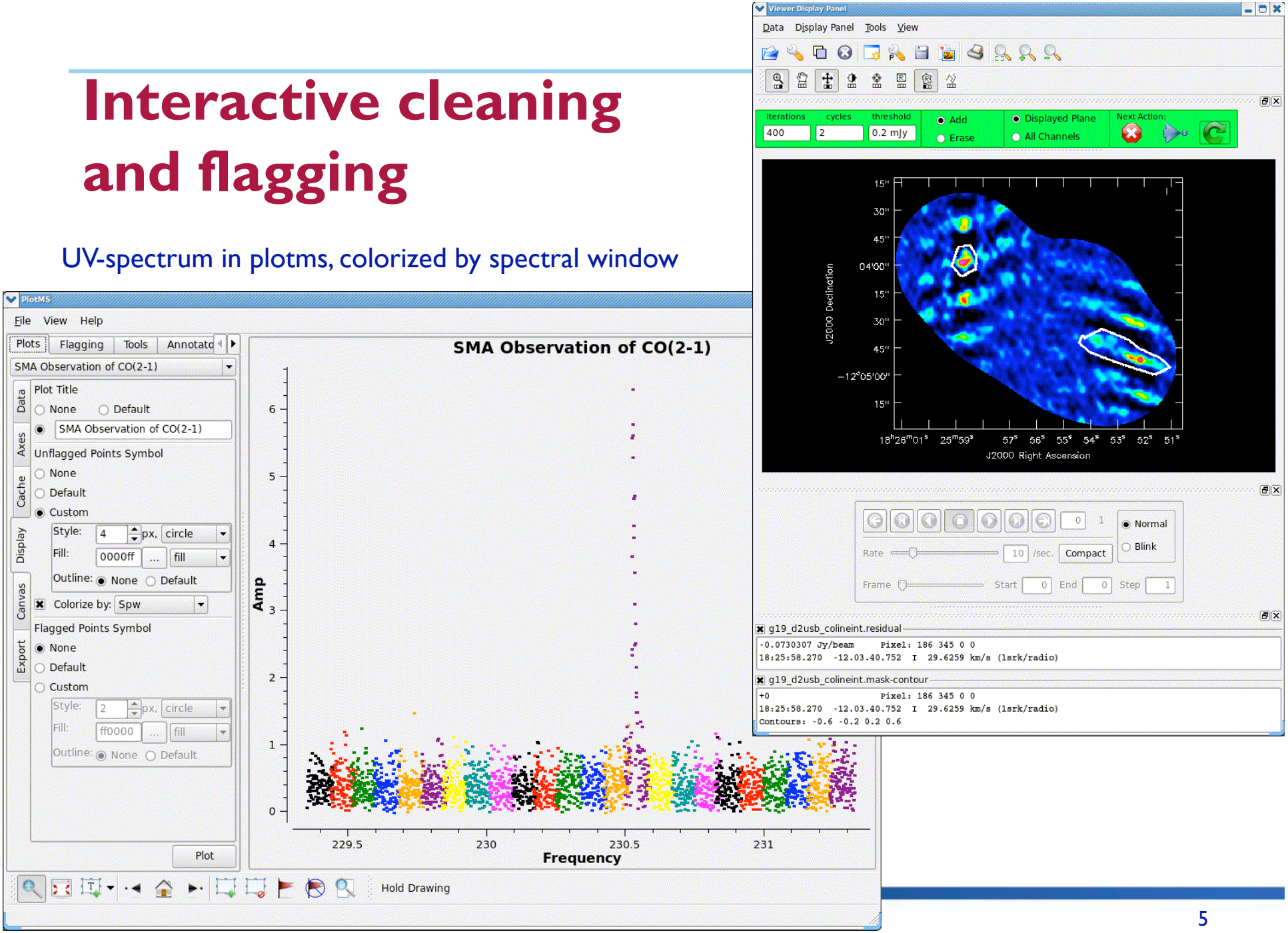
BIMA mosaic + 12m single dish CO(1-0) mosaic, This moment 0 map produced from joint 'feathered' image cube



NRAO

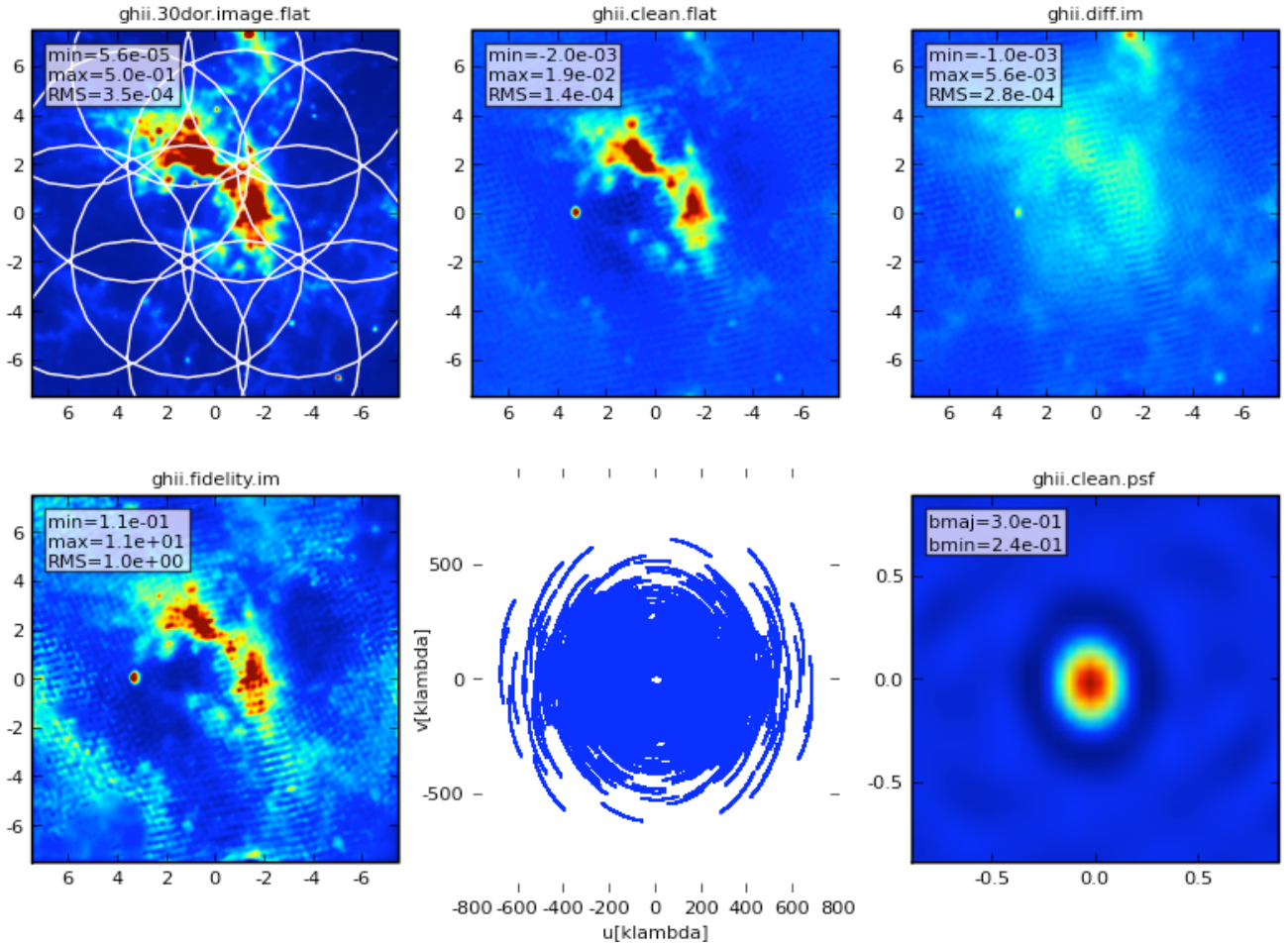
Interactive cleaning and flagging

UV-spectrum in plotms, colorized by spectral window



simdata: CASA Simulator

CASA



- Infrared image of 30Doradus used as input model
- This figure and the images and uv-data are all outputs of simdata

- Continuum and simple cubes
- Capability to create dummy header info for model image
- Add thermal noise using same ATM model as Telcal
- All configurations available including preliminary early science configs
- At tool level can add phase screen and cross pol
- Soon single dish simulation will be incorporated

NRAO

CASA Staff

- Group leader hired: Nick Elias (Oct. 2009)
- Development team
  - Currently ~12.5 FTEs with mix of scientific developers and programmers
  - Distribution of people: NRAO: 7.3, NAOJ: 2.7, ESO: 2.5
  - Two vacancies: Application (gui) developer, High performance computing specialist
- Overall Project management: Brian Glendenning (ALMA), Brian Butler (EVLA)
- NRAO Science Oversight
  - Juergen Ott (Project Scientist)
  - Crystal Brogan (ALMA subsystem scientist)
  - Steve Myers (EVLA subsystem scientist)
  - Ed Fomalont (E2E Scientist)
- Creation of ALMA CASA Coordination Group (ACCG)
  - Members from NOAJ – SD effort, NRAO, and JAO for feedback/priorities

Release Status and Usage

- Have had Beta (patch) releases every ~3-6 months since October 2007
  - Pretty much any recent linux flavor, Mac OSX leopard, snow leopard
  - Available to anyone after registration at http://my.nrao.edu
  - > 400 have downloaded so far

- Used every day in Chile for ALMA commissioning

- Being used heavily for EVLA science verification now, and by outside users for start of early science (March 1, 2010)

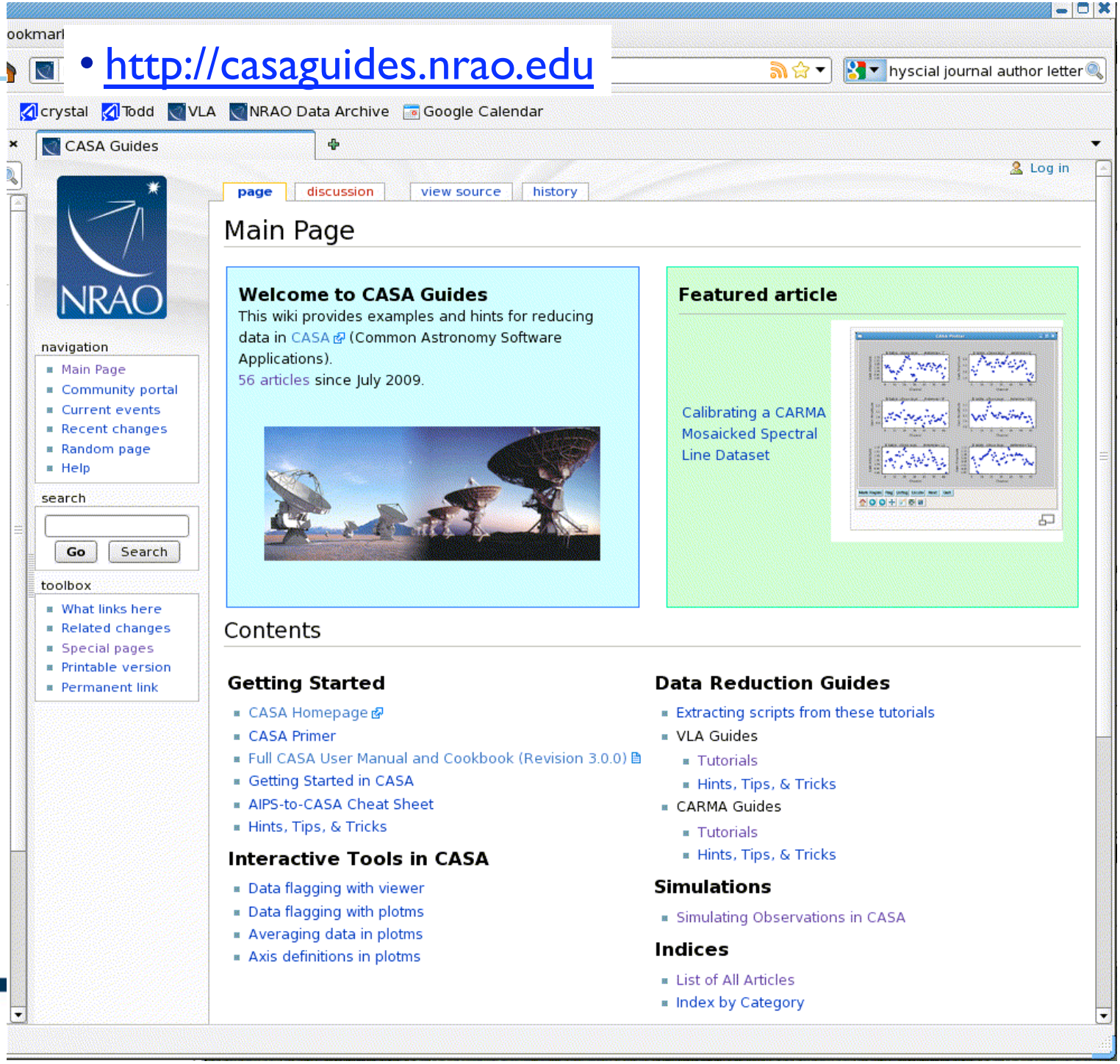
- Dec. 18, 2009 was first non-beta release (3.0.0) and the next patch (3.0.1) will be @March 15. Another patch (3.0.2) expected before the NRAO synthesis imaging workshop in June.

Tutorials

- Santiago tutorial for ALMA commissioning staff last year
- Other tutorials this past year: Garching (May 11-13), Hamilton, Canada (June 1-3), Bonn (Oct.), Taiwan (Feb)
- Coming up:
  - April, NAOJ
  - May, Miami AAS special session with talks/demos
  - June, NRAO Synthesis Imaging Workshop – all tutorials in CASA
  - Second Santiago tutorial in Spring or Summer (possibly by videocon)

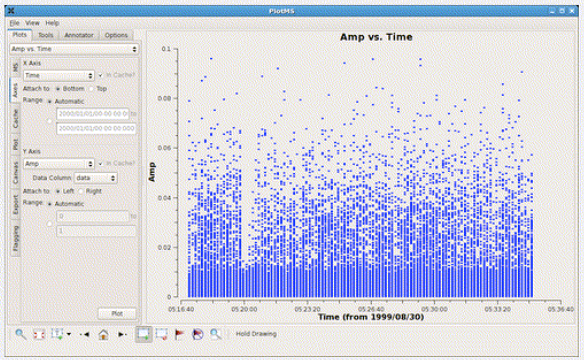
CASA Guides

- Uses mediawiki to enable fully annotated scripts
- Additional "guides" continue to be added



CASA Guides

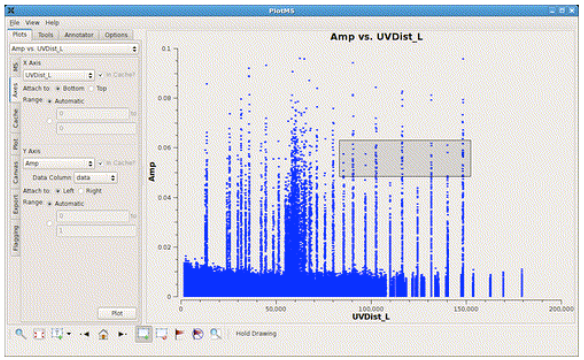
The figure at right shows the unflagged data for the first source (field = 0) after flagging the obviously discrepant points. Things look deceptively OK, but in fact there remain bad data from one antenna. The antenna contributed poor data for the entire observation of this source, and, since the problem is not isolated in time, it is difficult to see it in this projection.



Click to enlarge

Here are the same data reprojected onto baseline separations, (Axes)**X Axis** = **UVDist_L** (projected baseline separations in units of the observing wavelength). The misbehaving antenna shows up as spikes in these snapshot observations, because each baseline with that antenna spans only a narrow range of baseline separations. (A longer observation would produce broader spikes, because the projected baseline separations would span a greater range as they rotate with the earth under the source.)

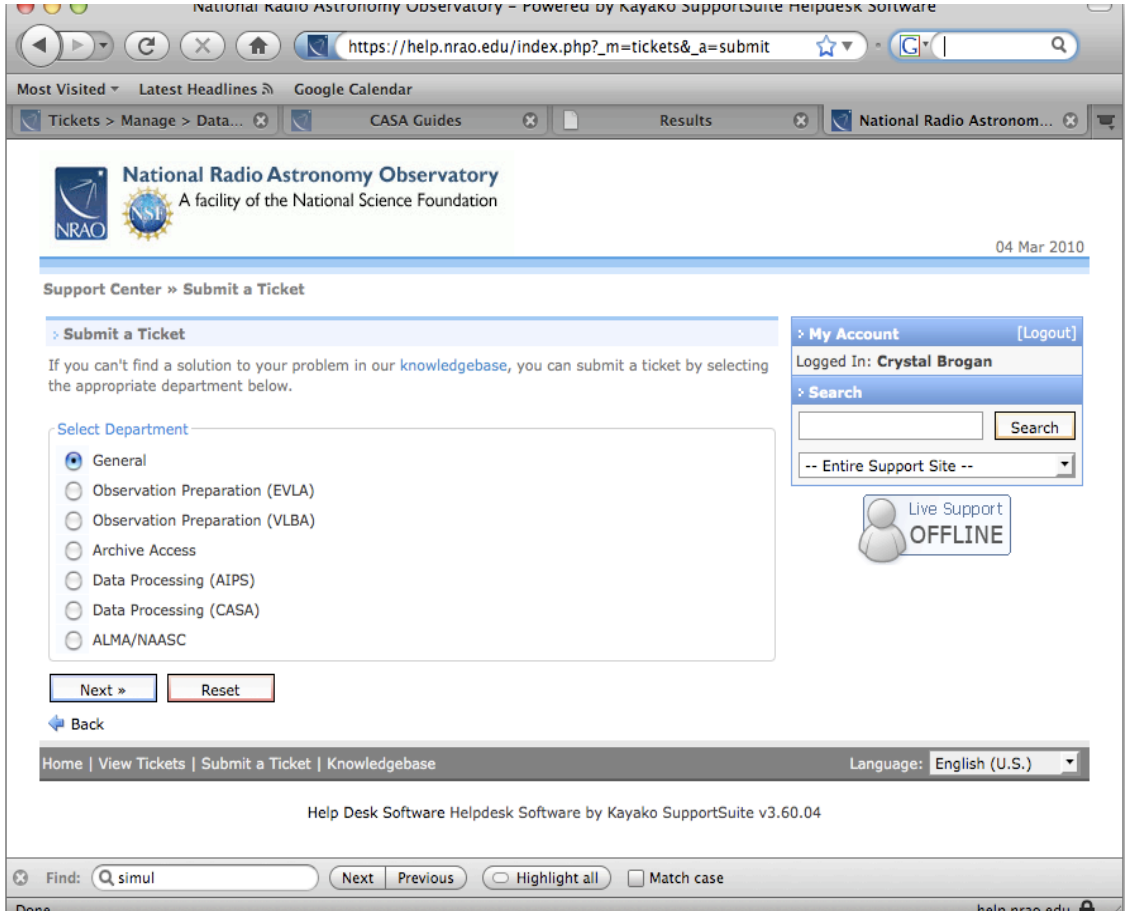
The idea now would be to highlight a subset of the discrepant data as shown in the figure and extend the flags to the common antenna of these baselines. At the time of this writing, this option is not available in the development build of plotms, but keep an eye on (Flagging)**Extend flags** = **Antenna**.



Click to enlarge

New CASA Helpdesk

- In mid-Feb. NRAO launched Kayako helpdesk at http://help.nrao.edu
- Kayako combines the utilities of managing tickets/ user support with a knowledge base
  - Herschel and Spitzer Science Centers
  - top candidate for the ALMA Helpdesk



Current CASA Strengths for ALMA/EVLA

- Full data import (e.g., complex correlator setups)
  - ALMA/EVLA raw data formats designed with CASA in mind
- Able to handle large datasets
- Wide-band imaging using Multiscale MFS
- W-projection imaging
- Non-linearized polarization calibration (for high dynamic range), frequency-dependent D terms
- Spline G (gain) and Bandpass solutions
- Data inspection/modification tools; scriptability in general

Calibration

- Standard gain & bandpass calibration
  - Sampled and Polynominal/Spline solutions available
  - Flux density reference scaling
  - Sampled baseline-based solution available
  - Solution normalization
  - Phase-only, *Amp-only options*
  - Auto-interpolation of flagged channels in bandpass
- Polarization calibration
  - Linearized instrumental polarization (D-terms) solutions available
  - *Channelized option for frequency-dependent instrumental polarization*
  - Optional solution for source polarization
  - Polarization position-angle solution support (for circular basis)

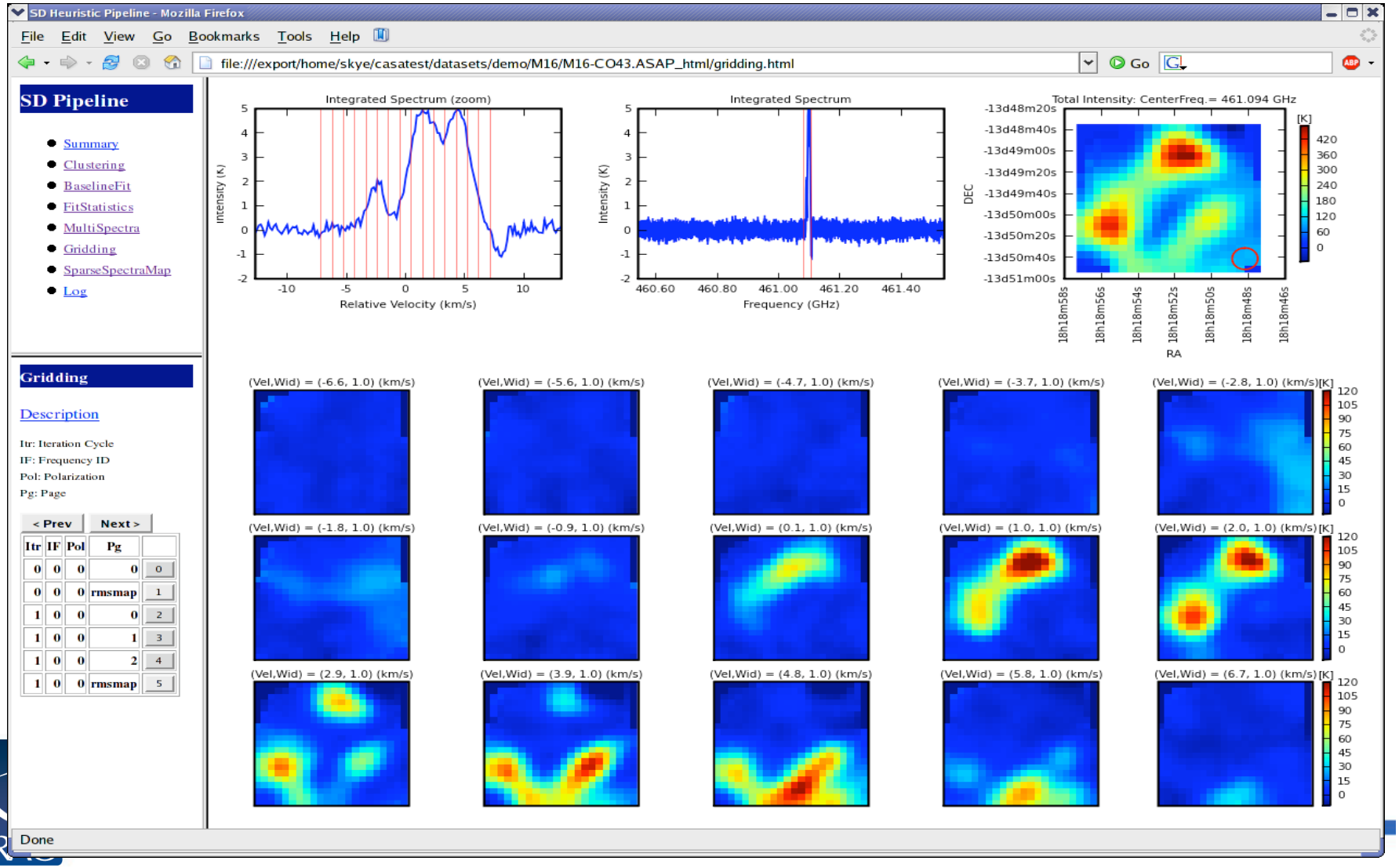
Imaging & Deconvolution

- Mosaic imaging
  - Joint deconvolution (Miriad style) and by *gridding convolution*
  - Mosaicing with heterogenous arrays (ALMA, CARMA)
- Widefield imaging: *W-projection* and faceting
  - W-projection more than 1 order of magnitude faster than faceting
- Multiple algorithms for single dish and interferometry combination
  - Feathering
  - Single Dish as a model for deconvolution
  - *True joint deconvolution using both visibility data and single dish data*
    - Requires data with well-calibrated weights between the single dish and interferometry data (ALMA), and testing
- *Full beam Stokes I, V imaging*
  - Targeted at friendly VLA users on a "shared risk" basis
- *Multiscale clean*
- MEM & NNLS (toolkit level only so-far)

Single Dish

- Based on ATNF ASAP package
- Calibration and flagging
  - Calibration of position–switching, frequency-switching spectral data
  - Calculate quotient spectra
  - Calibration of total power raster scan data
  - Opacity correction (based on input opacity value)
  - Flagging by channel or row
- Imaging
  - Imaging of total power and spectral line raster scan data, including solar-system objects: Uses interferometric Imager
- Data analysis
  - Baseline fitting and subtraction
  - Plotting and profile fitting
  - Basic vector arithmetic
  - Removal of scanning noise by "Press-out" and FFT filtering

Pipeline test based on CASA Single Dish Package



Performance

- For "small" to "intermediate" (1-10 GB) sized datasets CASA is comparable to other packages
  - Sometimes faster, sometimes slower; complex parameter space
- CASA's architecture was designed to allow parallelization to be introduced at several levels
  - Storage manager (I/O) through OpenMP through Python scripting
  - Also I/O is organized to minimize passes through the data
- Started: Terabyte initiative
  - Flag, calibrate, image 1 TB (raw data size) data = 10h of ***peak*** (~2012) data rate
    - Tests to date concentrating on initial (2010) large data sizes = 100 GB

Performance (2)

- Cluster (16 nodes, 128 cores) purchased, working on simulating the data and initial timing tests (joint ALMA/EVLA purchase)
- Initial end-to-end (flagging/calibration/imagine) testing of the data-parallel ("embarrassingly parallel") case
  - 100 GB ~ 3H on the cluster
- Collaboration with UVa parallelization group recently started
  - E.g., systematic profiling, CUDA sample implementation of gridding/degridding inner loop)
- Major question: nodes vs. cores (do we have enough FLOPS/IO to support many cores)

Testing ALMA/EVLA sized data sets is the important exercise!

Planned for the coming year…

- Support of ALMA commissioning needs
- Improvements needed for polarization calibration of linear feeds
- Improvements to calibration table plotting (incorporate into plotms)
- Planet models for use as resolved calibrators
- Splatalogue search capabilities and overplotting
- Viewer improvements (especially for spectral line plotting and analysis)
- Improvements to image analysis tasks
- Improvements to "TV" based flagging in the Viewer (on-the-fly spectral and time averaging)
- A CARMA miriad filler (through partnership with Peter Teuben at U. Maryland)
- Expanded and more modularized simulation capabilities, incorporate single dish simulation