2010 PASEO Meeting

July 15-16, 2010 – Socorro, NM

CASA: Common Astronomy Software Applications

Steven T. Myers (EVLA CASA Subsystem Scientist)

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

NRAO

EVLA Post-processing Philosophy

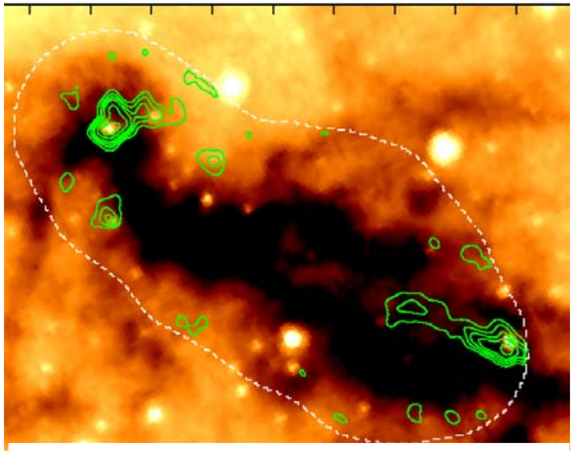
- Enable forefront science with the EVLA through basic and advanced software tools.
  - leverage ALMA co-development and use of CASA package
  - enable pipelines (e.g. for EVLA to produce calibrated data)
  - staged deployment and development, must be within our means
- Stakeholders : EVLA and ALMA
  - first: must support user to process own data (e.g. on workstation or small cluster)
  - second: support pipelines and HPC (e.g. large cluster) processing
  - third: support algorithm development
- Post-processing Software Requirements (2003, updated 2007,2008)
  - guides (but does not dictate) development and implementation

http://www.aoc.nrao.edu/evla/techdocs/computer/workdocs/EVLA_Data_Post-Processing_Software_Requirements_for_CASA.pdf

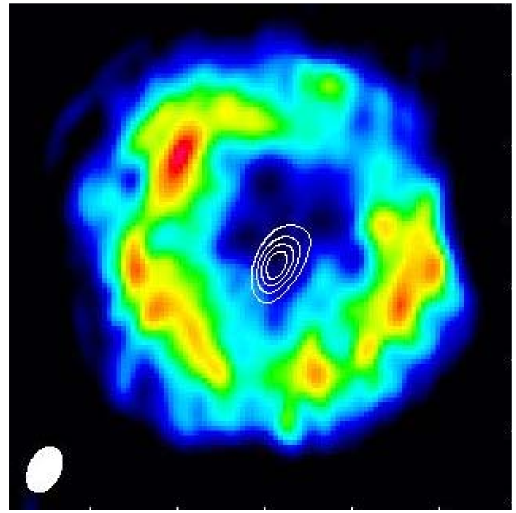
What is CASA? http://casa.nrao.edu

- CASA is the post-processing package for ALMA and EVLA
  - handles interferometric and single dish data
- “Toolkit” code are written in C++; interface & tasks in Python/IPython; GUI displays in Qt/matplotlib
- Fully scriptable (Python), with in-line and web-based help and scientist-written documentation (notably the user manual/cookbook)
- Telescope data (visibility and single-dish) are stored in a MeasurementSet (MS) with a asdm2ms filler (both ALMA & EVLA)
- Tasks for common post-processing operations
  - familiar task parameter input interface as well as call-by-function
- Toolkit-level functionality for manipulating/plotting core infrastructure data types (e.g., Images, Tables, Measures, …)
- Capability for users to “buildmytasks” in Python
  - e.g. used for *importevla* task

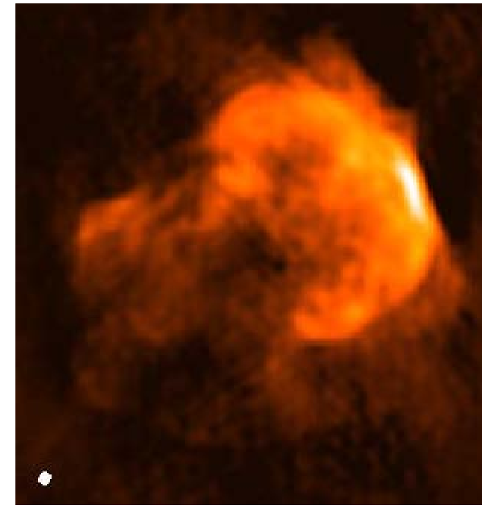
Diverse Data Reduced and Imaged in CASA…



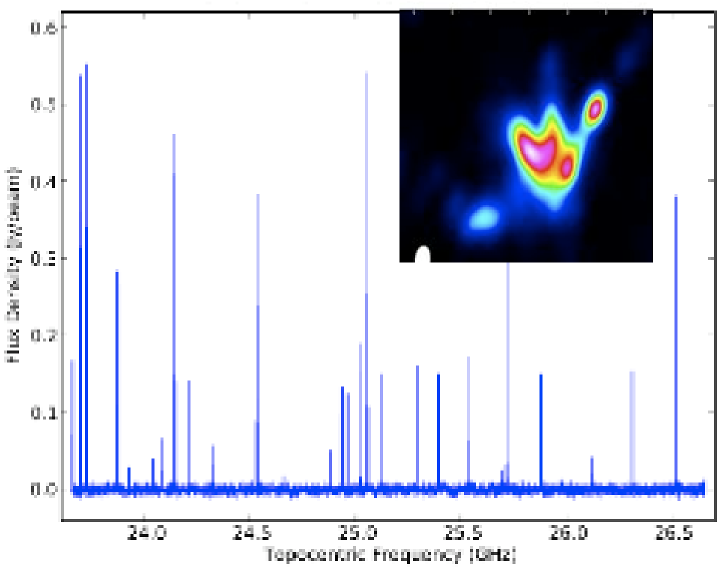
SMA ^{12}CO (2-1) mosaic toward IRDC G19.3+0.07



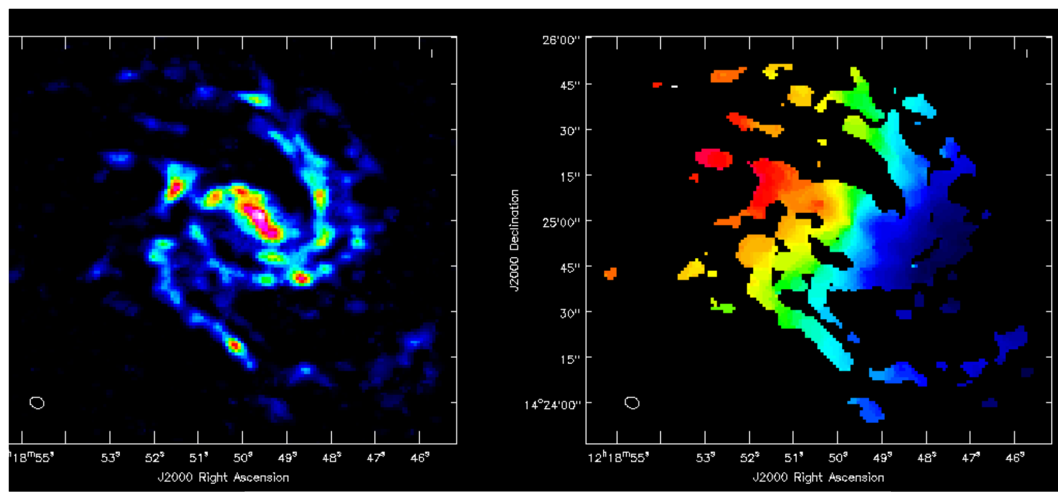
EVLA Ka-band of HC_3N in AGB star IRC+10216



EVLA 6cm mosaic of SNR: 3C391



EVLA demo science: Orion Hot Core this spectrum has 24k channels!



CARMA CO(1-0) mosaic of M99 (data courtesy STINGS team)

Import, Examination, Flagging

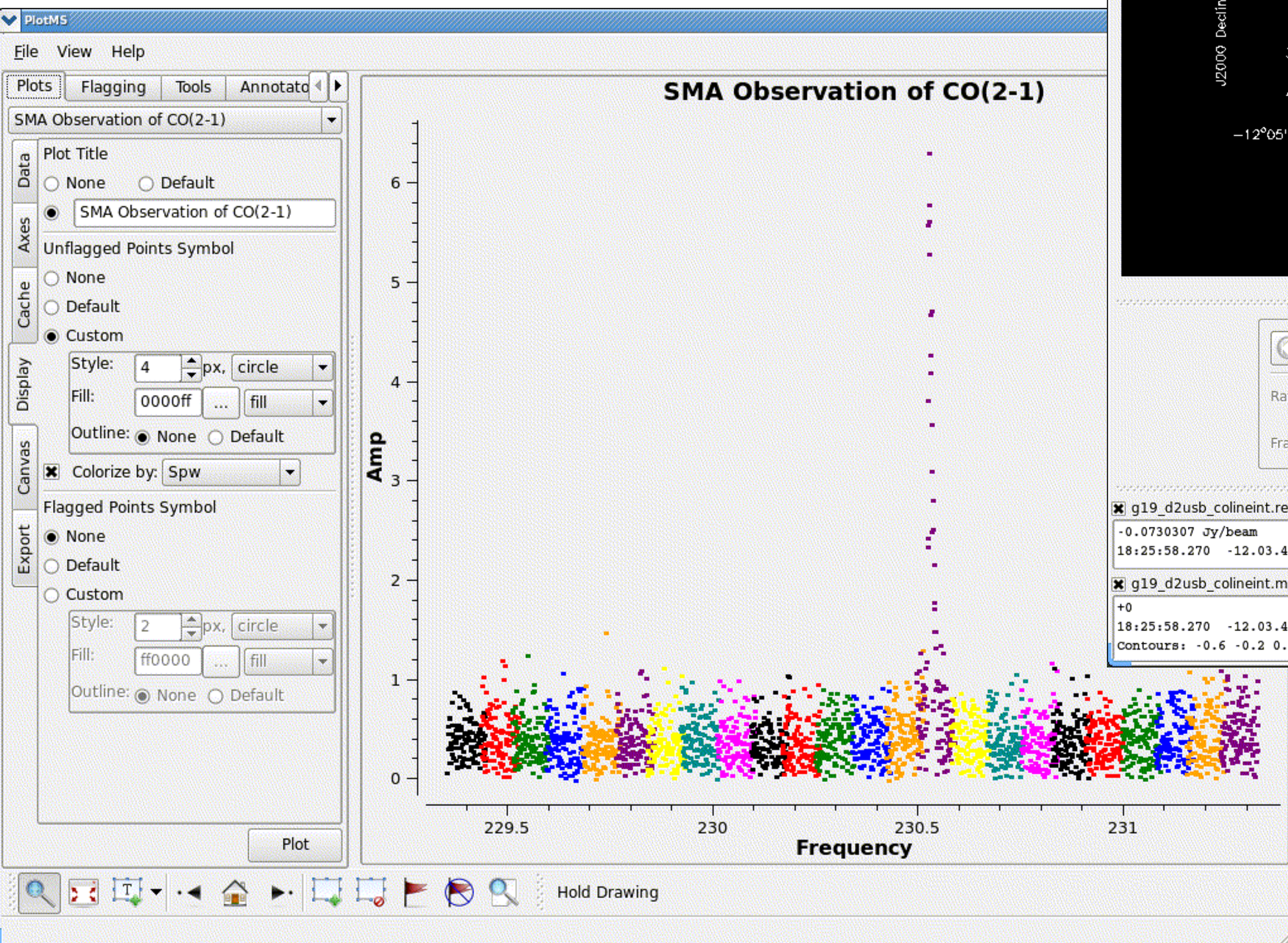
- Full ALMA/EVLA SDM data import (*importasdm*)
  - Raw SDM data formats were designed to be similar to MS
  - ALMA and EVLA use the same underlying format and filler
  - Custom task *importevla* to import and flag EVLA data
  - Importuvfits for data from other telescopes
  - Exportuvfits to AIPS etc. (some limitations)
- Interactive examination and flagging
  - X-Y plotting GUI plotms for visibilities and caltables (soon)
  - Raster (“TV”) display and flagging in *viewer*
- Non-interactive flagging
  - Task *flagdata* for simple flagging (manual, clip, shadow, “quack”)
  - Some experimental auto-flagging options (“rfi”, “autoflag”)
  - Auto-backup of flag columns of MS (access via *flagmanager*)

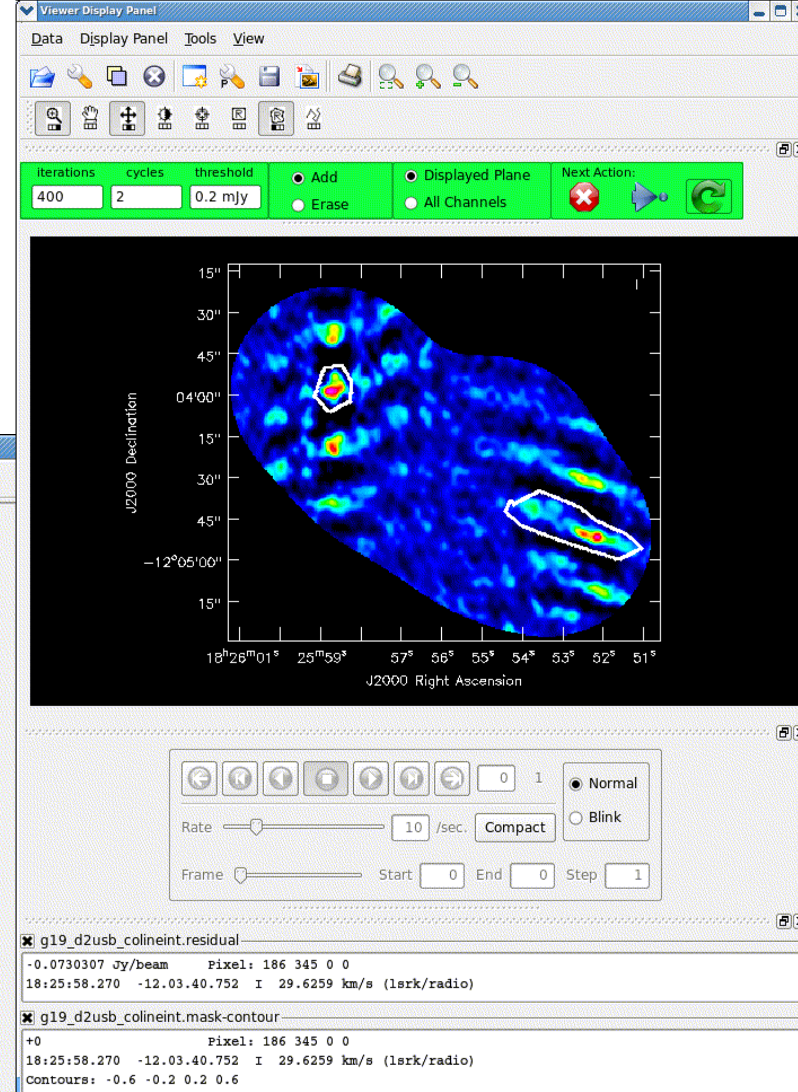
Calibration

- Application of a-priori calibration
  - antenna position, delays, gain curves, zenith optical depth
- Standard gain & bandpass calibration
  - Discrete and Polynominal/Spline solutions available
  - Flux density reference scaling (*setjy* and *fluxscale*)
  - Flexible mapping between spectral windows and/or fields
  - Amp+Phase, Phase, Amp solution modes, solution normalization
  - Auto-interpolation to flagged solutions (e.g. channels in bandpass)
- Polarization calibration
  - For circular RL basis, linear XY basis support under development
  - Linearized instrumental polarization (D-terms) solutions available
  - Channelized for frequency-dependent instrumental polarization
  - Optional solution for source polarization
  - Polarization position-angle solution support (cross-hand phase)

Interactive cleaning and flagging

UV-spectrum in plotms, colorized by spectral window





Imaging & Deconvolution

- Standard CLEAN and Multi-scale CLEAN
- Multi-frequency Synthesis Imaging
  - single term & Taylor series (spectral index,…) (Urvashi thesis)
- Mosaic imaging made simple
  - Joint deconvolution (Miriad style) and gridding convolution (new)
  - Mosaicing with heterogenous arrays (ALMA, CARMA)
- Widefield imaging: W-projection and faceting
  - W-projection more than 1 order of magnitude faster than faceting
  - Widefield can be combined with MFS and multi-scale
- Multiple algorithms for single dish and interferometry combination
- Experimental options:
  - RL “squint” correction & PA-dependent beam effects
  - pointing self-calibration
  - MEM & NNLS (toolkit level only so-far)

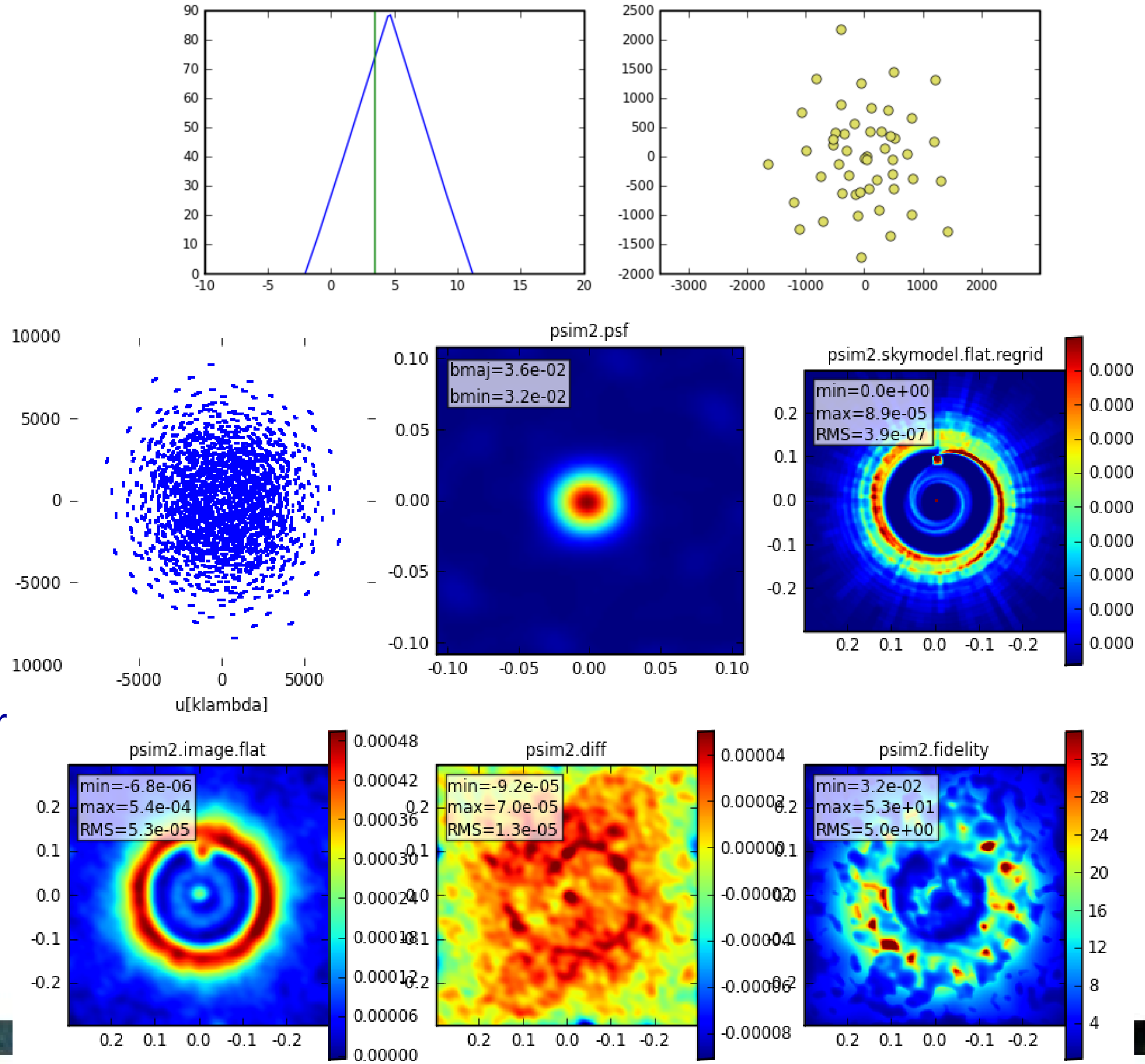
Analysis

- Image viewing in *viewer*
  - supports CASA images and FITS
  - standalone Qt application *casaviewer*
- Facilities for image manipulation
  - moment image calculation in *immoments*
  - *immath* task for combinations, polarization, spectral index
  - "Lattice Expression Language" in toolkit for advanced manipulation
  - other tasks (*specfit,* regridding, convolution) under development
- Access to image information and Python interface
  - can import or write Python functions
  - use of matplotlib
- Development of new analysis tools and tasks is a high priority
  - particularly in spectral line area (fitting, stacking, analysis)
  - user contributions via Python

CASA Simulator: "simdata"

o 1-stop task to simulate ALMA, EVLA, SMA, CARMA, ATCA, SKA

- Simulate continuum, simple spectral cubes
- Create coordinate system for model images
- calculate mosaic pointings (auto or manual)
- Optionally interleave a calibrator
- Simulate total power observations
- Add thermal noise and linear cross-polarization
- Re-image the data, interferometric + total power
- Analyze the difference in output and input images



Primary developer: Remy Indebetouw

NRAO

CASA Staff

- Group leader: Nick Elias (since Oct. 2009)
- Development team
  - Currently ~13.5 FTEs (scientific developers and programmers)
  - Distribution of people: NRAO: 8.3, NAOJ: 2.7, ESO: 2.5
  - Two recent vacancies: Application/GUI developer (just filled & started), High performance computing specialist (search underway)
- Overall Project management: Brian Glendenning (ALMA), Brian Butler (EVLA)
- NRAO CASA Science Steering Committee (CSSC)
  - Juergen Ott (Project Scientist)
  - Crystal Brogan (ALMA subsystem scientist)
  - Steve Myers (EVLA subsystem scientist)
  - Ed Fomalont (E2E Scientist)
- External Input:
  - ALMA/EVLA Commissioning teams, RSRO scientists, ASC/NAASC/ARC scientists, and scientific advisory committees (e.g. PASEO,ASAC,ANASAC)

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
- Being used heavily for EVLA science verification, RSRO visitors, and by outside users for start of early science since March 1, 2010
- Used every day in Chile for ALMA commissioning
- Dec. 18, 2009 was first non-beta release (3.0.0), patch 3.0.1 was released in April, the second patch (3.0.2) was released on June 14. The next major release 3.1.0 will be October 15.

NRAO

CASA Tutorials Around the World

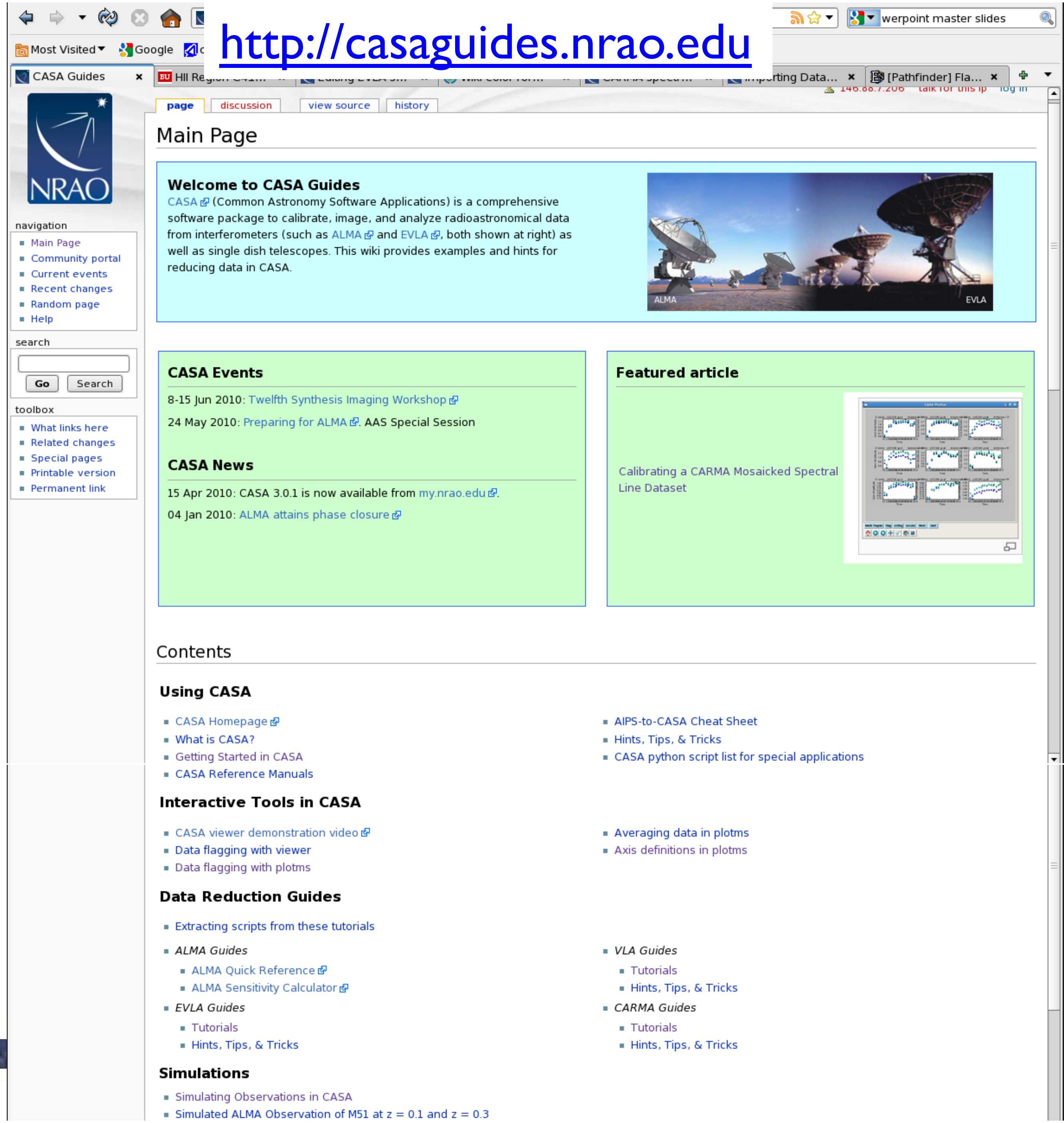
- Tutorials this past year:
  - Garching, Germany (May), Hamilton, Canada (June), Bonn, Germany (Oct.), Taiwan (Feb), NAOJ (April), Santiago, Chile (April)
  - May 24, Miami AAS special session with talks/demos
  - June 8-15, NRAO Synthesis Imaging Workshop (~150 students) – all first-day tutorials in CASA (except VLBA)
- Coming up:
  - July 19-20, Oxford, UK preceding the "Molecules in Galaxies" conference
  - Jan. 2011, Seattle AAS, demo/tutorials
  - Soon after ALMA early science decision (late fall) planning for NAASC workshops

NRAO

CASAguides

http://casaguides.nrao.edu

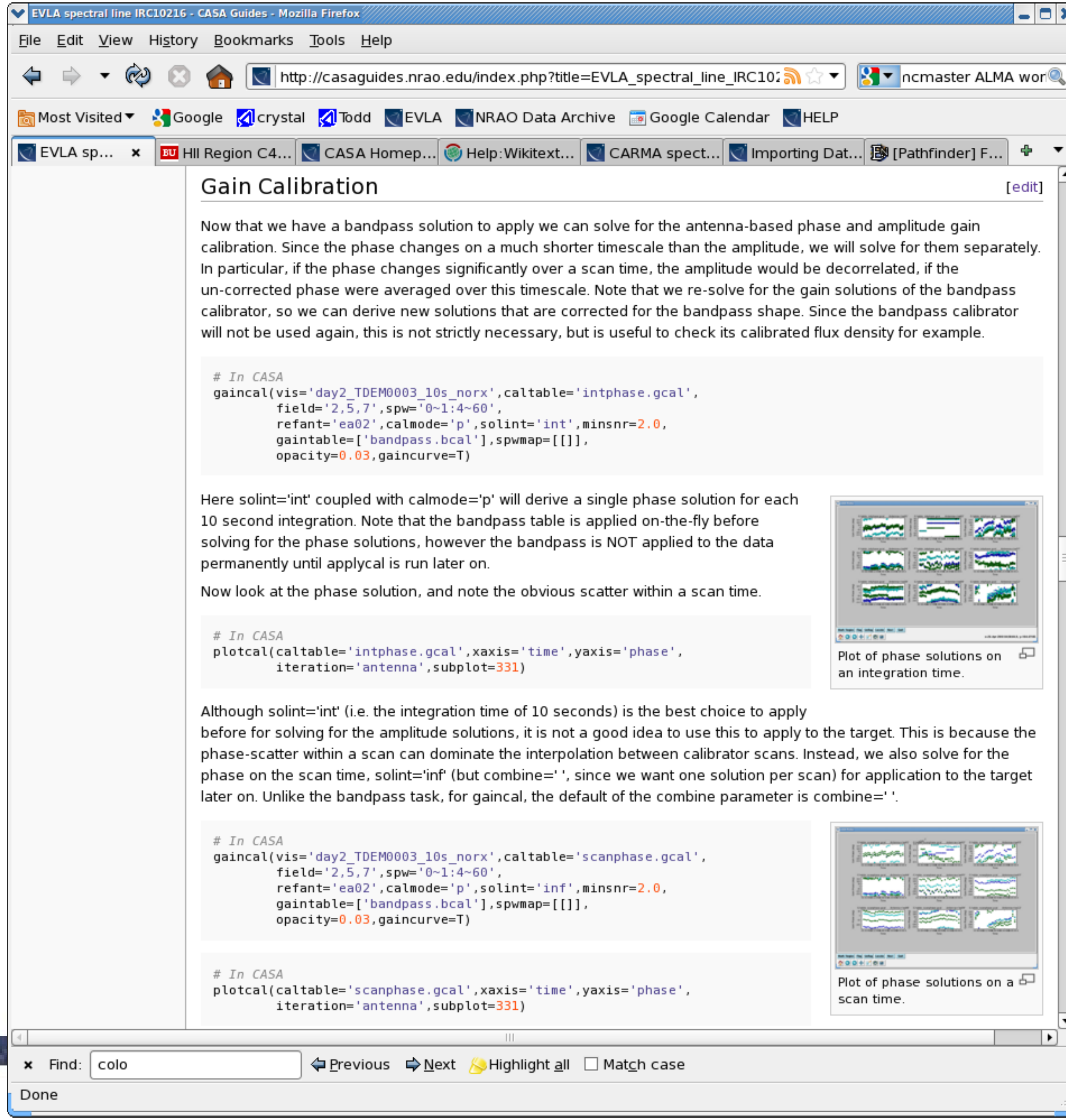
- Uses *mediawiki* to enable fully annotated scripts (that can be extracted and run)
- Additional "guides" continue to be added
- NRAO Synthesis Imaging summer school tutorials were successfully delivered this way (written by RSRO scientists)



CASAguides

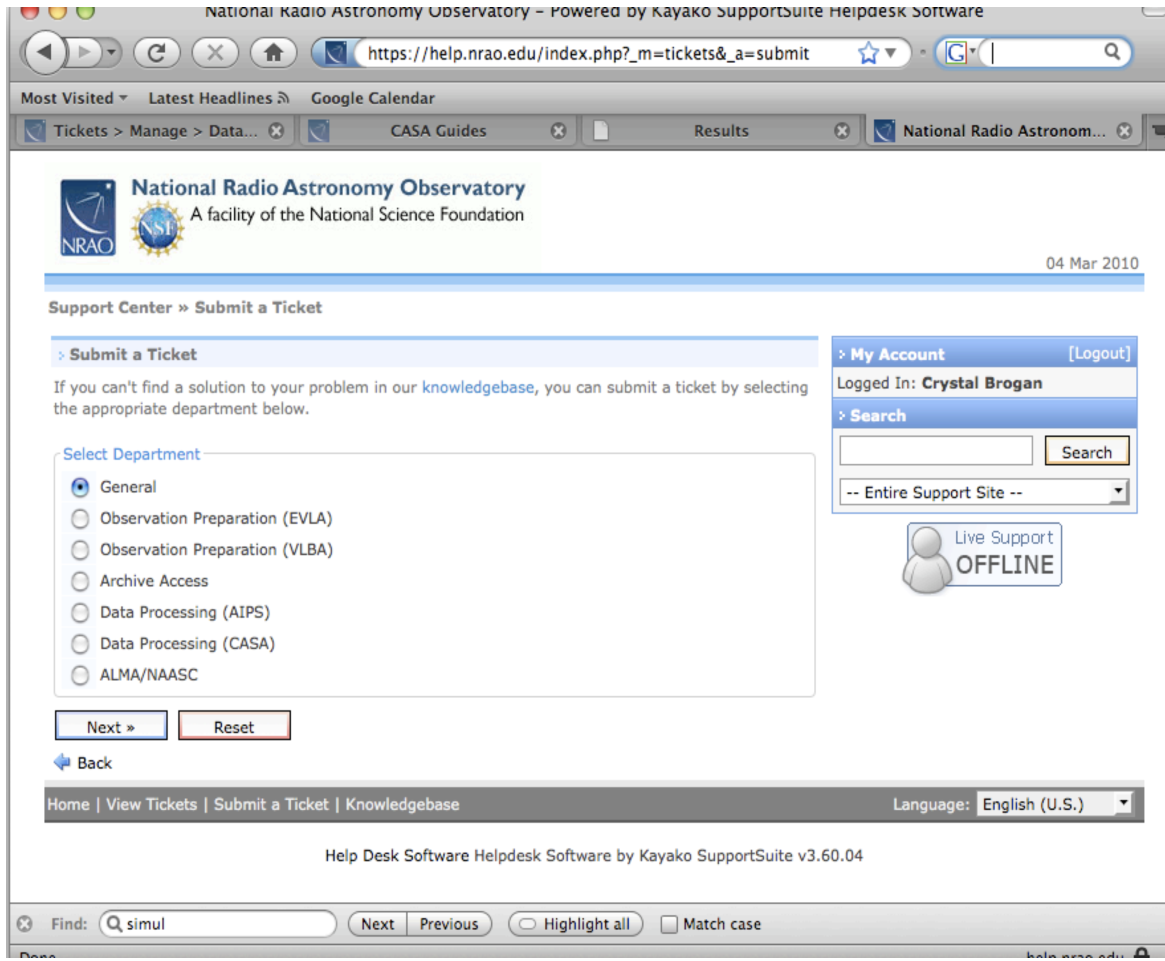
an example

- Uses *mediawiki* to enable fully annotated scripts (that can be extracted and run)
- Additional "guides" continue to be added
- NRAO Synthesis Imaging summer school tutorials were successfully delivered this way



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
- Since March 1 used for EVLA and CASA support



CASA Performance

- CASA currently has similar speed to other packages for ~ 10 GB datasets except for a few high nails being aggressively pursued (flagging, plotting)
- For any package the most expensive steps are: Flagging, Calibration (gain, bandpass, polarization, pointing, self-calibration), and Imaging
    - Channelization of data makes the problem “embarrassingly parallelizable”
    - However, particularly for imaging, the problem is I/O and not CPU limited making the problem trickier
- The release of multi-core CASA functionality will be staged.
    - CASA’s architecture has been written with parallelization in mind
    - Simple imaging (single field or simple mosaic <u>cube</u>) well progressed, expected for October release
    - Multi-core flagging and more imaging cases (multi-frequency synthesis continuum) expected June 2011
    - Successful hiring of HPC position critical
- Also pay attention to disk I/O (data rates > standard disk I/O speeds)
    - Investigating use of fast and/or parallel disk systems

Performance in Practice

- Size matters: currently CASA data is ~6 times larger than *miriad* and ~3 times larger than AIPS
  - Double precision 8 byte (Re,Im) floats (same as AIPS and 2x miriad) – essential for high dynamic range. CASA can use 4 byte instead
  - 3 scratch columns for intermediate calculations (AIPS: temporary scratch files on disk, miriad: mostly memory). CASA has high priority target to eliminate permanent scratch columns, reducing disk footprint to 1/3
- A test cluster with 16 nodes (128 cores) was purchased last year (joint EVLA/ALMA) to provide a test-bed for CASA parallelization
  - Recent imaging tests on a 0.1 TB dataset to make a 3600x3600x1024 cube on just one node (8 cores, 16 GB RAM) took ~17 hours
  - 6 hours peak RSRO data rate (75 MB/s) = 1.6 TB, so on 16 nodes, imaging takes ~ 17 hours on the test cluster. But the devil is in the details, and this is only imaging
  - For comparison, the test cluster was $70K. EVLA has ~$400K to purchase operations cluster (+NA ALMA has $250K). Thus the "real" pipeline clusters will be >5x more powerful than the test cluster (speed, memory, #nodes)
    - ➢ The problem is (potentially) within our means

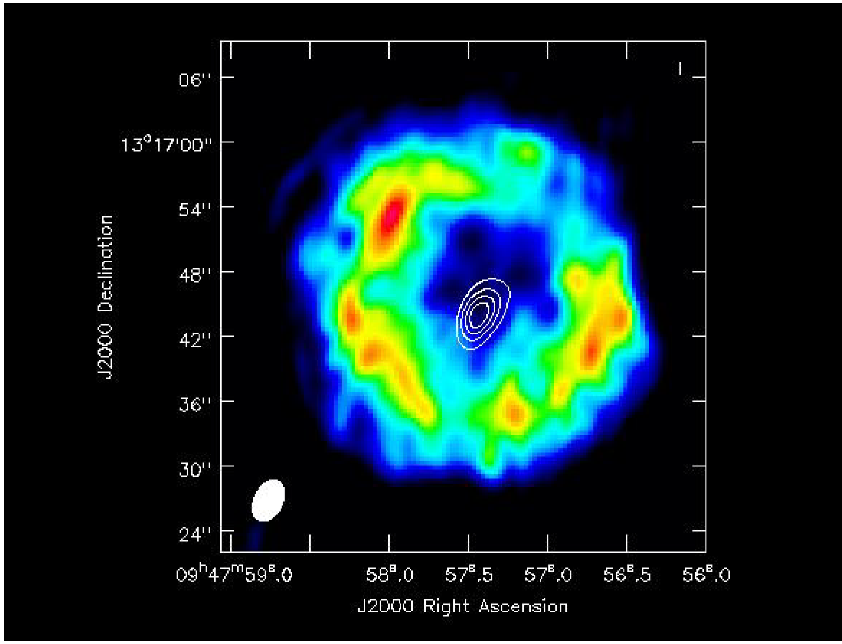
NRAO

Use Case: Basic Spectral Line

- Example Demo Science dataset: IRC+10216 mosaic at 36GHz

http://casaguides.nrao.edu/index.php?title=EVLA_Spectral_Line_Calibration_IRC+10216

- Dataset (OSRO-1, 26-Apr-2010) TDEM0003_sb1345754_1.55312.131578217595
  - 3hrs (15GB), 2 sub-bands x 64ch x 125kHz (HC_3N, SiO)
- Essential Steps:
  - Average to 10-sec (from 1-sec)
  - Examination and Flagging (in plotms)
  - Calibration: flux density, bandpass, gain
    - delays absorbed by bandpass
  - Calibration transfer (to target)
  - Imaging (single field) of spectral cube
    - standard clean and multi-scale clean
  - Analysis : moments, spectral profiles



HC_3N moment-0 (color) and continuum (contours)

Use Case: RSRO Spectral Line

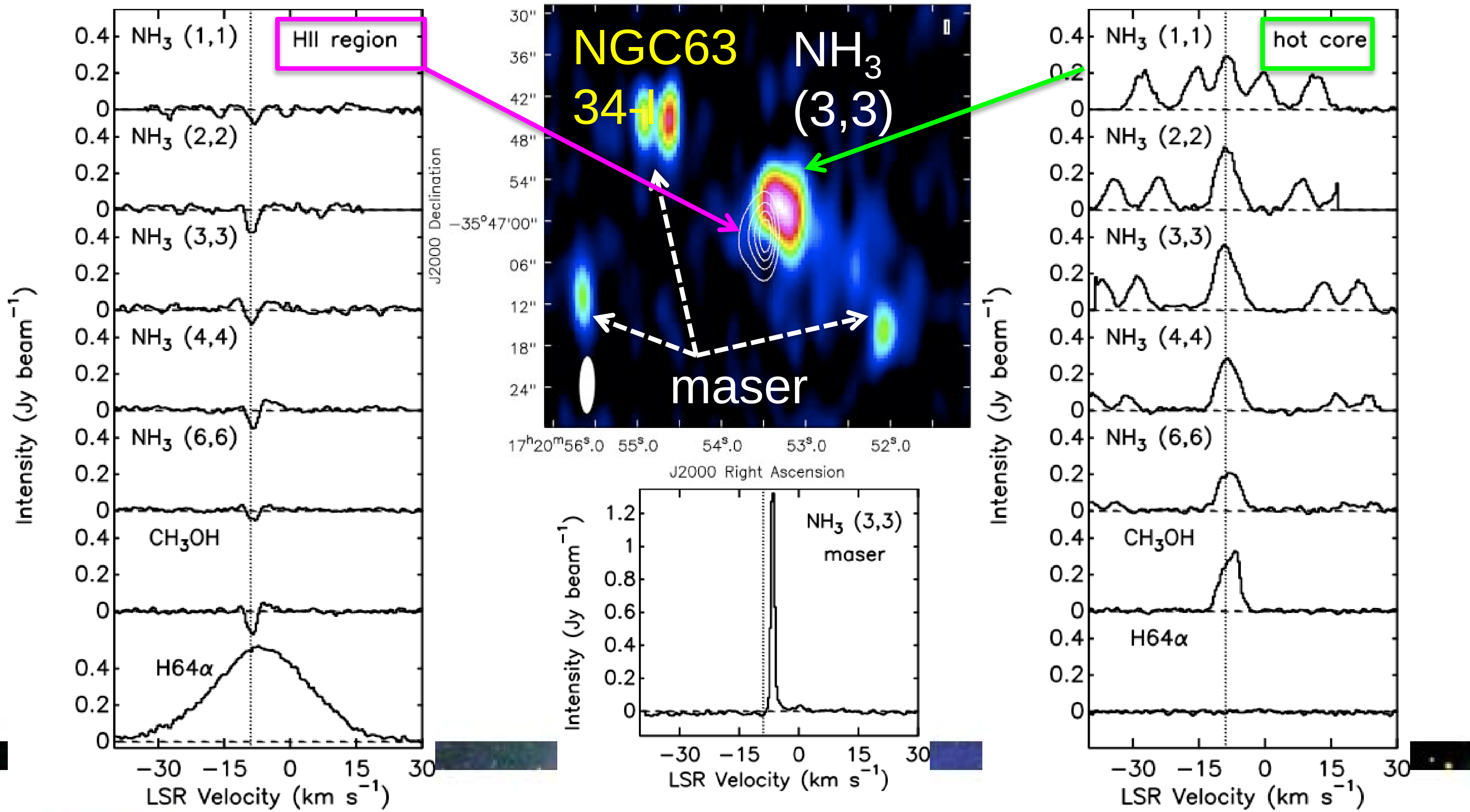
EVLA K-band Observations of massive young stellar objects in NGC6334-I

- 8 x 8 MHz sub-bands with 256 channels, RR only; also used referenced pointing: 10 minutes on source!!!!
- Test for RSRO project AB1346 (PI Crystal Brogan): “A Diagnostic K-band Survey of Massive Young Protostellar Objects” which will use 16 subbands

NRAO

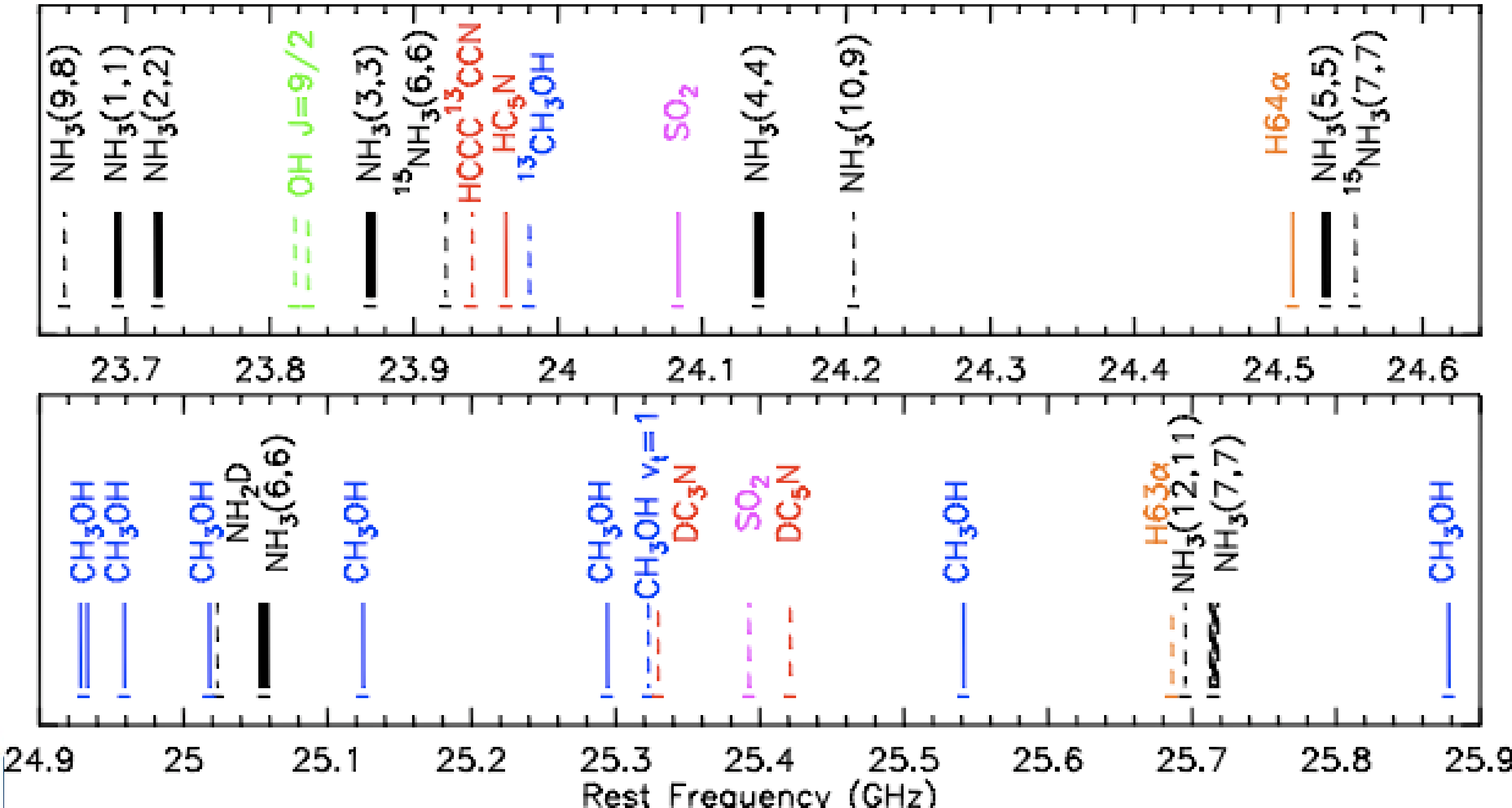
Use Case: RSRO Spectral Line

EVLA K-band Observations of massive young stellar objects in NGC6334-I



Use Case: RSRO Spectral Line

- RSRO (AB1346): A Diagnostic K-band Survey of 30 Massive Protostellar Objects



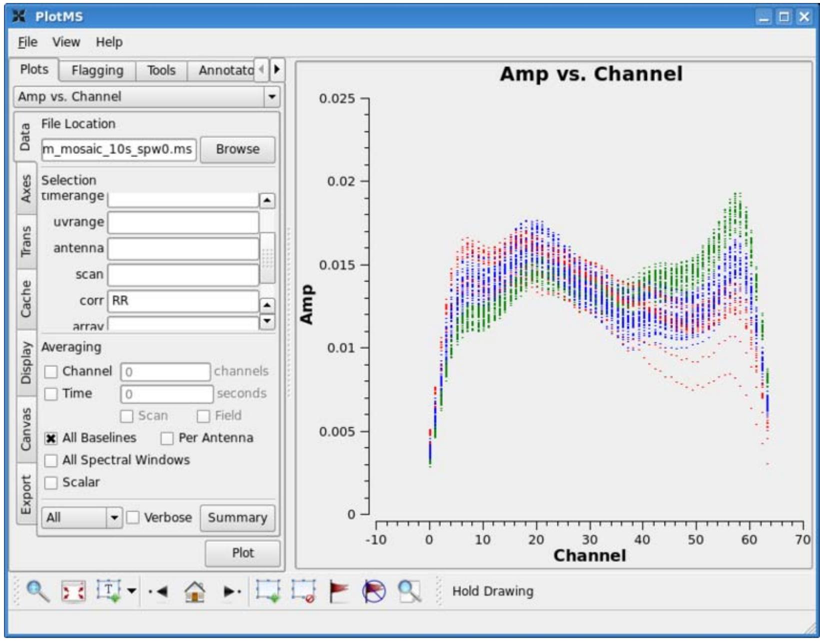
Use Case: RSRO Spectral Line

EVLA K-band Observations of massive young stellar objects in NGC6334-I

- 8 x 8 MHz sub-bands with 256 channels, RR only; also used referenced pointing: 10 minutes on source!!!!
- Test for RSRO project AB1346 (PI Crystal Brogan): "A Diagnostic K-band Survey of Massive Young Protostellar Objects" which will use 16 subbands
- In the "best case" scenario we will use 32 subbands (solid and dotted lines above) which includes a number of rare and Deuterated species
- In the "great case" scenario we will use 16 subbands (solid lines)
- Current test uses 8 of the sub-bands

Use Case: Basic Continuum Polarimetry

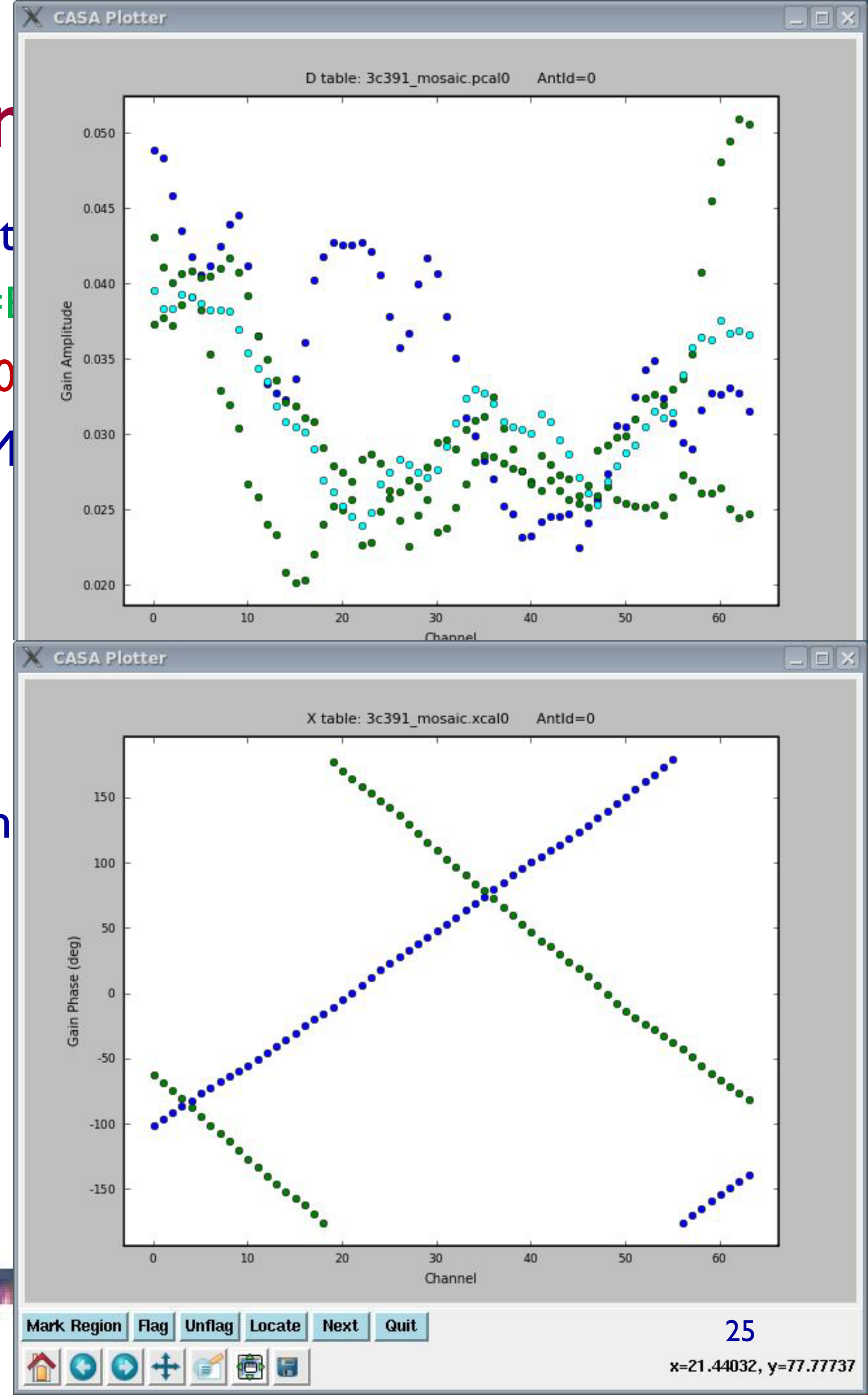
- Example Demo Science dataset: 3C391 at C-band
  http://casaguides.nrao.edu/index.php?title=EVLA_Continuum_Tutorial_3C391
- Dataset (OSRO-1, 24-Apr-2010) TDEM0001_sb1218006_1.55310.33439732639
  - 8hrs (40GB), 2 sub-bands x 64ch x 2MHz (4.54 and 7.44 GHz)
  - 7-field (hex-pattern) mosaic
- Essential Steps:
  - Average to 10-sec (from 1-sec)
  - Examination and Flagging (in *plotms*)
  - Calibration: flux density, bandpass, gain
    - delays absorbed by bandpass
  - Polarization Calibration
    - D-term (leakage) vs. frequency
    - R-L phase (angle) vs. frequency
  - Calibration transfer (to target)



Bandpass in *plotms* (amp vs. channel)

Use Case: Basic Contir

- Example Demo Science dataset: 3C391 at
    - http://casaguides.nrao.edu/index.php?title=I
- Dataset (OSRO-1, 24-Apr-2010) TDEM000
    - 8hrs (40GB), 2 sub-bands x 64ch x 2M
    - 7-field (hex-pattern) mosaic
- Essential Steps:
    - Average to 10-sec (from 1-sec)
    - Examination and Flagging (in *plotms*)
    - Calibration: flux density, bandpass, gain
        - delays absorbed by bandpass
    - Polarization Calibration
        - D-term (leakage) vs. frequency
        - R-L phase (angle) vs. frequency

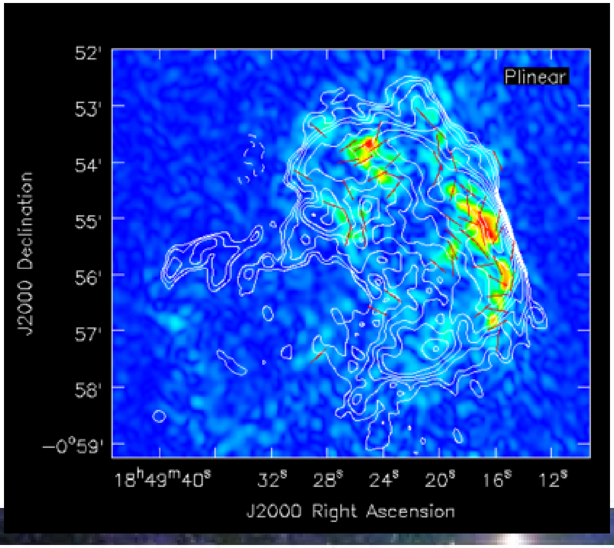


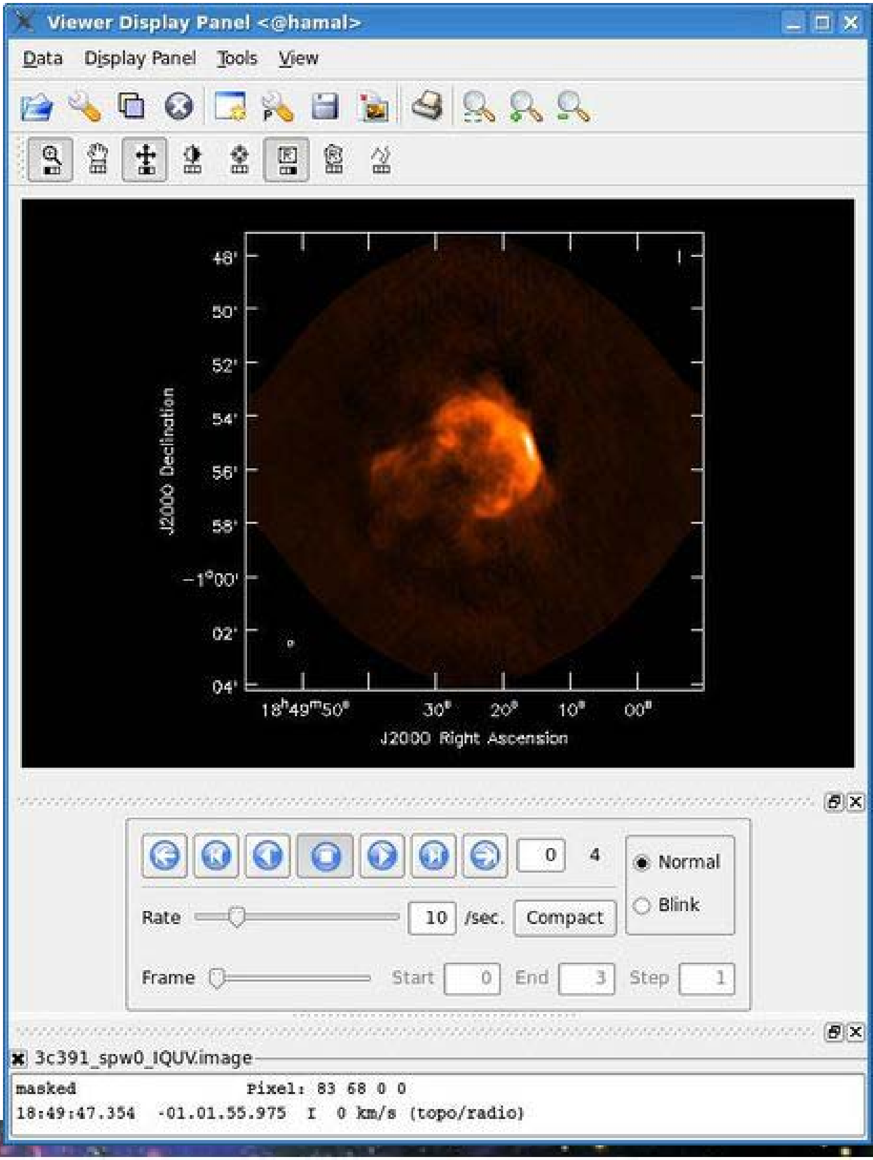
Df solution (top) and Xf solution (bottom)

NRAO

Use Case: Basic Continuum Polarimetry

- Essential Steps (continued):
  - Imaging (joint mosaic) MFS continuum
    - standard clean and multi-scale clean
    - IQUV imaging
  - Analysis : polarization
    - polarized intensity
    - polarization angle – 0.5atan(U/Q)





Key EVLA Development

- Priority – support of EVLA and ALMA commissioning needs
  - including Early Science, OSRO & RSRO
- Parallelization and cluster fine-tuning for imaging and flagging
  - use of EVLA cluster
- Evolving use and support of SDM
  - flagging tables & switched power (EVLA)
- Polarization calibration of linear feeds (ALMA, P/4-band EVLA)
- Improvements to interactive plotting & flagging in *plotms*
  - full incorporation of caltable plotting
  - support for switched power calibration & Tsys
- Improvements to "TV" based flagging in the *viewer*
  - on-the-fly spectral and time averaging
- Delay fitting & correction (incl. cross-hand, not global fringe-fitting)

NRAO

Key EVLA Development (continued)

- Planet tracking & models for use as resolved calibrators
- Use of atmospheric models (e.g. for optical depth)
- Calibration efficiency and usability improvements
  - use of source models without scratch column use
  - better management and mapping of calibration tables
- *Splatalogue* line-search capabilities and over-plotting
- Viewer improvements (especially for spectral line plotting, analysis)
- Imaging & deconvolution improvements:
  - mosaicing, high-dynamic range, wideband MFS + multi-scale
- Image analysis task improvements (spectral fitting, etc.)
  - current suite of analysis tasks is minimal, needs fleshing out
- Status Summary: CASA is handling current OSRO observing, RSRO D/C-config with some efficiency issues (data sizes), RSRO B/A-config will be more of a challenge (less averaging, big images)

NRAO

Challenges for EVLA processing

- Data rates and volumes
  - RSRO 75MB/s (2010-2011) → 1.6TB in 6hrs
- Wideband (1-8 GHz/pol)
  - large spectral cubes (thousands of channels)
  - multi-frequency synthesis (MFS) for continuum imaging
- Widefield (low-frequency single-field, high-frequency mosaicing)
  - "peeling" of bright sources in outer beam regions
  - direction-dependent corruption (polarized beam, ionosphere, pointing)
- High Dynamic Range imaging (10^4-10^7:1)
  - non-closing errors (e.g. due to polarization)
- Radio Frequency Interference (auto-flagging. correction)
- All of the above at the same time! <u>Avenues for mitigation:</u>
  - High-performance (cluster) computing and parallelization
  - New algorithms

NRAO

User Experiences

- Acceptance of CASA by AIPS & Miriad users has been slow
  - CASA focus has been for new users (e.g. Summer School)
  - young/new users seem to like it and catch on quickly
- AIPS is still supported, useable for most EVLA modes
  - used in commissioning & verification
  - new algorithm development (Greisen, Kogan, Cotton)
- CASA task interface
  - designed to be familiar to AIPS and Miriad users
  - full access to toolkit and use of Python (widely used & documented)
- Strengths & Weaknesses (real & perceived)
  - no known “fatal flaws” (see 2003 Technical Review)
  - lack of familiarity for veteran users (approach in different way)
  - its software! there are bugs, and currently gaps in functionality
    - working towards fulfilling user requirements (see documents)

Issues and Future Directions Summary

- CASA is now out of its “beta-testing” and into its use phase
  - being exercised daily for commissioning and science verification
  - growing user base (e.g. RSRO users)
  - focus is on EVLA and ALMA, but there are other interested users (SMA, CARMA, GMRT, e-MERLIN, other instruments)
- Focus shifting from filling gaps in functionality to “useability”
  - continually improving robustness and efficiency
  - hard to satisfy user preferences (e.g. interfaces, GUI look-and-feel)
- New Algorithms
  - must develop and implement new algorithms to enable the EVLA modes that make use of the main sensitivity gains (e.g. bandwidth)
- High Performance
  - must work efficiently in cluster environment for high data rate modes
  - defense in depth: MPI, OpenMP, simple task parallelization, I/O