CASA

EVLA Advisory Committee Meeting, March 19-20, 2009

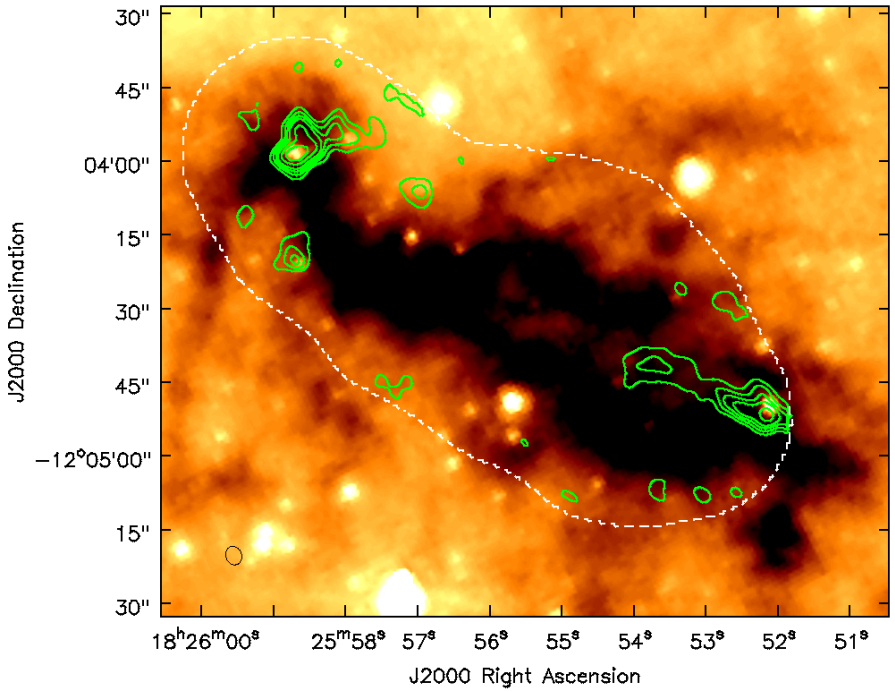
B.E. Glendenning

CASA Group Leader

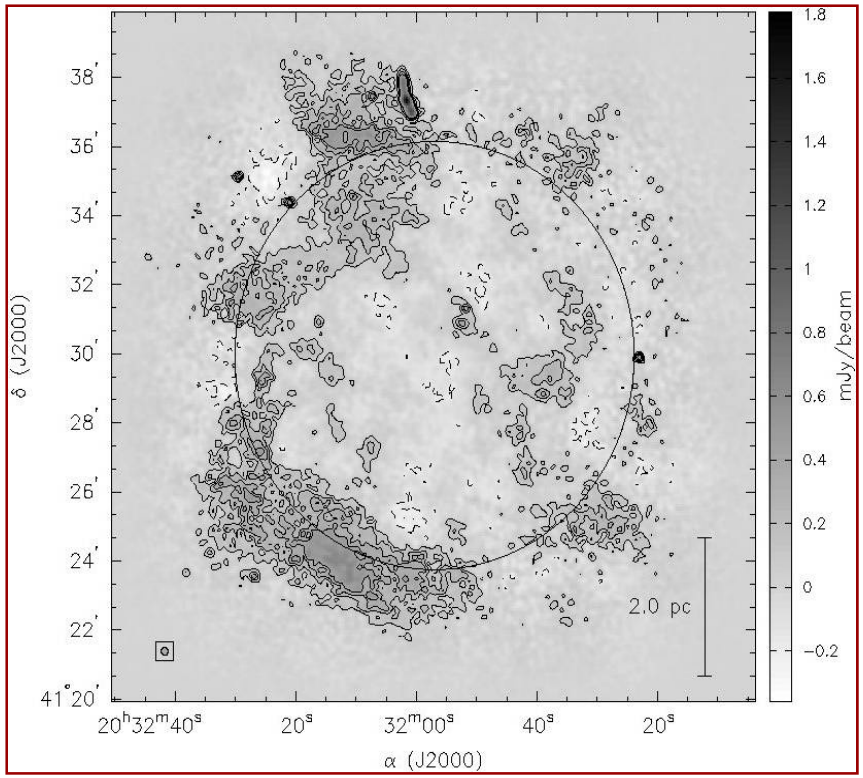
Atacama Large Millimeter/submillimeter Array
Expanded Very Large Array
Robert C. Byrd Green Bank Telescope
Very Long Baseline Array

NRAO

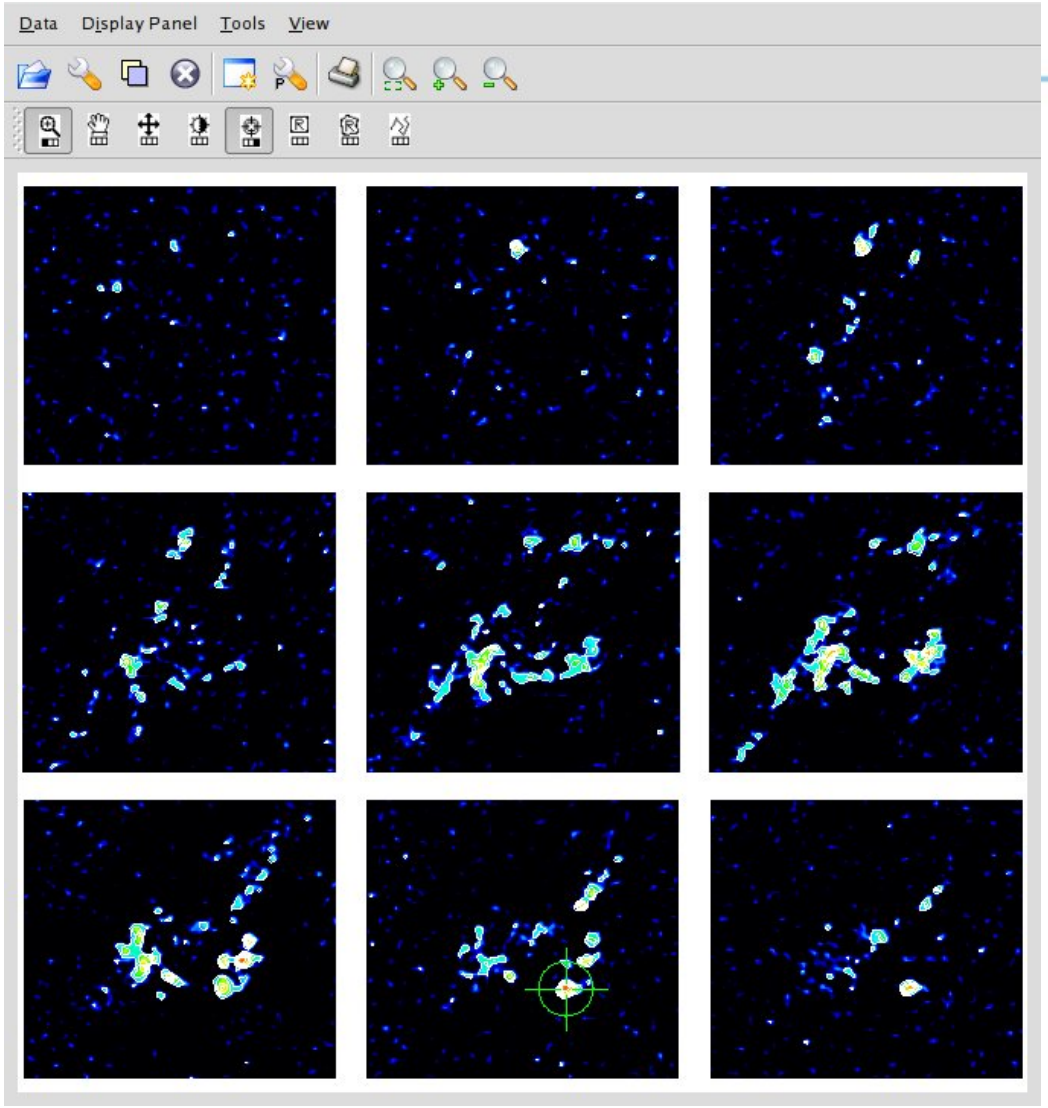
Production of Scientific Images in CASA Becoming Routine

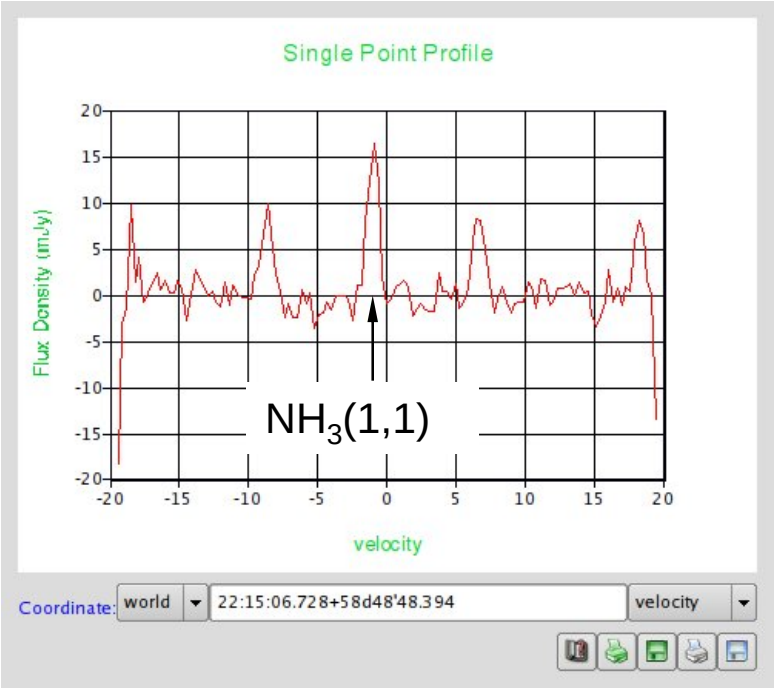


Green contours show SMA ^{12}CO (2-1) integrated intensity superposed on a GLIMPSE 8 µm image of the Infrared Dark Cloud (IRDC) G19.3+0.07. Six-pointing SMA mosaic imaged in CASA – calibration of SMA data coming soon. Brogan et al. (in prep).



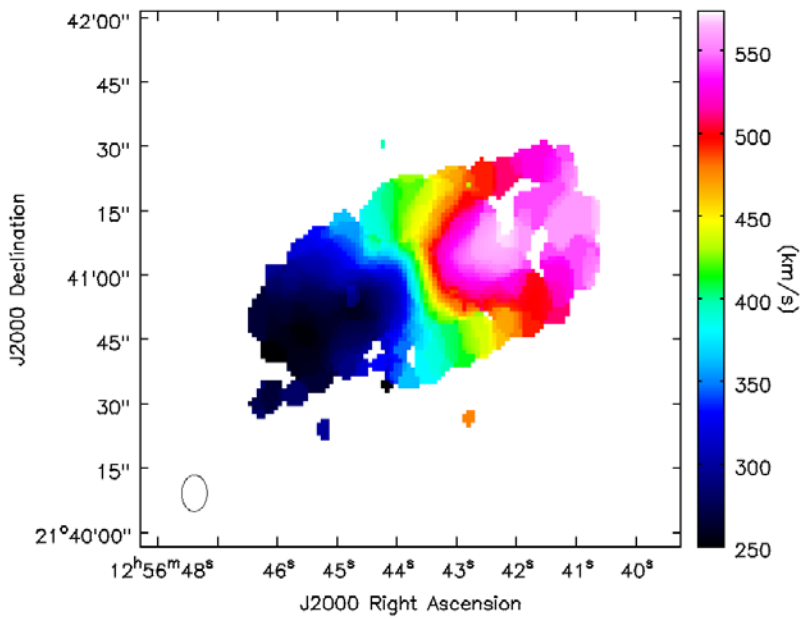
An extended radio counterpart of TeV J2032+4130 in the Cygnus OB association. VLA 3.6 cm continuum 5 point mosaic, D configuration, multi-scale clean. Butt et al. (2008)



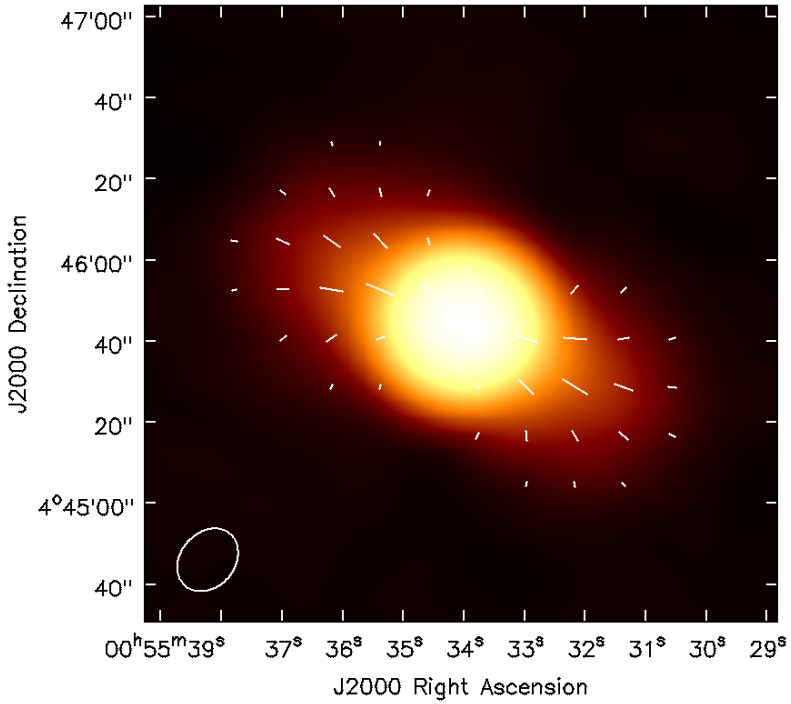


IRAS 22134 - a young ring cluster. VLA mosaic of $NH_3(1,1)$ and associated spectrum at one point in the ring. Main and hyperfine components are visible in the spectrum. Shepherd & Kumar 2008, ApJ, in prep.

Data Calibrated and Imaged in CASA Tutorials at NRAO Synthesis Imaging Workshop June 2008



CO(1—0) kinematics (moment 1) of the galaxy NGC4826 from the BIMA SONG survey (data originally published in Helfer, Thornley, Regan et al. 2003)

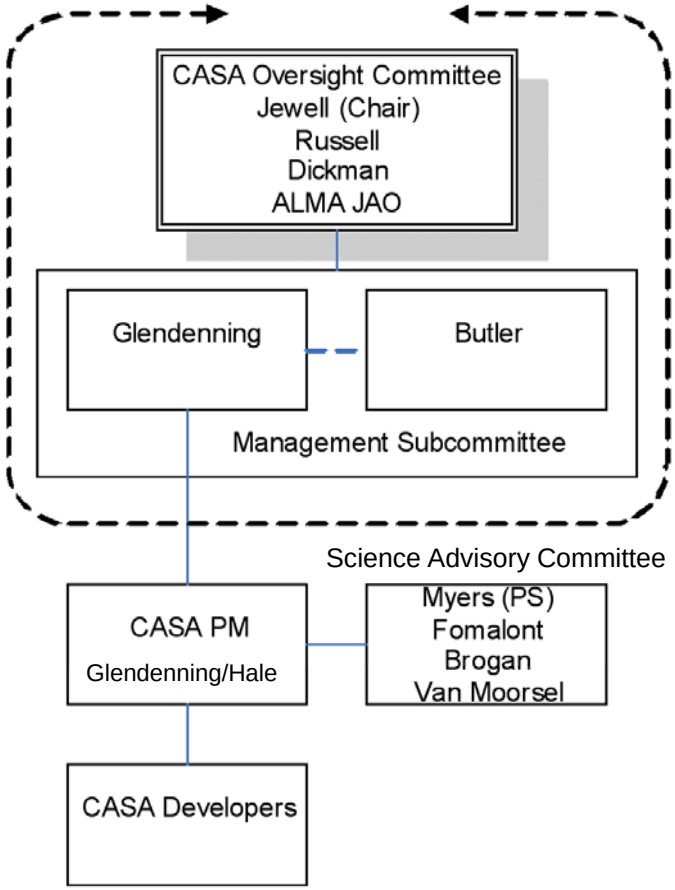


E-field vectors in Jupiter magnetosphere. Archival VLA 6 cm D-configuration full Stokes polarization data.

What is CASA?

- CASA is the baseline post-processing package for EVLA and ALMA data
- 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 cookbook)
- Telescope data (visibility and single-dish) are stored in a MeasurementSet (MS); a filler converts EVLA SDM+BDF 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

Organization



CASA Staff

– Group leader: B. Glendenning (Interim)

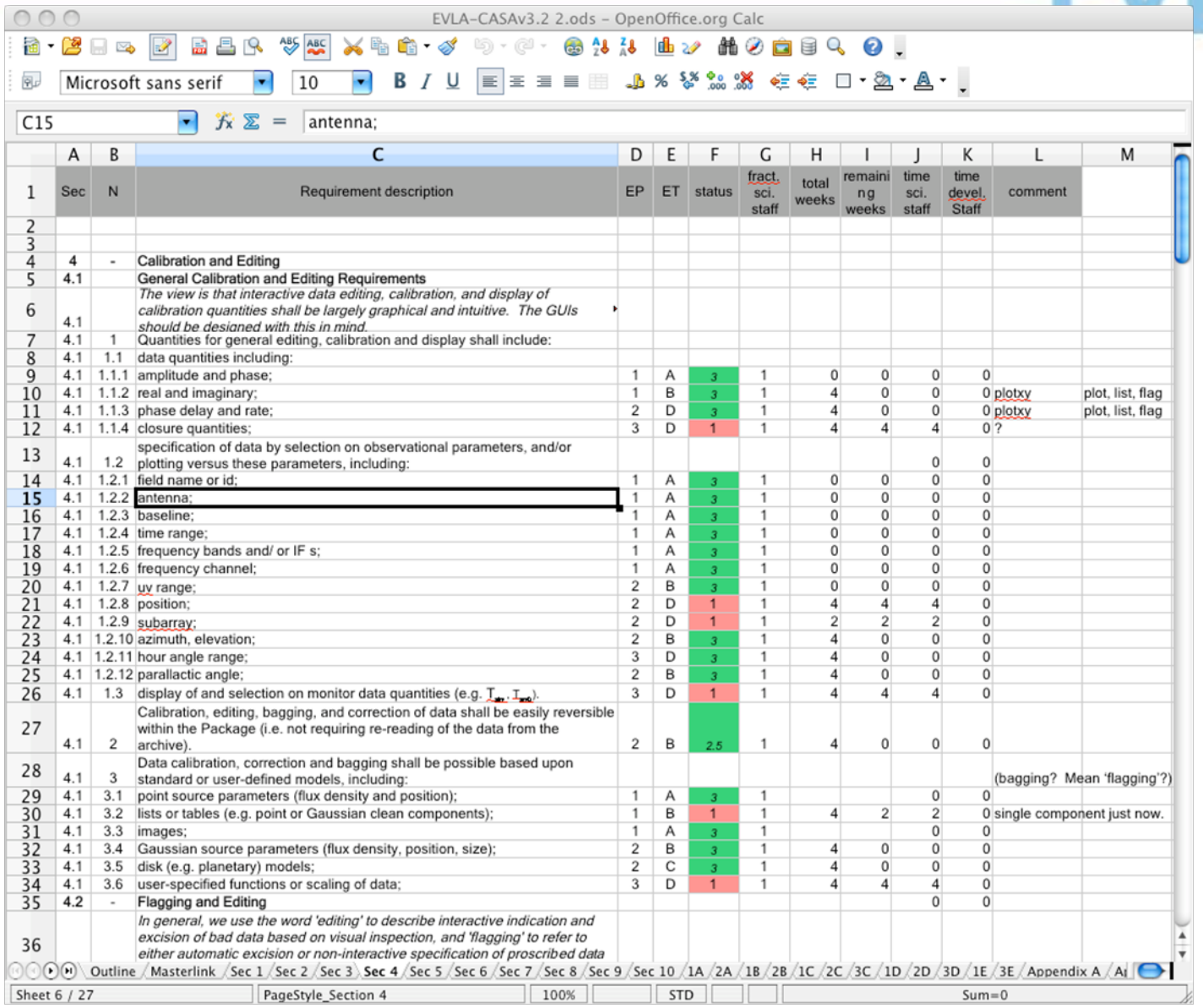
- Unsuccessful search and lack of full-time leader ***was*** a significant problem, added a deputy project manager (A. Hale) November 2008
- The situation is now satisfactory but not ideal, the search has restarted (2 interviews in April)

– Development team

- Significantly augmented by ALMA funding since the last EAC meeting
  - Better mix of scientist / CS developers
- Currently ~13.5 FTEs
- Distribution of people:
  - NRAO: 8.5 (including 1 vacancy), NAOJ: 2.2, ESO: 2, U Calgary: 1
  - Two term positions will end in 2009
- NRAO Additions 2009 (all start in the next 2-3 months)
  - 2 Scientific developers (D. Mehringer, T. Tsutsumi)
  - 1 Post-doc (R. Friesen)
  - 1 Summer Sabbatical (M. Thornley, Bucknell University)
  - At the end of 2009 will have 14.25 FTE working on CASA

EVLA Cost to Complete
(Based on van Moorsel, 2008)

- 36.5 FTE-y, 10.4 FTE
  - FTE-y based on requirements completion status and estimated effort for outstanding items, FTE based on EVLA project milestones
  - CASA is ready for OSRO
- Does not include all technical infrastructure & support activities
  - ~1.4x more staff approximately trades off
  - However entirety of lowest priority requirements is only marginally feasible
- Depends significantly on ALMA resources and overlap
  - Particular the addition of scientist/developers



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Sec	N	Requirement description	EP	ET	status	fract. sci. staff	total weeks	remaining weeks	time sci. staff	time devel. Staff	comment	
2													
3													
4	4	-	Calibration and Editing										
5	4.1		General Calibration and Editing Requirements										
6	4.1		*The view is that interactive data editing, calibration, and display of calibration quantities shall be largely graphical and intuitive. The GUIs should be designed with this in mind.*										
7	4.1	1	Quantities for general editing, calibration and display shall include:										
8	4.1	1.1	data quantities including:										
9	4.1	1.1.1	amplitude and phase;	1	A	3	1	0	0	0	0		
10	4.1	1.1.2	real and imaginary;	1	B	3	1	4	0	0	0	plotxy	plot, list, flag
11	4.1	1.1.3	phase delay and rate;	2	D	3	1	4	0	0	0	plotxy	plot, list, flag
12	4.1	1.1.4	closure quantities;	3	D	1	1	4	4	4	0	?	
13	4.1	1.2	specification of data by selection on observational parameters, and/or plotting versus these parameters, including:							0	0		
14	4.1	1.2.1	field name or id;	1	A	3	1	0	0	0	0		
15	4.1	1.2.2	antenna;	1	A	3	1	0	0	0	0		
16	4.1	1.2.3	baseline;	1	A	3	1	0	0	0	0		
17	4.1	1.2.4	time range;	1	A	3	1	0	0	0	0		
18	4.1	1.2.5	frequency bands and/ or IF s;	1	A	3	1	0	0	0	0		
19	4.1	1.2.6	frequency channel;	1	A	3	1	0	0	0	0		
20	4.1	1.2.7	uv range;	2	B	3	1	0	0	0	0		
21	4.1	1.2.8	position;	2	D	1	1	4	4	4	0		
22	4.1	1.2.9	subarray;	2	D	1	1	2	2	2	0		
23	4.1	1.2.10	azimuth, elevation;	2	B	3	1	4	0	0	0		
24	4.1	1.2.11	hour angle range;	3	D	3	1	4	0	0	0		
25	4.1	1.2.12	parallactic angle;	2	B	3	1	4	0	0	0		
26	4.1	1.3	display of and selection on monitor data quantities (e.g. T_{ant}, T_{sys}).	3	D	1	1	4	4	4	0		
27	4.1	2	Calibration, editing, bagging, and correction of data shall be easily reversible within the Package (i.e. not requiring re-reading of the data from the archive).	2	B	2.5	1	4	0	0	0		
28	4.1	3	Data calibration, correction and bagging shall be possible based upon standard or user-defined models, including:									(bagging? Mean 'flagging'?)	
29	4.1	3.1	point source parameters (flux density and position);	1	A	3	1			0	0		
30	4.1	3.2	lists or tables (e.g. point or Gaussian clean components);	1	B	1	1	4	2	2	0	single component just now.	
31	4.1	3.3	images;	1	A	3	1			0	0		
32	4.1	3.4	Gaussian source parameters (flux density, position, size);	2	B	3	1	4	0	0	0		
33	4.1	3.5	disk (e.g. planetary) models;	2	C	3	1	4	0	0	0		
34	4.1	3.6	user-specified functions or scaling of data;	3	D	1	1	4	4	4	0		
35	4.2	-	Flagging and Editing							0	0		
36			*In general, we use the word 'editing' to describe interactive indication and excision of bad data based on visual inspection, and 'flagging' to refer to either automatic excision or non-interactive specification of proscribed data*										

Current Status

- CASA is already a very usable data reduction package for the current VLA and OSRO
  - Focus of a 5-month study by a Scientific visitor (S. Palen)
- Have had Beta (patch) releases every ~3 months since October 2007
  - Initially restricted, now available after registration anyone (>200 people)
  - Tutorial at synthesis imaging summer school, ~50 students (positive feedback)
- Used every day for EVLA correlator data translation & at the ALMA Test Facility (ATF) (since shut down), ALMA Chile soon
- Generally very capable, although too much expertise is sometimes required
- Image analysis capabilities need to be exposed in user interface, bound to visualization

CASA Strengths for SRO

- Full data import (e.g., complex correlator setups)
- 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) solutions
- Low-level data inspection/modification tools; scriptability in general

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 raster single dish software*
    - 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)

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)

Calibration (2)

- Additional features
  - *Flexible combinations of data (over scan,field,spw) for solving ("fan in")*
  - *Flexible distribution of solutions to data ("fan out")*
  - Smoothing
  - Interpolation and accumulation (incremental)
  - Solution plotting, including interactive flagging

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

Performance (2)

- Started: Terabyte initiative (Bhatnagar/Rau talks)
  - Flag, calibrate, image 1 TB (raw data size) data = 10h of peak data
  - Cluster (16 nodes, 128 cores) purchased, working on simulating the data and initial timing tests (joint ALMA/EVLA purchase)
  - Some initial super-linear speedups for the embarrassingly parallel case!

Testing EVLA sized data sets is the important exercise!

Release 2.4 Features (1 June)

- Much improved (>10x faster) visibility GUI plotting/flagging tool
- Make scratch columns only when absolutely needed (often 3x less data on disk)
- Filler extension to follow updated raw data (correlator output) format definitions
- Flagging improvements (e.g., perform multiple flagging operations in a single pass through the data)
- Improved import/export of FITS files (UV & image)
- Visibility post-observing fixup task(recalculate UVWs, apply differential aberration, near-field corrections, …)
- Improved primary beam handling and channel interpolation in imaging
- Generalize ALMA simulator for EVLA simulation