EVLA Data Processing PDR

Post-processing

Athol Kemball, NRAO

Overview

- Introduction to AIPS++
- Planning for the EVLA within AIPS++
- EVLA post-processing science requirements
- Current capabilities in AIPS++

Introduction to AIPS++

- **A**stronomical **I**nformation **P**rocessing **S**ystem
- Next-generation data reduction package for radio astronomy
- Designed by a team of astronomers and programmers
- Developed by an international consortium of observatories
- Fifth public release available (v1.6) (December 2001)
- Sixth public release (v1.7) available in July 2002

http://aips2.nrao.edu

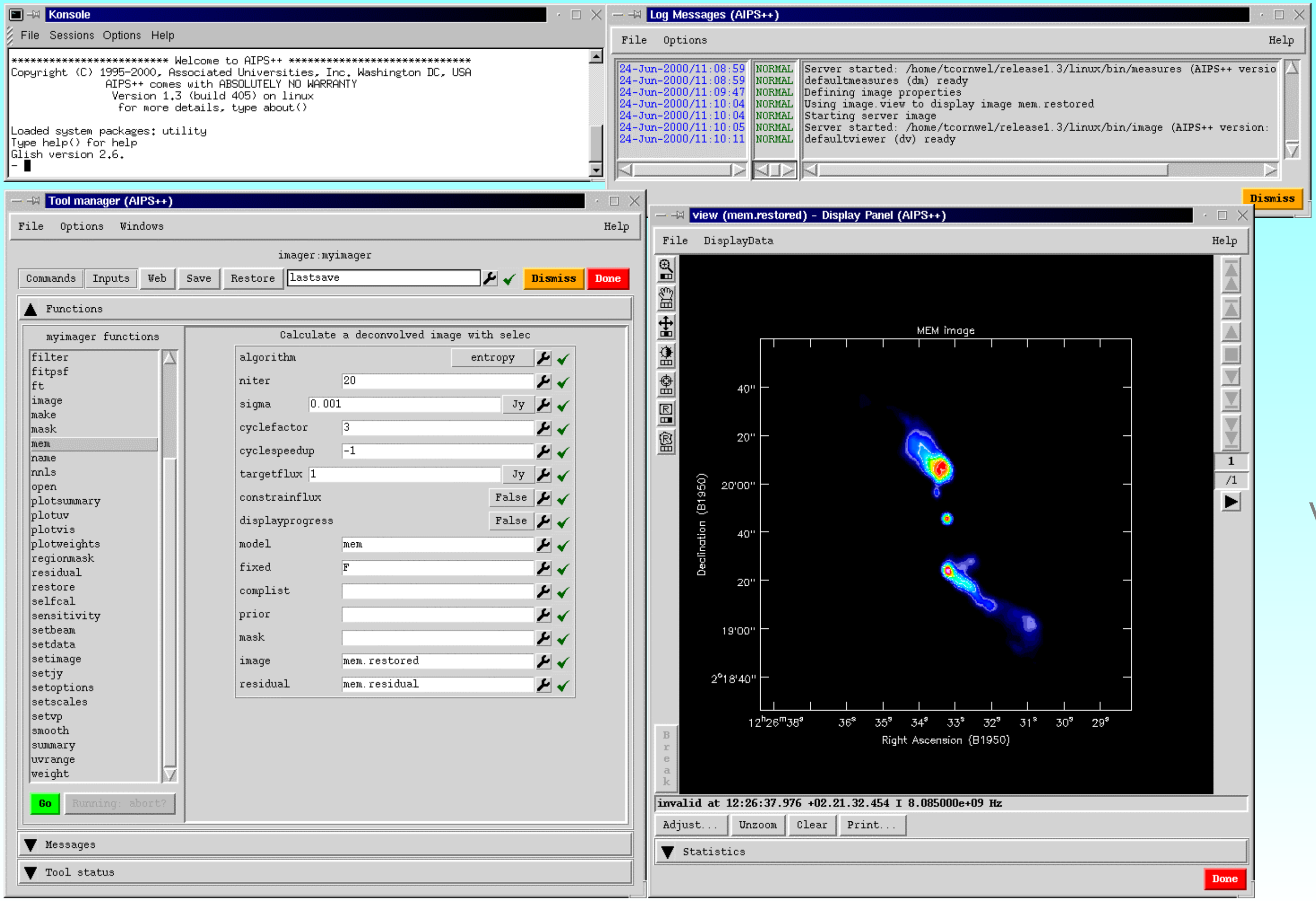
AIPS++ user interface

Glish

logger

Standard GUI

viewer



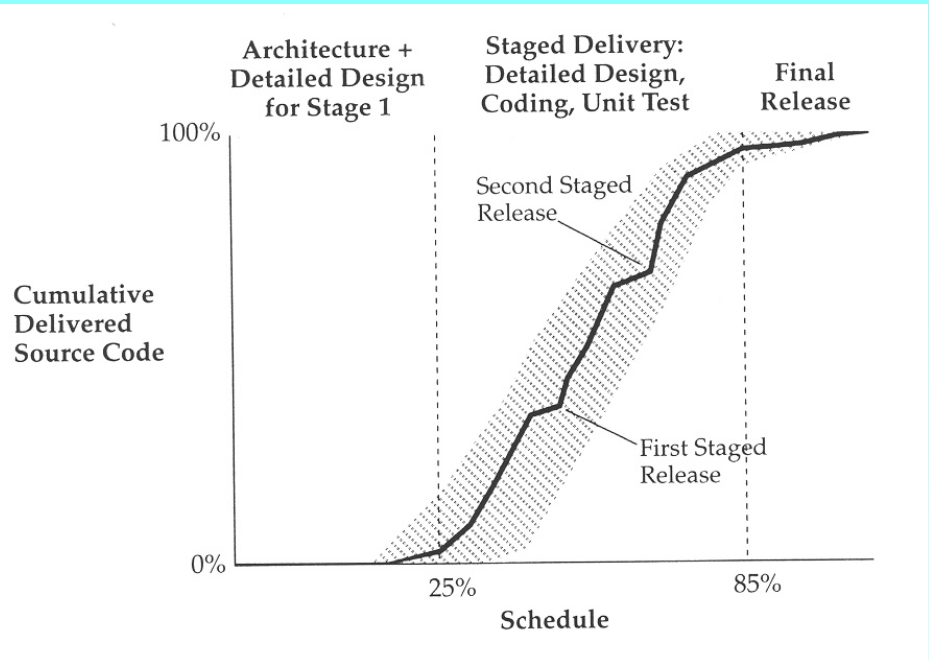
Overall AIPS++ objectives

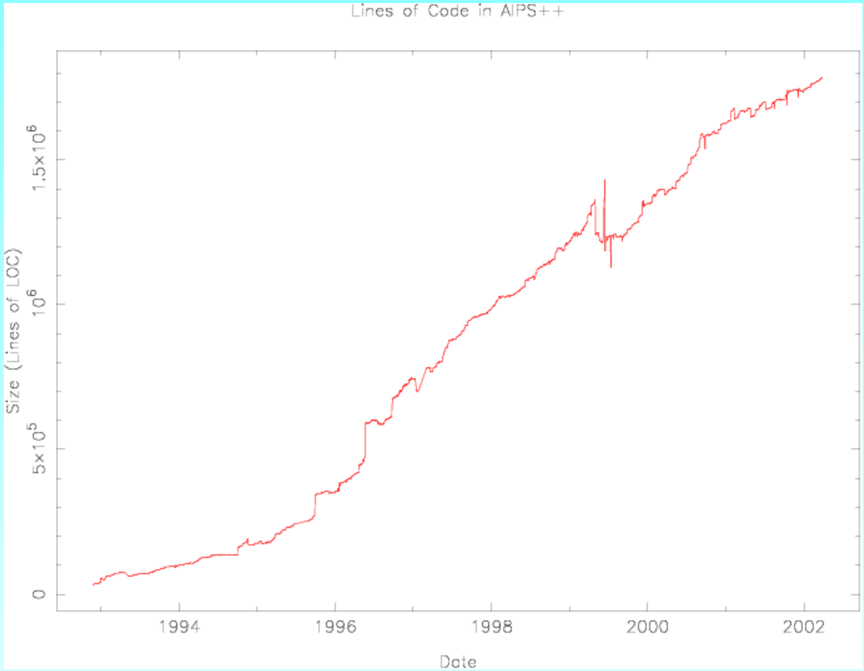
- As set in original scientific requirements specification:
  - Develop the next-generation data reduction package for radio-astronomy, as a successor package to existing systems
  - Support instrument-independent reduction
  - Provide support for a new generation of observing techniques and telescopes (e.g. ALMA, EVLA)
  - Provide a package that is extensible by astronomers, programmers and observatories alike
  - Provide support for near real-time and off-line reduction on an equal footing
  - Develop a toolkit targeted at radio astronomy but useable in a general astronomical context

Guiding design principles

- Scientific access at multiple levels
  - *High-level, integrated scientific applications*
  - *A toolkit of low-level capabilities*
- Scientific freedom for end-users
  - *Support custom reduction and data exploration through scripting using the toolkit*
  - *Strong scripting CLI binding the system together (Glish); C++ components bound in as distributed objects*
  - *Toolkit is applicable to general astronomy also*
- Instrument-independence
  - *Use of a generic data format, as well as calibration and imaging formalism*
- New approaches to data reduction
  - *Improved algorithms, support for new instruments and automated reduction, such as pipelines*

Current project phase





- Current AIPS++ size (1.8 MLOC, 1.2 MSLOC); effort ~110-140 FTE-years, cost ~$10m
- 1% of capital cost of target instruments
- Highly competitive code production: 13-15 KLOC per developer p.a.

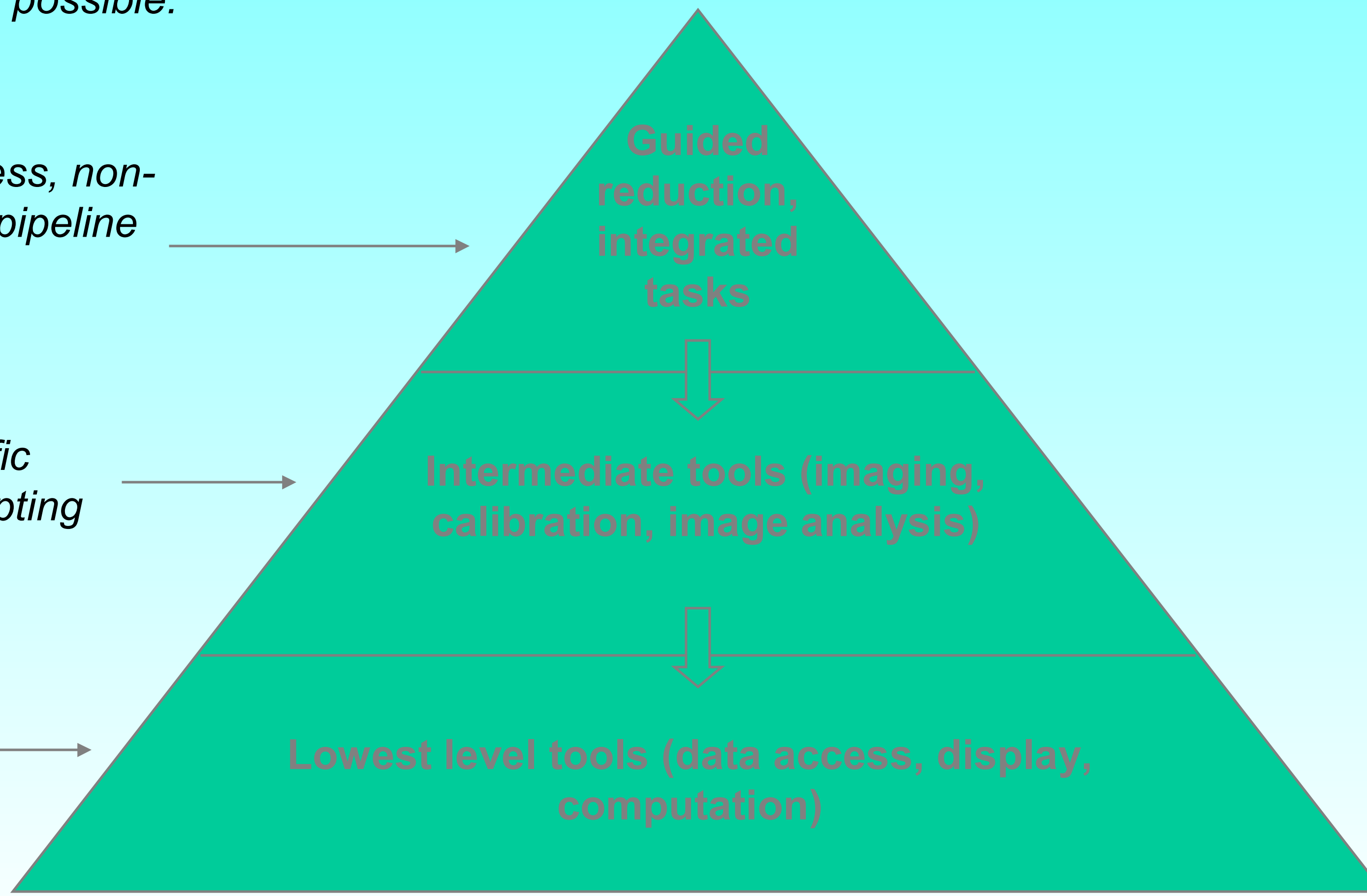
Applications architecture

Multi-level access possible:

Highest level access, non-specialist users, pipeline reduction

Custom scientific reduction and scripting

Basic astronomical scripting

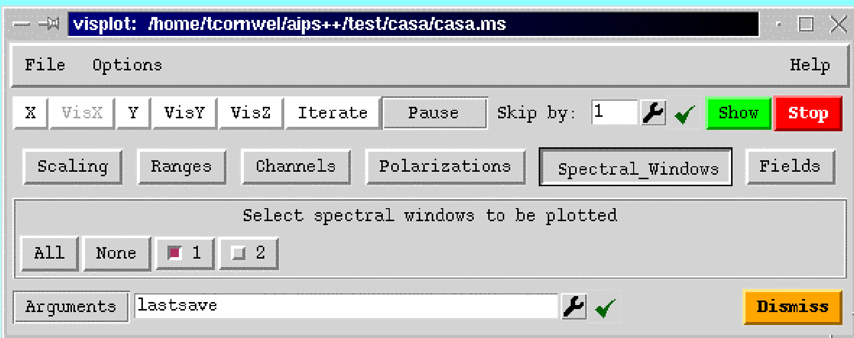


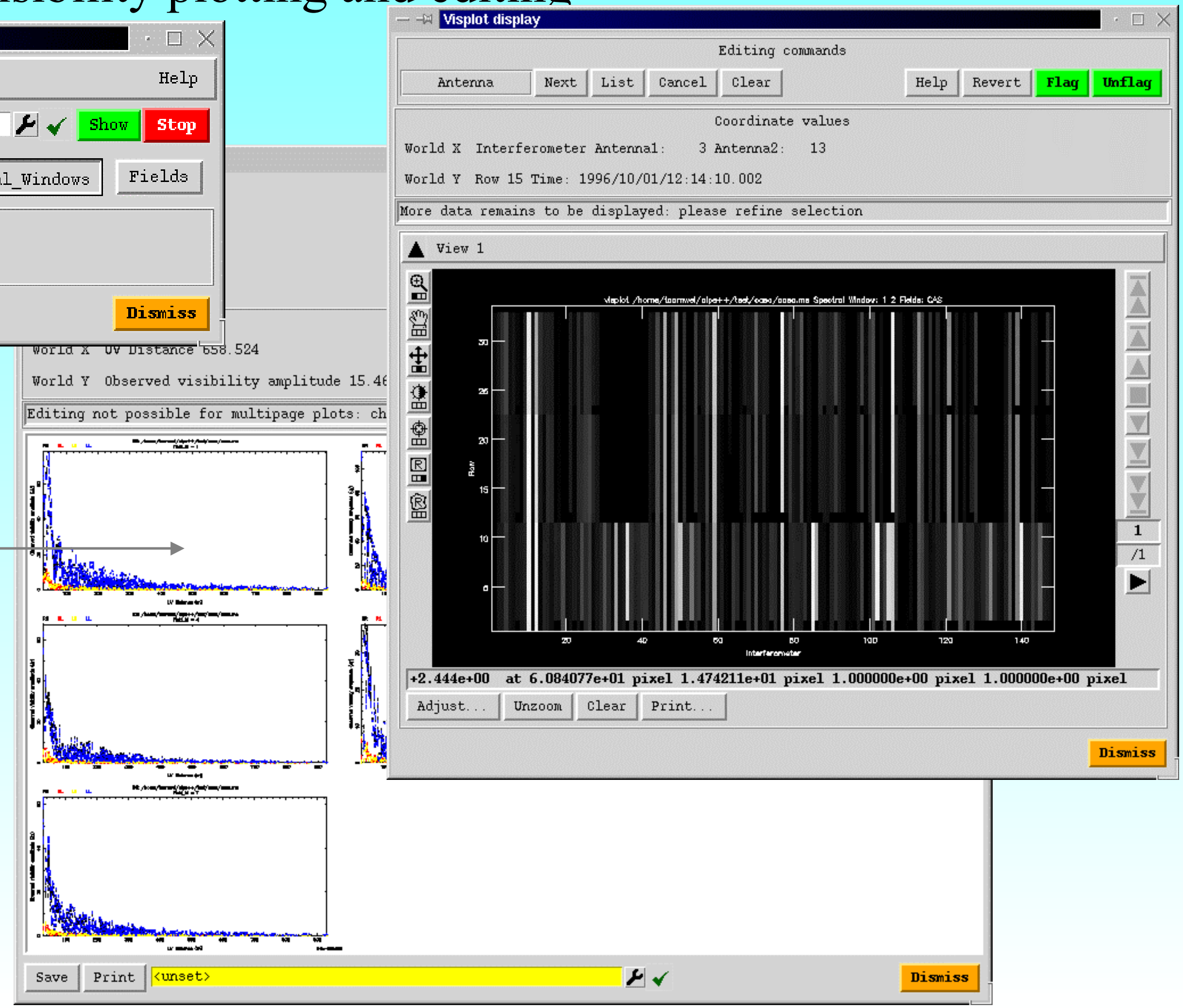
Synthesis completeness

- Data fillers
  - available for most consortium instruments and several archive and interchange data formats
- Editing and visualization
  - editing and visualization of visibility data
- Calibration
  - solvers for visibility-plane calibration effects; ability to apply image-plane effects.
- Imaging
  - a range of imaging contexts (mosaic, wide-field etc.) and deconvolution algorithms supported.
- Image analysis and visualization
  - capable image visualization and analysis tools

Data display and editing: msplot

- Interactive visibility plotting and editing

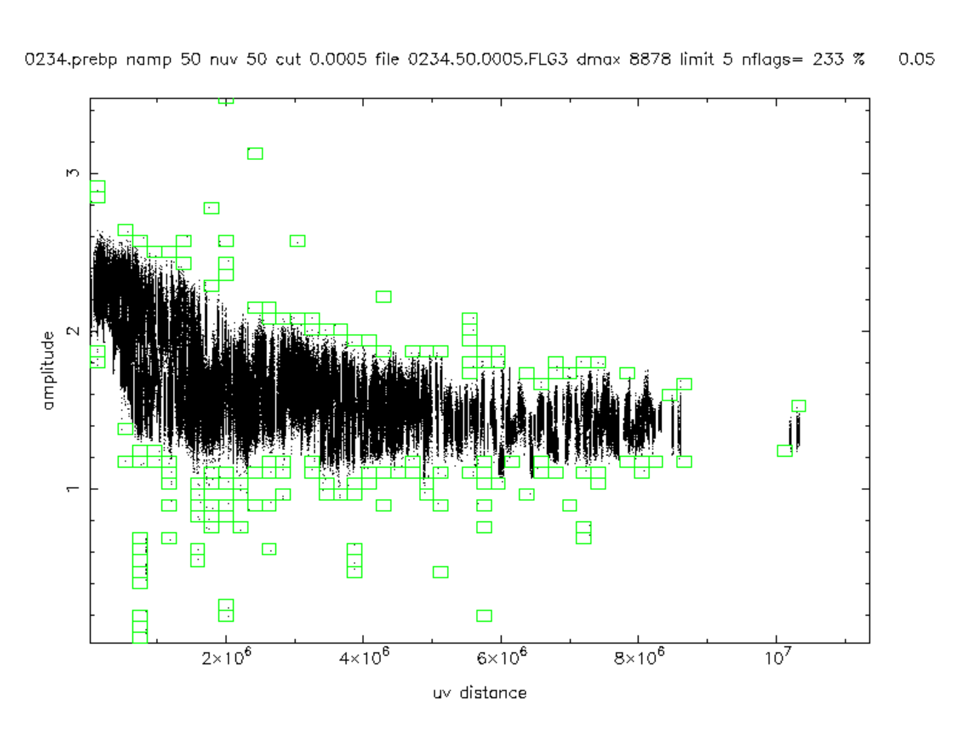




Many different types of plot

- *e.g.* Iterate over antennas for diagnosis of problems
- *e.g.* Iterate over fields for mosaic observations

Automated editing



Calibrator 0234+285, VLBA project BK31 (Kemball et al.)

- Vital for automated pipeline reduction
- Heuristics supported:
  - UV-plane binning (as left)
  - Median clip in time and frequency
  - Spectral rejection (spectral line baseline fitting)
  - Absolute clipping in a clip range

Calibration

- Visibility-plane components supported in *calibrater*
  - *P - parallactic angle correction (pre-computed).*
  - *C - polarization configuration (pre-computed).*
  - *G - electronic gain, solvable.*
  - *T - atmospheric correction, solvable.*
  - *D - instrumental polarization response, solvable.*
  - *B - bandpass response, solvable.*
  - *F - ionospheric correction, pre-computed from global, empirical model (PIM) (initial version).*
- Pre-computed, or solved using chi-squared computed from the Measurement Equation (ME).
- Pre-averaging, phase-only solutions, and reference antenna selection available in solver.

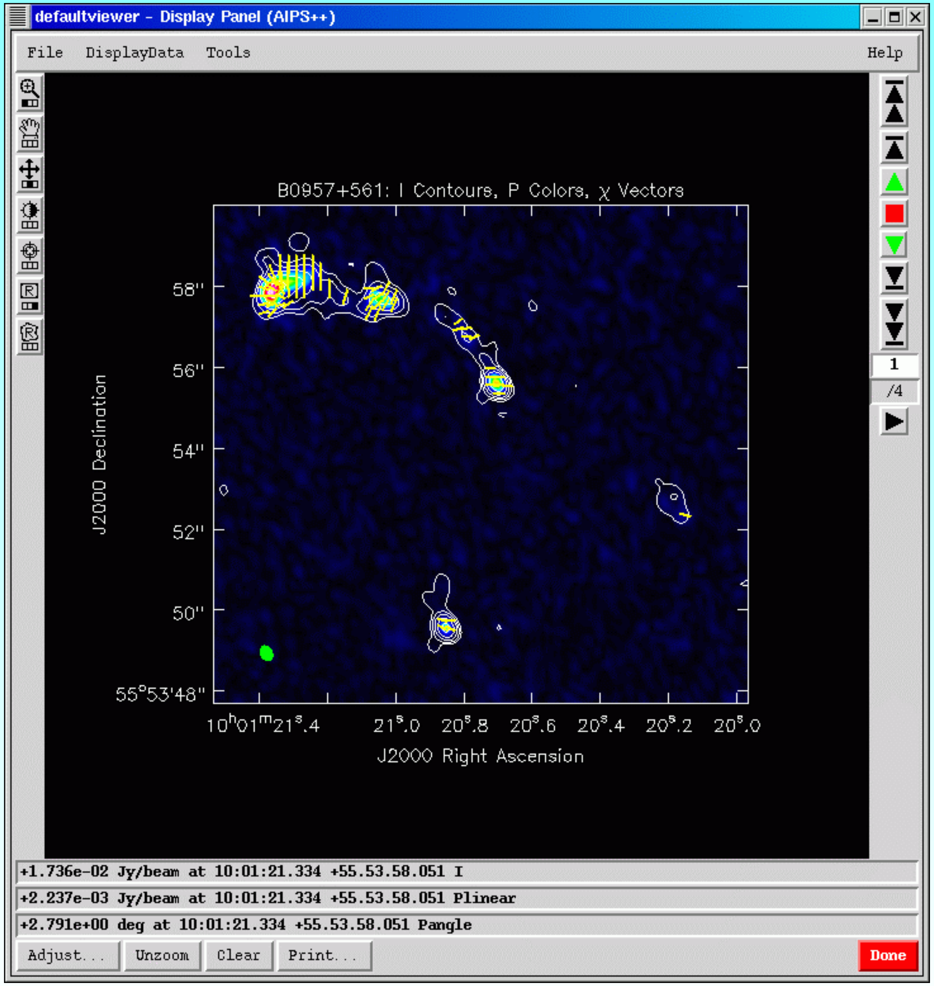
Data simulation

- Simulation of data from synthesis and single dish radio telescopes using the simulator tool.
- Complementary to imager and calibrater
- Two roles:
  - Provide synthetic data for testing AIPS++
  - Facility for telescope design
- Can construct MeasurementSet from description of an observation
- Can add noise and corruptions of various types
  - Based on HBS measurement equation

Imaging capabilities

- Imaging from synthesis and single dish data
  - Supports polarimetry, spectral-line, multiple fields, mosaicing, non-coplanar baselines (simultaneously)
  - Also single dish OTF, holography
  - Clean algorithms: Hogbom, Clark, Schwab-Cotton, Multi-scale
  - Incremental multi-field deconvolution
  - Non-Negative Least Squares and Maximum Entropy deconvolution
  - Supports imaging in a wide range of coordinate systems
  - Tracks moving objects
  - Discrete image component processing
  - Flexible in image size (2n not needed)
  - Novel "sort-less" visibility gridding algorithm
  - Advises on argument settings
  - User can "plug-in" customized (Glish) modules
  - Pixon deconvolution available in the image plane

Connected-element scientific completeness

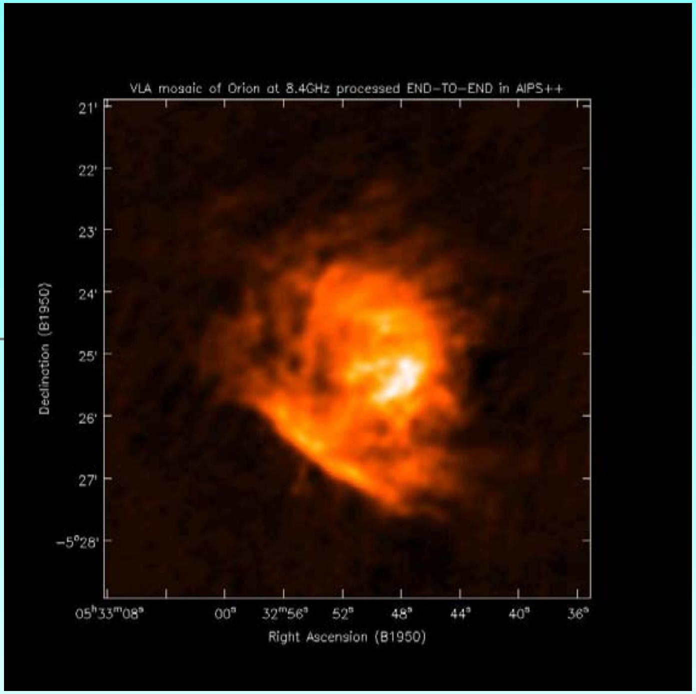


- Scientific completeness:
  - Thread completeness used in user documentation and recipes page
  - Split out usability issues explicitly

Mosaicing in AIPS++

- Mosaicing is standard processing
  - Data format supports mosaicing
  - Standard tool imager performs mosaicing
- Many primary beam models
  - e.g. can correct beam squint
- Mosaicwizard aids in processing
- But also easy to accomplish in glish

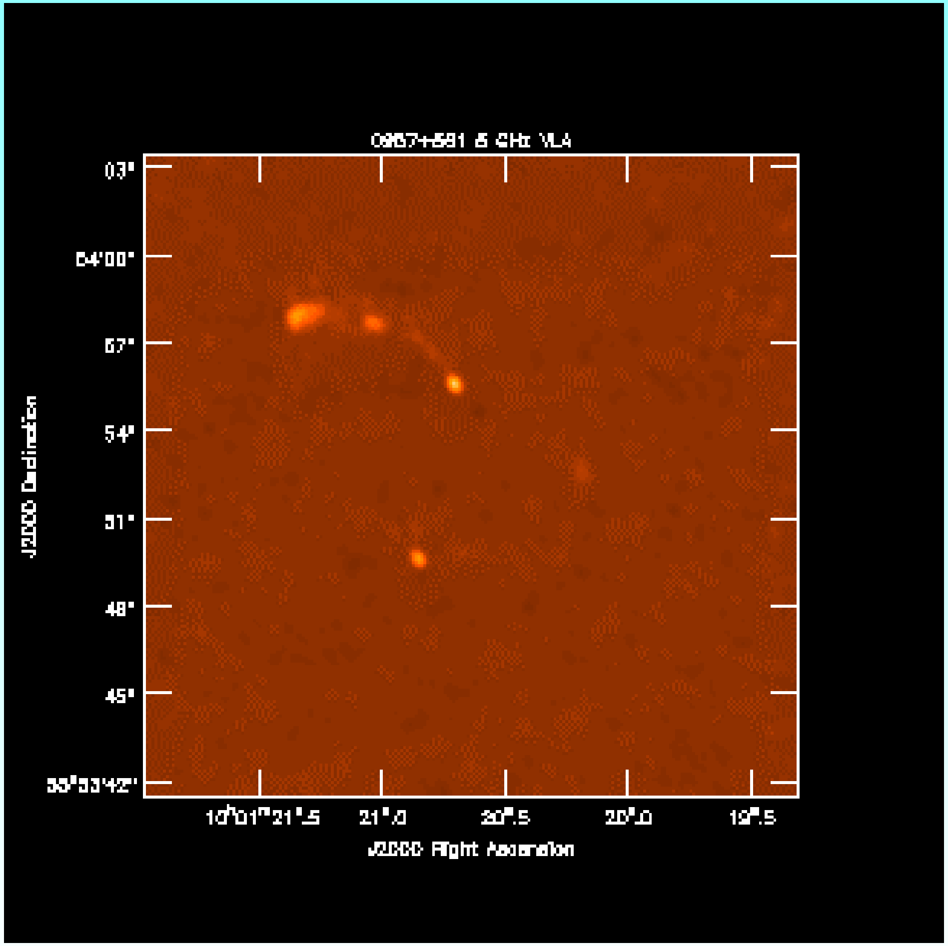
```
# make and deconvolve mosaic image
myimager:=imager(‘orion.ms’);
ok:=myimager.setimage(nx=300, ny=300, cellx=‘4.0arcsec’,
    celly=‘4.0arcsec’,
stokes=‘I’, spwid=[1, 2]);
ok:=myimager.setdata(spwid=[1, 2] , fieldid=3:11 , msselect='');
ok:=myimager.weight(type="briggs" , robust=-1);
ok:=myimager.setvp(dovp=T, dosquint=F);
ok:=myimager.mem(algorithm="mfentropy", niter=100, sigma=‘4mJy’,
    displayprogress=T, model="orion.mem");
myimager.done();
```



10 pointing VLA 8 GHz mosaic of Orion processed entirely in AIPS++

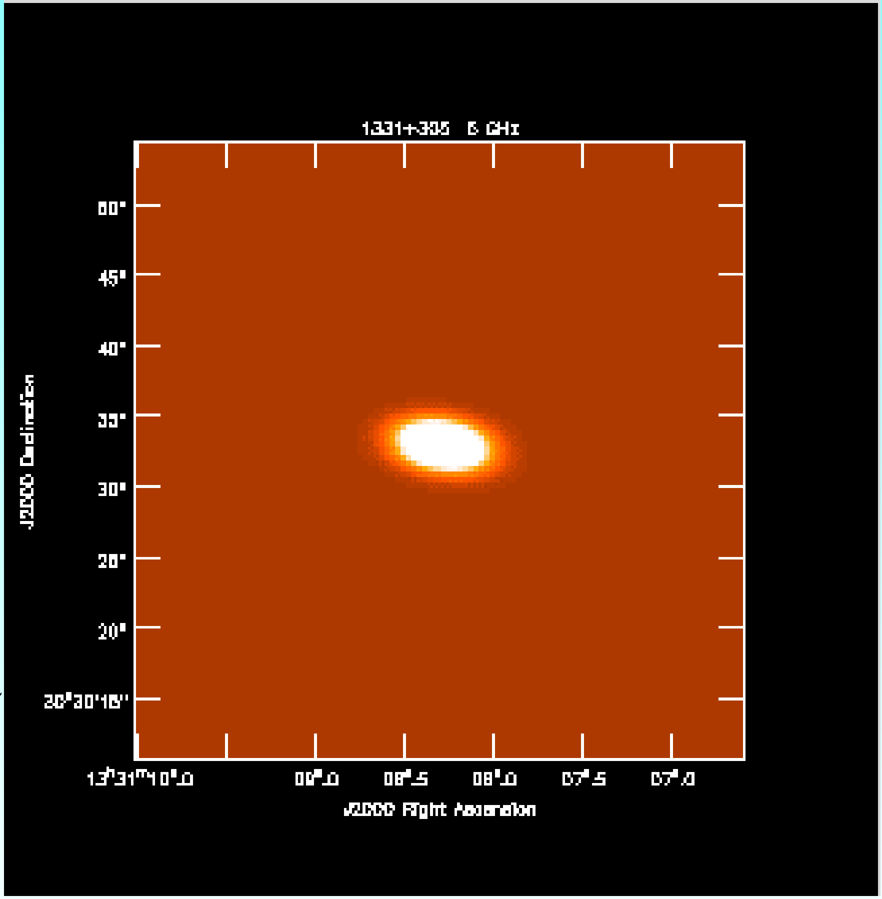
Continuum calibration and single-field imaging

- Project AP366:
  - Patnaik, Kemball et. al.
- 24-hour VLA observation in A-configuration of a sample of gravitational lenses
- Continuum imaging of 0957+561 at 5 GHz shown here
- Phase calibrator 0917+624.
- Amplitude calibrator 1331+305



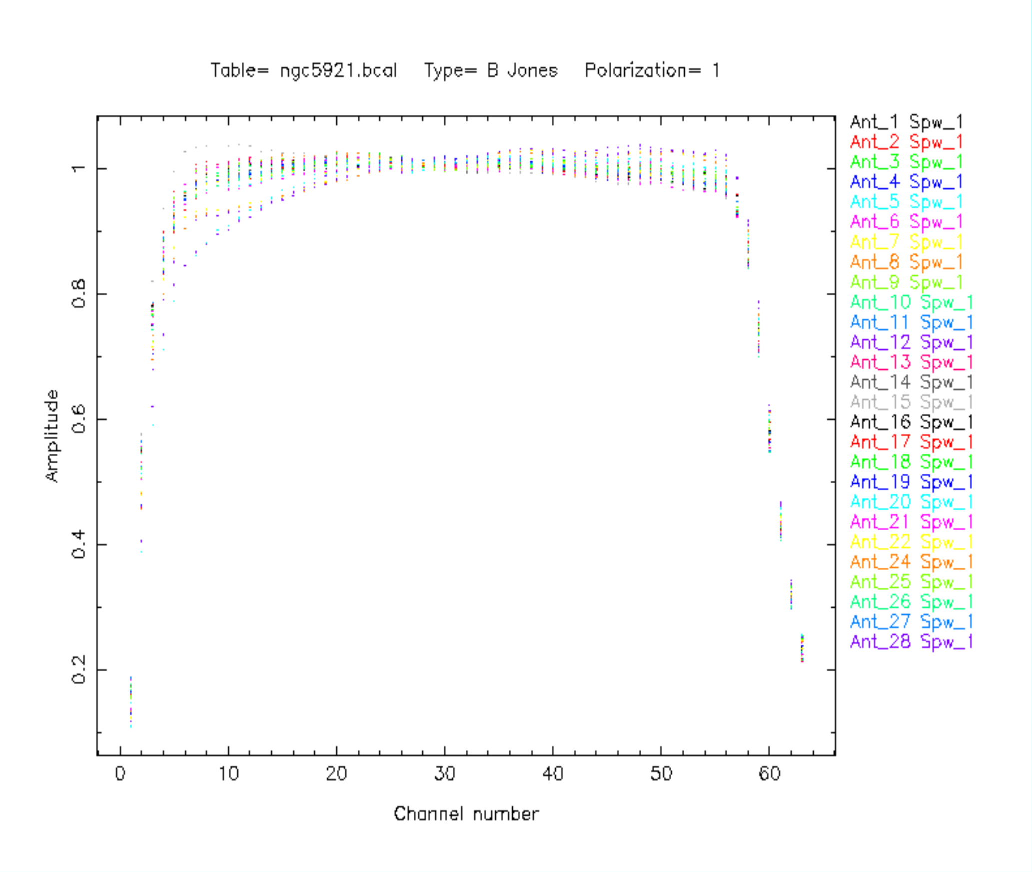
Continuum polarimetry

- Continuum polarimetry:
  - Solver for instrumental polarization response (D-terms)
  - Full second-order model for instrumental polarization.
  - D-terms can be time-variable
  - Supports (R,L) and (X,Y) data
  - Allows polarization self-calibration



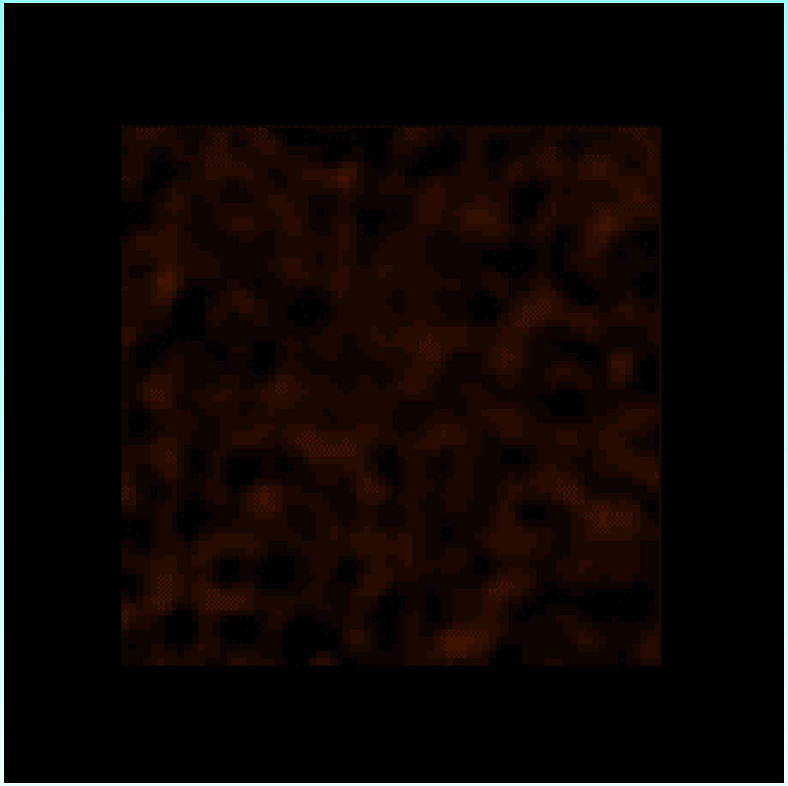
Spectral line calibration and imaging

- Spectral line reduction
  - Designated test dataset: HI observations of NGC 5921 in VLA D-configuration
  - Bandpass response solutions plotted



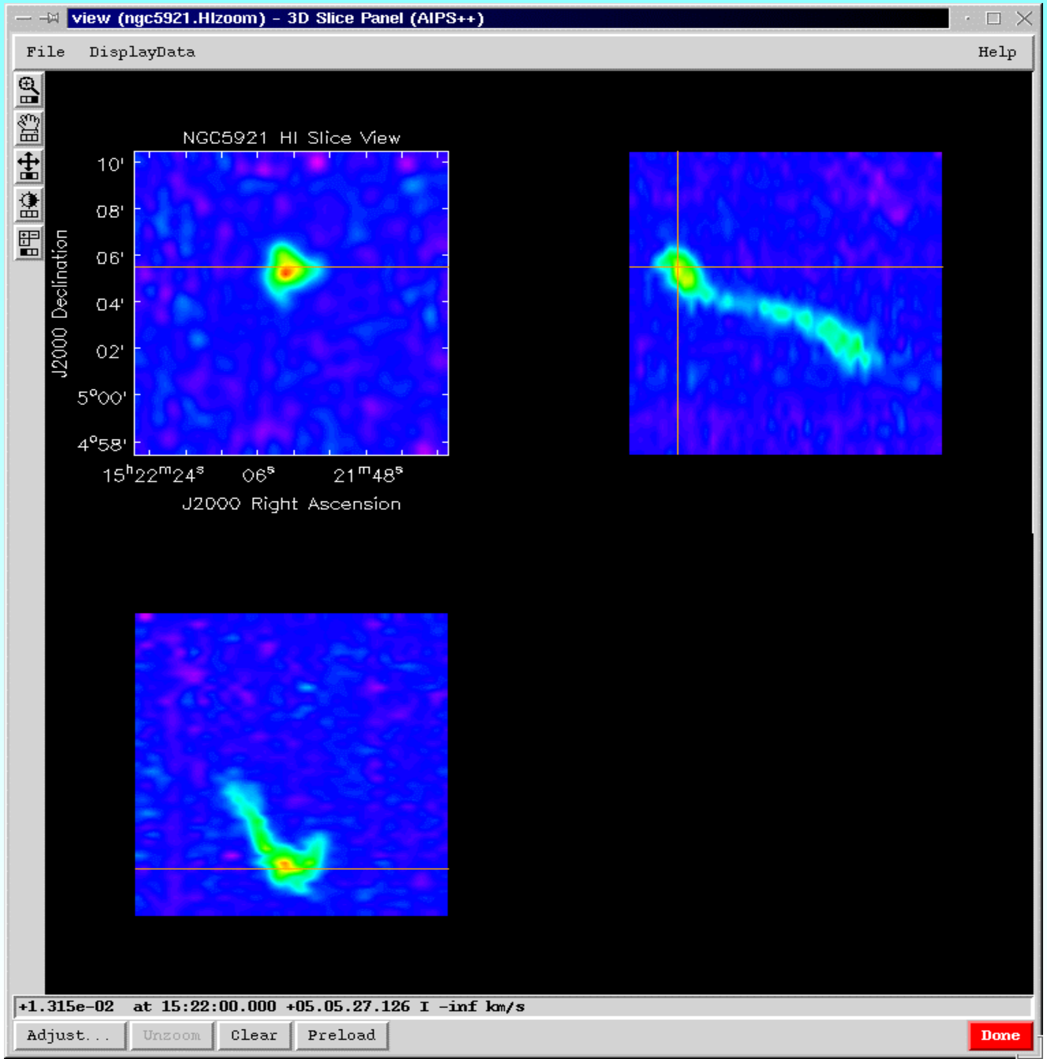
Spectral line calibration and imaging

- Spectral line reduction
  - Designated test dataset: HI observations of NGC 5921 in D-configuration
  - Calibrated and imaged, with map-plane continuum subtraction

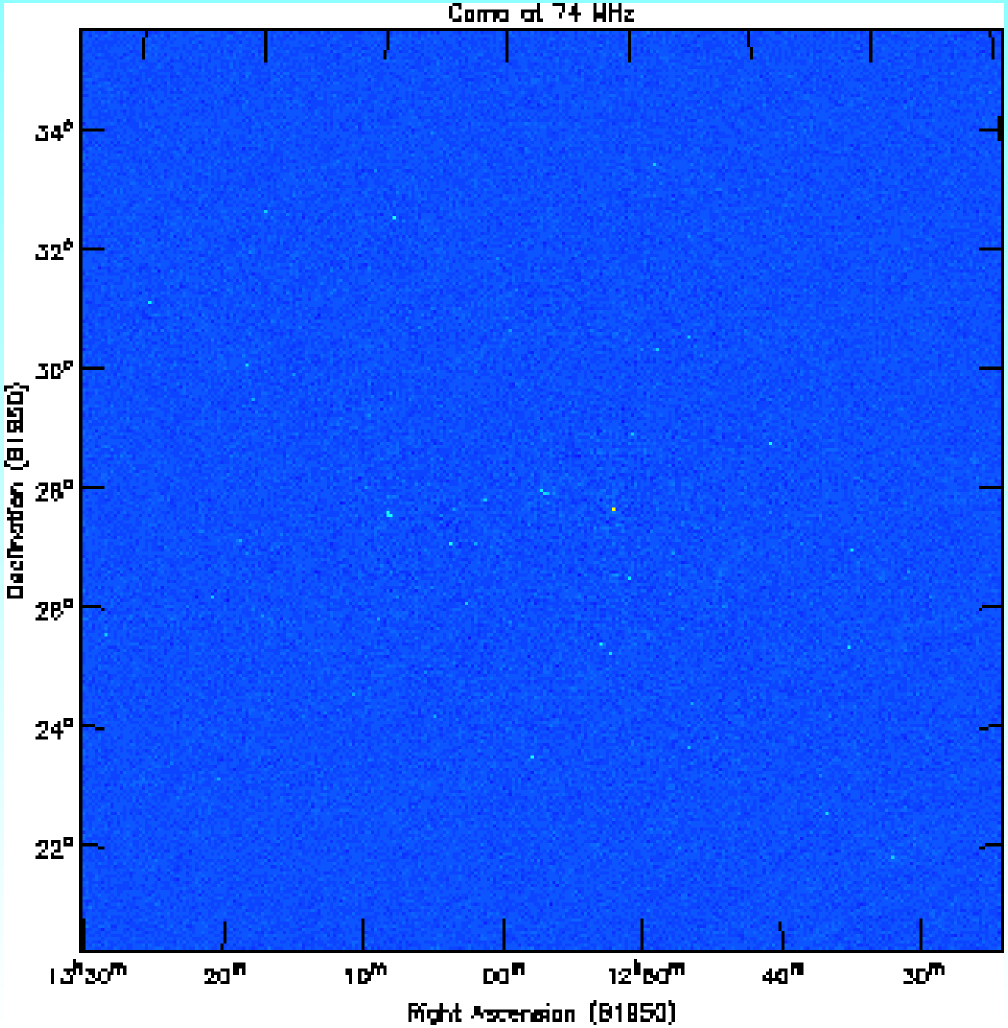


Spectral line calibration and imaging

- NGC 5921, HI VLA (designated test dataset)
  - Dec vs RA
  - Dec vs Frequency
  - Frequency vs RA



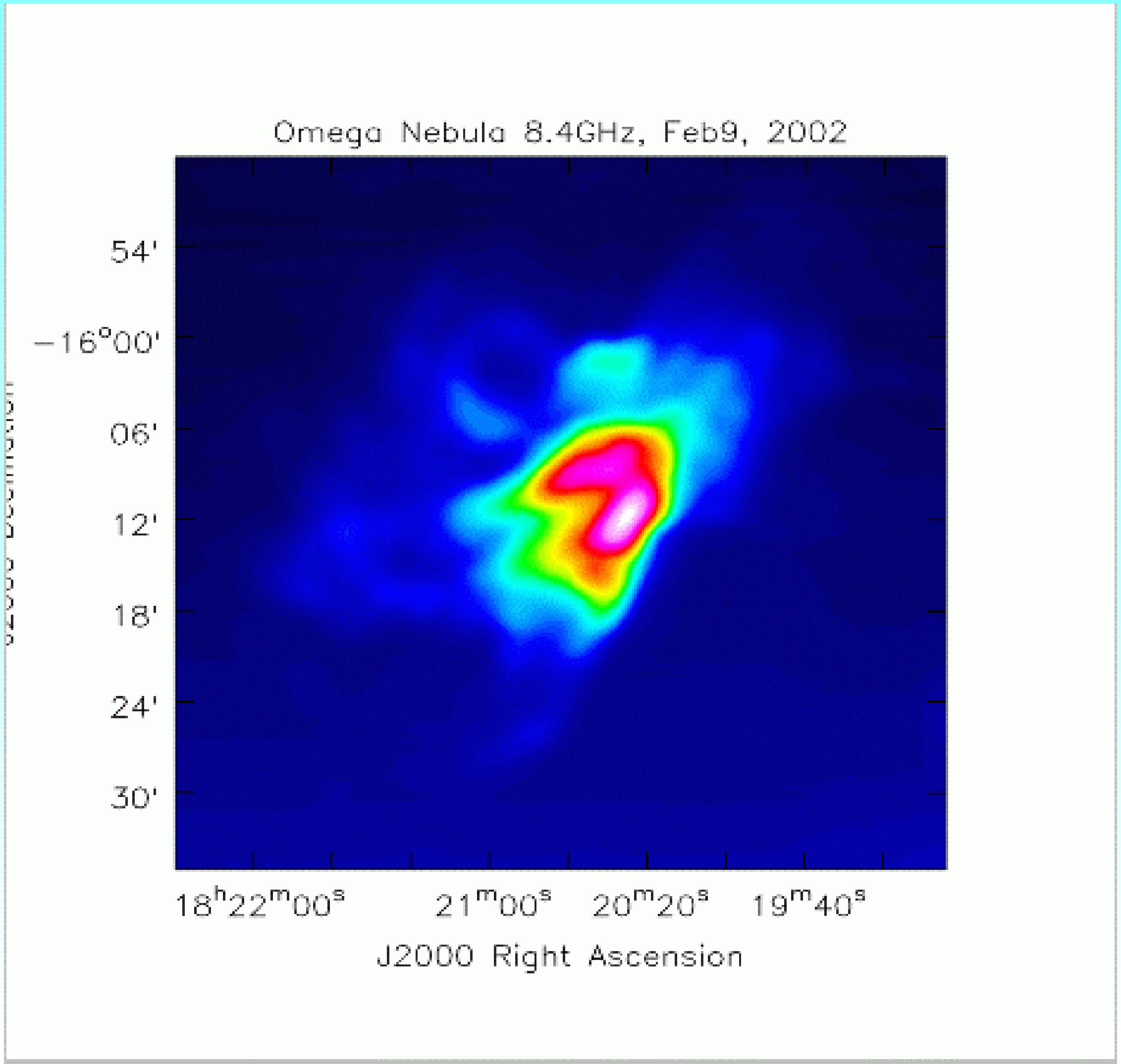
Wide-field imaging



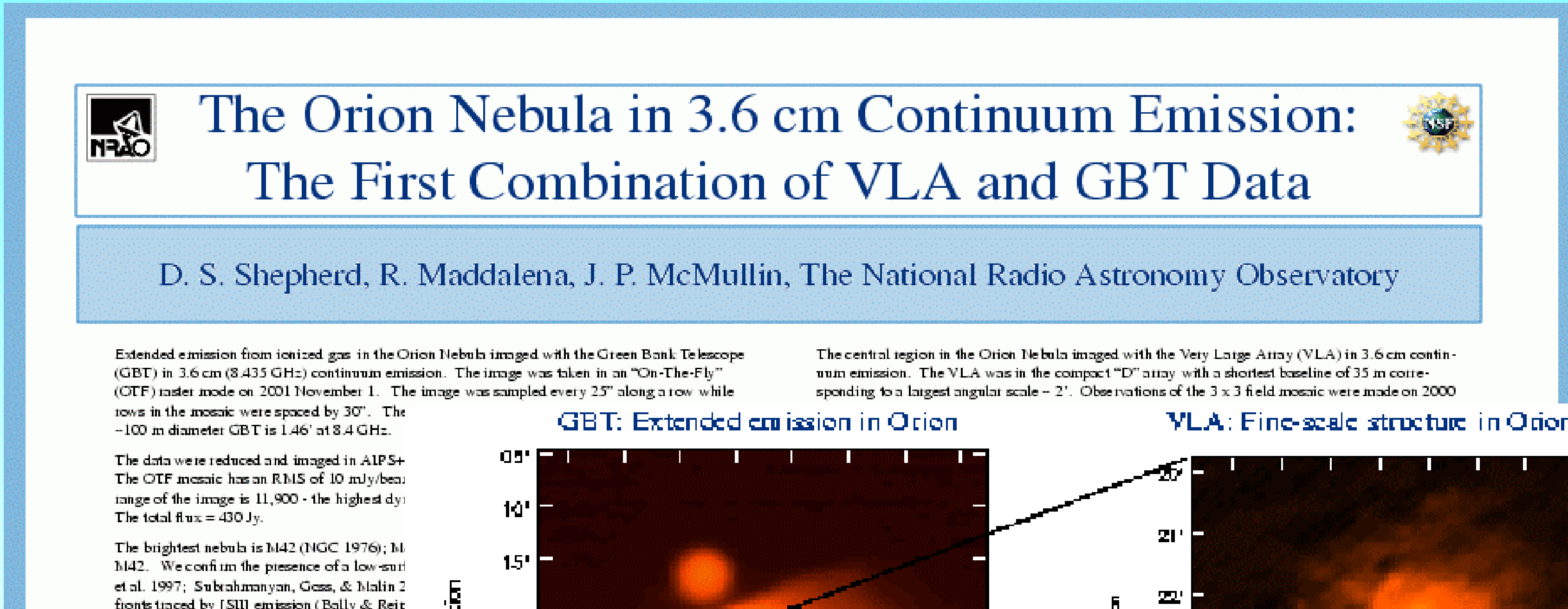
- Wide-field VLA Imaging
  - 90cm and 4m imaging
  - Performs cleaning and self-calibration simultaneously
  - Images multiple facets on a single tangent plane
- Written entirely in Glish using first tier tools
  - Auto-gui for user interface
  - imager, calibrater, viewer
- Parallelized

Single-dish imaging

- GUI, CLI, Plotter
- Analysis tools
- Data editing (flagging)
- Imaging
- Full data access
- Telescope specific calibration



Synthesis and single-dish imaging



The Orion Nebula in 3.6 cm Continuum Emission:
The First Combination of VLA and GBT Data

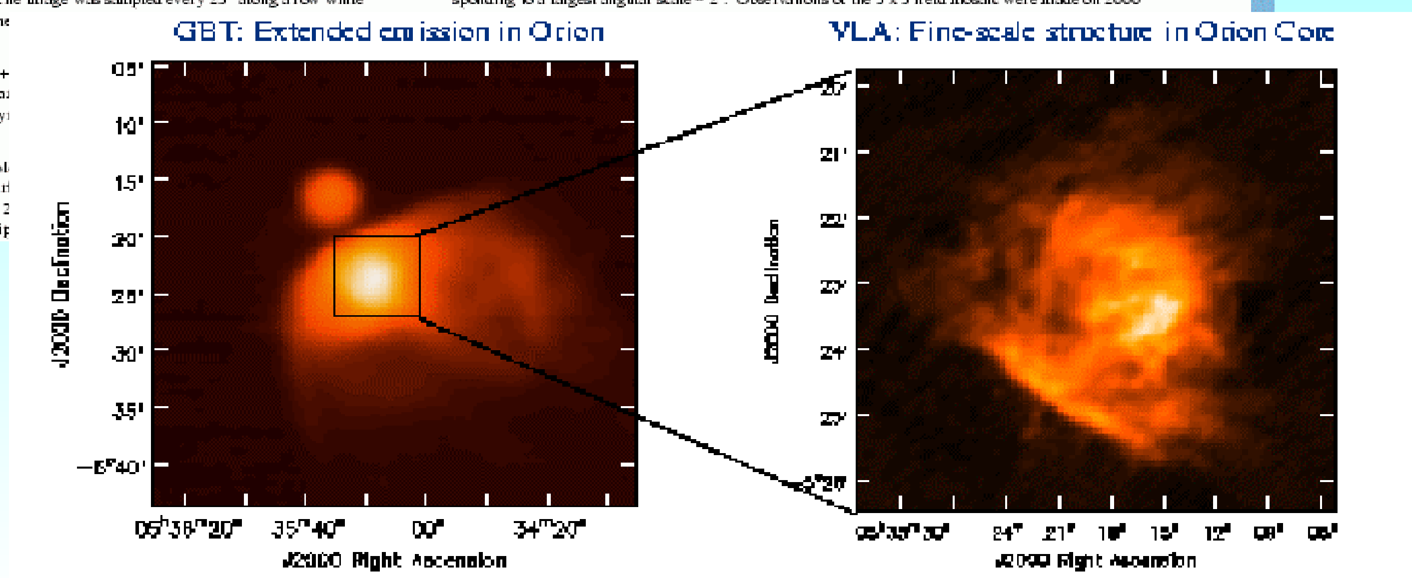
D. S. Shepherd, R. Maddalena, J. P. McMullin, The National Radio Astronomy Observatory

Extended emission from ionized gas in the Orion Nebula imaged with the Green Bank Telescope (GBT) in 3.6 cm (8.435 GHz) continuum emission. The image was taken in an "On-The-Fly" (OTF) raster mode on 2001 November 1. The image was sampled every 25" along a row while rows in the mosaic were spaced by 30". The ~100 m diameter GBT is 1.46' at 8.4 GHz.

The data were reduced and imaged in AIPS+ The OTF mosaic has an RMS of 10 mJy/bea range of the image is 11,900 - the highest dy The total flux = 430 Jy.

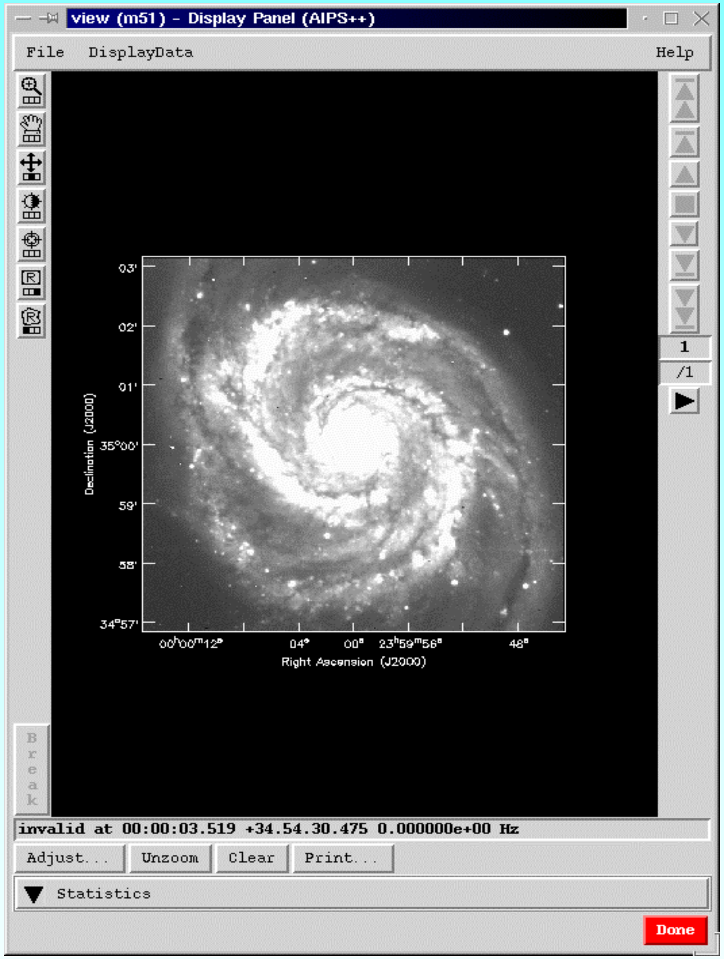
The brightest nebula is M42 (NGC 1976); M M42. We confirm the presence of a low-surf et al. 1997; Subrahmanyan, Goss, & Malin 2 fronts traced by [SII] emission (Bally & Reip

The central region in the Orion Nebula imaged with the Very Large Array (VLA) in 3.6 cm continuum emission. The VLA was in the compact "D" array with a shortest baseline of 35 m corresponding to a largest angular scale ~ 2'. Observations of the 3 x 3 field mosaic were made on 2000



Display Library

- Support for display of all AIPS++ data
  - Images, MeasurementSets, Tables, Catalogs etc.
- Programmable via Glish or C++



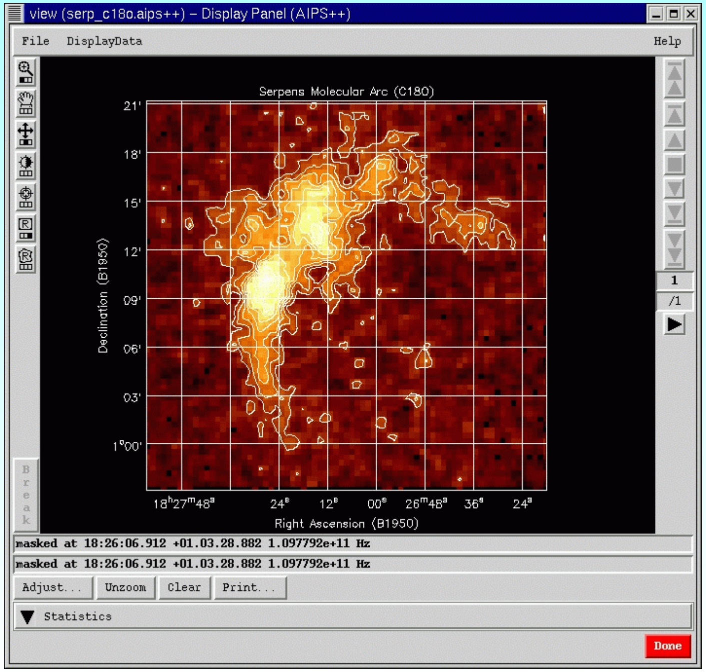
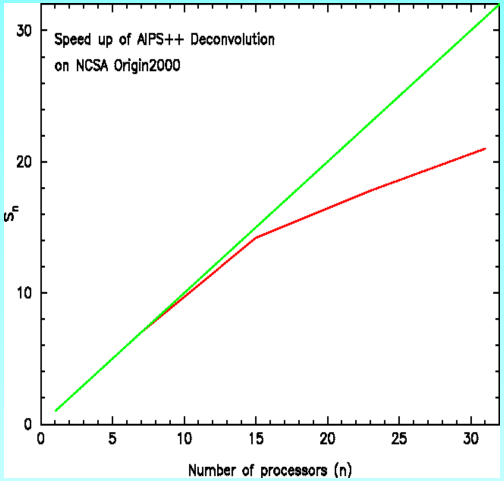


Image analysis

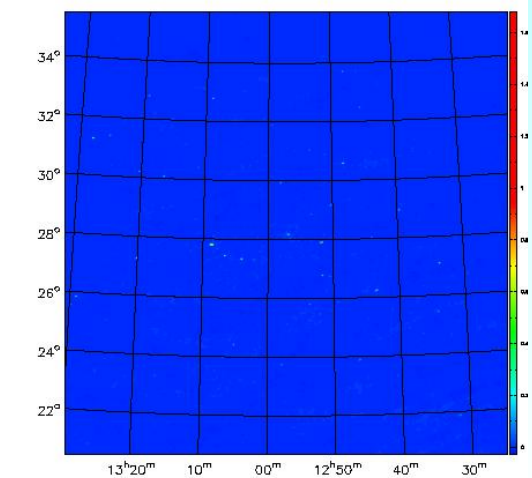
- Image operations
  - Very flexible use of “on-the-fly” images
- Convert to and from FITS
- Display, statistics, histograms, smoothing
- Numerous moment calculation methods
- Image polarization calculations
- Transfer image pixels to and from Glish
- Image calculator
  - Apply math operations to images, using a syntax based on the AIPS++ Lattice Expression Language
- Integrated with regions
  - Definition, conversion, math of regions in both pixel and world coordinates
- Coordinate system editing
- Fast, optimized image regridding
- Specialized mage polarimetry operations
- Component fitting

High-performance data reduction



<u>Advanced performance issues:</u>

- AIPS++ is a partner in the externally-funded NCSA Alliance Program (NSF)
- Many collaborations with high-performance computing groups and expertise:
- Have implemented a high-performance infrastructure in AIPS++ for parallelization; e.g. parallel wide-field imaging
- An important investment in expertise which will be important for EVLA



VLA observations of the Coma cluster (Perley et al); 225 imaging facets, 32 processors, speed-up factor ~20

Documentation

- Adopted standard commercial model
  - *Getting Results*
    - intent is to act as a cookbook, divided into chapters addressing specific scientific topics (e.g. data editing, calibration, imaging etc.).
    - consulted by the scientific end-user.
    - linked to Glish *recipes* and the *User Reference Manual*.
    - separate chapters for end-to-end processing per instrument also (e.g. VLA).
    - have recently assigned an editor for *GR* within AIPS++
  - *User Reference Manual*
    - authoritative guide for each synthesis tool and function.
    - consulted for the scientific end-user, particularly when developing scripts.

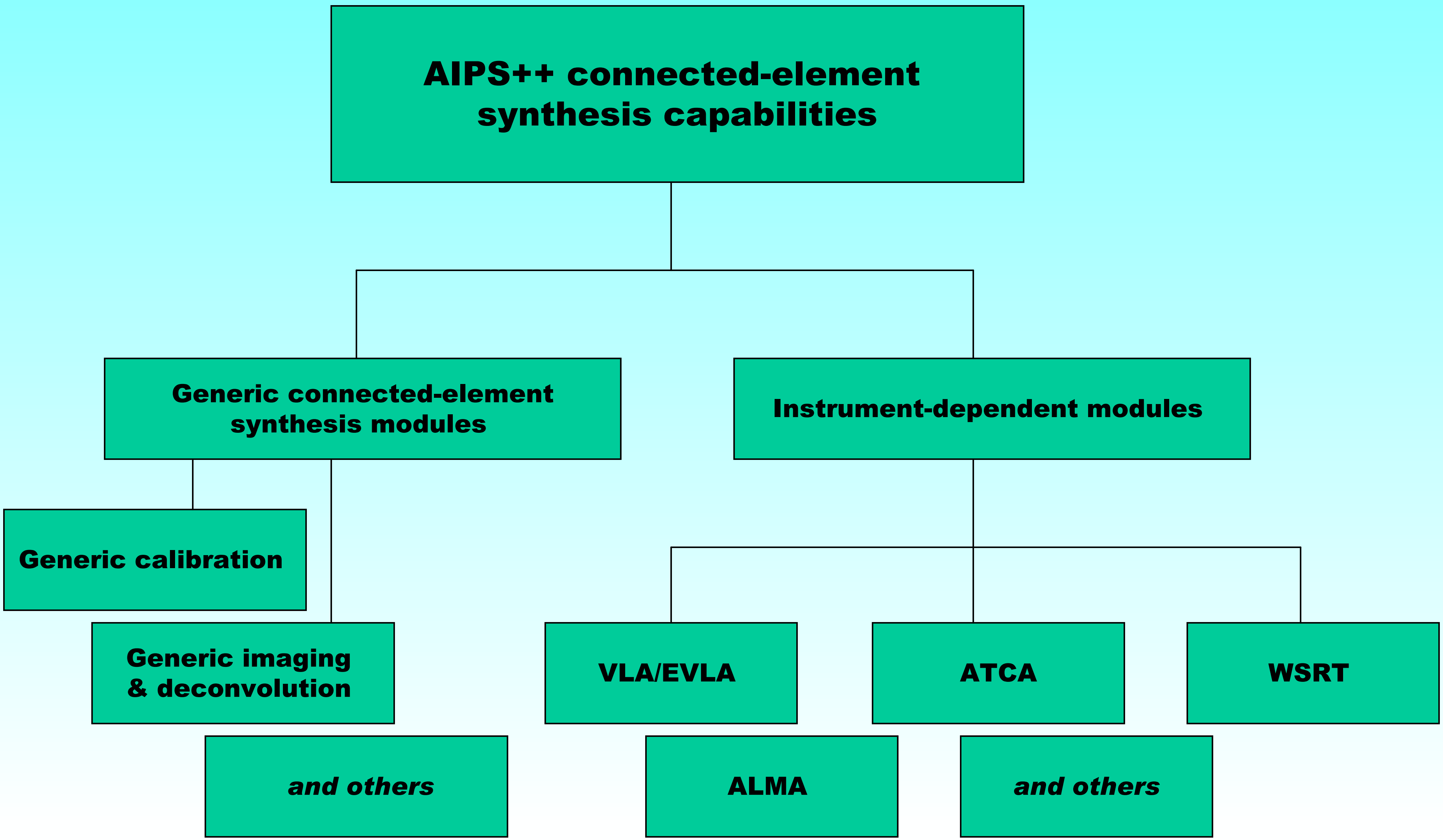
Use of AIPS++

Where?	Status	Comments
NRAO		
VLA	In use	Being taught at NRAO Synthesis Imaging school June 2002
GBT	In use	Only supported data reduction package
VLBA	Not in use	Staged deployment after users trained in VLA
ALMA	In development	Part of baseline plan, pipelines and post processing in AIPS++
EVLA	In development	Planned to provide all EVLA processing in AIPS++
e2e	In use	Foundation of e2e capabilities
US		
Arecibo	In use	Used partially within the Observatory
BIMA	In use	BIMA package for pipeline and post processing, not widely used yet
CARMA	Planned	Real-time system will write AIPS++ MeasurementSets
SMA	Under evaluation	Tested in commissioning and early data reduction (decision pending)
LOFAR	In use	Used for LOFAR simulation
NPOI	In use	Commissioning, data reduction
Consortium partners		
ATNF Parkes telescope	In use	Parkes Multibeam Project (since 1997)
ATCA	In use	Pipeline processing
ATNF RFI project	In use	RFI identification
WSRT	In use	Front-end processing only, pipeline in development
Non-US		
IRAM	Under evaluation	For ALMA-related test of applicability to PdB interferometer
JIVE	In use	For correlator backend (since 1998)
JCMT	In use	6 months testing
JCMT/ACSIS	In use	Parallel processing backend developed in AIPS++, awaiting deployment
Obs. De Paris	In use	ALMA simulator writes to AIPS++ MeasurementSets for processing
FARADAY	In development	Multi-beam single dish package
Amateur		
Patrick Wood	In use	LOBERT Observatory, Hinesville, GA

EVLA post-processing in AIPS++

- AIPS++ is instrument-independent by design and supports a range of connected-element interferometers at present
- Instrument-independence is achieved by:
  - A generic calibration and imaging framework (Hamaker, Bregman and Sault (1996) series)
  - A generic data format, compatible with the calibration and imaging framework
  - Support for instrument-dependent extensions where necessary

AIPS++ telescope support architecture



AIPS++ organization

AIPS++ Executive Committee
(Observatory Directors and/or Senior Management)
Chair: NRAO Director

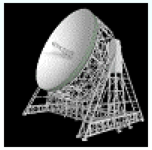
Observatory advisory committees

AIPS++ Project Management
Athol Kemball (NRAO)
Joe McMullin (NRAO)

Consortium user group (AUG)
(Scientific user representatives)
Chair: Tom Pauls (NRL)

Local site managers
(ASTRON: Jan Noordam; ATNF: Neil Killeen; BIMA/NCSA: Ray Plante; NRAL: Roger Noble; NRAO: project management)

Local site user groups
(ATNF, BIMA/NCSA, NRAO:NAUG)
Local scientific users



Planning for EVLA development within AIPS++

- EVLA post-processing requirements specification by science team
- Initial planning and resource allocation
- Iterative process in each 6-month AIPS++ cycle:
  - EVLA post-processing requirements refinement by science team
  - Planning and resource allocation
  - Design and implementation
  - User and system software testing

EVLA post-processing science requirements

- Have not yet been formally specified or prioritized
  - An important project science team activity
  - Best expressed as prioritized post-processing software requirements, per EVLA observing mode
- However, the requirements are reasonably well-understood, given existing post-processing experience with the current VLA
- Several informal documents exist
- User-specified and science-driven requirements are important however to ensure delivery of capabilities in the correct priority order and adequate resource allocation

EVLA post-processing requirements

- Overall science requirements need to be translated to incremental AIPS++ requirements in two broad areas:
  - Expansion of generic synthesis reduction support to meet EVLA science needs
  - EVLA-specific application or module development
- Available resources are approximately10 FTE from within AIPS++ in the period 2002-2009.
- An important activity will need to be algorithm evaluation and research (science team and AIPS++)

EVLA post-processing requirements

- EVLA post-processing requirements can be anticipated in the following key areas:
  - **Large data volumes:**
    - To disk: initial rate B_{max} = 25 MB/s (over 12 h = 1 TB)
    - Maximum aggregate correlator rate is ~1.8 GB/s
      - Burst-mode to memory for short intervals could be possible in principle at higher fractions of the maximum aggregate correlator rate
    - If iteration over 1 TB must take place within 0.1-1.0% of elapsed observe time, the required I/O bandwidths to the post-processing applications are 2-25 GB/s
    - Spectra with 16,384 to 262,144 frequency channels:
      - 3 sec averaging and 8192 channels is 80 MB/s (exceeds B_{max})
      - 1 sec averaging and 16,384 channels is 240 MB/s (exceeds B_{max})

EVLA post-processing requirements

- **RFI mitigation:**
  - RFI will be a major factor in defining the achievable dynamic range. Mitigation will need to occur in post-processing also despite earlier layered defenses:
    - ***Automated editing***
      - » Data volumes preclude routine interactive data editing
      - » Will require generalized automated editing for efficient and robust outlier detection
    - ***RFI identification***
      - » Can identify against catalog of known interfering sources
      - » Can identify by signature (e.g. movement through field; closure phase signature)
      - » Can at times identify using general automated editing
      - » Very short time integration intervals will be required to optimize this
    - ***RFI removal***
      - » Can remove if RFI signature is identified
      - » Blanket flagging in other cases

EVLA post-processing requirements

– **Routine high-fidelity imaging:**
  - More accurate treatment of calibration effects in the visibility- and image-plane (Perley et al):
    - – ***Image-plane effects: pointing***
      - » Scan locus will move smoothly within 3 arcsec but may have unknown offset per antenna of up to 10 arcsec in absolute sense
      - » Q-band primary beam is 60 arcsec
      - » Interferometric solvers for remaining pointing offsets after reference pointing using strong point sources in the field
    - – ***Image-plane effects: primary beam shape***
      - » Primary beam shapes vary by antenna, elevation and frequency across band
      - » Will need to apply a-priori primary beams corrections known from strong source observations
      - » May need secondary solvers to refine these models from compact sources embedded in the target field
      - » Solve and apply for time- & freq-dependent primary beam shapes

EVLA post-processing requirements

- Routine high-fidelity imaging:
  - More accurate treatment of calibration effects in the visibility- and image-plane:
    - ***Image-plane effects: primary beam phase***
      - » Phase across the antenna beam is not constant
      - » Primary effect is a slope caused by an offset in the feed direction plus other secondary effects.
      - » May be time- and elevation-dependent
      - » May need inferometric solvers for beam phase using embedded compact sources to refine a-priori beam phase models
    - ***Image-plane effects: primary beam polarization***
      - » Overall polarization D-term fidelity will likely need to be ~0.01% to achieve 1 microJy thermal limit in Stokes I (scaling argument Rick Perley)
      - » Will need to solve for spatial D-term dependence both from calibration scans and refinements from compact sources embedded in the target field

EVLA post-processing requirements

- **Routine high-fidelity imaging:**
  - More accurate treatment of calibration effects in the visibility- and image-plane:
    - ***Visibility-plane effects: bandpass calibration***
      - » Will need high-fidelity bandpass calibration for weak Zeeman measurements (e.g. Zeeman splitting of HII recombination lines and weak absorption)
    - ***Visibility-plane effects: polarization calibration***
      - » High-fidelity D-term solve and apply
      - » Support functional dependence on frequency and time/elevation
      - » Full second-order polarization model
      - » More advanced ionosphere models
    - ***Imaging: non-coplanar baselines***
      - » Will be required for many bands at lower end of 1-50 GHz range
      - » Hundreds of fields will be needed for high-fidelity imaging
      - » Non-isoplanatic gain solvers will be required for reasonable fidelity
    - ***Imaging: high-fidelity deconvolution***
      - » At 10^6:1 dynamic range, high-fidelity deconvolution of extended emission will be necessary
      - » Will likely require adaptive multi-scale deconvolution (e.g. pixons)
      - » MFS over large frequency ranges

Scale of EVLA data processing

- Peak initial data rate out of correlator backend ~ 25 MB/s
- Total data volume for Peak 8-hr observation ~ 700GB
- Floating point operations per float ~ 100 - 10000
- Peak compute rate ~ 5Tflop
- Average/Peak computing load ~ 0.1
- Average compute rate ~ 0.5Tflop
- Turnaround for 8-hr peak observation ~ 40 minutes
- Average/Peak data volume ~ 0.1
- Data for Average 8-hr observation ~ 70GB
- Data for Average 1-yr ~ 80TB

Scale of processing

- Assume Moore's Law holds to 2009
  - Moore himself believes this……..
- Scale:
  - Desktop (2009) will handle many projects
  - Larger projects require ~ 10 – 20 processor parallel machine at NRAO
    - ~ $100K - $200K (2009)
  - Archive ~ 50TB per year
    - ~ $50K - $100K (2009)
- Comparable to scale of processing for ALMA

EVLA post-processing challenge problems

Fast-slew mosaicing	~10ms data sampling rate. Remove sliding primary beam.
Full bandwidth synthesis	Deconvolve wide bandwidths while accounting for spectral index, polarization, rotation measures, opacity, *etc.*
Full-beam high-fidelity polarization imaging	Correction of time- and angle-dependent beam polarization.
High fidelity imaging	Image and deconvolve at ~ 10^7. Currently about ~ 100 away from this in best possible cases.
Wide-angle full-beam imaging	Huge images, fast data sampling rates, many imaging facets to accommodate non-coplanar baselines
RFI mitigation	Removal of RFI post-correlation – requires high data rates

EVLA post-processing trial WBS

- **Visibility data handling (1 FTE):**
  - MS concatenation improvements (P1, 0.3 FTE)
  - Sub-band alignment and calibration (P2, 0.3 FTE)
  - Parallelized I/O visibility data access (P2, 0.3 FTE)
- **Data editing (2.0 FTE):**
  - Improved automated outlier detection algorithms (P1, 0.5 FTE)
  - Improved RFI identification algorithms (P1, 0.75 FTE)
  - RFI excision (P2, 0.75 FTE)
- **uv-plane calibration (1 FTE):**
  - Time-variable polzn D-term solve and apply (P1, 0.1 FTE)
  - Second-order D-term solve and apply (P1, 0.1 FTE)
  - Improved bandpass solvers (P2, 0.3 FTE)
  - Improved ionosphere models (P2, 0.5 FTE)

EVLA post-processing trial WBS

- **P1 = trial priority one; P2 = trial priority two**
- **Image-plane calibration (2.5 FTE):**
  - Apply a-priori primary beam models (arbitrary symmetry) (P1, 0.25 FTE)
  - Non-isoplanatic gain solve and apply (P2, 0.75 FTE)
  - Primary beam shape solve (P2, 0.5 FTE)
  - Primary beam offset solve and apply (P2, 0.5 FTE)
  - Primary beam polarization solve and apply (P2, 0.5 FTE)
- **Imaging (2.0 FTE):**
  - Optimized multi-scale CLEAN (P1, 0.25 FTE)
  - Full pixon deconvolution using visibility data (P2, 0.25 FTE)
  - Optimized GBT-EVLA combination (P1, 0.25 FTE)
  - OTF mosaicing (P2, 0.75 FTE)
  - MFS over large frequency range (P1, 0.5 FTE)

EVLA post-processing trial WBS

- **Image analysis (0.5 FTE):**
  - Improved automated component finders (P1, 0.5 FTE)
- **Visualization (1 FTE):**
  - Large data file support (> 1 TB) (P1, 0.5 FTE)
  - Visualization of spectra with large number of channels (P1, 0.5 FTE)
- **Simulation (0.5 FTE):**
  - Addition of error models of interest to the EVLA (P1, 0.5 FTE)
- **Algorithm test and evaluation (2 FTE):**
  - Sofware testing and algorithm performance evaluation (P1, 2 FTE)

EVLA post-processing trial WBS

- **Documentation (1.0 FTE):**
  - Documentation development for EVLA post-processing software (P1, 1 FTE)
- **Commissioning support (1.0 FTE):**
  - Support EVLA commissioning (P1, 1 FTE)
- **High-performance data reduction (1.5 FTE):**
  - Parallelized mosaicing (P2, 0.5 FTE)
  - Optimization of existing parallelized algorithms for cluster architectures (P2, 0.5 FTE)
  - Burst-mode reduction (e.g. solar) (P2, 0.5 FTE)

EVLA post-processing trial WBS

– Visibility data handling:	1 FTE
– Data editing / RFI mitigation	2 FTE
– uv-plane calibration	1 FTE
– Image-plane calibration	2.5 FTE
– Imaging	2.0 FTE
– Image analysis	0.5 FTE
– Visualization	1 FTE
– Simulation	0.5 FTE
– Test and algorithm evaluation	2.0 FTE
– Documentation	1 FTE
– Commissioning support	1 FTE
– High-performance data reduction	1.5 FTE
Total:	**16 FTE**

EVLA-specific: *10 FTE*

Generic: *6 FTE*

Possible post-processing project risk factors

- Inadequate resources:
  - ***Mitigation: scope control***
- Unconstrained requirements management:
  - Non-prioritized or unspecified scientific requirements
  - ***Mitigation: early requirements specification by project science team; active continuing involvement in requirements refinement***
- Weak scientific contract:
  - Insufficient involvement of scientific staff in algorithm development, evaluation and testing
  - ***Mitigation: active involvement with EVLA Project Scientist and science team; iterative 6-month AIPS++ release cycle gives an opportunity for scientific feedback on EVLA development; assigned FTE contribution from scientific staff for testing and algorithm evaluation.***