NRAO Algorithms R&D Priorities

S. Bhatnagar

Socorro, 5-6 March 2013

NRAO CASA and CASA based Pipelines Review

CASA (Common Astronomy Software Applications) by NRAO CASA Group in collaboration with ALMA Integrated Computing Team staff at ESO and NAOJ

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

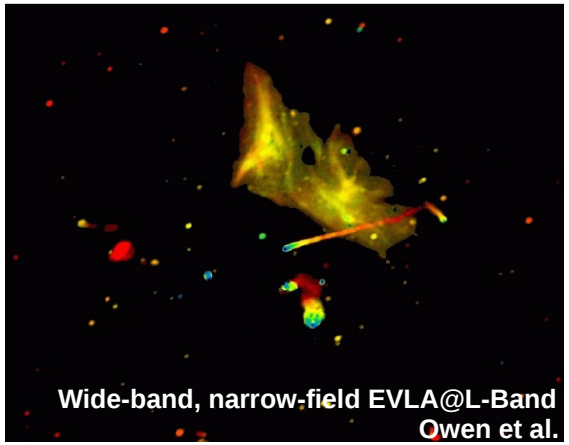
Introduction

- Developed R&D priorities for new algorithms required to achieve the project goals for imaging with EVLA and ALMA (frequency range covered: 1 GHz – few 100 GHz!)

- Identified and characterized the calibration and imaging problems in collaboration with the relevant scientific staff in the projects *(memos listed at the end)*
- Prioritization based on
  - Strongest effects that would limit the imaging performance
  - Common prerequisite techniques for both telescopes
  - Require long term focus in R&D
  - Analysis of accepted proposals
- This presentation only about CASA-related ARD
  - ~0.8FTE: Do the entire algorithm/software-development/user support cycle
- Develop/analyze wide range of solutions to minimize risk
- Initial investigations, research and development
  - Documented as technical reports/scientific papers/progress reports (list on the last slide)

Introduction: The highest nails

- Developed a Plan Document (2009)
  - By it's very nature, it is a scientific/technical document. Reviewed by competent scientists
  - The plan is being executed. Status Document (2012)
- Deconvolution errors higher than thermal noise limits
- Data flagging
- Effects due to wide bandwidths
  - Sky-domain and instrumental effects
- Effects of time-varying Primary Beams (PB)
- Heterogeneous array
  - Variable antenna aperture illumination, pointing errors
- Wide-field effects in wide-area imaging



We are ~here

  - Non co-planarity, PB effects in mosaic + combination of both
- Full polarization imaging
- Computing & data I/O costs, memory foot-print issues
- Deployment on HPC platforms

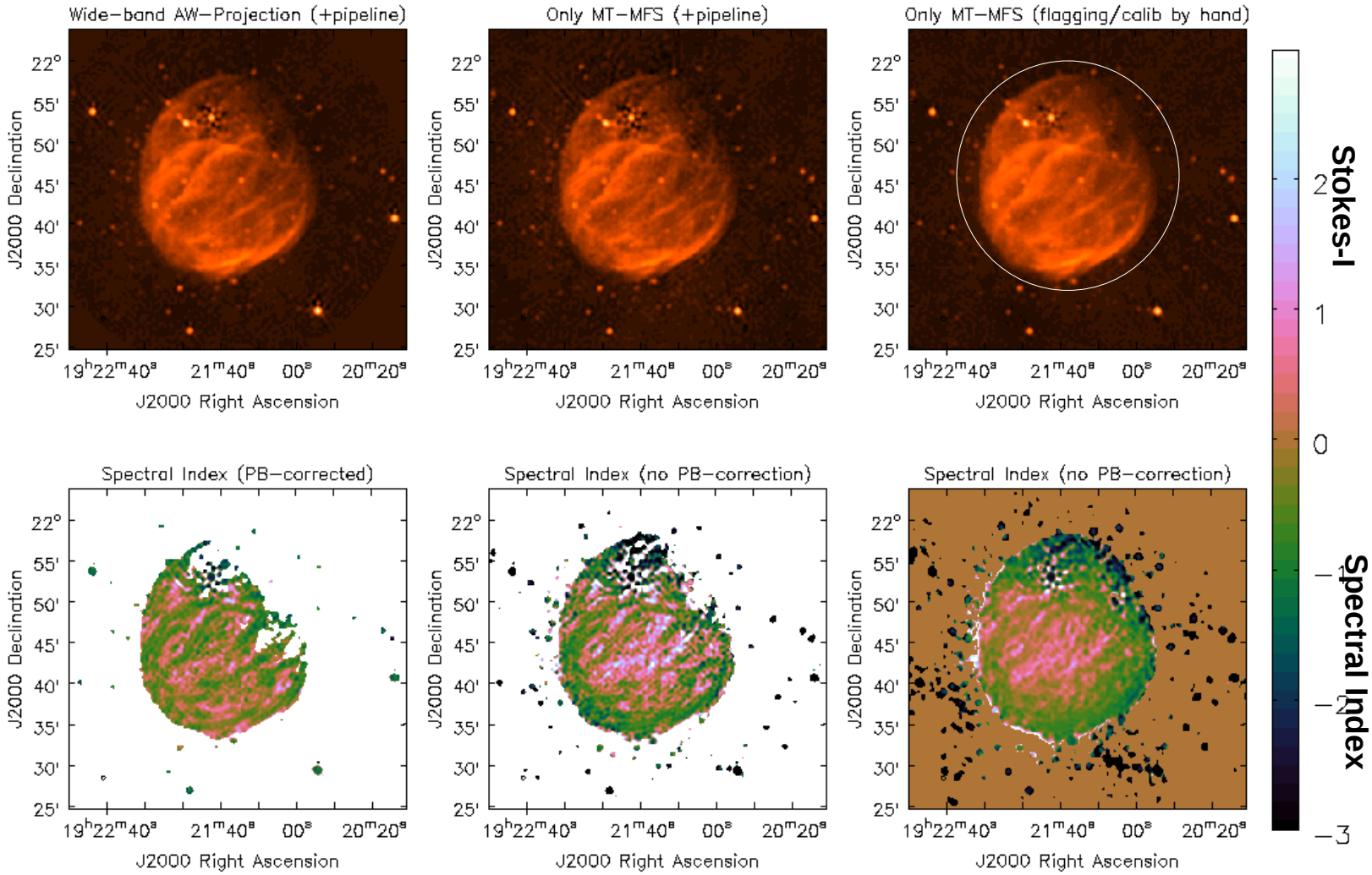
Status

- Almost on-track w.r.t. the Plan Document
- Sky spectral index + scale-sensitive deconvolution
  - MS-MFS (Rau et al. PhD Thesis, scientific papers)
  - Developed in CASA. Also available in software at other observatories
- Time-varying PB effects
  - Narrow-band A-Projection, W-Projection

    Developed in CASA. Also available in software at other observatories
- Data flagging
  - rflag (Griesen) and tfcrop (Rau): developed in AIPS and CASA
- Full-field wide-band imaging [In progress]

  We are here!
  - Wide-band A-Projection + W-Projection
  - Combine with MS-MFS to account for sky and instrumental frequency and time dependence
  - Being deployed in CASA: Internal scientific testing in progress

Progress: An example



2012... 2011 2010 (ApJL)

(Higher spectral index around the edge of the SNR (2010) due to PB-effects)

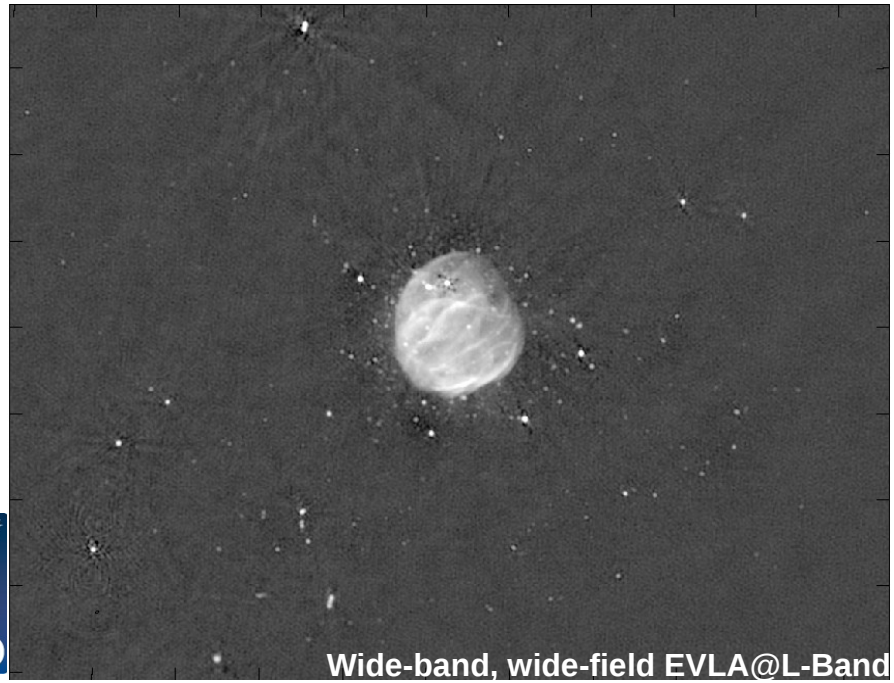
[Courtesy: Rau]

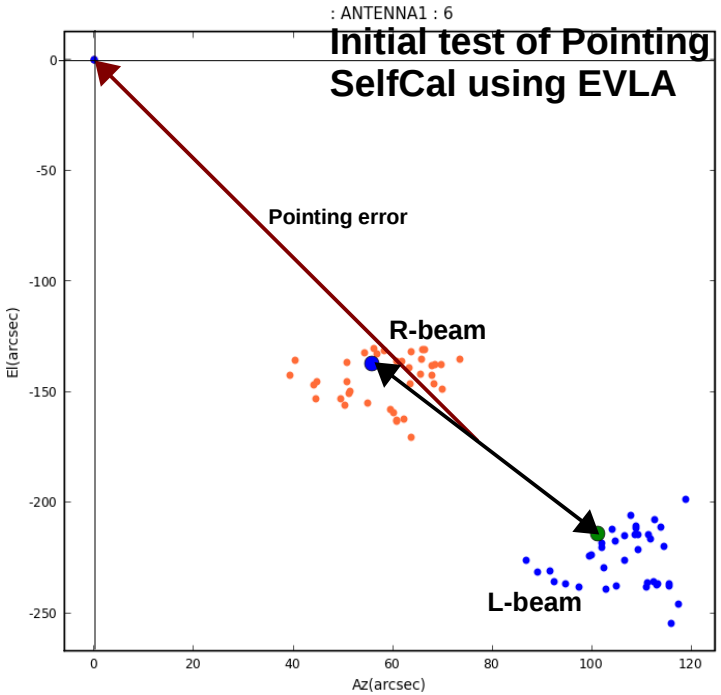
Status

- Heterogeneous array
  - Available in CASA. Handle varying PB size (not shape)
- Mosaicking (ALMA)
  - Linear and non-linear without proper PB corrections available in CASA
  - Mosaicking with full heterogeneous PB correction: As part of full-field imaging work [In progress]
- Low frequencies (LCS-band @ EVLA): [In progress]
  - Narrow-band AW-Projection
    - Combined with wide-band effects as part of full-field imaging work
- RM-synthesis (image-plane algorithm)
  - New algorithms needed

Status

- Simulations to assess the effects of random pointing errors
- Pointing SelfCal prototype (ALMA driven)
- Scale sensitive deconvolution approaches investigated
  - MS-Clean, Asp-Clean, CS approach
- Various approaches to RFI identification, removal
  - E.g. Eigenvalue analysis





Status/Issues

- HPC and new algorithm development: Compute-i/o-memory
  - Develop algorithms with higher compute-to-I/O ratio which scale with the projected computing and i/o technology development
  - Minimize the memory foot-print, Minimize serial operations
- Major vs. Minor Cycle cost in imaging
  - With Projection algorithms, imaging requires a single pass through the data
  - Cost of deconvolution: In general higher for spectral-line imaging
    - May need algorithmic work here
- Resource conflict: New development vs. maintain advanced code
- Software development for R&D in conflict with production software needs
  - R&D needs code stability, Longer term focus
  - Sometimes require large scale code re-factoring (without show-case “new functionality”)

Path forward

- Revisit priorities and update the Plan Document
- Consolidate implementation + evaluate imaging and run-time performance of full-field wide-band imaging
  - Mosaicking with full heterogeneous PB support
- Full-pol. Imaging
  - Classical RM-Synthesis is insufficient
- Investigate the magnitude of errors due to terms ignored so far and resulting limits on imaging performance
  - Pointing errors, Elevation dependence,...
- Post-imaging analysis
  - Image cubes: Can they be combined with appropriate deconvolution algorithms? Should they be combined?
- Auto-boxing:
  - Multiple solutions available (Griesen, Cotton, Golap). Input provided to CASA management

Path forward/Conclusions

- SD+Interferometric imaging
  - Characterize problem, if any, in the current approach
- System Integration of component algorithms
  - Simultaneously account for multiple Direction Dependent Effects
- Develop cost-performance equation based on scientific goals?
  - Evaluate limits of traditional, hybrid and modern approaches
  - Generic algorithms/software vs. solutions tuned for specific scientific goals
- Engage + utilize current interest in private sector/other groups?
  - Advanced algorithms require higher computing using significantly larger data volume
    - Current collaboration with nVIDIA R&D group for use of GPU computing
    - Collaborations with LOFAR (for AA-Imaging, heterogeneous array) and GMRT (high sensitivity low frequency issues)

Path forward/Conclusions

- Exploit NRAO domain expertise for developing next-gen solutions?
  - EVLA and ALMA as SKA-pathfinders
    - Exposes most of the problems that next-gen telescopes face, but in a “controlled” environment
    - Develop and test algorithms for Direction Dependent Errors, measure computing costs, test imaging preformance, etc.
    - Real vs. simulated data: Use EVLA/ALMA at near-SKA data rates
  - Test the limits and cost-performance estimates using software solutions and large-scale HPC (Parallel Computing platforms, GPU clusters, multi-threading,...)
- Invest in developing thread safe code?: Currently very hard to exploit multi-core, multi-CPU h/w
  - Need reliable support from underlying libraries, support code
    - CASACore is not thread-safe
- Streaming-data processing
  - The difficult part: Fundamentally iterative calibration and imaging algorithms

List of algorithms/References

- Wide-field effects
  - W-Projection (Cornwell et al., IEEE), A-Projection (Bhatnagar et al., A&A), Non-linear Mosaic, Linear mosaic, A-Projection for AA-telescopes (Tasse et al., A&A)
  - Traditional approaches: Partial image domain corrections, faceting (Smirnov et al.)
- Wide-band effects
  - Imaging (sky) domain: MFS/MS-MFS (Rau et al, Thesis, A&A, IEEE)
  - Instrumental: Wide-band A-Projection, AW-Projection (in prep. ApJ)
  - Hybrid approach: Cube-imaging + partial image domain corrections
- Full-field wide-band imaging
  - AW-Projection + MS-MFS (in prep. ApJ)
  - Wide-band mosaicking, full heterogeneous array support
- Technical reports / Progress Reports:
  - Memo on the projected algorithms R&D needed.
  - EVLA Memos 84 (pointing selfcal), 132/133 (HPC), 139 (Plan document, 2009), Progress status & Plan (2012)
  - Internal white paper on algorithms R&D, 2012

The National Radio Astronomy Observatory is a facility of the National Science Foundation
operated under cooperative agreement by Associated Universities, Inc.

www.nrao.edu • science.nrao.edu