

# EVLA Project Book Chapter 12

## Post-Processing - Software

Bryan Butler & Steve Myers

2004-Dec-09

### 1 Introduction

It is clear that the post-processing needs of the EVLA will far exceed those of the VLA (or any other current interferometer). The salient features of the EVLA that will drive the post-processing software development are the capabilities to produce data covering: (i) a large instantaneous bandwidth (2:1 at the lowest frequency bands), and (ii) a large number of spectral channels (flexibly arranged in multiple spectral windows). These two features result in high data rates ( $> 5$  MB/s) as a matter of course, with the WIDAR correlator capable of producing much higher rates (up to 350 GB/s!), and also therefore with the prospect of dealing with large ( $> 1$  TB) datasets. Therefore, calibration and imaging of the data from the EVLA will present problems that must be solved by a combination of a scalable post-processing package, algorithmic improvements (including parallelization), and processing and I/O speed gains. This is also true for ALMA (which also has a large correlator), which must be kept in mind, especially given that one of the high level goals of NRAO software is to have common software across the observatory (as much as possible).

The package of choice for post-processing reduction (and in fact also for pipeline reduction) of EVLA data is AIPS++. This decision is based on several factors:

- AIPS++ is scalable and poised to take advantage of parallel processing (Golap et al. 2001; Young & Roberts 2000; Roberts 1996);
- Post-processing manpower at NRAO is concentrated much more heavily in AIPS++ than in any other package (AIPS, e.g.);
- ALMA has chosen AIPS++ for its post-processing needs (see documentation from the ALMA Computing IPT CDR2 which occurred June 7-8, 2004).

We note, however, that this is a decision concerning the final (when full EVLA operations begin) post-processing package, and there is an additional issue of transitioning to this end, in which AIPS plays a major role.

### 2 Requirements and Use Cases

The detailed requirements for the post-processing package for EVLA are listed in Myers et al. (2003), which was adapted from a similar ALMA document (Myers et al. 2002). This document describes the requirements in the general post-processing area as well as the default image archive pipeline (note that the ALMA pipeline requirements are contained in Lucas et al. [2003]). Requirements for the quick look and calibration pipelines for EVLA are contained in Butler et al. (2003). These two documents (Myers et al. 2003; Butler et al. 2003) constitute the complete set of requirements for EVLA post-processing. These documents are intended to be updated as needed, by the appropriate groups of people (the authors, mostly).

Use cases are being written by members of the scientific staff, and incorporated in the design as needed. In general, use cases are produced as requested by the AIPS++ group, when needed. NRAO staff scientists are intimately involved with the AIPS++ group, with direct and frequent contact encouraged, in order to best facilitate the flow of communication between the two groups (astronomers and programmers).

## 3 Common Software for NRAO Telescopes

There are two main components of the common NRAO models which affect the post-processing: the Project Model, and the Science Data Model. It is the responsibility of the NRAO e2e Oversight Committee to come up with the detailed descriptions of these global models, while each of the projects must determine in which way they deviate (most of the models are constructed as a base or core model, with the intention that each telescope defines extensions to that core model). At this point, only draft versions of the descriptions of the above models have been produced by the e2e Oversight Committee, which are available via the NRAO wiki.

### 3.1 Project Model

The Project Model contains the information on observational intent and the structure of the observing program. Depending on how the overall structure of the EVLA software develops, it could be considered that this is unimportant to the post-processing software itself, as the necessary information could be passed along with the data in some way. But it will likely be necessary for the post-processing software to access the information in the Project Model directly (presumably it will be stored in some database which can be queried).

### 3.2 Science Data Model

This model defines the content of science data products stored in the Science Archive or exported to the user community. All post-processing is defined in terms of the SDM. At the end of the Observing process data is produced which is conformant to the SDM. Data is stored in the Science Archive in this form, and is exported to the user in an Export Data Format, which is a mapping of the SDM to some external representation such as FITS or XML. An Export Data Format will be specified that will be used to convert input received by DCAF into a form suitable for storage in the Science Archive and use by data reduction systems. Active efforts are underway to specify the format in conjunction with the needs of ALMA and GBT. The current thinking is that there will be a core NRAO SDM, with extensions for each of the particular telescopes. The core SDM will be adapted from the current ALMA Science Data Model (Viallefond & Lucas 2004). A document describing necessary EVLA extensions has been written (Benson & Fomalont 2004).

The current plan is that the EVLA Computing Division will provide a tool which will convert from the EVLA Export Data Format (which is the physical implementation of the logical structure contained in the Science Data Format) to measurement sets, for importing into AIPS++.

## 4 Data Reduction Pipelines

A large fraction of data collected by the EVLA will be taken in one of a handful of standard observing modes, for example, low frequency continuum, high frequency continuum, HI (neutral hydrogen) spectral line, or polarization. Given the considerable experience in reducing data taken in similar kinds of modes with the VLA, it is reasonable to assume that reduction of this type of data can be completely automated. The post-processing package (see section 5), when combined with some information collected in the observation preparation stage, during actual observing, and with some *heuristics* (rules for what to do given certain situations) should be sufficient to complete such automatic reductions. This does imply, however, that certain critical parts of the post-processing package which are not currently in existence are implemented and robust, for example automatic flagging of bad data, and imaging of wide-bandwidth ratio and wide-field data. There are three types of pipelines which will be utilized in the software for all NRAO telescopes, including the EVLA: the *calibration pipeline*; the *quick-look pipeline*; and the *default image archive pipeline*.

ALMA is devoting significant effort into investigations of pipeline reduction, and we intend to capitalize on that. In fact, the current plan is to borrow almost whole-cloth the infrastructure developed for ALMA and incorporate it as the EVLA pipelines. There will be additional items to be developed for EVLA (notably RFI flagging or excision, imaging and calibration details, and the heuristics for reducing the data [to be supplied by scientific staff and implemented by the AIPS++ group]).

## 4.1 Calibration pipeline

There are a number of types of observations of calibrators which will be used by the monitor and control system to set instrumental parameters necessary for observing. Among these are antenna location determination, focus determination, and reference pointing determination. In fact, for every calibrator observed, a quick reduction of the data will provide the current atmospheric conditions as well as other information. It is thus intended that all calibrators will pass through the calibration pipeline, and the results be made available to the monitor and control system. This pipeline will be a part of the TelCal subsystem.

## 4.2 Quick-look pipeline

During observations of a source, it is instructive to obtain an initial data reduction to see whether the instrument is behaving as expected, and in order to guide further observations (in the case of manual or interactive observing). Such an initial data reduction might use the default stored calibration parameters for things like bandpass and polarization calibration, and would not do detailed deconvolutions of the raw images. This is the job of the quick-look pipeline. Even though it is quite simple, there are a few parameters which should be able to be tuned for the quick-look pipeline, for example, when looking at maser sources, it may be only interesting to look at a small subset of the total number of available channels in order to see if the maser emission is present. The astronomer should also be able to set the frequency of how often the quick-look pipeline is run, or be able to initiate it manually (assuming that the reduction can actually be done in a relatively short amount of time).

## 4.3 Default image archive pipeline

It is the intent of the project that one of the primary data products of the EVLA be a standard image archive. This will be a valuable resource for future astronomers (“data-miners”), and will provide a consistent record of the images produced from the data taken by the EVLA. In addition to images, this archive would also contain intermediate products such as calibrated visibility data, and pipeline diagnostic information.

The development of a pipeline that can produce robust, scientifically viable images for an instrument such as the EVLA will be a technical challenge, and constitutes a high-risk task for the project. This pipeline will rely upon the software and algorithms developed for the general post-processing problem, as well as detailed heuristical knowledge of the processing of EVLA data, plus development of a pipeline technical framework. As ALMA is also developing a similar pipeline, we plan to co-develop as much of the capability as is possible, and concentrate what additional resources we have on the EVLA-specific tasks.

It is expected that this archive should only be initiated once for each project, although if the processing is not excessive it can be repeated if more accurate calibrations become available. Since it must be consistent across projects, there should be very few (if any) selectable parameters for this pipeline. Of course, for complicated projects which do not fit into one of the standard observing modes or models, there will be no guarantee that this pipeline will produce meaningful results. One of the difficulties is then estimating exactly what the quality of the images in the default image archive is. Various measures of image quality can be constructed, but in practice it is extremely difficult to apply them to images taken across a very wide range of the possible observing mode parameter space.

Introduction of such a pipeline into EVLA operations has a number of consequences (which we are in the process of investigating). For example, the extent to which the pipeline is automatic or requires extensive human intervention affects the staffing profile. These concerns will be placed under study, and will enter into the Pipeline Design.

# 5 AIPS++ for EVLA Post-Processing

For all data which cannot be reliably reduced via a pipeline, or for astronomers who wish to modify or extend what is done within the pipeline, there must be a post-processing software package capable of performing all steps necessary to turn the measured visibilities into final image cubes. For the VLA, several packages have been used for data editing and calibration over the years, but for nearly the entire lifetime

of the VLA, AIPS (Greisen 2003) has been the primary software package for data processing. For some time now, however, AIPS++ (Cornwell & Wieringa 1997) has been developed as the modern successor to AIPS. Currently, as described in the introduction above, AIPS++ is considered the primary package for the calibration and imaging of data from ALMA and EVLA. However, we note that any package that satisfies all of the requirements for EVLA post-processing (Myers et al. 2003) can be used for these purposes. Currently, the AIPS++ software is being developed with the Scientific Software Group (SSG) at NRAO. There is another committee looking at the longer-term issues involved with data reduction for the observatory, which is providing input to the SSG.

## 5.1 Algorithm development

The developments needed for EVLA (and future interferometers in general) can be generally grouped into three categories: doing things faster, e.g. parallelization [Golap et al. 2001; Young & Roberts 2000; Roberts 1996] or grid computing (Berriman et al. 2003); doing things that we are already doing, but better, such as multi-scale CLEAN (Bhatnagar & Cornwell 2003); and doing completely new things. Specifically, in the second and third of these categories for the EVLA there are a number of important developments needed which are common with ALMA:

- automatic flagging;
- tropospheric corrections, including “phase screen” derivation (Carilli & Holdaway 1999; Cai & Cornwell 2004);
- exotic imaging algorithms (pixons [Cornwell 2001; Puetter & Yahil 1999], wavelets (Maisinger et al. 2004), other Bayesian methods, etc...);
- combination of single-dish and interferometric data (Stanimirovic 2002);
- wide-field full-polarization imaging (Cornwell 2003) especially with different primary beams per antenna and time variable primary beams (Cornwell 2002) (some of this is common with ALMA, some unique to the EVLA).

There are also a number which are unique (at least mostly) to the EVLA:

- wide-field, wide-bandwidth imaging (Cornwell et al. 2003; Cornwell 2002);
- RFI excision and subtraction (Perley & Cornwell 2003);
- ionospheric corrections, including “phase screen” derivation (Smirnov 2000).

The intent is for scientists and programmers to work on these algorithmic developments in concert, since input from both is necessary to solve these very difficult problems.

## 6 Testing and Acceptance

A detailed testing and acceptance plan is currently being developed, given constraints on current levels which can be supported (which is highly dependent on the total amount of scientific staff support available). The Project Scientist for Software is responsible for the ultimate acceptance of the software, in consultation with the overall Project Scientist. The committee of staff scientists will certainly contribute critically here as well. It is imagined that the testing and acceptance program will mimic that of ALMA in many ways, utilizing the NAUG and outside testers extensively.

## 7 Deliverables

The EVLA project, in consultation with AOC operations, determines the dates at which certain AIPS++ EVLA capabilities should be in place. A plan containing the schedule of deliverables is currently being developed.

## 7.1 Development Plan

The AIPS++ group takes the timeline presented above, and confronts it with available manpower to derive the development plan. This plan has both long-term and short-term components. These must then be iterated with the EVLA project.

## References

- Benson, J., & E. Fomalont, 2004, EVLA/VLBA Requirements for the ASDM
- Berriman, G.B., D. Curkendall, J.C. Good, J.C. Jacob, D.S. Katz, M. Kong, S. Monkewitz, R. Moore, T.A. Prince & R.E. Williams, 2003, Architecture for access to a compute intensive image mosaic service in the NVO, in *Virtual Observatories*, ed. A.S. Szalay, Proc. SPIE, 4846, 91
- Bhatnagar, S., & T.J. Cornwell, 2003, Scale-sensitive (multiscale) deconvolution of astronomical interferometric images, in *Astronomical Adaptive Optics Systems and Applications*, eds. R.K. Tyson & M. Lloyd-Hart, Proc. SPIE, 5169, 331
- Butler, B., C. Brogan, C. Chandler, B. Clark, S. Myers, P. Napier, F. Owen, R. Perley, & M. Rupen, 2003, EVLA e2e Science Software Requirements, EVLA Computing Memo 26
- Cai, C., & T. Cornwell, 2004, Parameterized Phase Screen Modelling for the VLA, in preparation
- Carilli, C.L., & M.A. Holdaway, 1999, Tropospheric phase calibration in millimeter interferometry, *Radio Sci.*, 34, 817
- Cornwell, T., & M. Wieringa, 1997, The Design and Implementation of Synthesis Calibration and Imaging in AIPS++, in *ADASS VI*, eds. G. Hunt & H.E. Payne, ASP Conference Series 125, 10
- Cornwell, T., 2001, Pixon imaging in AIPS++, AIPS++ Note 242
- Cornwell, T., 2002, High dynamic range radio synthesis imaging with time variable primary beams, Presented at 2002 URSI meeting, Boulder
- Cornwell, T.J., 2003, Full Primary Beam Stokes I, Q, U, V Imaging, EVLA Memo 62
- Cornwell, T.J., K. Golap, & S. Bhatnagar, 2003, W Projection: A New Algorithm for Non-Coplanar Baselines, EVLA Memo 67
- Golap, K., A. Kembal, T. Cornwell, & W. Young, 2001, Parallelization of Widefield Imaging in AIPS++, in *ADASS X*, eds. F.R. Harnden, Jr., F.A. Primini, & H.E. Payne, ASP Conference Series 238, 408
- Greisen, E.W., 2003, AIPS, the VLA, and the VLBA, in *Information Handling in Astronomy - Historical Vistas*, ed. A. Heck, Kluwer, 109
- Lucas, R., and 18 others, 2003, ALMA Science Software Requirements and Use Cases, ALMA Computing Memo 11
- Maisinger, K., M.P. Hobson, & A.N. Lasenby, 2004, Maximum-entropy image reconstruction using wavelets, *MNRAS*, 347, 339
- Myers, S., B. Butler, C. Chandler, B. Clark, F. Owen, M. Rupen, C. Brogan, R. Perley, & P. Napier, 2003, DRAFT - EVLA Data Post-Processing Software Requirements, EVLA Computing Memo 28
- Myers, S., and 18 others, 2002, ALMA Offline Data Processing Requirements, ALMA Computing Memo 18
- Perley, R., & T. Cornwell, 2003, Removing RFI Through Astronomical Image Processing, EVLA Memo 61
- Puetter, R.C., & A. Yahil, 1997, The Pixon Method of Image Reconstruction, in *ADASS VIII*, eds. D.M. Mehringer, R.L. Plante, & D. A. Roberts, ASP Conference Series, 172, 307
- Roberts, D., 1996, Parallelization of AIPS++, AIPS++ Note 203
- Smirnov, O., 2000, Ionospheric Corrections in AIPS++, AIPS++ Note 235

- Stanimirovic, S., 2002, Short-Spacings Correction from the Single-Dish Perspective, in *Single-Dish Radio Astronomy: Techniques and Applications*, eds. S. Stanimirovic, D. Altschuler, P. Goldsmith, & C. Salter, ASP Conference Series, 278, 375
- Viallefond, F., & R. Lucas, 2004, ALMA Export Data Format v2.0, ALMA Document ALMA-70.00.00.00-004-A-SPE
- Young, W., & D. Roberts, 2000, Parallelization Test Results, AIPS++ Note 232