EVLA Memo 122: EVLA Requirements for Postprocessing Algorithms and Computing

Frazer Owen

1/7/2008

1 Introduction

The approach NRAO takes toward postprocessing for EVLA (and ALMA) is critical to the success of the instrument as well as the number of astronomers who can make use of the new instrument. At one extreme NRAO could simply give the output of the correlator backend to individual astronomers to turn into publishable results somehow. Some of our most expert users have suggested such a mode of operation. For some projects for our most elite users this might work. For other projects it is not clear any existing group could solve some of the issues listed below on their own. The other extreme would be that NRAO delivers the reduced results for almost all projects to users, who primarily analyze the results and write papers. In practice we don't have the time even if we had unlimited resources to get to this latter mode of operation by the end of the construction project. What we absolutely need is somewhere in between these extremes in the early days of EVLA operation. However, in order for most of our users to be able to make use of the EVLA for the research goals we used to justify the project, we need to provide postprocessing resources. These include 1) a set of basic algorithms in some generally available software package and 2) the necessary computing system to allow all our users to reduce, image and analyze data for projects which are not possible on their local hardware alone.

2 Algorithms Needed For EVLA

As we get closer to having to have the EVLA hardware available including the WIDAR correlator, many, perhaps most, of the scientific investigations need significant algorithmic development to be possible. It seems likely that the initial hardware system will be capable of providing the data before we are ready to support many key capabilities. In many cases we probably need data to begin flowing before we can converge on the details of the algorithms. The algorithms will need to be tailored to the realities of the hardware and the environment. Thus commissioning and algorithm development often need to go together. However, given the length of the (incomplete) list I give below, it seems clear not everything of importance can be tackled immediately. Thus we will need to prioritize. Nonetheless we need to define a basic suite which we can finish in a time not too much longer than the construction project.

The nature of the work required for these algorithms is creative at the in the early stages. Thus it is hard to promise that we can reach the goal of completing algorithm X on date Y. However, we have a telescope project which requires certain functionality in a finite time to be successful. Thus we need to balance the creative and the pragmatic. If we can't reach the ultimate solution when we desire it, we need to have realistic fall-back positions which will allow research to advance. It also seems clear we will need algorithm development for a long time after construction as part of operations model if we are ever to reach EVLA's potential.

Below I list work we need in some broad priority categories. For each algorithm I give a description followed by the year and quarter when the algorithmic work described should be made available to users in order to keep up with the overall VLA to EVLA transition plan.

2.1 What we can do now:

For EVLA, we believe we know how to make spectral line images assuming we have the data storage and computing cycles needed for the current algorithms (but see section 3). Thus emission or absorption line spectral imaging and line searches can be carried out with algorithms we have now. Whether we can effectively use these images for science with the current software is less clear but we can likely do something with such data as long as the dynamic range requirements are not too high and interference is not severe. Perhaps we can also do some crude detection experiments in the continuum at the high frequencies by adding up narrower band images in unconfused areas of the sky and assuming interference if not a problem. There certainly are other special cases but beyond these sorts of experiments and without artificially limiting the bandwidth analyzed, we need new algorithms. The discussion below lists the algorithms in approximate priority order and gives a target date for the algorithms to be available to the community at large, not just a few insiders.

2.2 The two big, hard ones

These two types of algorithms are particularly important and challenging. They need to addressed with the highest priority.

2.2.1 Editing of uv data

Almost every discussion of EVLA projects starts with the assumption that we will automatically remove all the interference from the dataset before going forward. Simple algorithms, some experimental, exist which can do an adequate job for some VLA datasets. For difficult cases, one falls back on flagging by hand using a TV display of the uvdata channel by channel. These procedures seem inadequate for EVLA. We need a major, creative, probably sophisticated, mathematical approach to deal the auto-editing of the data. These techniques probably need to be developed starting with the first prototype WIDAR data from the EVLA.

2009 Q4

2.2.2 Wide-band, widefield imaging:

The EVLA was sold on its much higher continuum sensitivity which it achieves primarily by using a much larger instantaneous bandwidth, up to an octave and at least 30%. This will require imaging with bands over which the spectra of most sources change significantly and where the monochromatic synthesized beam and primary beam change by a factor of up to four in area over the observed frequency range. Furthermore, a typical observed field will contain many sources, which unless they are taken into account, will limit the sensitivity, especially at longer wavelengths. Making noise-limited, full band images will require algorithms which address these effects. Dynamic range of $10^5 : 1$ is required. One can divide the complexity of what needs to be done into three levels, none of which can be handled by NRAO's publicly available software now.

1. **Narrow-field imaging of single, unconfused sources**

   Here existing algorithms which solve for the spectral index as well as the total intensity image can probably be added to CASA and/or AIPS. However, work is involved in doing that.

   2009 Q1

2. **Narrow-field imaging of single sources with confusing sources**

   This case will be very common since the increase in continuum sensitivity of EVLA make this case the norm for the majority of the cases, except perhaps in simple regions of the sky at the highest frequencies. Ideas exist for how to handle this problem (e.g. Urvashi Rao's PhD thesis project). However, not only does any new algorithm need to be developed and tested but it needs to migrate into publicly available packages to be useful.

   2009 Q4

3. **Wide-field, Wide-band imaging of the full primary beam**

   This is the holy grail. To my knowledge general ideas exist but no clear, detailed proposal has been made to achieve this goal. If we could just jump to this solution that would be best but seems unrealistic.

   We probably need to work on all three projects in parallel to achieve the best solution as a function of time.

   2011 Q1

2.3 Other Imaging and Calibration Algorithms

We also need work in a number of other areas which will be important for EVLA to reach its goals. In some of the cases below these algorithms may be as the critical as those in section 2.2 but that is not absolutely clear. It is clear that the full potential of EVLA science proposed will not be realized without most of them. In some cases below we may be able to design a single algorithm which solves the problem. In others we need to plan for continuing development beyond the end of the construction project.

2.3.1 Pointing/Primary Beam Corrections

Many routine observation, maybe most with EVLA are likely to be limited by pointing. Pointing self-calibration algorithms are under development and need to be finished.

2011 Q1

2.3.2 Polarization Calibration

EVLA offers the potential for new approaches to polarization calibration which need to be mathematically developed and tied to the realities of the EVLA hardware as we learn them using WIDAR. Polarization self-calibration was also demonstrated by Mark Holdaway many years ago but needs to developed further and added to the available tasks. Many experiments also require full field polarization calibration which needs further development.

2011 Q1

2.3.3 High Dynamic Range Imaging

To reach its potential very high dynamic range images need to be produced ($10^6 - 10^7 : 1$ This involves a variety of potential corrections/calibrations (e.g. correction of the parallel hands for instrumental polarization, correction for the primary beam and beam squint as a function of parallactic angle). This requires experimentation and development.

2012 Q1

2.3.4 Maximum Entropy and other Deconvolution Algorithms

We rely on Clean primarily for imaging. We need to continue to study other approaches. We have maximum entropy (MEM) and related algorithms in AIPS and CASA but their convergence algorithms are not ideal and need to be improved to be generally useful. MEM could become the default way much of our complex imaging is done if the program was improved. Other sophisticated algorithms might also become the norm with adequate development.

2011 Q1

2.3.5 Auto-boxing for Clean

Bill Cotton has developed an auto-boxing routine in OBIT which needs to migrate to other packages and to continue to be developed. With the size and the number of frequency channels with EVLA, this is a necessity for correct imaging.

2009 Q4

2.3.6 Mosaicking/ Short spacings

EVLA needs continued development of mosaicking as does ALMA. In particular, we need to be able to combine GBT data with EVLA data routinely to make images.

2011 Q1

2.3.7 Ionospheric effects

At low frequencies the ionosphere will limit our imaging and we need to add better techniques, even at 20cm or higher frequencies to produce the best images. At 90cm and 4m (where WIDAR should facilitate significant new science), we need algorithms which take into account the general distortion and de-focusing. Such development is naturally led by the new low frequency arrays under development (e.g. LWA, LOFAR) but we need to stay involved and add their techniques to EVLA to make fully use of the WIDAR correlator in these bands.

2012 Q1

2.4 Science Reduction Algorithms

These algorithms are required in order to extract scientific results from the basic images produced with the algorithms discussed above.

2.4.1 Rotation Measure Imaging

One major goal of EVLA is the study of the magnetic universe for which possibly the most important tool is full field rotation measure imaging. Only very limited algorithms exist in AIPS for this purpose base on using a small number of frequencies. A new algorithm is needed for the EVLA case with its very large fractional bandwidths.

2011 Q1

2.4.2 3D source Cataloging

We have a good 2D source cataloging program in AIPS but with EVLA we need a 3D program for spectral line cube searches. We also want to be able to determine the spectral properties of continuum sources over the full primary beam (e.g. spectral index, free-free absorption, spectral curvature).

2012 Q1

2.4.3 3D Visualization

We need better ways of looking at a data cube with many channels. Perhaps here we need to adapt existing techniques to EVLA. Obviously this is an ALMA issue as well.

2012 Q4

2.4.4 Specialized Algorithms

Several new types of observations will be possible with the VLA which will need fairly straight-forward algorithmic work. For example, stacking of several recombination lines in a single EVLA band will allow big sensitivity increases for this type of science. However, the differences in the line-widths in frequency will require appropriate transformation of each line into velocity space before they are combined. It is fairly clear what needs to be done but software is needed to make this type of experiment straight-forward. The date given for this task is the end of the construction project, although one would hope some of these algorithms would be available sooner as observing pressure dictates.

2012 Q4

2.5 Requirements

The e2e division, the EVLA project and the AOC operations staff need to come to an agreement as to how to address these needs for new algorithms practically. The availability of these algorithms will affect the plan for first EVLA science and perhaps end-game strategies for EVLA hardware deployment.

What we require by April 1 2008 is a plan with the number of required FTEs and a schedule for work on each algorithm to meet the goals outlined above. The funding for these positions also needs to be addressed so that the plan is realistic. If this plan extends beyond FY09 then we also need to modify the plans for first science accordingly. What we require first and foremost is a realistic, defensible plan.

3 EVLA Computing Requirements

One of the important requirements for EVLA success is to have the computing capacity available to do the necessary postprocessing (imaging, calibration, analysis). The EVLA data rates will be 100-1000 times the VLA and the calibration and imaging algorithms are by nature more extensive and therefore more costly in computing resources. Even straight-forward spectral line imaging will require more computer resources than are available on most astronomer's desktops. Besides computer speed, IO is also likely a major issue and may require a nonstandard solution. It seems likely that parallel processing will be required and probably some sort of cluster computing. More disk storage and memory will likely be needed than most users will have at their home institutions, at least in the early days of EVLA. The configuration of the CPUs, data storage and memory may need to be tuned to our problem The existing postprocessing packages do not make direct use of parallel computing and may not be designed to take best advantage of memory and/or scratch disk space. All these issues need to be studied and solutions defined within the short period we have before EVLA is ready for science. Thus there are three related requirements for successful EVLA computing:

1. **To introduce parallel computing techniques into its existing and future postprocessing;**

2. **To specify and obtain the necessary hardware to address the EVLA processing demands as a function of time;**

3. **To develop an operations plan for the use of the new hardware by NRAO and/or the user community to produce the needed data products in order to achieve the scientific goals of the EVLA observations.**

All this requires a detailed plan with dates and personnel requirements from now until some time after the end of the construction project. The e2e division, the EVLA project and AOC operations staff need to work together to define the details of a plan. What is required is a pragmatic plan by April 1, 2008 with personnel, dates and funding identified to address the three requirements listed above. Given the short time-frame available some work on these issues probably needs to begin immediately. This plan should not slow down the VLA-EVLA transition. In any case the plan needs to take into account and be merged with the VLA-EVLA Transition plan as outlined in the EVLA project book.