

Data processing software for radio astronomy

Huib Jan van Langevelde¹

JIVE Dwingeloo & Sterrewacht Leiden
Postbus 2, 7990 AA Dwingeloo, the Netherlands
E-mail: langevelde@jive.nl

In this paper ongoing software projects among the European radio astronomy observatories are reviewed. In particular, I report on the progress in the ALBUS project which aims to enhance the data products, as well as the tools for radio astronomy data processing. In particular the capabilities of ParselTongue will be discussed. This is a programmable Python interface to classic AIPS that is now publicly available. The options for work on future data reduction packages will be discussed as well. A new effort is the FABRIC project which is part of the EC-funded collaboration EXPReS. This prepares for the next generation of e-VLBI and includes a pilot project on distributed correlation, implementing the correlation in software on standard computing environments, employing the Grid.

The 8th European VLBI Network Symposium
September 26-29, 2006
Toruń, Poland

¹ Speaker, on behalf of the participants in ALBUS & FABRIC

1. Introduction

The user appreciation of data from radio telescopes is intimately related to the software tools made available for the data processing. The data reduction package provides the primary interface between the user and the data. After a period of relative little activity in this area, there is now an effort to produce new user software tools and enhancement of the data product in the European radio astronomy community. This EC-funded project, ALBUS (for Advanced Long Baseline User Software), has established new expertise in various areas. Moreover, it has produced a first user tool in the form of ParseiTongue, a Python scripting interface for AIPS.

Besides this collaboration on software in the user domain, there is also an increasingly important role for software in data acquisition, as more functionality gets implemented in a flexible manner on commodity hardware. In order to explore the possibilities for faster data acquisition and transportation on commodity hardware, as well as software correlation, several EVN institutes collaborate in FABRIC (Future Arrays of Broad-band Radio-telescopes on Internet Computing).

1.1 The need for new products and tools

Modern computing components are transforming the way VLBI is done on the input end (disk recording systems, e-VLBI), but they are also enabling new scientific user products at the output end. An example of this is the enhancement of the correlator output bandwidth using modern computing equipment in order to increase the field of view. In traditional VLBI, a very small (arc-second) patch around a source of interest is imaged. Upgrades to the EVN correlator currently allow output speeds of 6MB/s with 48MB/s planned, which allows full sensitivity on arc-minute scales. Also, the e-MERLIN array will yield a wide-band correlation leading to a similar explosion in data rates. Within the ALBUS project, the problems that these increased data rates pose are addressed by supplying the users with new tools. More information on this project, which is carried out in the framework of RadioNet, can be obtained from <http://www.radionet-eu.org/rnwiki/ALBUSHWikiPage>. To deal with the data rates from the above example there is a work package to make the EVN archive accessible in an interactive way, allowing the user to request specific products from the densely sampled data in the archive, selecting in frequency, *uv*-coverage or phase centre.

1.1.1 Calibration enhancements

Besides the planned upgrade of archive tools, three work packages are concentrating on improving the calibration of radio interferometry. The first has focused on the data product from the EVN. It has enabled system temperature data to be appended directly to the data product and has improved the so-called van Vleck correction. In addition, there is now also software to deal with phase-cal signals in the EVN MkIV data processor at JIVE. These tones can now be detected by the correlator hardware and be used to align the phases in independent sub-bands (*Figure 1*).

There is also an effort at JIVE to improve the ionospheric calibration with a specific aim to improve the EVN phase-referencing performance at 18cm. The project is currently (2006) in its scientific evaluation phase, in which different methods for estimating the total electron content are being compared. The methods that are being tested include estimates from local GPS measurements, as well as global distributions, often augmented with theoretical models. It is not trivial to feed these estimates into the currently used VLBI calibration schemes. This has, however, successfully been implemented in ParselTongue (see below). Some success with improved ionospheric calibration can be reported, but it appears difficult to find a method that leads to robust improvements in all cases.

A similar project at MPI Bonn focuses on the troposphere, using initially local water vapour radiometers [8].

1.1.2 Massive data processing

Three work packages handle the data explosion from new telescopes. Clearly it is necessary to use parallel computing to process the data products on a timescale that allows interaction with the scientific user. However, the image processing tasks in radio astronomy are very data intensive, with few operations per byte, making the usual message passing systems very inefficient for the distribution of the data. This field was researched at ASTRON (<http://www.radionet-eu.org/rnwiki/EvaluationParallelization>) and it has been concluded that such problems are best addressed by ‘task’ parallelization. Such an approach follows the data structure closely and distributes the workload and data to individual nodes. A high level scripting environment could be used to create the workflow for such approaches.

The specific problem associated with wide-band imaging is being addressed in Jodrell Bank. Following the Sault-Wieringa [9] wide-band cleaning algorithm approach, data is being processed, solving for the first and second order derivative of the spectral index of each source. Wide-field imaging is being addressed at JIVE; several aspect of the problem have been reported in this conference [1][5].

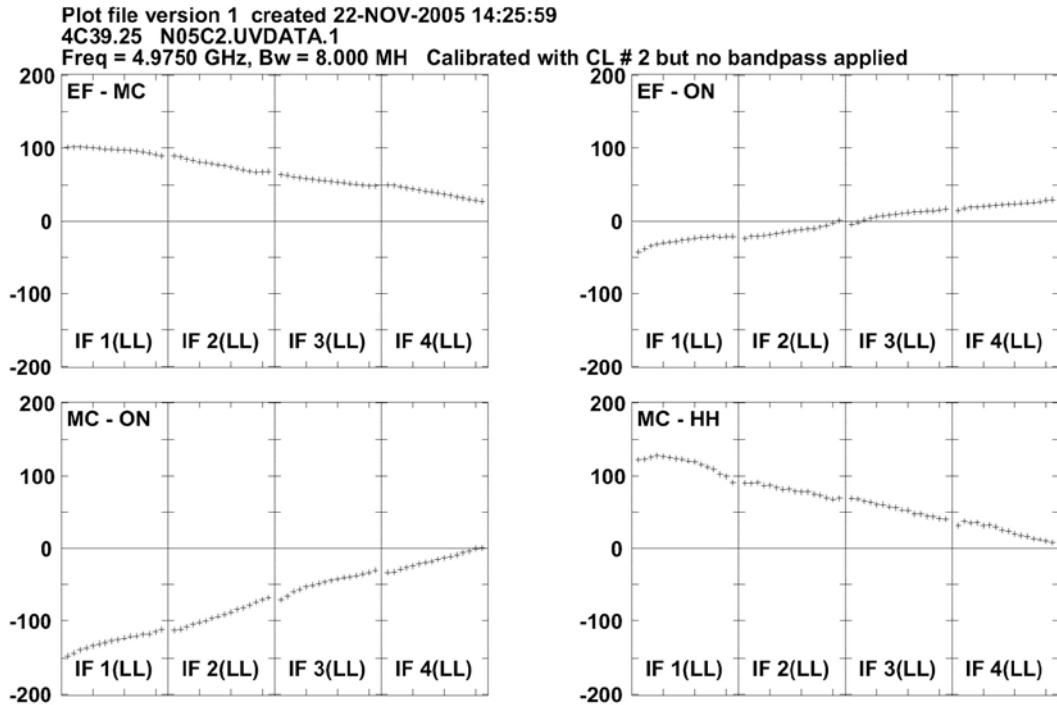
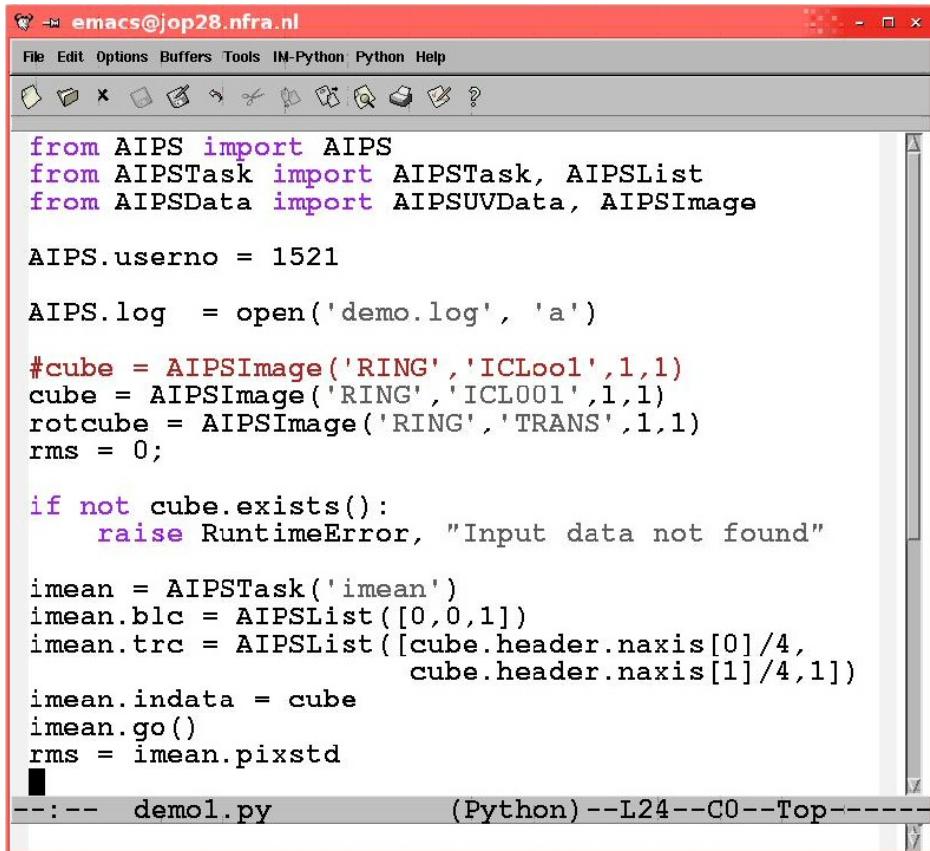


Figure 1; Example of phase-tone calibration; after application of the measured phase values the phases align across different bands.

1.1.3 ParselTongue

From the very beginning of the ALBUS project, the coding environment of this software project has been a concern. It was decided in 2005 to develop a Python interface to classic AIPS. As AIPS is the only package that deals with VLBI calibration, it has a large user base in our community. It was felt that this tool, named ParselTongue [4], could provide us with tools to access the calibration information and improve it, and at the same time produce a high-level scripting tool for distributed computing, following task parallelization.

Moreover, it would serve as a means to distribute the ALBUS software to the user domain. The implementation of ParselTongue proved to be relatively simple, thanks to the existing effort in Obit (<http://www.cv.nrao.edu/~bcotton/Obit.html>), which allows one to access data structures in AIPS catalogues in a modern computer language. This provides direct bindings to the data structures and calibration tables. Another feature is the ability to start existing AIPS tasks from the Python environment (*Figure 2*).



```

emacs@jop28.nfra.nl
File Edit Options Buffers Tools IM-Python Python Help
from AIPS import AIPS
from AIPSTask import AIPSTask, AIPSLList
from AIPSData import AIPSUVData, AIPSIImage

AIPS.userno = 1521

AIPS.log = open('demo.log', 'a')

#cube = AIPSIImage('RING','ICL001',1,1)
cube = AIPSIImage('RING','ICL001',1,1)
rotcube = AIPSIImage('RING','TRANS',1,1)
rms = 0;

if not cube.exists():
    raise RuntimeError, "Input data not found"

imean = AIPSTask('imean')
imean.blc = AIPSLList([0,0,1])
imean.trc = AIPSLList([cube.header.naxis[0]/4,
                      cube.header.naxis[1]/4,1])
imean.indata = cube
imean.go()
rms = imean.pixstd

```

--:-- demo1.py (Python) --L24--C0--Top-----

Figure 2; Example of a very small ParselTongue script. This example determines the rms noise in the first quadrant of the first plane in an image cube.

Not only does this serve the specific needs of the ALBUS project, but it also provides any interested user with a modern AIPS scripting environment. Besides the ability to automate large data reduction steps, and the direct access to the calibration, this has also proven extremely useful for controlling tasks that need to connect to the outside world, for example archives or web sites. There is now a significant user base for ParselTongue, as is demonstrated by its use in several conference contributions [1][3][5][7].

ParselTongue is available from <http://www.radionet-eu.org/rnwiki/ParselTongue>.

1.2 Future of data processing software

Recently, discussions have begun on the question whether a follow-up of these activities should be considered. The future of cm radio data reduction depends on the success of CASA (formerly aips++), which is now focused on the ALMA data handling. The system being developed for LOFAR is based on similar libraries. Moreover, both systems concentrate on pipeline processing and delivering calibrated products and images.

It is felt that, even when these systems mature, there will be a need to be able to do calibration and special processing in the user domain. The user acceptance that current software has gained will be hard to defeat. It should be researched how much of the currently

implemented algorithms can be ported to the new software platforms and what level of interoperability can be achieved among these packages. There is a large role for Python as a common user interface in this scheme, as it has been adopted by almost all packages.

2. Distributed correlation

As the capacity of off-the-shelf computing continues to grow, various groups are looking at software correlation (e.g. [2]). Software correlation offers advantages in flexibility and scalability over hardware correlation. Although it is not necessarily the optimal solution for the next generation EVN correlator, this option is being researched as part of FABRIC. This is a research activity in the EC-funded project EXPReS, which has as a primary goal to turn e-VLBI into an operational instrument.

2.1 The FABRIC project

FABRIC looks beyond connecting the telescopes with existing Mk5 data acquisition systems to the EVN data processor that has sixteen 1 Gb/s inputs [10]. It concentrates on two parts; the first is upgrading the data acquisition and transport means beyond 1 Gb/s. This is carried out in conjunction with e-MERLIN and LOFAR, in order to study the link between private and public networks. The second part of the project focuses on distributed correlation, as a means to find the computing power necessary to implement VLBI correlation in software.

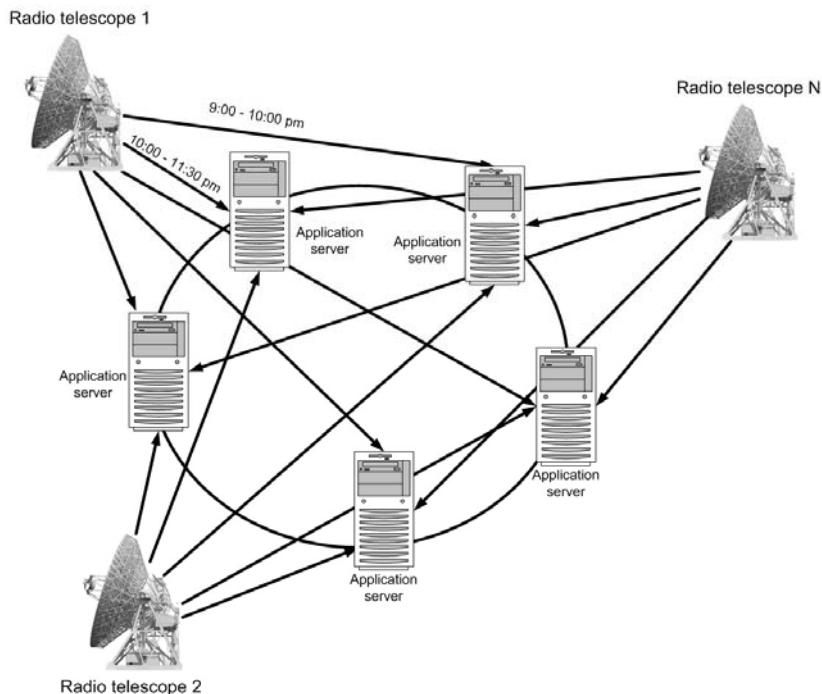


Figure 3; Schematic view of distributed correlation

There are advantages for doing correlation on standard floating point processors rather than special purpose 2-bit hardware. The correlation on such "software correlators" can be more precise, yielding a better signal-to-noise. It can also be more flexible, arriving at arbitrary resolution for parts of the spectrum. This has recently been shown by JIVE scientists tracking the descent of the Huygens probe on Titan with VLBI [6].

The deliverable for FABRIC is the ability to correlate a small-scale (5 telescope) project with limited bandwidth. Although this will not suffice to replace the operational EVN correlator, there may be astronomical applications for such a network. Because the software correlator can be much more flexible than the current hardware, one can consider extreme spectral resolution or pulsar gating. Another application is a continuously available VLBI network for monitoring purposes.

2.2 Options for Grid correlation

The basic scheme being pursued is shown in Figure 3. Rather than all data being collected in a single node, the data is being packaged and sent to distributed compute nodes. In principle, there are many topologies possible to route the data from telescopes to compute nodes. One could organise fixed baselines to be processed on certain nodes, or slice the data in frequency. Analysis shows that time slicing is in fact preferred as it minimizes the internet traffic and is the most easily scalable.

This distributed correlation scheme fits relatively nicely with the concept of the Grid. The Grid allows users to find flops (floating point operations) for their (scientific) problems, by offering access to cluster computers at, for example, university computer centres. Authentication protocols and queuing mechanisms are being developed for the Grid, as well as standard interfaces for distributing software and data.

Within FABRIC, the Poznań Supercomputer and Networking Center (PSNC, Poland) brings the expertise to work with Grid computing. Work has started to encapsulate the software correlator developed at JIVE into the Virtual Lab environment (*Figure 4*).

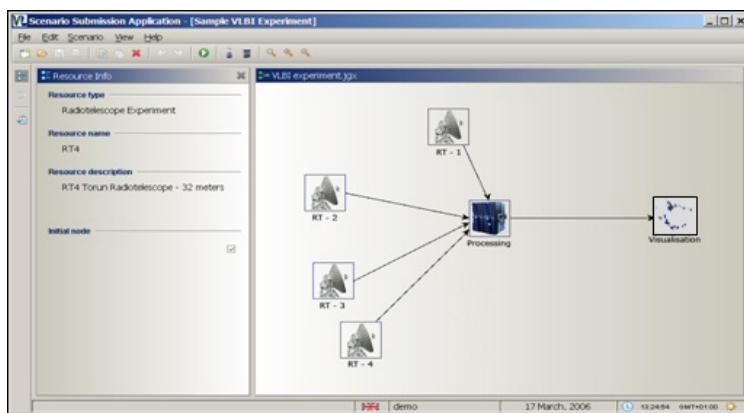


Figure 4; An example of how a workflow manager for e-VLBI with distributed correlation may look.

Acknowledgement

ALBUS is part of the RadioNet project. This work has benefited from research funding from the European Community's sixth Framework Programme under RadioNet R113CT 2003 5058187. FABRIC is part of the EXPReS project. EXPReS is partially funded by the European Commission (IST FP6 e-Infrastructure initiative – Communication Network Development) under contract number 026642 EXPReS.

References

- [1] S. Bourke, van Langevelde H.J., *Analysing very large datasets with ParselTongue*, in proceedings of the 8th EVN Symposium, PoS(8thEVN)066.
- [2] A. Deller, *Distributed FX software correlation for e-VLBI*, in proceedings of the 8th EVN Symposium, PoS(8thEVN)057.
- [3] D. Fenech, *Supernova remnants in the central starburst region of M82*, in proceedings of the 8th EVN Symposium, PoS(8thEVN)032.
- [4] M. Kettenis, van Langevelde H.J., C. Reynolds, B. Cotton, *ParselTongue: AIPS Talking Python*, in proceedings of ADASS XV, 2006 ASP Conference Series, Vol. 351 p497
- [5] E. Lenc, *Living life on the edge – wide-field VLBI at 90 cm!*, in proceedings of the 8th EVN Symposium, PoS(8thEVN)079.
- [6] S. Pogrebenko et al., in Proceedings of *Planetary Probe Atmospheric Entry and Descent Trajectory Analysis and Science*, 2004, ESA SP-544, Noordwijk, p197
- [7] A. Richards, *On-demand radio imaging and access to calibrated data for all astronomer*, in proceedings of the 8th EVN Symposium, PoS(8thEVN)056.
- [8] A. Roy, *Phase correction of VLBI with water vapour radiometry*, in proceedings of the 8th EVN Symposium, PoS(8thEVN)058.
- [9] R.J. Sault, M.H. Wieringa, *Multi-frequency synthesis techniques in radio interferometric imaging*, 1994 A&AS 108, 585
- [10] R.T. Schilizzi, W. Aldrich, B. Anderson, A. Bos, R.M. Campbell, J. Canaris, R. Capallo, J.L. Casse, A. Cattani, J. Goodman, H.J. van Langevelde, A. Maccafferri, R. Millenaar, R.G. Noble, F. Olnon, S.M. Parsley, C. Phillips, S.V. Pogrebenko, D. Smythe, A. Szomoru, H. Verkouter, A. Whitney, "The EVN MarkIV VLBI Data Processor", 2001, Experimental Astronomy 12, 49-67