The Low-Frequency Radio Universe
ASP Conference Series, Vol. 407, © 2009
D. J. Saikia, D. A. Green, Y. Gupta, and T. Venturi, eds.

Calibrating LOFAR using the Black Board Selfcal System

V. N. Pandey,[1,2] J. E. van Zwieten,[3,2] A. G. de Bruyn,[2,1] and R. Nijboer[2]

[1]*Kapteyn Astronomical Institute, Groningen, The Netherlands*
[2]*ASTRON, Dwingeloo, The Netherlands*
[3]*Science & Technology B.V., Delft, The Netherlands*

Abstract. The Black Board SelfCal (BBS) system is designed as the final processing system to carry out the calibration of LOFAR in an efficient way[1]. In this paper we give a brief description of its architectural and software design including its distributed computing approach. A confusion limited deep all sky image (from 38−62 MHz) by calibrating LOFAR test data with the BBS suite is shown as a sample result. The present status and future directions of development of BBS suite are also touched upon.

1. Introduction

LOFAR (Low Frequency Array, see http://www.lofar.org) is first of a new generation of innovative large low frequency radio telescopes under construction in the Netherlands with stations to be distributed over several European countries (Falcke, H. D. et al. 2007). It is a low frequency telescope which observes in two frequency ranges 20−90 MHz (Low Band Antennae-LBA) and 110−240 MHz (High Band Antennae-HBA). The Stations consist of phased array of orthogonal dipole pairs which allow simultaneous observations in multiple directions (beams) across wide parts of the sky. An IBM Blue Gene/P supercomputer works as the software correlator for unprecedented amount of sampled data.

Calibrating LOFAR requires '3rd generation calibration' which can deal with instrumental effects which vary not only with time and frequency but also viewing direction. This challenge is compounded by increased RFI, troublesome ionospheric behaviour (refraction and Faraday rotation), wide field of view, complicated instrumental polarization as well as formidable processing requirements due to large streams of data. One of the primary objectives of LOFAR includes attempts to image diffuse redshifted HI from the epoch of reionization (EoR), which makes the calibration requirements exceptionally stringent. Thus a robust calibration framework is required for which BBS system has been designed. In this short paper, we briefly describe software design aspects of BBS, a recent sample result and ongoing as well as future developments under progress.

2. BlackBoard System Overview

The Black Board Selfcal (BBS) system derives its name from the black board design pattern (Buschmann et al. 1996). The central concept of this pattern is a pool of independent processes that operate on shared memory (the *black board*). These processes do not call each other directly and there is no pre-determined sequence in which they are activated. Instead, a central control process examines the black board and decides what is to be done next depending on the current state.

[1]Although BBS is mainly developed for LOFAR, it may also be used to calibrate other instruments once their specific algorithms are plugged in.

The current implementation of BBS does not strictly adhere to the black board pattern: Some information *is* exchanged between processes directly because it is not necessary to share it amongst *all* processes, or simply because the amount of information is too large to justify routing it via shared memory.

2.1. Components

The BBS system is split into three components: `control`, `kernel`, and `solver`. The `control` component is responsible for coordinating the execution of the calibration *strategy*. A strategy is an ordered list of processing steps that need to be executed to transform raw visibility data into calibrated visibility data.

The `kernel` component operates on visibility data. It implements the *full measurement equation including polarization* Hamaker et al. (1996) which models the response of an interferometer given a description of the sky, the environment, and the interferometer. Using the measurement equation the `kernel` can simulate visibilities, subtract sources, correct visibilities for a given reference direction, and solve for model parameters.

The `solver` component is used when the visibility data of several subbands has to be combined to fit model parameters. A `solver` process co-operates with a group of `kernel` processes (a *calibration group*). It receives sets of equations generated by the `kernel` processes, which it merges and solves. The updated values for the model parameters are sent back to the `kernel` processes. This is an iterative process that continues until a stop criterion is met.

2.2. Control

The calibration strategy is executed such that disk I/O is minimized. Each subband is processed in chunks, because a single subband is still too large to be kept in main memory. After a chunk has been loaded into memory, all the steps of the calibration strategy are executed on the chunk before the next chunk is read. Thus, the raw visibility data will be read only once and the calibrated visibility data will be written only once at the end of the strategy.

2.3. Communication

The communication paths used during a typical calibration run is organised as shown in Figure 1. As mentioned earlier, the visibility data is split along the frequency axis into separate *subbands*. Each subband is stored on a local disk of one of the offline processing nodes and is processed by a separate `kernel` process. Groups of `kernel` processes are connected to `solver` processes. All processes are connected to the black board. The `control` process examines the black board and post commands for the other processes to execute.

The black board is essentially used as *shared memory* that retains the state of the distributed calibration process. It is implemented as a relational database. This allows us to take advantage of the features of the database management system, e.g. support for concurrent access, locking, and transactions. Futhermore, it allows easy monitoring of the current state by external tools.

A potential risk of the design is that the black board becomes a bottleneck, because all processes need to access it. Moreover, access will typically be (quasi-) concurrent, because all `kernel` processes perform the same operations on approximately equal amounts of data. This is why only low volume information (the control state) is kept on the black board. High volume information is exchanged between the processes directly. A related concern is that processes may have to poll the black board to check for updates. This depends on the implementation of the black board component. Most modern database management systems, for instance, support asynchronous notification of client processes thus avoiding the need for polling.

Most of the processing can be performed on each subband independently. This is desirable because it avoids the need for communication between `kernel` and `solver` processes. Should it prove necessary to consider the visibility data from multiple subbands together, e.g. when using a weak calibrator source, then the set of `kernel` processes can be partitioned into several calibration groups.

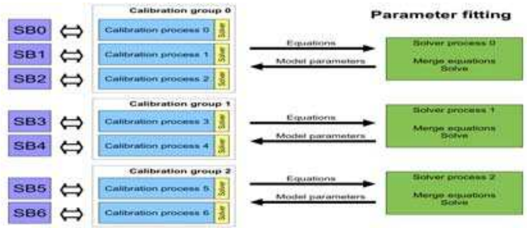


Figure 1. Schematic to show working calibration implementation including when solving for calibration parameters for which data across many subbands has to be considered together (e.g. solving for Global bandpass or calibration using weak sources). Multiple calibration groups, each comprising of the subbands required for its calibration are formed and calibrated independently in a distributed way.

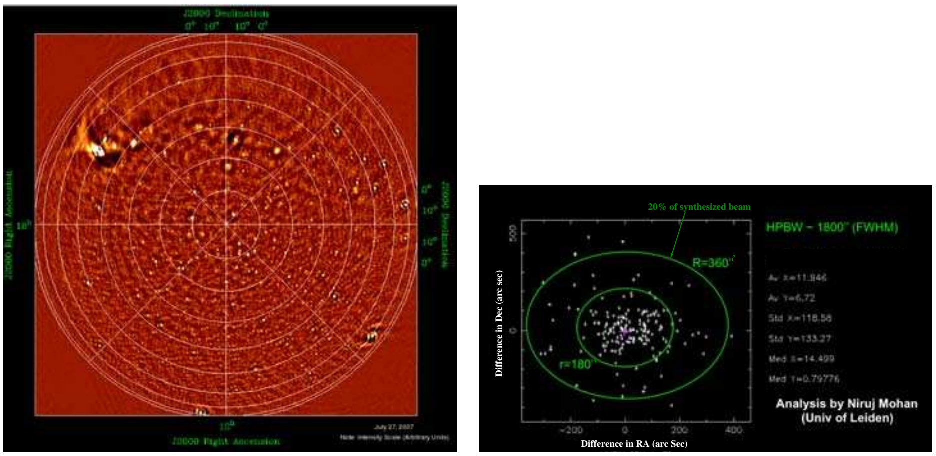
All the visibility data within a single calibration group will be considered together, independent of other calibration groups. Each `kernel` process within a calibration group generates a set of equations that encodes the difference between the observed visibilities and the simulated visibilities that were computed using the measurement equation. Each `kernel` process sends its set of equations to the calibration groups' designated `solver` process. The `solver` process merges the different sets and computes a solution. The updated values of the model parameters are then send to all the `kernel` processes in the calibration group.

In the extreme case where all `kernel` processes form a single calibration group, the `solver` process can clearly become a bottleneck. It remains to be seen how often this case will occur.

3. Recent Results

The data stream from the stations is fed into the super computer Blue Gene-P which carries out various operations like reception, transpose, correlation, beam forming, de-dispersion etc. during the online processing. The visibility data is stored on the short term storage section for offline processing. The main steps during offline processing include Pre-Processing, Calibration, Imaging and Source extraction. The Pre-Processing is carried out by software suite called Default Pre-Processing Pipeline (DPPP). DPPP Carries out all pre-Calibration common default tasks which include RFI detection and mitigation, determination and correction for phases due to different clocks and cable lengths at stations, correcting for the band-shape, compression of data along time and frequency as required/specified, all in an efficient way. This is followed by calibration by BBS system in a distributed way. After this Imaging is carried out by Master Worker Imager (MWI), also in a distributed way. This may be followed by Source extraction or other steps depending upon the specific requirements of processing under consideration.

As a sample result an all sky image is shown in Figure 2(a) produced with data from the test station CS-1 (Core station 1). These are first of a series of initial deep all sky wide field (full hemisphere centered on North Celestial Pole (NCP)) dirty LOFAR/CS1 images with about $30'$ resolution, showing more than 500 radio sources and produced by calibrating the data with the Black Board Self-calibration System. The image is made using a day of observations carried out in 36 sub-bands (each 0.15 MHz wide, total bandwidth 5 MHz) in the frequency range 38−62 MHz with 16 single LBA dipole pairs phased at NCP and spread to have a maximum baseline of 485 m. The data processing steps include initial flagging, solving simultaneously for the complex gains (J Jones) in the direction of the two brightest sources Cas A and Cygnus A. The contribution of these two sources were subsequently subtracted from the visibilities and the residuals were corrected for the complex gain in the direction of Cas A. Subsequently AIPS++ imager was used (with the W-projection algorithm) to transform the calibrated visibilities into



(a) A deep all sky wide field (full hemisphere centered on NCP) dirty LOFAR/CS1 images with about 30' resolution, showing more than 500 radio sources. The sidelobes of PSF of strong sources like Tycho, Taurus, Virgo etc. are evident. The Sun present in the right bottom corner can also be easily noticed. The intensity color scale for the images is in arbitrary units.

(b) Difference in RA and DEC positions of radio sources detected in two LOFAR images on different days. The mean offsets in both directions is less than 1% of the FWHM of synthesized beam.

Figure 2. All sky image, radio sources and their positional accuracy.

all sky images. The images have not been corrected for the primary beam of the individual dipoles. During the entire data processing only the projected baselines longer than 70 m were used. The positional accuracy of radio sources on two different days is shown in Figure 2(b). Their positions have also been independently confirmed by comparing with 3CR, 4C, and NVSS catalogues. Similar results for the HBA data have also been obtained.

A few approximations during processing include averaging the visibilities within each sub-band before calibration and assuming same flux densities for two point source sky model consisting of Cas A and Cygnus A over the entire frequency range. The image shows that the source Cas A has been subtracted quite satisfactorily but Cygnus A residuals are still there to a significant level (about 1% of original). This is most likely due to the fact that we also used data when Cygnus A is very close to the horizon where its signal to noise ratio becomes very low for accurate simultaneous calibration in two directions. This residual is also the most probable reason for apparently slightly less number of sources seen in the top half of the image compared to the bottom half.

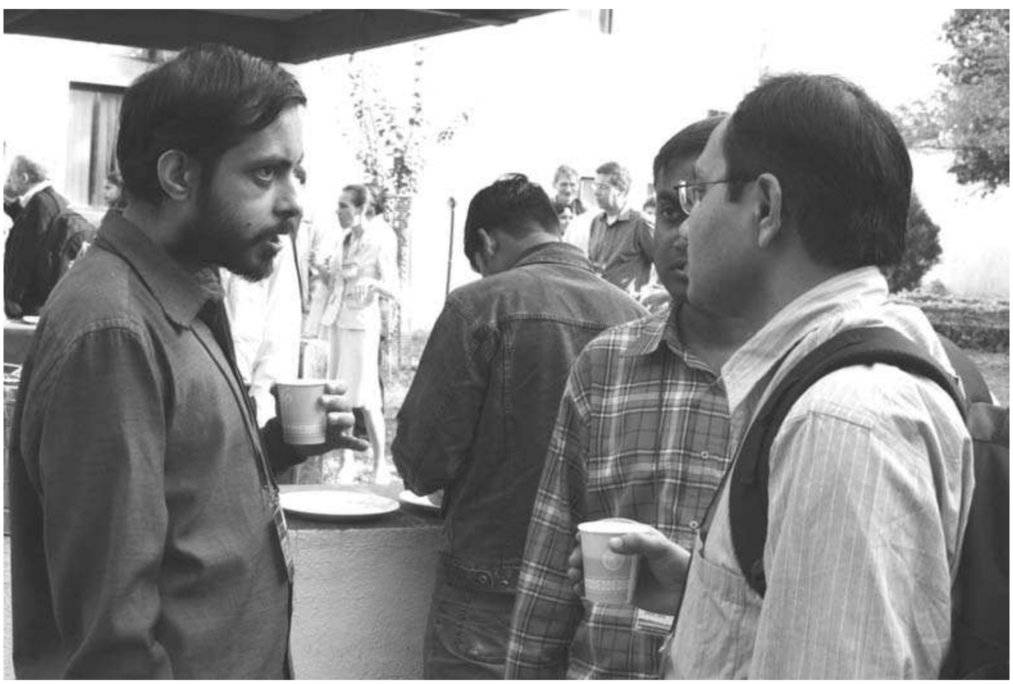
4. Conclusions

BBS has been successfully used to calibrate LOFAR data in a distributed way. Deep all sky images with good positional accuracies of radio sources reveal the success of the entire LOFAR processing pipeline. The BBS system is being further developed including addition and testing of various new functionalities like instrumental beams, ionospheric calibration, source models etc.. There is still much to be learned regarding its scaling and performance on bigger data sets and multi-core processors. Once a reasonable number of LOFAR stations are up and running as expected in the next few months, we will further improve on our understanding of how to optimally calibrate LOFAR data. This may and probably will yield new insights that will have an impact on the system design. Optimization and an evolving system design are difficult to be carried out together and thus may take a couple of cycles.

Acknowledgments. We would like to thank various people involved in the LOFAR project including people in the offline processing team for their help and contributions.

References

Buschmann, F., Meunier, R., Rohnert, H., Sommerlad, P., & Stal, M.. 1996, Pattern-Oriented Software Architecture, (New-York: John Wiley & Sons)
Falcke, et al. 2007, Highlights of Astronomy, 14, 386
Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&AS, 117, 137



Subhashis Roy, Amitesh Omar and Vishambar Pandey