LOFAR: THE FIRST OF A NEW GENERATION OF RADIO TELESCOPES

Marco de Vos (devos@astron.nl)

ASTRON, P.O. Box 2, 7990 AA Dwingeloo, The Netherlands

ABSTRACT

The Low Frequency Array (LOFAR) opens a previously largely unexplored frequency domain for challenging radio-astronomical research. At the 10 to 240 MHz operating frequencies of this radio telescope it is feasible to employ very large numbers of simple, all-sky antennas with wide-band early digitization. This means that almost the full signal processing chain can be realized in (embedded) software. This approach makes it possible to deal with earth-based radio signals in effective and novel ways. The signal processing challenges in LOFAR are manifold, since the ultimate dynamic range in astronomical images depends on the quality of the full chain of operations that combines tenthousands of antenna signals into a single multi-channel image cube, while correcting for a large variety of instrumental and environmental effects.

1. INTRODUCTION

The conventional design paradigm for radio telescopes employs large parabolic dish antennas with special, often cooled, front-end receivers. RF signals are usually transported over relatively long distances to a central location where signals are digitized and further processed. This paradigm does not scale well towards huge collecting areas. The scientific drivers for a new generation of radio telescopes argue for roughly two orders of magnitude increased sensitivity which cannot be obtained by conventional large parabolic dishes at any realistic cost. To reduce total cost as well as to optimize flexibility it is unavoidable to move to designs consisting of large numbers of relatively small antennas, where most signal processing functions are realized in digital electronics and software[1].

This approach still allows for the use of a large number of small parabolic dishes, optionally with early digitization right after the front-end. It should be noted that early digitization is also being integrated in existing radio telescopes, in particular in the Expanded VLA[2], thereby allowing high performance signal transport over optical fiber and significantly increasing signal processing flexibility.

At the lower radio frequencies, it becomes increasingly attractive to replace the parabolic dishes by phased arrays of omni-directional antennas, thereby replacing the beamforming function of the reflector with an electronic equivalent. Depending on the frequency range, beamforming can be done hierarchically in both the RF and digital domains or entirely in the digital domain, bringing the majority of the signal conditioning and handling in the realm of digital signal processing.

This paradigm shift has many important implications that go far beyond the initial cost issue. Foremost, the use of robust wideband receivers in combination with powerful and flexible signal processing techniques seems to be the best approach to handle the ever-increasing levels of man-made interference (RFI). Furthermore, the use of omni-directional antennas allows for instantaneous imaging of very wide fields. This offers great opportunities for the study of variable phenomena on the sky, and for the detection of transient events occurring at very short timescales such as cosmic rays.

LOFAR is the first full-scale instrument that has been designed according to this new design paradigm. The LOFAR band is covered by two antenna types (10-80 MHz and 120-240 MHz respectively, avoiding the FM band). LOFAR will consist of 77 stations each with 100 dual-polarized antennas of both types. Antenna signals are combined in phased array mode, forming beams at one or many directions on the sky. Signals are transported over a glass fiber network to central processing systems, where the phased array beams can be combined in several ways. For imaging, stations are combined through correlation, thus forming an aperture synthesis array.

The combination of very large datarates, high levels of man-made interference and severe phase distortions due to ionospheric turbulence puts extreme challenges on the signal processing and calibration of LOFAR. The very large datarates force LOFAR to calibrate data in a streaming processing model. This requires new programming models, an increased focus on signal integrity and the introduction of system health management concepts.

The paradigm shift has been extended further by positioning LOFAR as a Wide Area Sensor Network, allowing the processing concepts and infrastructure to be reused by other applications. Geophysical sensors and sensors for research in precision agriculture have been added to the RF-sensors for astronomy. LOFAR as a multi-disciplinary research infrastructure has been funded by the Dutch government and is scheduled for completion in 2008, with initial operations starting in 2006.

The remainder of this article summarizes the scientific justification for the instrument and gives an overview of the system architecture, identifying the main signal processing challenges.

2. KEY SCIENCE PROJECTS

LOFAR is a new radio telescope that will open a high-resolution window on the Universe at frequencies just above the ionospheric cutoff. It will operate from 10 MHz to 240 MHz and provide sensitivities and angular resolutions that are at least two orders of magnitude greater than previous instruments. LOFAR will survey the sky for the first time at these frequencies at high angular resolution. In addition, it will be optimized for the detection and study of transient radio signals from cosmic sources and will make a major effort to observe signatures from the epoch of cosmic re-ionization. Solar and space plasma physics as well as upper atmosphere research will be important application areas, while the study of gamma ray, very high energy cosmic ray and neutrino events in the Earth‘s atmosphere may provide important complementary information to satellite-based instruments and ground-based optical Çerenkov detector arrays.

The design of LOFAR is driven by four fundamental astrophysical drivers or Key Science Projects[3]:

- The search for redshifted 21cm line emission from the Epoch of Reionization (EoR), a very early phase in the evolution of the universe;
- Extragalactic surveys and their exploitation to study the formation and evolution of clusters, galaxies and black holes;
- Transient sources and their association with high energy objects such as gamma ray bursts;
- Cosmic ray showers and their exploitation to study the origin of ultra-high energy cosmic rays.

The Key Science Projects differ significantly in their signal processing requirements.

The large extragalactic surveys operate LOFAR in much the same way as existing imaging instruments, combining phased array stations in an aperture synthesis array. Multi-beaming can be used to speed up the survey, but is not essential for its success. Since datarates are an order of magnitude higher than from current instruments, unattended processing and calibration is necessary requiring advanced quality monitoring. Only final dataproducts need to be exported and archived.

The search for the EoR requires the instantaneous imaging of a large fraction of the sky over a large band. This requires a large amount of processing resources for the correlation in the central processing systems and in fact determined the dimension of the central processor in terms of processing power and bandwidth. The expected signature of the EoR is extremely weak. Detection in the presence of overwhelming foreground emission from both diffuse galactic structures and extra-galactic sources will be extremely challenging. Careful control of the data processing techniques and the ability to reprocess the original data with improved algorithms and models is therefore essential. This means that in contrast to the standard procedure, original data will be stored for a longer period of time (~ 2PByte).

The study of transient sources has completely different requirements. Here the key issue is to monitor as large a fraction of the sky for changes that have an astronomical origin. One possible approach is to generate and compare series of all-sky snapshot images. Since snapshots have low signal to noise, they require reliable algorithms separating man-made interference from celestial events. Alternatively, LOFAR can be configured as a large phased array, covering the sky with a large collection of pencil beams. This “Tied Array” mode is also useful to observe time-variable objects like pulsars. The analysis of the data thus obtained often requires going back to the original time domain data. This mode therefore relies on linear time-invariant signal processing, putting strong constraints on especially the first digital filters in the system.

LOFAR offers unique capabilities to follow-up on external detections. The instrument can be pointed to any position of the visible sky on timescales of 10s (very small compared to the time needed to reposition a mechanical dish). The transient detection capabilities have been further extended by buffering raw antenna data for a brief period of time (1…10 s depending on the buffered bandwidth). An internal or external trigger can be used to freeze these buffers, after which they can be read out for further processing, pointing e.g. to a different position on the sky than the telescope was observing at when the trigger event occurred.

The data buffers are also used for the detection of cosmic ray showers[4]. This detection can run entirely in parallel to other observations. Embedded processors close to the buffers continuously monitor the antenna signals. When an event is detected in one or several antenna signals, all buffers in the station are frozen and the data corresponding to the time of the event (~ms) are transported to the central processing systems for analysis and archiving.

3. LOFAR TOP-LEVEL ARCHITECTURE

The LOFAR top-level architecture has been defined in the system level Architectural Design Document[5] and was adjusted to the new boundary conditions implied by funding in a Phase 1 Baseline definition[6] (both available from www.lofar.org). The Phase 1 baseline geographically consists of a Compact Core area and 45 Remote Stations. In the Core area, with 2km diameter, there will be 32 stations. The Table below gives an overview of the installed sensors and datarates and processing power.

The Remote Stations are configured in a scale-free distribution, consisting of nine rings with increasing distances. This configuration combines the properties of good snapshot capabilities and optimal full spatial coverage. The maximum distance between stations is approximately 100 km, resulting in a 6 arcsec FWHM beamsize at 150 MHz.

LOFAR	For each Sensor Field	Phase 1 2008
Sensor Fields (including Central Core)		77
Sensors		
"LF" antennas (10-80 MHz)	100	7700
"HF" antenna tiles (120-240 MHz)	100	7700
Geophysical vibration sensors (geophones)	18	1386
Geophysical microbarometers (infrasound)	3	231
Other sensors	8	616
Datarates		
Total digitized datarate from sensors	0.5 Tb/s	37 Tb/s
Datarate over LOFAR Backbone	10 Gb/s	0.8 Tb/s
Outgoing datarate over SURFnet6		40 Gb/s
European datarate over Géant		20 Gb/s
Installed Processing Power		
Total processing power		160 Tops/s
Distributed at Sensor Fields	1.5 Tops/s	116 Tops/s
Central Processor (including BlueGene)		43 Tflop/s
BlueGene		33 Tflop/s

5.1. Station processing systems

LOFAR uses specially developed integrated receivers and digital processing systems[7]. For the receiver a wideband direct digital conversion architecture is adopted. This reduces the number of analog devices in the signal path. The A/D converter converts the analog signal into a 12 bit digital signal at a sampling rate of 200 MHz. Depending on the required band, different Nyquist zones are used. To allow for more flexibility in the selection of frequency bands, a sample frequency of 160 MHz can be chosen as well. No mixers are required, greatly simplifying the design and improving the overall performance. Robustness against man-made interference has been a main design driver for the receiver. Several integrated filter stages are implemented to suppress out of band noise contributions. The 12-bit digitization is matched to the interference environment above 30 MHz. Below 30 MHz observing is still possible, but only for a limited fraction of the day.

After A/D conversion, the band is split into 512 equidistant subbands. The negative part of the original spectrum is omitted, i.e. the real input signal is from here on represented by complex signals. Each subband signal is decimated with a factor of 1024 after filtering. This first filter is a critical element in the processing chain and has been highly optimized[7,8].

Selected subbands are combined in a distributed beamformer. The number of subbands, and thus the maximum subband width has been matched to the beamsquint criterion. Independent beamformers are used for each subband, such that either a broad band is covered for a single direction on the sky, or a larger number of directions are observed at a reduced bandwidth. Independently, the digitized signal are be buffered for ~1 sec for Cosmic Ray detection and Transient Processing.

Each Remote Station delivers a single dual polarization beam at 32 MHz, or 8 dual polarization beams at 4 MHz or any combination in between. The resulting output data rate is 2 Gb/s. Since 0.5 Tb/s is generated digitally, the station level processing reduces the datarate by a factor of 250. Each station has 1.5 Top/s in embedded processing, mainly in FPGAs and by far dominated by the subband filter.

Data are framed and time-tagged based on GPS receivers and rubidium clocks at station level. The combination of a GPS receiver and a rubidium clock allows for <1ns time accuracy at station level. This allows maintaining synchronization within the system even if data transport is not fully deterministic.

5.2. Central processing systems

At the central site, a massive data processing facility is needed for the coherent processing of the signals from all antennas[9]. Here the aperture array is synthesized and data can be processed for multiple scientific applications simultaneously. The aim of the central processor architecture is to provide first of all enough processing power, data transport bandwidth and storage capacity to allow for continuous operation of the observation facility. Secondly, the central processor will be built as a signal-processing cluster aimed at multiple types of data processing.

The total input data rate to the central processor is approximately 256 Gb/s. The input data have to be buffered, creating a sliding window of 30 seconds worth of historical data that can also be used for specific analysis of transient events. The complete data flow has to be routed towards on-line processing pipelines that perform observation specific data transformations on the streams. These routing implements an all-to-all

communication pattern on the full 256 Gb/s data stream programmed in software to maximize flexibility.
For imaging applications, all correlations for all station-station combinations have to be calculated within the data stream, increasing the data stream to 10 Tb/s. This increase in datarate is compensated by a time averaging operation. The integrated data is stored temporarily (order days) for further processing, in particular calibration (see 5.3). The total processing power needed for the on-line stream processing is in the order of 30 Tflop/s. Another 10 Tflop/s is available for other processing tasks.

The central processor has been designed as a hybrid cluster computer. In the core of the processor is an IBM BlueGene/L system on which critical processing tasks like correlation are performed. Research is ongoing to implement the streaming processing model on the BG/L system, extending its capabilities to a broader range of applications[10].

5.3. Distributed end-user processing

Typical end-user processing tasks are removal of interference, calibration and formation of final dataproducts. This standard approach no longer holds for LOFAR, the challenges for end-user processing being twofold. Foremost new concepts and algorithms are needed to remove the distortions of a pathological ionosphere[11] and severe man-made interference[12]. Secondly, these algorithms have to be implemented in a robust streaming data model[9], given the high datarates involved. LOFAR has adopted a processing model in which final data products for routine observations are formed in a highly automated mode with user interaction at well-defined moments. Later processing steps can take place at the central processor or at local resources, in a Grid-like processing model.

6. CONCLUDING REMARKS

From a signal processing point of view, LOFAR can be seen as an extreme example of a classical data stream processing system [14]. The condition that the data in the stream can be accessed only a limited number of times is extended beyond the limited-memory case to 250 TB temporary storage of correlated data, giving rise to interesting new research topics. LOFAR is an ideal test bed for alternative signal processing schemes like continuous queries[15], and will be used to advance model based system health management[16].

The performance of the calibration and RFI mitigation algorithms will determine the ultimate success of LOFAR. Experiments with an Initial Test Station consisting of 64 antennas prove that the station-level effects can be calibrated down to the theoretical limit[13]. System level calibratability so far has been confirmed only through simulations and analysis[11,12].

The radio astronomical community has long had a somewhat *ad hoc* approach to signal processing models. Joint research on LOFAR[17] has proven already that a more rigorous and formal signal processing approach is likely to be the key in getting the ultimate performance out of the next generation of radio telescopes.

11. REFERENCES

[1] Bregman, J.D., “Concept Design for a Low Frequency Array”, Proceedings of SPIE conf. 4015, 2000.
[2] Perley, R.A., “The Very Large Array Expansion Project,” in SPIE Proc. 4015, pp 2-7, 2000.
[3] van Haarlem, M.P., *et al.*, Development and Commissioning of LOFAR Astronomy, ASTRON, 2004.
[4] Horneffer, A., Falcke, H., Kampert, K.H., “LOPES - Detecting Radio Emission from Cosmic Ray Air Showers,” in Proc. of the 6th European VLBI Network Symposium, p. 23, MPIfR Bonn, Germany, 2002.
[5] de Vos, C.M., (editor), “LOFAR Architectural Design Document,” LOFAR-ASTRON-ADD-006, v3.1, 2003.
[6] de Vos, C.M., “LOFAR Phase 1 Baseline,” LOFAR-ASTRON-MEM-145, v2.1, 2004.
[7] Gunst, A.W., Kant, G.W.,”LOFAR Remote Station Baseline”, LOFAR-ASTRON-MEM-142, v1.0, 2004.
[8] Alliot, S., Soudani, M., Bregman, J.D., “Comparison of filters with poly-phase structure applied to large embedded systems for telescopes”, in Proc. IEEE Benelux Signal Processing Symposium, 2002.
[9] van der Schaaf, K., Bregman, J.D., de Vos, C.M., “Hybrid Cluster Computing Hardware and Software in the LOFAR Radio Telescope,” in Proc. of the IEEE conference on Parallel and Distributed Processing Techniques and Applications, pp. 695-701, Las Vegas, 2003.
[10] van der Schaaf, K.S., “Correlation, self-calibration and all-sky imaging integrated on the LOFAR processing platform”, Experimental Astronomy, special SKA issue, 2005 *in press*.
[11] Noordam, J.E., “LOFAR Calibration Challenges”, Proc. SPIE 5489, pp. 817-825, Glasgow, 2004.
[12] Boonstra, A.J, “Calibration and interference mitigation for future radio telescopes,” these proceedings, 2005.
[13] Wijnholds, S.J., Bregman, J.D., Boonstra, A.J., “Sky Noise Limited Snapshot Imaging in the Presence of RFI with LOFAR’s Initial Test Station”, Experimental Astronomy, special issue on SKA conference, 2005 *in press*.
[14] Henzinger, M., *et al.,* “Computing on data streams,” in *TR1998-011 Compaq System Research Center*, 1998.
[15] Koparanova, M.G., Risch, T., “High-Performance GRID Stream Database Manager for Scienti.c Data,” in Across Grids 2003, LNCS 2970, pp. 86–92, 2004.
[16] Bos, A. “Model-based health tracking,” in Proc. IEEE Systems Readiness Technology Conference AUTOTESCON, Valley Forge, 2001.
[17] Boonstra, A.J., van der Tol, S., “Spatial filtering of interfering signals at the initial LOFAR phased array test station”, Radio Science, special issue on interference mitigation in radio astronomy, 2005 *in press.*