

# The LOFAR Telescope: System Architecture and Signal Processing

M. de Vos, A.W. Gunst, R. Nijboer

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

## Abstract

The Low Frequency Array (LOFAR) is a large distributed radio telescope, consisting of phased array antenna stations that are combined in an aperture synthesis array. Antenna stations consist of many simple, omni-directional antennas. Flexible station based signal processing allows for trading bandwidth against instantaneous sky coverage. Central processing implements a software correlator, which can be reconfigured as a full tied array beamformer, and on-line calibration functions to handle the large data streams produced by the system. The key science programs for LOFAR challenge the technical specifications in several directions, which resulted in a highly reconfigurable architecture.

This paper describes the LOFAR system design, the configuration and the signal processing chain. LOFAR has been developed by ASTRON and a consortium of universities and industrial partners. The instrument is currently being deployed in the Netherlands. Additional stations are being built in several other European countries. The telescope is considered an important pathfinder for the Square Kilometer Array (SKA) in demonstrating the potential of (sparse) aperture arrays, in developing solutions to major calibration issues that are directly applicable to the SKA, and in paving the way for the mass-production and operations of such large distributed radio telescope systems.

**Keywords** Radio astronomy; Phased Array; Aperture Synthesis; Multi-beaming; Low Frequency astronomy;

## 1. INTRODUCTION

LOFAR was originally conceived as a large survey telescope for the frequencies below  $\sim 100$  MHz. At these frequencies, phased array technologies had been widely used. In order to make the next step in sensitivity and spatial resolution, it became necessary to develop a coherent system that was at least an order of magnitude larger than existing facilities [1]. A major driver for this was the need to accurately calibrate the ionosphere, which otherwise severely limits the dynamic range in images. This driving requirement became even more important with the realization that an extension of LOFAR to higher frequencies (110 – 240 MHz) would make it possible to detect the weak signature of a critical phase in the evolution of the universe: the Epoch of Reionisation (EoR).

With the choice of omni-directional dipole antennas and early digitization, it was soon realized that the instrument would be an extremely powerful all-sky monitor for detecting transient events and pulsars. This stretched the specification of data rates and response times. The capability of the instrument for detecting Ultra-High Energy Cosmic Rays (UHECRs) could be further enhanced by adding local memory and fast local data analysis. This complex set of driving requirements [2], that challenged the technical specifications of the instrument in various directions, could only be accommodated by adopting a highly flexible system design. The basic starting point of the system remained the concept of low-cost, non-moving, omni-directional antenna elements, carefully configured at station and array level, to allow for a cost-optimised data transport and processing architecture.

## 2. SYSTEM LEVEL CONSIDERATIONS

Theoretically, the maximum amount of flexibility is obtained by digitizing the electrical antenna signals at an early stage and transporting them all to a central location for storage and processing. While this is feasible for systems with a few antennas, it soon becomes completely unrealistic for the large number of antennas needed to reach the required sensitivity levels. The cost-performance trade-offs involved are simplified by configuration requirements. Array calibration and the imaging needs of the EoR and all-sky monitoring, all require a dense antenna area in the centre of the array. Theoretical studies [3] and experiments with initial stations proved that it is not necessary to have identical beamshapes for all stations. This led to a natural division between the LOFAR Core (situated near the village of Exloo) and Remote Stations, each independently optimized in terms of number of antennas and data output. The increasing interest from other European institutes made it possible to add long-baseline capabilities to LOFAR. Table 1 summarizes the antenna types and numbers.

Antenna	Core	Remote	European
Low Band Antenna (LBA) 15-80 MHz	96	96	96
High Band Antenna (HBA) 110-240 MHz	48 tiles	48 tiles	96 tiles

**Table 1: Number of LOFAR antennas per station (refer to text for details).**

Freq (MHz)	$\lambda$ (m)	Station Diameter (D) (m)	Station Beam FWHM (deg)	Field of View (sqr.deg.)	Resolution L = 2 km (arcsec)	Resolution L = 10 km (arcsec)	Resolution L = 80 km (arcsec)
15	20.0	81.3	18.3	263	1650	330	41.3
30	10.0	81.3	9.16	65.8	825	165	20.6
45	6.67	32.3	15.4	186	550	110	13.8
60	5.00	32.3	11.6	105	413	82.5	10.3
75	4.00	32.3	9.24	67.0	330	66.0	8.25
120	2.50	30.8 / 41.1	6.06 / 4.54	28.8 / 16.2	206	41.3	5.16
150	2.00	30.8 / 41.1	4.84 / 3.63	18.4 / 10.3	165	33.0	4.13
180	1.67	30.8 / 41.1	4.04 / 3.02	12.8 / 7.18	138	27.5	3.44
210	1.43	30.8 / 41.1	3.46 / 2.59	9.40 / 5.28	118	23.6	2.95
240	1.25	30.8 / 41.1	3.03 / 2.27	7.20 / 4.04	103	20.6	2.58

**Table 2** Primary beam and resolution of LOFAR for various frequencies in the range of the two antenna types. Below 40 MHz the LBA outer configuration is used. Above 40 MHz the LBA inner configuration is used. The area covered by HBA tiles is slightly different for the Core (D=30.8 m) and Remote Stations (D=41.1 m), values for both configurations are quoted. Resolutions are given for three different maximum baselines (L).

The stations in the Netherlands have 48 dual polarized signal chains, each with two low-band inputs (differing only in filter configuration) and a high-band input. The European stations will typically have 96 dual polarized signal chains. The Low Band Antenna (LBA) is a dual polarized dipole. In the Netherlands stations, half of the 96 LBAs are connected to each low-band input. The configuration in the field of each set of 48 LBAs is optimized for a different frequency range (see also Table 2). In the European stations all 96 LBAs are connected to a single low-band input. The HBAs are compound elements, each tile consisting of 16 dual polarized dipoles and an dual polarized analogue beamformer.

The raw data stream for a station is determined by the number of antennas and the digitization bit-depth. The latter follows from the RFI environment of the sites. For LOFAR a 12 bit digitization scheme has been adopted, resulting in 230 Gbit/s raw data at a 200 MHz sample rate. In trading data transport cost against maximum instantaneous sky coverage and bandwidth, a reconfigurable processing scheme has been defined. For Remote Stations the outgoing data rate is set such that configurations are allowed ranging from a single beam at 32 MHz to eight beams at 4 MHz each. This satisfies the needs of most science cases, within reasonable cost for the network. Flexibility usually comes at a price in terms of processing overhead. To minimize this overhead, LOFAR uses atomic data quantities containing data for a 200 kHz band for a single beam (so called “beamlets”) in the signal processing.

Freq (MHz)	NL-Core (kJy)	NL-Remote (kJy)
15	483	483
30	89	89
45	48	48
60	32	32
75	51	51
120	3.6	1.8
150	2.8	1.4
180	3.2	1.6
210	3.7	1.8
240	4.1	2.0

**Table 3:** System equivalent flux densities (SEFD) of the LOFAR Core and the Remote Stations within the Netherlands. System equivalent flux densities are an estimated average over elevation using an average sky temperature for a single polarization. The SEFD is a way to express the system temperature, which includes contributions from receiver noise, feed losses, spillover, atmospheric emission, galactic background and cosmic background. The SEFD is defined as the flux density of a source that would deliver the same amount of power. The SEFD thus takes into account the efficiency and the collecting area of the antenna, plus the system noise.

Freq. (MHz)	$\lambda$ (m)	$\Delta S_{13+7}$ (mJy)	$\Delta S_{18+18}$ (mJy)
15	20.0	201	110
30	10.0	37	20
45	6.67	20	11
60	5.00	13	7.2
75	4.00	21	12
120	2.50	0.74	0.41
150	2.00	0.58	0.32
180	1.67	0.67	0.37
210	1.43	0.76	0.42
240	1.25	0.84	0.46

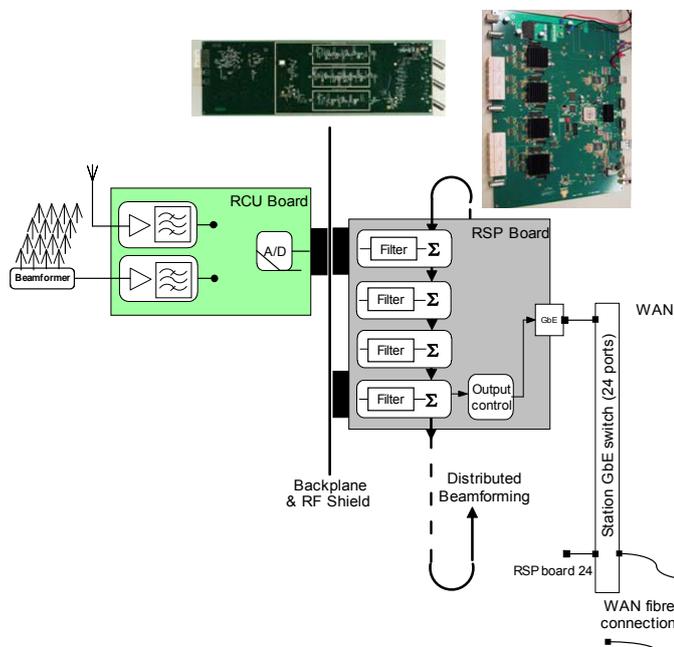
**Table 4:** LOFAR array sensitivity for 2 polarizations, 1h integration time, 4 MHz bandwidth (3.57 MHz effective bandwidth), and an imaging noise increase by a factor of 1.3. Two different values are given: 13 Core + 7 Remote Stations and 18 Core + 18 Remote Stations. The effect of station tapering is not included.

A further trade-off has to be made after combining all stations in the central processing systems. The amount of data resulting after correlation or tied-array beamforming is too large to be kept indefinitely in general. Balancing the processing and storage requirements for the various science cases, averaged correlated data are kept available for semi-on-line processing for at least a week. A dedicated processing cluster will be available for calibration and imaging. The amount of processing power in this cluster will be sufficient to allow for timely reduction of data to more manageable volumes. There are exceptions to this scheme: some science projects will keep unprocessed data to allow later reprocessing.

Throughout the system, such trade-offs have been made to maximize the scientific capabilities and flexibilities within the available budget. Table 2 shows the primary beams and resolution of the currently planned array. Table 3 gives an overview of the resulting sensitivity per stations. Table 4 presents the sensitivity of the array.

### 3. LOFAR ANTENNA STATIONS

In the LOFAR stations electromagnetic signals are received by multiple dipoles. At station level the signals from all of these dipoles are combined by beam forming to reduce the data rate and processing required. The main station architecture is depicted in Figure 1.



**Figure 1:** The main station architecture, showing the antennas, the receiver (RCU board, green), the digital signal processing (RSP board, grey) and the wide-area network (WAN) interface.

#### A. Antennas

LOFAR operates in the 15 to 240 MHz range, excluding the 87–108 MHz FM radio band. Since the sensitivity range spans five octaves, two types of antennas were developed: the Low-Band Antenna (LBA), optimized for 30–80 MHz, and the High-Band Antenna (HBA), optimized for 120–240 MHz.

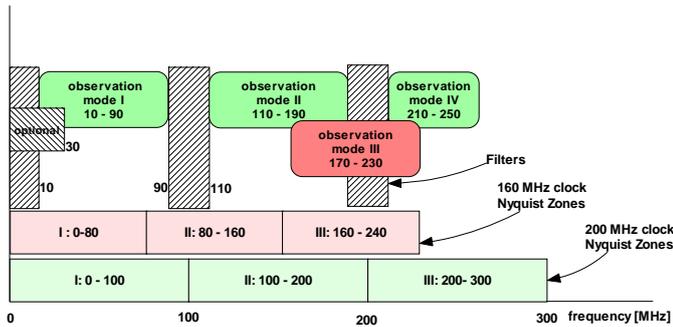
Within a station, the LBAs are placed in a pseudo-random way, with exponentially increasing distances [4]. The diameter of an LBA field is approximately 81 m. The HBA tiles will be installed as a dense and regular array with a size of about 41 m in diameter for the remote stations. In the core stations, the HBA field will be split into two arrays with a diameter of 31 m each.

Each LBA consists of one dipole per polarization. Each HBA is a compound element (“tile”) where 16 dual polarized antenna elements are combined using analog beamforming per polarization. In this way the effective area is extended at minimal cost. The analog beamforming is done close to the antenna elements to minimize the number of cables. The decrease in field of view is outweighed by the increase in sensitivity. The size of the tile (4x4 dual-polarized dipoles) is such that the effective area of LBA and HBA are roughly equal.

After pre-filtering, amplification and, for the HBA, beamforming, the signals from LBA and HBA are transported over coax cables to station cabinets.

#### B. Receiver

The LOFAR receiver has three input sections connecting to a single A/D converter. Two input sections are configured for the LBA and HBA as described above. The third input section can accommodate an LBA but has a more relaxed filter at lower frequencies, to accommodate observations below 30 MHz. A wide-band direct digital conversion architecture is used in the receiver. This removes the need for mixing stages. A stable and coherent A/D clock is derived through a hierarchical scheme from a GPS conditioned Rubidium clock. A large bandwidth span is covered by sampling signals in several Nyquist zones. The maximum sampling rate is 200 MHz. To create overlap between the Nyquist zones, a sample frequency of 160 MHz can be chosen as well. The Nyquist zones I to III of the A/D converter with a sample frequency of 200 MHz and 160 MHz respectively are depicted in Figure 2. After selecting an antenna, the signal is filtered with one of the four available observing bands. After filtering, the signal is amplified and filtered again to reduce the out-of-band noise contribution (anti-aliasing). A pre-amplifier in front of the A/D converter converts the single-ended signal into a differential signal prior to A/D conversion.



**Figure 2:** The available modes in the LOFAR receiver, selecting one of three Nyquist zones for the two available clock rates.

### C. Digital Processing

Three parameters are available to reduce the data rate: bandwidth, instantaneous sky coverage and bit-depth. The main function of the station level digital signal processing is to balance bits, beams and bandwidth for the various science cases of LOFAR. The main trade-off is between the number of subbands and the number of station beams. Reducing the bit-depth (at the cost of data-loss in case of sudden increases in RFI) is being studied as an option.

The first step in the processing chain is the formation of subbands. A filterbank in the stations separates the total band into 512 equidistant subbands, resulting in bands of 195 kHz (200 MHz sample clock) or 156 kHz (160 MHz sample clock). Station level filtering is implemented in a polyphase filterbank programmed on Field Programmable Gate Arrays (FPGA). A subset of the 512 subbands can be selected for further processing. Selected subbands can be arbitrarily distributed over the band and will add up to a total of 32 MHz. For pulsar studies, options for near-perfect reconstruction of time series are available.

The bandwidth of the subbands is chosen such that a phase shift beamformer can be used. To avoid beam squint, bands should be smaller than  $\sim 200$  kHz. To form beams, the antenna signals are combined in a complex weighted sum for each selected subband. Each subband gets its own phase shift and all subbands are treated independently of each other. In this way, the number of pointings on the sky can be exchanged against the bandwidth per pointing. Configurations ranging from a single dual-polarized beam of 32 MHz to a maximum of eight dual-polarized beams of 4 MHz can thus be formed.

The number of beams is limited by the processing power of the Local Control Unit, which has to calculate the weights each second, for each direction on the sky. The complex weights are also used to correct for gain and phase differences in all the individual analog signal paths. The gain and phase differences are

determined by a station calibration algorithm [5], which runs parallel to observations. The full array correlation matrix of all dipoles in the stations is calculated for one subband each second as input to this calibration. Each second, another subband is selected, so that the station calibration algorithm loops over the complete band in about 512 seconds. The array correlation matrix can also be used for RFI detection and mitigation [6].

The collection of beamformed subbands (“beamlets”) is transported to the central processing systems over dedicated optical Gigabit Ethernet connections. Core stations and the majority of the Remote Stations in the Netherlands are connected to a Concentrator Node close to the Core. From there they are further transported over the LOFAR backbone fiber to the Central Systems, which are deployed at the University of Groningen. Some LOFAR stations near Groningen have a direct connection to the Central Systems. European stations are connected over European research networks and dedicated links.

## 4. LOFAR CENTRAL SYSTEMS

The LOFAR Central Systems have three components:

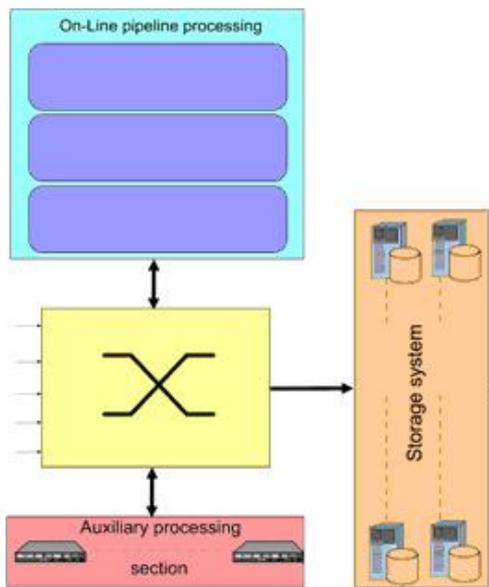
- Streaming data processing (filtering, correlation, tied-array beamforming)
- Semi-online processing (calibration, imaging)
- Operations (Monitoring & Control, Specification & Scheduling and System Health Management)

These functions are all implemented in software that is deployed on a variety of hardware platforms. A buffer storage acts as an interface between the first two categories which are briefly described below.

### A. Streaming processing functions

Figure 3 shows the top-level architecture of the streaming processing functions. A switching network is used to bring the corresponding beamlets from different stations together. The on-line pipeline processing implements the correlation and beamforming functions. Currently these functions are deployed on an IBM BG/P supercomputer. The results from the on-line processing are routed to the intermediate storage.

LOFAR uses a distributed FX-correlator. The first stage of the “F” part of the correlator is implemented at the stations in the subband filter. The first stage of the transpose function is implemented in the switching network. The input section of the on-line pipeline further separates the subbands in 1 kHz channels. The final stage of the transpose is implemented in the internal networks of the BG/P[7]. Finally the “X” part of the correlator is implemented on the CPUs of the BG/P. This distributed implementation makes maximum usage of the qualities of the various (mostly off-the-shelf) hardware components[8].



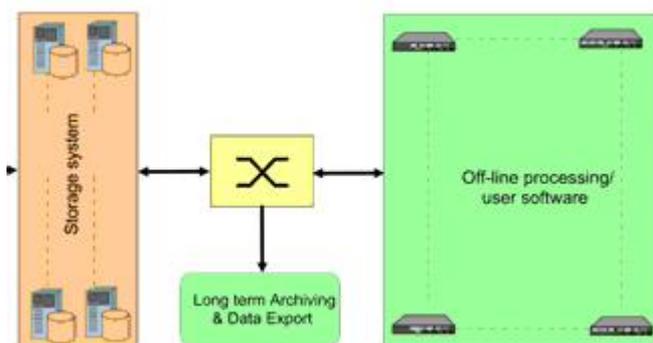
**Figure 3:** The streaming processing section of the LOFAR Central Systems, with the buffer storage.

The streaming data processing can be reconfigured into beamforming mode, and additional pre-calibration and flagging functions can be easily integrated into it.

The input data rate of the streaming processing section is 25,000 TB/day. This can be reduced to 250 TB/day after integration of time. These data are stored in the buffer storage, which will eventually be a Petabyte system to allow for storage of the data over several days.

### B. Semi-online processing

Figure 4 gives the top level architecture of the semi-online processing systems. The term semi-online is used because the limited availability of the storage requires these processing steps to take place in a time constrained manner.



**Figure 4:** The semi-online processing systems, with the buffer storage that forms the interface with the streaming processing.

Data are read from the buffer storage through a second switching network, to allow efficient routing to nodes in the processing cluster. This processing typically connects seamlessly to the streaming processing pipelines.

The main functions for the semi-online processing are calibration and imaging. The huge data rates, the time-dependent beamshapes of the aperture array stations and the effects of the ionosphere bring significant challenges to LOFAR calibration. LOFAR calibration is a joint estimation problem for instrumental parameters, environmental (e.g. ionospheric) parameters, and parameters for celestial sources. At its heart lies the “Measurement Equation” which is used to model the observed data [9]. A detailed description of all steps involved [10] and a signal processing data model with Cramer-Rao lower bound analysis [3] can be found elsewhere.

The first step of LOFAR calibration consists of removing bad data points. Then the contaminating contribution of a couple of very strong sources (like CasA, CygA, TauA, VirA) that enter through the station beam sidelobes are removed. Since modelling the station beam sidelobes is not feasible due to the large number of parameters involved, the combined effect of the sources and the instrumental effects has to be estimated and subtracted from the data. Once the interfering signals have been removed from the data, the data are further integrated, based on the required field of view and maximum baseline. Averaging over time is limited by the coherence time of the ionosphere.

The remaining calibration steps are performed iteratively in a process called the “Major Cycle”. First, the instrumental parameters, the environmental parameters, and parameters for the brightest celestial sources are estimated, using the visibility data. Then, after the brightest sources have been removed from the visibility data, an image is constructed. Finally, the parameters (position, flux, shape) of the faint celestial sources are estimated using the image data. Since not all parameters are estimated jointly, the Major Cycle will be traversed a number of times in order to iteratively refine the estimates. After initial operation of the LOFAR instrument, the parameters for the strongest sources will be known. Thereafter, the strongest sources in the field of view can be used to estimate ionospheric parameters, instrumental parameters, and to refine the estimate for the station beams that is available from the station calibration.

The direction dependent estimation of ionospheric parameters is the most challenging part of this estimation problem. One visibility sample contains the combined contribution from all sources in the sky. Since LOFAR has a large instantaneous sky coverage, the contribution from different sources is distorted by

different ionospheric and beam effects. Facet imaging is a well known technique to overcome smearing effects that are due to the fact that the baselines are non-coplanar [11], especially when combined with the w-projection algorithm [12] to ensure the maximum facet size is not restricted by the effect of non-coplanar baselines. The facet size will only be determined by the variability of the station beam and the ionosphere, and thus be far smaller than the total field of view, which allows the data to be further integrated in both time and frequency [13]. However, since for every facet a new set of integrated, corrected data are constructed the total amount of data will be approximately the same as for the original observed data at full resolution. Once the image is produced, source finding and source extraction algorithms will be used to estimate source parameters for the faint sources and refine the parameter estimates for the bright sources. This results in an updated source model and a new cycle of the Major Cycle is entered.

Most processing pipelines further extend this calibration approach with dedicated functions for e.g. source identification or feature extraction.

### B. Operations

Operational functions are implemented on computer nodes both at the central processing location in Groningen and in the Radio Observatory Control Room of ASTRON in Dwingeloo, see Figure 5. These functions include modules for specification and scheduling, a hierarchical monitoring and control system and innovative system health monitoring software [14]. Together these allow for flexible and reliable operations of a distributed instrument. This novel operations concept will also serve as a valuable pathfinder for the SKA.



**Figure 5:** The LOFAR and WSRT control room at ASTRON in Dwingeloo. Both instruments are operated and supported remotely from this location.

## 5. CONCLUSIONS

System level trade-offs have received continuous attention during the R&D phase of LOFAR. As a result, the instrument has been highly optimized for cost versus performance. Such system level trade-offs are

of increasing importance as radio telescopes become more complex, require more flexibility in configuration and operational modes, and are deployed in widely distributed remote locations.

ASTRON used a phased approach for the LOFAR development, where a series of prototype stations were installed to verify system concepts, demonstrate hardware performance and solve calibration issues as early as possible. This phased approach has been very successful [15] and allowed the project to accommodate new insights and requirements.

The roll-out of LOFAR is well underway. The central area of the Core is being prepared (see Figure 6) and remote stations are being deployed. Commissioning activities have been planned. Software development is planned to continue well into the operational phase to ensure maximum performance is reached. Interest in LOFAR has grown significantly over the past years. Several European countries have secured funding for LOFAR stations or are preparing proposals. These long-baseline stations further increase the imaging capabilities and scientific potential of the instrument.



**Figure 6:** The central area of the LOFAR Core near the village of Exloo in the Netherlands. The inner circle has a diameter of 350 m, the total core area is approximately 2 km in diameter. In the lower left corner, part of a test station is visible. The black rectangles are High Band Antenna tiles. Some Low Band Antennas are only visible from the lack of vegetation around their groundplanes.

## ACKNOWLEDGEMENTS

LOFAR is funded by the Dutch government through the BSIK program for interdisciplinary research and improvement of the knowledge infrastructure. Additional funding is provided through the European Regional Development Fund (EFRO) and the innovation program EZ/KOMPAS of the Collaboration of the Northern Provinces (SNN). ASTRON is part of the Netherlands Organization for Scientific Research, NWO.

## 6. REFERENCES

- [1] J.D.Bregman, "Concept design for a low frequency array", Proc. SPIE Vol. 4015, pp. 19-33, 2000.
- [2] A. de Bruyn, R. Fender, J. Kuijpers, G. Miley, R. Ramachandran, H. Röttgering, B. Stappers, M. Weygaert, and M. Haarlem, "Exploring the Universe with the Low Frequency Array, A Scientific Case," September 2002, <http://www.lofar.org/PDF/NL-CASE-1.0.pdf>.
- [3] S. van der Tol, B. Jeffs, and A. van der Veen, "Self Calibration for the LOFAR Radio Astronomical Array," IEEE Tr. Signal Processing, vol. 55, no. 9, pp. 4497–4510, September 2007.
- [4] W.A. van Cappellen, S.J. Wijnholds, and J.D. Bregman, "Sparse antenna array configurations in large aperture synthesis radio telescopes", 3rd European Radar Conference, 2006. September 2006.
- [5] S.J. Wijnholds, A.J. van der Veen, "Multisource Self-Calibration for Sensor Arrays", IEEE Tr. Signal Processing, accepted, 2009.
- [6] A. J. Boonstra, S. J. Wijnholds, S. van der Tol, and B. Jeffs, "Calibration, Sensitivity and RFI Mitigation Requirements for LOFAR," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Philadelphia (PA): IEEE, March 2005, pp. V–869–872.
- [7] K. Iskra, J.W. Romein, K. Yoshii, P. Beckman, "ZOID: I/O-Forwarding Infrastructure for Petascale Architectures", ACM SIGPLAN Symposium on Principles and Practice on Parallel Programming (PPoPP'08), February 2008.
- [8] J.W. Romein, P.C. Broekema, E. van Meijeren, K. van der Schaaf, W.H. Zwart, "Astronomical Real-Time Streaming Signal Processing on a Blue Gene/L Supercomputer", ACM Symposium on Parallel Algorithms and Architectures (SPAA'06), July 2006.
- [9] J.P. Hamaker, J.D. Bregman, R.J. Sault, "Understanding Radio Polarimetry, I Mathematical foundations", Astronomy & Astrophysics Suppl. Ser. 117, 137-147, May 1996.
- [10] R. J. Nijboer and J. E. Noordam, "LOFAR Calibration," in Astromical Data Analysis Software and Systems XVI, ser. ASP Conference Series, R. A. Shaw, F. Hill, and D. J. Bell, Eds., vol. 376, 2007, pp. 237 – 240.
- [11] G. B. Taylor, C. L. Carilli, and R. Perley, Eds., Synthesis Imaging in Radio Astronomy II, ser. ASP Conference Series, vol. 180, 1999.
- [12] T. J. Cornwell, K. Golap, and S. Bhatnagar, "The non-coplanar baselines effect in radio interferometry: The W projection algorithm." in IEEE Journal of Selected Topics in Signal Processing, 2008.
- [13] H.T. Intema, S. van der Tol, W.D. Cotton, A.S. Cohen, I.M. van Bemmelen and H.J.A. Rottgering, "Ionospheric Calibration of Low Frequency Radio Interferometric Observations using the Peeling Scheme, I. Method Description and First Results," submitted to Astronomy & Astrophysics.
- [14] R. Overeem *et al.*, "LOFAR Monitoring & Control System", Experimental Astronomy, special issue on LOFAR, *in preparation*.
- [15] S.J. Wijnholds, Jaap D. Bregman, Albert-Jan Boonstra, "Sky Noise Limited Snapshot Imaging in the Presence of RFI with LOFAR's Initial Test Station", Experimental Astronomy, V17, pp35-42, 2004