Overview of the LOFAR signal processing architecture

André W. Gunst, Ronald Nijboer, and John W. Romein

***Abstract*—LOFAR is the first of a new generation of phased-array radio telescopes, that combines the signals from many thousands of simple, omni-directional antennas, rather than from expensive dishes. Its revolutionary design and unprecedented size enables observations in the hardly-explored 10–250 MHz frequency range, and allows the study of a vast amount of new science cases.**

This paper describes the LOFAR signal processing chain from the stations, where the signals are received to the Central Processors. The Central processing is split in real-time correlation and off-line calibration and imaging.

I. Introduction

In the Netherlands a LO(w) Frequency ARray (LOFAR) is developed for radio astronomy optimized for the frequency band from 30–240 MHz. LOFAR is the first large scale radio telescope based fully on the phased array technique. This prevents the use of moving large constructions and enables multi-beaming. The total collecting area of LOFAR is achieved by many small dipoles which are grouped in stations to reduce the data rate to an acceptable level. The main reduction is achieved by selecting only a part of the sky by using the phased array technique. The number of stations installed in the Netherlands will be at least 36. Half of the number of stations will be core stations and the other half remote stations. The main difference between them is that the core stations can be split up in two independent arrays delivering the two fold of the remote station bandwidth.

- For the low frequencies involved in LOFAR traditional telescopes would be very large and hence costly
- Pointing can be done electronically, without using moveable parts and hence saving on maintenance costs
- Pointing is enabled in multiple directions at the same time
- Operational flexibility is enhanced significantly (e.g. rapid switching between observations is possible

Since the concept of LOFAR is so different compared with the traditional radio telescopes, the astronomical science that can be done with it is completely different as well. Despite the new concept the processing steps of a radio telescope remain the same. The general data path for an aperture synthesis array is depicted in Figure 1.

In LOFAR, the detector is composed of multiple sensors. Since these sensors convert electromagnetic radiation into electronic signals, the sensors are further referred to as antennas. Ideally, the number of detectors equals the number

A. Gunst, R. Nijboer, and J.W. Romein are with ASTRON, Dwingeloo, The Netherlands, www.astron.nl

Manuscript received January 31, 2008; revised .

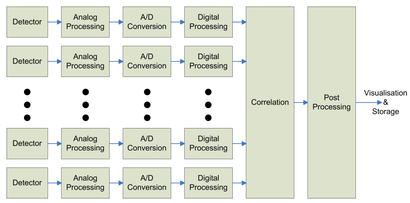


Fig. 1. LOFAR remote station architecture.

of antennas to accommodate all sky imaging. However, cost of data transport and processing power limits the number of detectors which can be afforded for a feasible design with a reasonable price. The analog processing shown in Figure 1 covers the (low noise) amplification, filtering, analog signal transport and further signal conditioning functions before the signal is converted into the digital domain (Analog/Digital (A/D) conversion block). From there, the signals are digitally conditioned before entering the correlator. Typical operations in LOFAR digital processing are frequency selection, beam forming, delay tracking and fringe stopping. In the correlator, all signals are correlated with each other to form the cross correlation matrix. Furthermore, the correlation results are calibrated for instrumental and environmental effects. Additionally, known sources are subtracted to enhance the dynamic range. Another post processing task is to transform the correlation products into an image.

For LOFAR, the sensors should be distributed over a large area to achieve an angular resolution of arcsec accuracy with an acceptable UV coverage. All data coming from the sensors should come together in the correlator. So, on the one hand the instrument should be distributed over a large area, while on the other hand all data should come together in a central location. To balance the hardware and operational costs between (1) the equipment in the field, (2) the transport network and (3) the volume of the central systems, multiple antennas (96) are grouped in so called stations. Within such a station the information of all individual antennas is weighted and summed. Such an array of antennas is often called a phased array. By using this technique a spatial selection on the sky is made, which reduces the instantaneous Field Of View (FOV) of each station. The main reason for this is to reduce the total data stream to the correlator.

For at least two Key Science Projects (KSPs) a larger FOV is required in the core of the instrument (2 km in diameter) than at extended ranges. Since there is only a limited distance to bridge from the core to the central processing location, more beams, requiring more bandwidth, can be generated in the

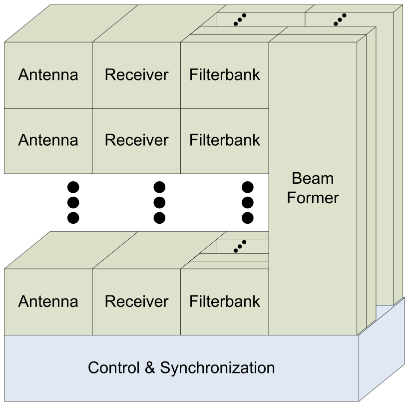


Fig. 2. LOFAR remote station architecture.

core. Hence, in LOFAR two types of stations are distinguished: the remote stations and the core stations. The main difference between them is that the core stations can be split up in two independent arrays delivering the two fold of the remote station bandwidth.

The number of stations installed in the Netherlands will be at least 36. Half of the number of stations will be core stations and the other half remote stations.

All data of the stations is transported to one central place via a Wide Area Network (WAN) and using owned and leased light paths. In Groningen all station data is processed. The processing is done by off the shelf hardware, varying from a supercomputer to clusters of computers. The processing done will accomodate for several pipelines, amongst them are pipelines for imaging modes for applications like the Epoch Of Reionization (EOR), surveys and transients and tied-array beamforming for applications like pulsars. However, also more sophisticated processing will be done to enable various energy cosmic ray mode pipelines of different intensities.

The heart of LOFAR will be installed in the Northern part of the Netherlands. The maximum baseline of LOFAR including only the Dutch stations is about 100 km. Since, also other European institutes show interest and are building LOFAR stations the maximal baseline of LOFAR will be extended, possible to 1000 km.

II. Station processing

In the LOFAR stations the electromagnetic signals of interest are received by multiple dipoles to achieve the sensitivity requirements. At station levels 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 2. Each subsystem will be discussed in the next subsections.

A. Antennas

The operating frequency range of LOFAR is from 10 MHz to 240 MHz, while the antennas should be optimized for the range 30–80 MHz and 120–240 MHz. Since, the optimized bandwidth of the operating frequency range spans 8 octaves at least two types of antennas are necessary in order to fulfill the sensitivity requirements. Therefore, two types of antennas are developed: the Low Band Antenna (LBA) and the High Band Antenna (HBA). To accommodate science below 30 MHz an extra provision is made for a third antenna, also referred to as the Low Band Low (LBL) antenna. In that context the LBA is also referred to as Low Band High (LBH) antenna. All antennas are designed for two polarizations.

In total 48 LBL, 48 LBH and 48 HBA tiles will be installed in a station. The HBA is organized in tiles, wherein 16 antenne elements are combined via analog beamforming to yield a comparable effective area for both the low band and high band antennas. Analog beamforming for the HBAs is used to combine the 16 antenna elements at tile level, which saves on costs. This processing step is done locally near the tile to reduce the number of cables necessary to connect to the central station location where the receivers are installed.

Both the low band and combined high band antenna signals are pre-filtered and amplified near the antenna prior to transportation over coaxial cables to a central location within a station. The low band array will be a randomized and exponentially space tapered configuration [1] with a size of about 85 m in diameter, while the HBA array will be installed as a dense and more regular array.

B. Receiver

In interferometry it is important to keep the signal paths equal in (electrical) characteristics (because in fact differences between signals received are measured). This also applies to the signals before beamforming. Any difference in gain or phase introduced prior to the beamforming operation will degrade the signal to noise ratio (here defining ”signal” as the signal of interest, the sky noise, and the ”noise” as the noise generated by the system). For these reasons early sampling and digitization is preferred and therefore done prior to beamforming in the LOFAR stations (an exception are the HBA arrays, where an analog beamformer stage is used as well for cost reasons).

For the receiver a wideband direct digital conversion architecture is adopted. This reduces the number of analog devices used in the signal path. The maximum sampling rate is 200 MHz, which is sufficient to directly convert the analog signals. To fill the gaps in 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 3.

Since the LOFAR stations are installed in civilized areas the dynamic range of the A/D converter must be sufficient to handle the Radio Frequency Interference (RFI) signals in the band of interest. Hence, the A/D converter converts the analog signal into a 12 bit digital signal.

The antennas in the field are all connected via coaxial cables to the rest of the station hardware which is localized in the centre of each station. The three types of antennas are connected to the receiver, which selects one out of these three antennas. After selecting an antenna, the signal is filtered with one of the integrated filters. These filters select one of the four available observing bands. After filtering, the signal is amplified and filtered again to reduce the out of band noise

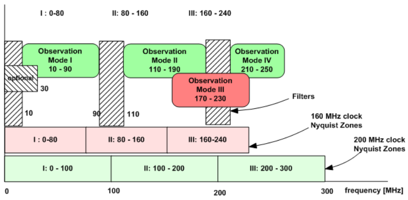


Fig. 3. The supported modes in the LOFAR receiver based on the available Nyquist zones.

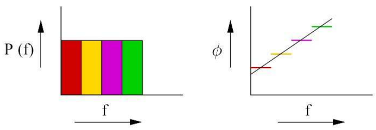


Fig. 4. Illustration of 4 subbands and the error which is introduced by approximating the time delays by phase shifts per subband (the black line on the right hand side is the ideal phase).

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.

C. Digital Processing

To form a phased array at station level, the analog antenna signals must be delayed and added which results in a beam on the sky. Moreover the beamformer should be able to track sources on the sky and be flexible in exchanging beams for bandwidth.

The beamformer can be implemented by using true time delays or by applying phase shifts on narrow subbands. The time resolution required for using true time delays is smaller than the time resolution available (one over 200 MHz). On the contrary phase shifts can be applied only if the subband width is narrow enough. The error which is made at the edges of each subband is shown in Figure 4, since the phase is frequency dependent and only one phase can be set per subband. The choice between both depends also on the frequency resolution required further down the stream.

Since, the correlator in the LOFAR system is an FX correlator as is explained in Section III a certain frequency resolution is required for that (order 1 kHz). If the beamformer is implemented by phase shifts, another frequency resolution (order 200 kHz) is required as well which is determined by the error made at the edges of each subband. So, since a frequency resolution of order 1 kHz is required anyway it was chosen to implement the station beamforming with phase shifts after a first stage filterbank which realizes a frequency resolution sufficient for the beamforming operation. A second stage filterbank will make an even higher frequency resolution which is required before the correlator.

The reason to split the filter banks instead of using one is because the first stage filter bank operates at antenna level, while the second stage filter bank operates on station beams (which are a factor 48 smaller). Since no extra significant data reduction will be done after the second stage filterbank, that functionality will be implemented in the central systems.

The first stage filter bank in the stations splits up the total band into 512 equidistant subbands resulting in order 200 kHz subbands. After the filtering operation, a subset of the subbands can be selected. The selected subbands can be arbitrary over the band and will add up to in total 32 MHz. This bandwidth is matched to the current capacity of the central processor.

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 are treated independent of each other. In this way the number of pointings on the sky can be exchanged against the bandwidth per pointing, i.e. a user can choose between 1 beam of 32 MHz to a maximal of 8 beams of 4 MHz. This is limited by the processing power of the Local Control Unit (LCU) which is responsible to calculate the weights each second, given a certain direction on the sky.

The weights applied in the beamformer have a phase component and a gain component. Both 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 [2] which runs online with the observations. As an input to the station calibration algorithm the full cross correlation matrix of all dipoles in the stations is calculated for one subband each second. Each second another subband can be selected, so that the station calibration algorithm can tune over the complete band in about 512 seconds. Additionally the cross correlation algorithm will be used for Radio Frequency Interference (RFI) detection as well [3].

The station output signal is coded in 16 bit complex samples and will be in total 2.2 Gbps. Each second a timestamp is added to the output stream because the signal is partly transported over existing infrastructure. The protocol used is UDP (User Datagram Protocol), because it is not possble to do re-transmitts of packets. In the central processing facility the station signals will be synchronized with each other as is discussed in Section **??**.

In parallel with the data stream as discussed, a user can also store the raw data or (part of) the filtered data. This is especially done for the transients application and the cosmic ray applications. For those applications data need to be stored and freezed in a transient buffer, when for example a cosmic ray event is taking place. Freezing of the buffer content can be controlled by internal or external triggers. The internal trigger algorithms are running in parallel with the data stream. The stored data or selections thereof can be sent to the central processing facility for further processing.

max 54 subband/RSP board ==¿ wil ik eigenlijk niet doen omdat ik het niet over implementatie heb ==¿ discuss

Correlator choice: In classical radio telescopes an XF correlator was generally used, meaning that first the correlation and integration of the signals was done in time domain (X) where after the Fourier transform (F) was accomplished to get a cross power spectrum out of the correlator. This is still an economically attractive technique for radio telescopes with a limited number of antennas

(input signals to the correlator). However, for LOFAR an FX correlator (first Fourier transform and then correlating the resulting channels) is favorable in terms of processing at the expense of data transport (the signals must be regrouped per channel instead of per antenna, resulting in a transpose operation). Using only a Fourier transform in the FX correlator leads to a significant amount of leakage between the channels. Therefore it was chosen to use filter banks before the correlator. This architecture is also known as an HFX (Hybrid FX correlator) architecture.

III. Central Processing: the Correlator

The station data are sent over a dedicated Wide-Area Network to the Central Processor for further processing. Figure 5 shows the processing steps that take place at the different computer clusters within the Central Processor. The Central Processor is divided in an online (real time) and offline part. The online part reduces the data volume to an amount that can be stored on a PetaByte-sized storage system, which provides space for three to five days of data. After the observation is finished, the data are further processed (see Section IV).

The main task in the online part is to correlate all data [4]. LOFAR uses a six-rack IBM Blue Gene/L supercomputer for this purpose, unlike traditional telescopes that typically use customized hardware. The desire for a flexible and reconfigurable instrument demands a *software* solution, but the data rates and processing requirements compel a supercomputer.

The Blue Gene/L provides 34 TFLOPS peak performance and has a fast internal interconnect: the 3-D torus. A total of 768 Gigabit Ethernet interfaces are available for external I/O. Each processor is extended by two double-precision floating point units that are very suitable for signal processing, since a variety of operations on complex numbers are natively supported. The Blue Gene/L is surrounded by an input cluster and a storage cluster.

The input section receives all station data that are sent via the wide-area network. We use the UDP datagram protocol over the WAN, which is an unreliable protocol. Using a reliable protocol like TCP would complicate the design of the processing boards at the stations significantly, because TCP requires bi-directional communication (unlike UDP) and needs buffering of large amounts of transmitted data (to be able to retransmit lost packets). Since in practice few (less than 0.001%) packets are lost and occasional loss of data does not harm the astronomical quality of the data, we chose for UDP. The input section handles duplicated, missing, and out-of-order UDP packets. Lost data are appropriately flagged as being ”invalid”, and the remainder of the processing pipeline handles this accordingly. The data are received into a circular buffer, that holds the most recent six seconds of data.

The buffer serves three purposes. First, it is used to synchronize the station data, since the WAN links exhibit different travel times for different stations. Second, it provides some headroom to recover from small hiccups in the remainder of the processing pipeline, without data loss. Third, the buffer is used to form beams; the bulk of the delay is introduced by shifting the buffer’s read pointer by an entire amount of samples (the remaining delay is corrected by a phase shift later).

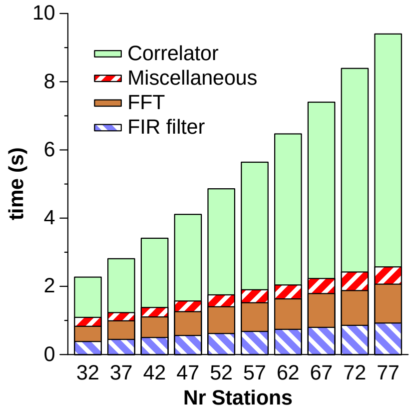


Fig. 7. Execution times of the filters for different numbers of stations
.

Each input node receives data from up to 54 subbands from a single RSP board of one station. Unfortunately, this data distribution is not suitable for the correlator, because to correlate a subband, the data from all stations are needed. Also, we need hundreds of CPUs to correlate 54 subbands. Thus, the next step is to redistribute all data over the CPUs, using a fast interconnect. Initially, we did this on a separate input cluster connected by a high-speed infiniband network, but we found that the internal 3-D torus network within the Blue Gene/L can do this job much faster [5].

...

Next in the pipeline is the second-stage PolyPhase (PPF) filter, that splits each 195 KHz (resp. 156 KHz) subband into 256 channels of 763 Hz (resp. 610 Hz) wide. Splitting the subbands into narrow frequency channels allows flagging of narrow-band RFI without much data loss (see Section IV-B). The PPF filter consists of 256 Finite Impulse Filters (FIR) filters and a Fast Fourier Transform (see Figure 6). Each FIR filter is a 16-tap band pass filter. The incoming samples are round-robin distributed over the FIR filters; the outputs are fourier transformed. To achieve optimal performance, both the FIR filter and the FFT are implemented in assembly (although we maintain an equivalent C++ version for debugging and portability purposes).

After the subbands are split into narrow channels, the remainder of the delays are compensated, by shifting the phase of each sample. The correction factor depends on time and frequency. The delays are computed exactly for the beginning and ending of each integration period (typically: one second), and interpolated both in time and (channel) frequency such that the phase of each sample is corrected by an accurate factor.

After the phase correction, the data are correlated, by multiplying the samples of each station with the complex conjugate of all other stations. To reduce the output data rate, the correlations are typically integrated over one second. Since the cross correlation of station A and B is the complex conjugate of station B and A, we compute it only once. Autocorrelations are computed as well, but treated separately since they require only half the amount of computations. The computational

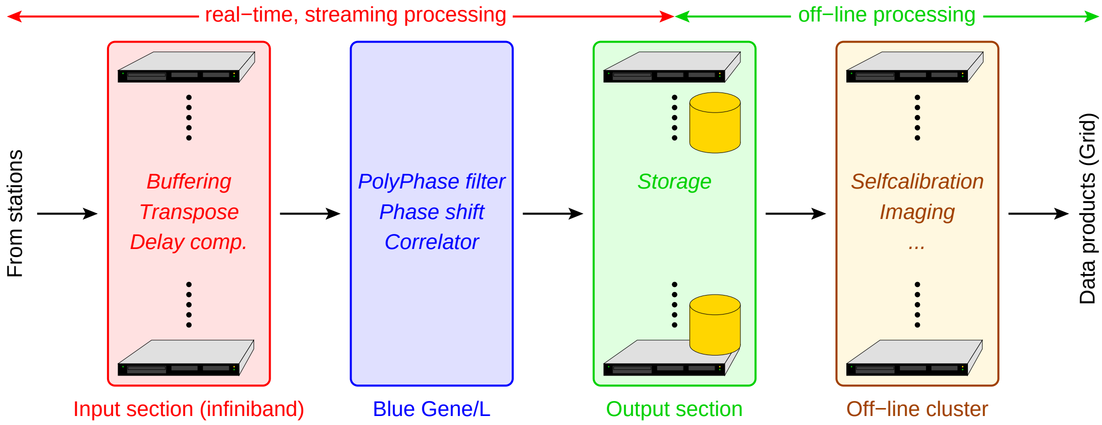


Fig. 5. Clusters in the Central Processor.

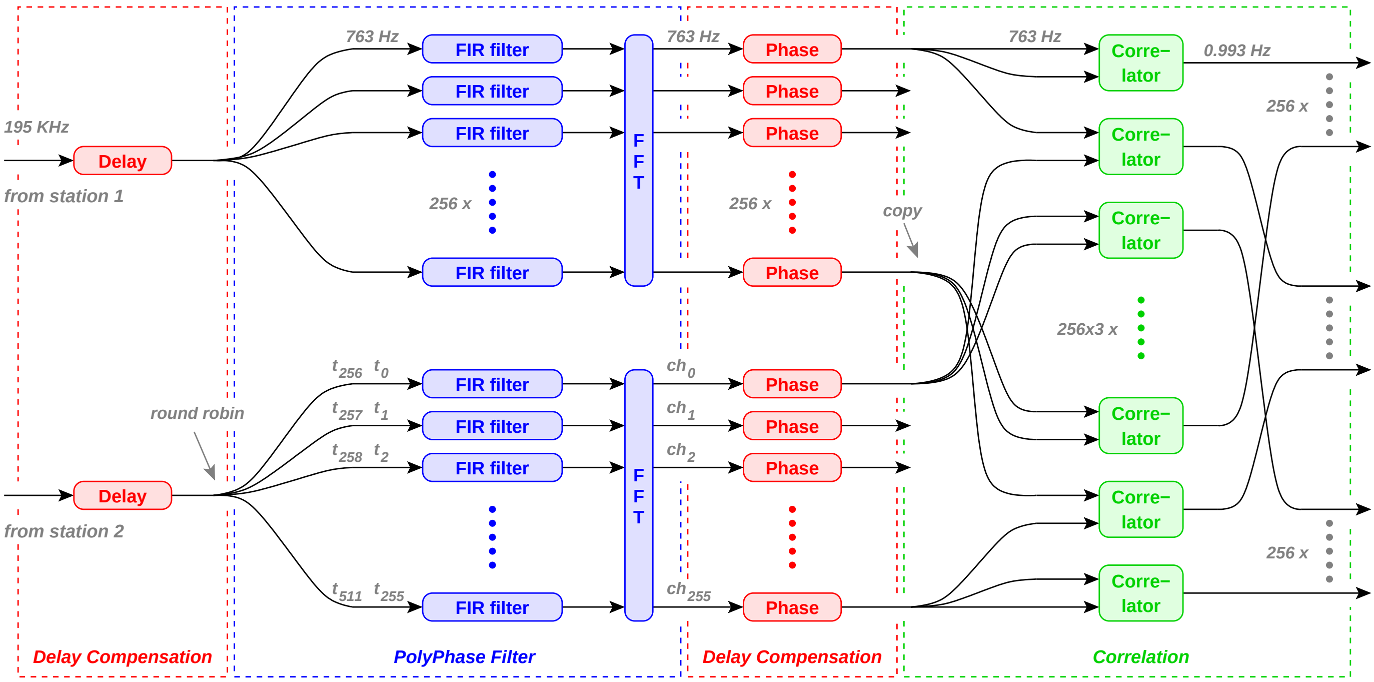


Fig. 6. Real-time filters on the Central Processor. For simplicity, the figure shows two stations, one subband, one polarization. The numbers are valid for the 200 MHz mode.

requirements of the correlator are squared in the number of stations, and dominate the total online processing demands (see Figure 7). The correlator, which is also implemented in assembly, is extremely efficient: it achieves 98% of the floating-point peak performance [4].

Storage section

The PPF implicitly converts the 16-bit complex number to 32-bit floating-point numbers, since the BG/L performs floating point computations faster than integer operations. Also, the phase correction and the correlations are done in floating point. Internally, the BG/L only supports double-precision arithmetic, but the high precision is not necessary for our application.

…

Maybe this paragraph should be moved to the IEEE Computer paper. An example that illustrates the benefits of a flexible software solution was the ease with which we could move the second-stage PolyPhase Filter, which was originally designed to run on FPGAs at the stations, to the Blue Gene/L. Once we saw that we could obtain really high performance from the BG/L in practice and recognized that sufficient computational power was left, a considerable cost reduction was achieved by removing the PPF FPGAs from the station design.

…

Although the BG/L is computationally very efficient, streaming the station data into the machine at the required data rates turned out to be a major problem, despite the BG/L's atypical high number of I/O interfaces. For each 16 BG/L compute cores, there is one *I/O node* that has one external gigabit-Ethernet interface and transparently handles all I/O calls initiated by its associated compute cores (see Figure **??**).

We found that the stock network system software was not particularly optimized for high-throughput I/O, and that the obtained bandwidths were far from what theoretically should be achievable.

The dissatisfaction about the performance and about the I/O model in general led to a joint effort to redesign the entire network software infrastructure, and resulted in a new environment called *ZOID* [5]. ZOID does not only yield better performance, but it is much more flexible, since it allows application code to be run on the I/O node. With ZOID, we were able to move the receipt of the station data from the input cluster nodes to the BG/L I/O nodes, so that the station data are sent directly through the WAN into the BG/L. Not having to build a separate input cluster results in an estimated cost saving of €700,000.

IV. Central Processing: Calibration

The off-line processing of LOFAR data has to deal with a number of challenges [6], [7]. First of all, the data volumes are huge and being able to process the data in finite time dictates the design of the data processing chain. Second, compared to traditional steel dishes, the phased array station beams are far more variable (in time, in frequency, as well as over the different stations), they yield a higher degree of instrumental polarization, and they have relatively high sidelobes. All these issues complicate the processing of the data. Especially, since a high dynamic range must be reached.

The third category of challenges lies in the sky itself. At the low frequencies where LOFAR will observe there will be very bright sources so that a high dynamic range and, hence, a high accuracy is needed to see the faint background sources. The sky will also be filled with a large number of sources, giving rise to confusion. Last, but not least, the Earths ionosphere seriously defocusses the images.

An introduction to signal processing for radio astronomical arrays can be found in [8], [9], [3]. With LOFAR we enter a new regime in radio astronomical data processing. The challenges imply that for LOFAR we have to reconsider exisiting processing strategies and algorithms and develop new strategies and algorithms. The off-line processing therefore remains a work in progress of which we give an overview of the current status.

A. Processing large data volumes

The total amount of data that is produced is detemined by the total number of stations that are used in the observation. This number is still uncertain and will be different in the Low Band and the High Band, since the High Band stations in the Core will be split in 2 half stations. Using 20 Core stations and 20 Remote stations as a working example the correlator generates between 0.28 Gbyte/s for the LBA Core Array in 200 MHz sampling mode and 3.1 Gbyte/s for the HBA Full Array in 160 MHz sampling mode. After a typical observation of 4 hours between 3.3 Tbyte and 35 Tbyte will be collected, again depending on the exact mode of operation.

Since a permanent data storage is not part of the LOFAR telescope these data volumes have to be processed near real time. Fortunately, the non-imaging LOFAR applications are not so data intense, so that for every 1 hour of observation we may have up to, say, 4 hours to further process the data. With this in mind data I/O becomes a serious problem. Obviously the data needs to be processed in a parallelized and distributed way minimizing the I/O that is needed [10], [11].

Data can be distributed over a large number of processing nodes in a number of ways. Distribution over baselines is not very suitable for imaging, where data from all baselines must be combined to produce an image. Distribution over time has the disadvantage that up to sevaral Gbytes/s have to be sent to a single processing node. Frequency, therefore, seems to be the best way. This distribution scheme matches with the design of the correlator. It is also a convenient scheme for the imager, where images are created per (combined) frequency channel.

A consequence of distribution over frequency is that in the self-calibration step solver equations from different compute nodes may need to be combined. The combining of solver equations, however, involves far less data then the underlying observed visibility data.

Even though the processing of the data will be done on a large cluster of computers, the total amount of data can be such that we expect the quality of the final result to be processing limited. This means that for all the algorithms we have to weight accuracy against the amount of Flops needed. It also means that the LOFAR instrument can be improved by upgrading the processing cluster in the future.

B. Processing steps

LOFAR calibration is a joint estimation problem for both instrumental parameters and source parameters. At its heart lies the “Measurement Equation” that is used to model the observed data [12]. A signal processing data model and a Cramer-Rao lower bound analysis are given in [13]. The latter paper also provides a good introduction to the signal processing aspects of LOFAR Self-Calibration.

The final LOFAR calibration strategy is still under development. However, we foresee that the following steps and iterations will be part of it. The first step consists of removing bad data points, which are due to e.g. Radio Frequency Interference (RFI). After this step the contaminating contribution of a couple of very strong sources (like CasA, CygA, TauA, VirA) that enter through the station beam sidelobes needs to be removed. Since modelling the station beam sidelobes is infeasible 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 are removed from the data, the data may be further integrated. The high resolution in frequency is only needed for removing RFI. The final resolution is determined by bandwidth smearing requirements [14]. In the frequency direction the data may be reduced by a factor of 3 to 10, depending on the size of the array used for the observation **CHECK**. In principle the data may also be integrated along the time axis. Here, however, we have to make sure that the effect of the ionosphere remains constant over a time sample.

The maximal reduction factor determined by time-average smearing ranges from 3 tot 10, again depending on array size **CHECK**.

Next an iterative loop, dubbed the "Major Cycle", is entered where we first estimate instrumental and source parameters using the visibility data, then image the data, and finally refine the estimation of the source parameters using image data. Since not all parameters are estimated jointly, the Major Cycle will be traversed a number of times [7].

After initial operation of the LOFAR instrument the parameters for the strongest sources will be known. From then on the strongest sources can be used in every observation to estimate ionospheric parameters, instrumental parameters, and to refine the estimate for the station beams that is available from the station calibration.

In [13] it is shown that the unconstrained direction dependent calibration problem is ambiguous. The authors, however, present three physical contraints to get an unambigous solution:

1) use a calibrated subarray to calibrate the rest of the array,
2) use assumptions on the structural dependence of a certain corrupting effect, e.g. the ionosphere,
3) use polynomial smoothing on larger time / frequency domains.

In the first approach the LOFAR core is calibrated first, where use is made of the fact that the core stations all share the same ionosphere. This is a simpleer problem. Van der Tol et al. show that in this case the remote stations can be calibrated, provided the number of calibration sources is less then the number of core stations [13], [15].

In the second approach, use is made of the fact that the effect of the ionosphere has a predictable frequency dependence [15]. The number of parameters that need to be estimated may be further reduced by using suitable base functions for the spatial dependence of the ionosphere. The use of Karhunen-Loeve base functions seems very promosing in this respect [16].

In the third approach, multiple samples iin frequency and time are combined in a joint estimation, where the time and frequency dependence is modelled by e.g. polynomials and in this way the number of parameters that need to be estimated is reduced from 1 per individual sample to the polynomial coefficients for all samples together. In [13] it is reported however that this approach needs good initial estimates, since the continuous phase polynomial is ambiguous to integer multiples of 2π.

For LOFAR all three approaches will used and they will be combined with the so-called "Peeling" approach [6], [13].

...**Peeling** ...

The sky image is the Fourier transform of the visibility domain. Due to the fact that the visibility domain is only discretely sampled, sources in the sky image are convolved with a Point Spread Function (PSF). The contribution from sources that generate PSF far sidelobes that are higher than the image noise level should be subtracted from the visibility data. Using the solutions to the parameter estimation problem on the visibility data the contributions from the strongest sources are removed from the visibility data. The remaining residual visibility data is then corrected and imaged.

One visibility sample is the summation of contributions from all sources in the sky. Since LOFAR has a large Field of View (FoV), the contribution from different sources is distorted by different ionospheric and beam effects. When imaging the visibility data, however, it is only possible to correct the data for one direction in the sky. This would mean that the image would be sharp for the direction of correction and the image quality would degrade outwards. To overcome this problem LOFAR images will be made in facets, where we can correct the data for the center of each facet.

Facet imaging is a well known technique to overcome the problem related to the so-called "w-term", which are due to the fact that the baselines are non-coplanar [14]. However, the non-coplanar baseline problem is better solved by the w-projection algorithm [17]. Therefore, we will apply the w-projection technique per facet and the facet size will only be determined by the variability of the station beam and the ionosphere.

Since we correct the data per facet, this means multiplying the total amount of data with the number of facets. Fortunately, the facet size will be far smaller then the total FoV. This allows us to shift the data to the center of the facet and then integrate the data in both time and frequency. Hence, the total amount of data will be more or less the same.

Once the image is produced, source finding and extraction algorithms may be used to estimate source parameters. This would then lead to an updated source model and we are ready to enter a new cycle of the Major Cycle.

By sampling the data in each iteration of the Major Cycle, doubling the sampling density in every cycle, and using only the full resolution data in the last cycle, we effectively have not more then twice the I/O that is needed for the full resolution data. We expect that will seriously inprove the total speed of the processing.

V. Current state and roll-out planning

Currently [18], four partially-built stations are functional: 3 stations with 16 LBAs and 1 station with 48 LBAs. To create more baselines and achieve better UV coverage, the stations each can be split into four *microstations*. This yields 16 microstations, which are treated the same as real stations in the online and offline processing pipelines.

A consequence of quadrupling the number of stations is that the bandwidth is reduced to 36 subbands of 195 KHz or 48 subbands of 156 KHz. Alternatively, the station with 48 LBA can be split into 12 microstations, so that together with the other 3×4 microstations a total of 24 microstations can be formed, but WAN restrictions limit the bandwidth to 12 resp. 16 subbands.

Figure 8 shows a series of images there were made from data using the LOFAR CS1 configuration. 16 microstations, each consisting of a single dipole with essentially an all sky FoV, were used. The images are centered on the North Celestial Pole and contain 48 hours and about 20 subbands of data. First the data is flagged for RFI and an image of the flagged-only data is shown on the left ("observed").

The following calibration is performed in two steps. In the first step, a point source model is used for both CasA and

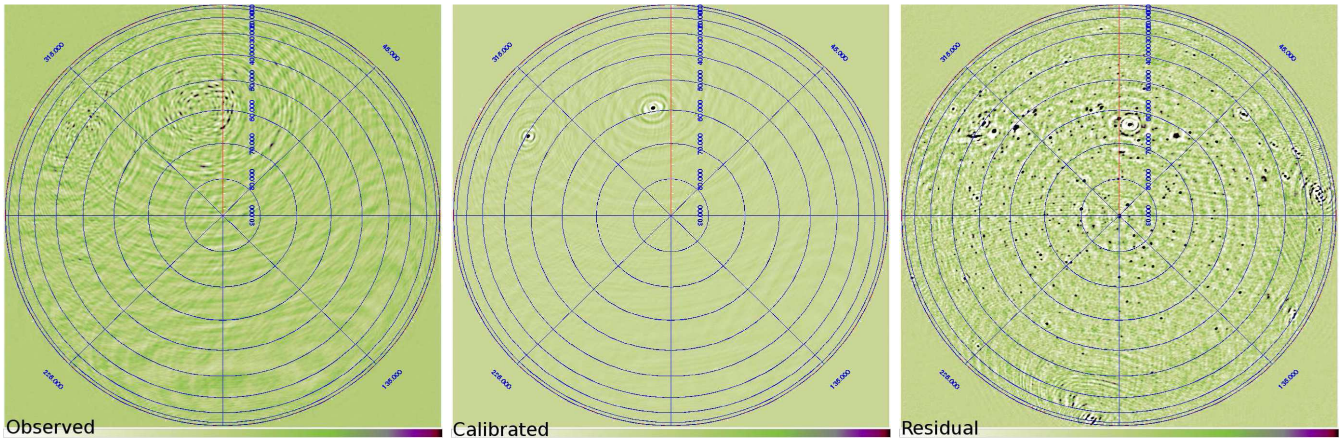


Fig. 8. Images from the LOFAR CS1 configuration using 48 hours and about 20 subbands of data. Observed: an image of the flagged, non-calibrated data. Calibrated: an image of the flagged, calibrated data showing CasA and CygA. Residual: an image of the flagged, calibrated data where CasA and CygA are removed from the data. Images courtesy of S.B. Yatawatta.

CygA, both at 20000 Jy flux and no polarization. An analytical beam shape is used and we solve for a single complex gain for the whole sky. In this way an estimate for the instrumental complex gains (due to e.g. clock drifts) and ionospheric phase differences is obtained. After correcting the data a second step is performed, where we estimate a complex gain in both the direction of CasA and CygA. In this second step no assumptions on the beam are made. For the middle image (“calibrated”) the data is corrected using the estimates for the direction of CasA. In this middle image CasA and CygA can be clearly seen as point sources.

CasA and CygA completely overshine the background sources, since they are at least 50 times stronger then the average background source. After subtracting the contributions from CasA and CygA from the data the other sources become visible. This is shown in the right panel (“residual”), where now some hundred other sources are visible.

The HBA units are currently being commissioned. A full LBA station in Effelsberg, Germany is also operational, and will soon be connected via a dedicated wide-area link to the Central Processor.

In the course of this year 18 full stations (13 core + 5 remote) will be produced and installed in the field. Also the WAN infrastructure and Central Processor facility will be ready in the end of 2008 to handle the data of the 18 full stations and a couple of international stations as well. In the year thereafter another 18 stations will be produced and installed in the field.

Construction of the full stations is planned as follows. ***”station” should be consistent with HBA-core ”double station”*** The first 18 full stations (13 core + 5 remote) ***to be confirmed*** including the WAN links to the Central Processor will be operational by the end of 2008, and the remaining 18 stations will be built in the course of 2009. Meanwhile, construction of international stations will continue. The Blue Gene/L is capable of handling all foreseen future data rates.

VI. Conclusion

The conclusion goes here.

VII.

Acknowledgment

The authors would like to thank the LOFAR team. ASTRON is a NWO institute.

LOFAR is funded by the Dutch government in the BSIK programme for interdisciplinary research for improvements of the knowledge infrastructure. Additional funding is provided by the European Union, European Regional Development Fund (EFRO) and by the “Samenwerkingsverband Noord-Nederland,” EZ/KOMPAS.

References

[1] W. A. van Cappellen, S. J. Wijnholds, and J. D. Bregman, “Sparse Antenna Array Configurations in Large Aperture Synthesis Radio Telescopes,” in *European Radar Conference (EuRAD)*, September 2006.

[2] S. J. Wijnholds and A. J. Boonstra, “A Multisource Calibration Method for Phased Array Radio Telescopes,” in *4th IEEE workshop on Sensor Array and Multichannel Processing (SAM)*, Waltham (MA), July 2006.

[3] 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.

[4] J. W. Romein, P. C. Broekema, E. van Meijeren, K. van der Schaaf, and W. H. Zwart, “Astronomical Real-Time Streaing Signal Processing on a Blue Gene/L Supercomputer,” in *ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, Cambridge, MA, July 2006.

[5] K. Iskra, J. W. Romein, K. Yoshii, and P. Beckman, “ZOID: I/O-Forwarding Infrastructure for Petascale Architectures,” in *ACM SIGPLAN Symposium on Principles and Practice on Parallel Programming (PPoPP)*, Salt Lake City, UT, February 2008.

[6] J. E. Noordam, “LOFAR calibration challenges,” in *Proceedings of the SPIE*, vol. 5489, October 2004, pp. 817 – 825.

[7] 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.

[8] A. J. van der Veen, A. Leshem, and A. J. Boonstra, “Signal processing for radio astronomical arrays,” in *Sensor Array and Multichannel (SAM) technical committee on the IEEE Signal Processing Society*, Sitges, Barcelona, Spain, July 2004.

[9] ——, “Array signal processing for radio astronomy,” *Experimental Astronomy*, vol. 17, June 2004.

[10] G. M. Loose, “LOFAR Self-Calibration using a Blackboard Software architecture,” in *Astromical Data Analysis Software and Systems XVII*, ser. ASP Conference Series, J. Lewis, R. Argyle, P. Bunclarck, D. Evans, and E. Gonzales-Solares, Eds., 2008, in press.

[11] G. van Diepen, "Distributed processing of future radio astronomical observations," in *Astromical Data Analysis Software and Systems XVII*, ser. ASP Conference Series, J. Lewis, R. Argyle, P. Bunclarck, D. Evans, and E. Gonzales-Solares, Eds., 2008, in press.

[12] J. P. Hamaker, J. D. Bregman, and R. J. Sault, "Understanding radio polarimetry. I. Mathematical foundations." *Astron. Astrophys. Suppl. Ser.*, vol. 117, pp. 137 – 147, May 1996.

[13] 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.

[14] G. B. Taylor, C. L. Carilli, and R. Perley, Eds., *Synthesis Imaging in Radio Astronomy II*, ser. ASP Conference Series, vol. 180, 1999.

[15] S. van der Tol, B. Jeffs, and A. van der Veen, "Calibration of a large distributed low frequency radio astronomical array (lofar)," in *Proc. of the European Signal Processing Conference (EUSIPCO 2005)*, Antalya, Turkey, September 2005.

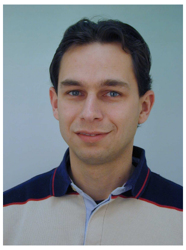
[16] S. van der Tol and A. van der Veen, "Ionospheric calibration for the LOFAR radio telescope," in *Proc. IEEE Int. Symp. Signals, Circuits, Systems*. Iasi (RO): IEEE, July 2007, pp. 457–460, iSBN 1-4244-0968-3.

[17] T. J. Cornwell, K. Golap, and S. Bhatnagar, "W projection: A New Algorithm for Wide Field Imaging with Radio Synthesis Arrays," in *Astromical Data Analysis Software and Systems XIV*, ser. ASP Conference Series, P. L. Shopbell, M. C. Britton, and R. Ebert, Eds., vol. 347, 2005, pp. 86 – 90.

[18] A. Gunst, K. van der Schaaf, and M. J. Bentum, "Core Station 1 - The first LOFAR station," in *SPS-DARTS 2006*, Antwerp, Belgium, March 2006.

André W. Gunst Biography text here.

Ronald Nijboer Blah blah blah.



John W. Romein John W. Romein is senior system engineer/researcher high-performance computing at ASTRON, where he is responsible for the online data processing of LOFAR. He obtained his Ph.D. on distributed board-game playing at the Vrije Universiteit, Amsterdam. As a postdoc, he solved the game of Awari using a large computer cluster, and did research on parallel algorithms for bio-informatics. His research interests include high-performance computing, parallel algorithms, networks, programming languages, and compiler construction.