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 10–240 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 antennas in the field to the central processing. The central processing is split in real-time processing and offline calibration and imaging.

I. INTRODUCTION

LOFAR is an acronym for *LO*w *F*requency *AR*ray, a phased-array radio telescope operating in the 10 to 240 MHz frequency range. Its design breaks radically with conventional telescopes: rather than using large, expensive dishes, LOFAR is built as a distributed sensor network of simple antenna receivers [1].

This new design has important advantages. First, obtaining sufficient sensitivity at these low frequencies with traditional dishes requires prohibitively large dish sizes, while the costs of a sensor network are modest. Second, pointing is done electronically, and can be changed instantaneously. Third, multiple observations, even of different types, can be handled simultaneously. Forth, LOFAR is more flexible, because most of the processing is done in software, rather than using customized hardware.

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

Since the concept of LOFAR is so different from traditional radio telescopes, new astronomical science can be done with it. First, LOFAR should be able to detect the *Epoch of Reionization (EoR)*, the time that the first star galaxies and quasars were formed. Second, LOFAR's large Field of View (FoV) makes it uniquely suited to study *all-sky surveys* and *transient sources*. A galactic event triggers the dump of raw antenna data that can be examined afterward. Third, LOFAR offers a unique possibility in particle astrophysics for studying the origin of high-energy *cosmic rays*. Neither the source, nor the physical process that accelerates such particles is known. Forth, it is expected that LOFAR will find many new *pulsars*,

A. W. Gunst, R. Nijboer, and J.W. Romein are with ASTRON, Dwingeloo, The Netherlands, www.astron.nl
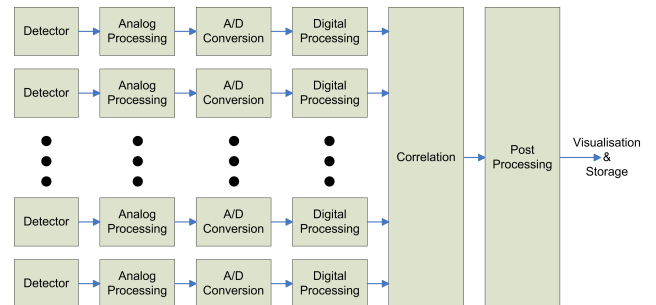
Fig. 1. Datapath of an aperture synthesis array.

that can only be observed in LOFAR's low frequency range. For a more extensive description of the astronomical aspects of the LOFAR system, see de Bruyn et. al. [2].

Despite the new concept the overall processing steps of a radio telescope remain the same. The general data path for an aperture synthesis array is depicted in Figure 1.

LOFAR is composed of multiple antennas, structured in a hierarchical way, to limit the costs of data transport and processing costs. 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 filtering, frequency selection, and beamforming. 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 in the post processing stage. Additionally, known sources are subtracted to enhance the dynamic range. Another post processing task is to transform the correlation products into an image.

todo: The LOFAR antennas are distributed over a large area to achieve an angular resolution of arcseconds accuracy with an acceptable spatial coverage. All data coming from the antennas comes together in the correlator. So, on the one hand the instrument is 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 the equipment in the field, the transport network and the volume of the central systems, multiple antennas (48) 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. As a
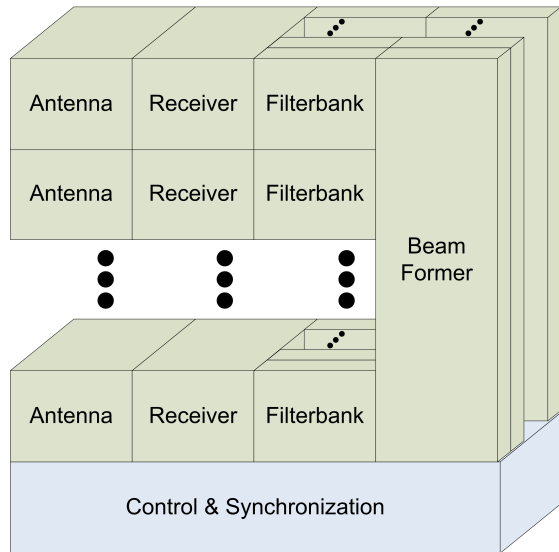
Fig. 2. LOFAR station architecture. The third dimension represents the subbands made in the filterbank.

consequence a spatial selection on the sky is made, which reduces the instantaneous FoV of each station. So the main reason for having stations is to reduce the total data stream to the correlator.

Andre: introduceer het begrip "central core" + een mooie aansluiting op de volgende paragraaf These applications only use the central core of the instrument, of which the diameter is 2 km. Since the distance over the Wide-Area Network (WAN) between the core and the central processing location is limited, higher bandwidths can be afforded from the core stations than from the remote stations. The extra bandwidth can be used to split a core station into two independent arrays, effectively doubling the number of stations in the core. At least 36 Dutch stations will be built, eighteen of which are in the core.

All station data are transported to the central processing location via a WAN, which uses owned and leased light paths. At a central location, all station data are combined and processed. The processing is done by a supercomputer, surrounded by clusters of off-the-shelf computers. The processing will accommodate several pipelines: for imaging modes, for tied-array beamforming, and for more specialized modes.

The remainder of this paper decribes the station processing, the real-time correlator, the offline postprocessing, and a description of the current state.

II. STATION PROCESSING

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 2.

A. Antennas

korter: The operating frequency range of LOFAR is from 10 to 240 MHz, while the antennas are optimized for the 30–80 MHz and the 120–240 MHz ranges, skipping the FM radio band. Since the optimized bandwidth of the operating frequency range spans 5 octaves, at least two types of antennas are necessary to fulfill the sensitivity requirements. Therefore, two types of antennas were 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, 96 LBAs and 48 HBAs will be installed in a station.

Each LBA consists of one dipole per polarization, but each HBA is organized as a tile, wherein 16 antenna elements are combined using analog beamforming, to yield a comparable effective area for both the LBAs and HBAs. The use of a digital beamformer for the HBA tiles would be prohibitively expensive. The beamforming is done locally near the tile, to reduce the number of cables to the central station location, where the receivers are installed in cabinets. Both the LBA and HBA tile signals are pre-filtered and amplified near the antenna, prior to transportation over coaxial cables to the cabinets.

Within a station, the LBAs are placed in a randomized way, with exponentially increasing distances [3]. The diameter of an LBA field is approximately 85 m. The HBA tiles will be installed as a dense and regular array with a size of about 50 m in diameter for the remote stations. In the core stations, the HBA field will be split in two arrays with a diameter of 38 m each.

B. Receiver

In interferometry it is important to keep the signal paths equal in (electrical) characteristics (because 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 wide-band 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 Figure 3.

Since the LOFAR stations are installed in populated areas, the dynamic range of the A/D converter must be sufficient to handle the Radio Frequency Interference (RFI) signals in the bands of interest. Hence, the A/D converter converts the analog signal into a 12-bit digital signal.
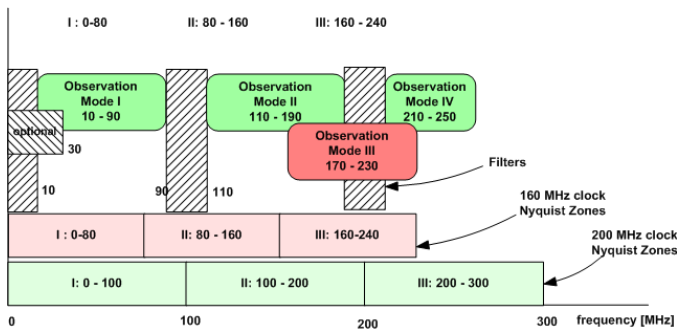
Fig. 3. The supported modes in the LOFAR receiver based on the available Nyquist zones (the bands indicated are larger than the optimized range).
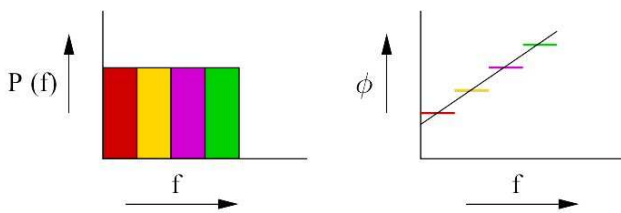

Fig. 4. Illustration of 4 subbands and the error which is introduced by approximating the time delays by phase shifts per subband (the black line on the right hand side is the ideal phase). The left picture shows the power spectrum, while the right picture shows the phase as function of frequency.

The three types of antennas are all connected via coaxial cables to the receiver, which selects one out of the three antenna inputs (LBL, LBH, and HBA). 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 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 are delayed and added, which results in a beam on the sky. Moreover, the beamformer is able to track sources on the sky and can exchange beams for bandwidth.

A 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 other hand, 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 approaches depends also on the frequency resolution required further down the stream.

The correlator in the LOFAR system is an FX correlator as is explained in Section III. The correlator uses a frequency resolution of less than 1 kHz. This frequency resolution is sufficient for a beamformer implemented by phase shifts.

However, implementing a filterbank with this resolution for each antenna signal path is extremely expensive. For a phase shift beamformer, a frequency resolution of order 200 kHz is sufficient, which is determined by the error made at the edges of each subband. Hence, it was chosen to use a first stage filter bank which operates at antenna level, to result in a frequency resolution sufficient for the phase shift beamforming. The remainder of the required frequency resolution before the correlator is achieved by a second-stage filter bank, which operates on station beams (which are a factor 48 smaller than the number of antennas). Since no extra significant data reduction is done after the second stage filterbank, that functionality is implemented in the central systems.

The first-stage filter bank in the stations splits up the total band into 512 equidistant subbands, resulting in 195 kHz subbands for the 200 MHz sample frequency and 156 kHz for the 160 MHz sample frequency. The filter bank is efficiently implemented as a Poly-Phase Filter bank (PPF) on Field Programmable Gate Arrays (FPGA).

The observer selects a subset of the 512 subbands from the first-stage PPF. The selected subbands can be arbitrary over the band and will add up to a total of 32 MHz. The capacity of the central correlator matches this bandwidth.

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, i.e. a user can choose between 1 beam of 32 MHz to a maximum of 8 beams of 4 MHz. The number of beams is limited by the processing power of the Local Control Unit, which has 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 [4], 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 is selected, so that the station calibration algorithm loops over the complete band in about 512 seconds. Additionally, the cross correlation algorithm will be used for Radio Frequency Interference (RFI) detection as well [5]. ***Maar hier doen we nog niets mee.***

III. Central processing: the correlator

The station data are sent over a dedicated WAN 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).
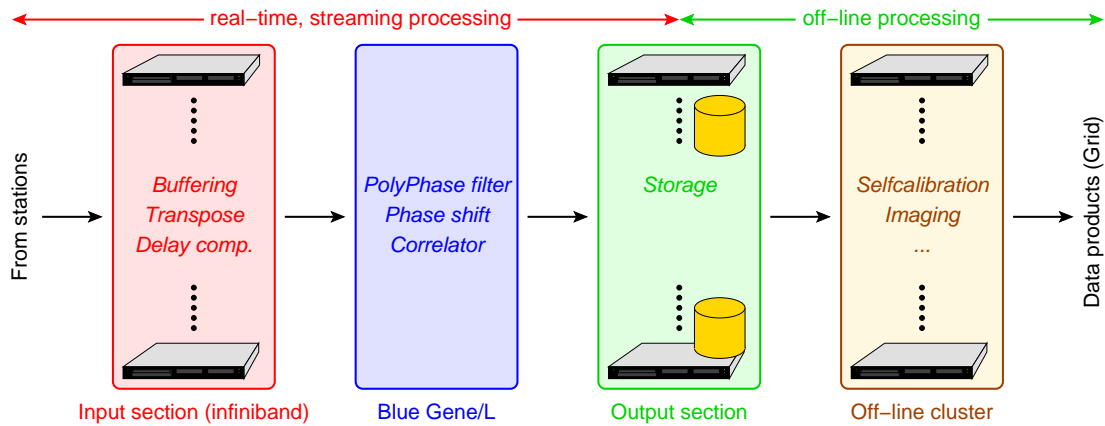
Fig. 5. Clusters in the central processor.

In the standard imaging mode, the main task in the online part is to correlate all data [6]. Classical radio telescopes use an XF correlator, meaning that first the correlation and integration of the signals is done in time domain (X), after which the Fourier transform (F) is 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 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 requirements, at the expense of data transport (the signals must be regrouped per channel instead of per antenna, resulting in a collective transpose operation).

LOFAR uses a 12,288-core IBM Blue Gene/L (BG/L) supercomputer for the real-time data processing, 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 BG/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 BG/L is surrounded by an input cluster and a storage cluster.

The input section (see Figure 5) receives all station data that are sent via the WAN. LOFAR uses 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, UDP is used. 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, since a wave from a celestial source is received by the stations at different times, the buffer is used to compensate for the time difference. The delay depends on the observation direction and the station positions, and changes continuously due to the rotation of the earth. The bulk of the delay is corrected by shifting the read pointer of the buffer by an entire amount of samples (the remaining delay is corrected by a phase shift later). Second, the buffer is used to synchronize the station data, since data from different stations have different travel times over the WAN links. Third, it provides some headroom to recover from small hiccups in the remainder of the processing pipeline, without data loss.

Each input node receives all subband data of one station. Unfortunately, this data distribution is not suitable for the correlator, because the correlator needs the data of one subband and all stations. Also, hundreds of CPUs are necessary to correlate all subbands. Thus, the next step is to redistribute (transpose) all data over the CPUs, using a fast interconnect.

Currently, the input buffer and the transpose are performed on a small-scale input cluster, but in a prototype implementation, the input cluster is bypassed and its functionality is moved to the BG/L. A redesign of the network system software was required to be able to receive the station data directly on the BG/L, to achieve sufficient external I/O bandwidth, and to increase flexibility [7]. The internal 3-D torus network is used for the transpose, and is very well able to sustain the required switching rates. Not having to build a full input cluster when more LOFAR stations are added yields a substantial cost saving.

All compute-intensive operations are performed on the BG/L. The first compute-intensive operation is the second-stage Poly-Phase Filter (PPF), 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 Response (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. Using a Fourier
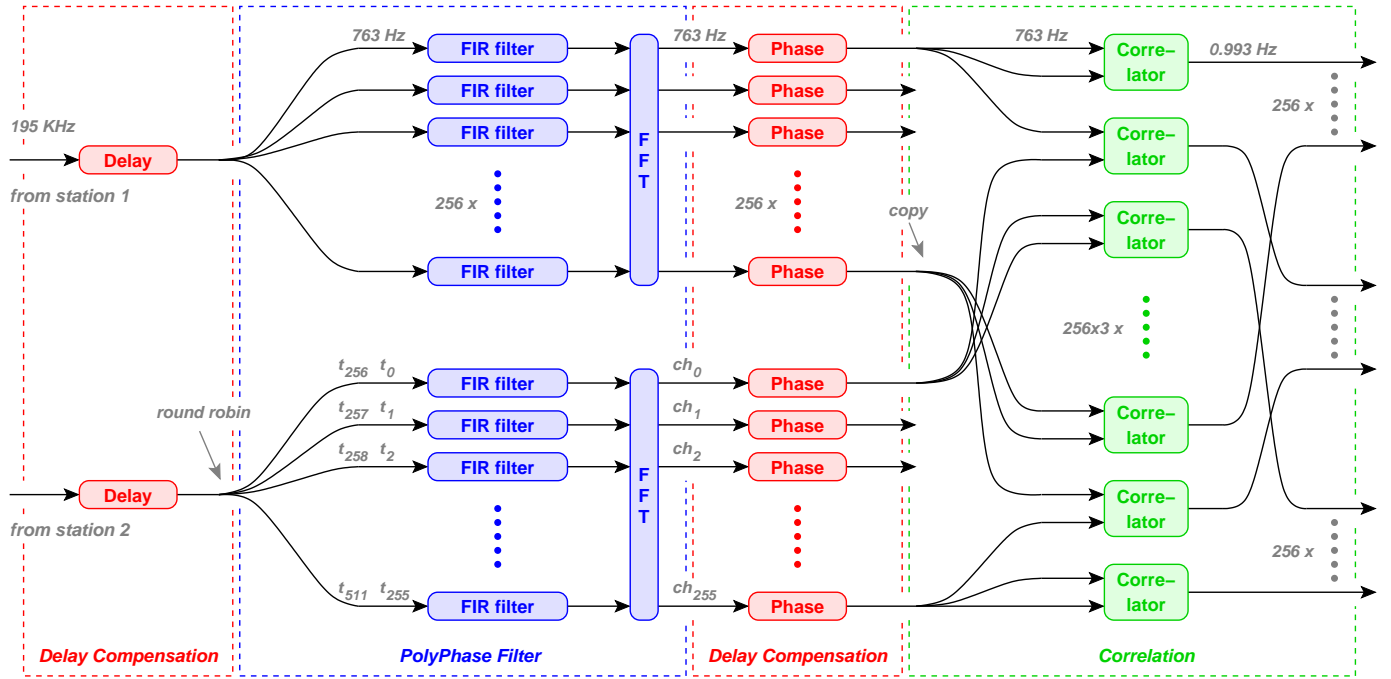
Fig. 6. Real-time filters on the central processor. For simplicity, the figure shows two stations, one subband, one polarization. The numbers are valid for the 200 MHz mode.

transform only would lead to a significant amount of leakage between the channels, therefore filter banks are used. To achieve optimal performance, both the FIR filter and the FFT are programmed in assembly. *even naar kijken* The FIR filters implicitly convert the 16+16-bit complex integer numbers from the stations to 32+32-bit floating-point numbers, since the BG/L performs floating point computations faster than integer operations. Double (64-bit) precision is not required for these filters.

twee bijzinnen; gebruik van "we" An example that illustrates the benefits of a flexible software solution, was the ease with which the second-stage PPF, which was originally designed to run on FPGAs at the stations, could be moved to the BG/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 second-stage PPF from the station design.

After the subbands are split into narrow channels, the remainder of the delays are compensated, by shifting the phase of each sample, also known as fringe stopping. 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, only the former is computed. Autocorrelations are computed as well, but treated separately
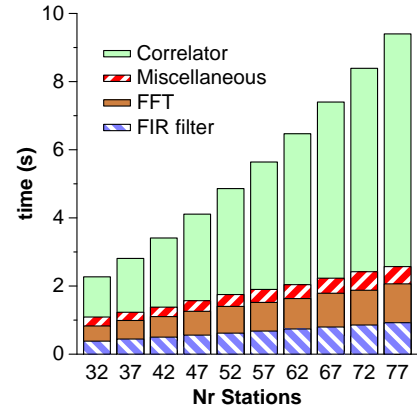

Fig. 7. Execution times to process one second of subband data on one processor core, for different numbers of stations. Multiple cores are necessary to process one subband in real time; consecutive seconds of subband data are round-robin distributed over the cores.

since they require only half the amount of computations. The computational requirements of the correlator grow quadratically with 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 [6].

The correlated data are sent to another cluster, the storage section, and are written to disk. This is where the real-time pipeline ends. Eventually, the storage section will be able to store about one Petabyte of data, so that after an observation, several days are available to calibrate and image the data.

IV. CENTRAL PROCESSING: CALIBRATION

The offline processing of LOFAR data has to deal with a number of challenges [8], [9]. First of all, the data volumes are huge. 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 have a high degree of instrumental polarization that varies with scan angle, 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 observes there are 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.

Ronald: 3x "we" An introduction to signal processing for radio astronomical arrays can be found in [10], [11], [5]. 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. Therefore, the offline processing is still 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 determined by the total number of stations that are used in the observation. This number depends on the particular mode of observation. The correlator produces a data stream of the order of a few Gbyte/s, which yields of the order of several tens of Tbytes of data after a typical observation of four hours. 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 approximately 4 hours are available to further process the data offline. With this in mind data I/O becomes an issue. Obviously the data needs to be processed in a parallelized and distributed way minimizing the I/O that is needed [12], [13].

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 allowing estimation of parameters using data that is distributed over several nodes.

The combining of solver equations, however, involves far less data I/O then the underlying observed visibility data.

Ronald: 2x "we" 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, environmental parameters, and source parameters. At its heart lies the "Measurement Equation" that is used to model the observed data [14]. A detailed description of all steps involved can be found in [15]. A signal processing data model and a Cramer-Rao lower bound analysis are given in [16]. The latter paper also provides a good introduction to the signal processing aspects of LOFAR Self-Calibration.

The current LOFAR calibration strategy consists of the following steps. The first step consists of removing bad data points, which are due to e.g. 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 is further integrated. The final resolution is determined by bandwidth and time-average smearing requirements that follow from the desired FoV and the maximum baseline [17]. In the frequency direction the data may be reduced by maximal one order of magnitude. In principle the data is also integrated along the time axis. Here, however, the effect of the ionosphere has to remain constant over the integration period. The maximal reduction factor determined by time-average smearing is also maximally one order of magnitude.

Next an iterative loop, dubbed the "Major Cycle", is entered where we first estimate instrumental, environmental, 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 in order to iteratively refine the estimates [9].

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. It is the direction dependent estimation of ionospheric parameters that is the most challenging part of this estimation problem.

In [16] it is shown that the unconstrained direction dependent calibration problem is ambiguous. However, three physical contraints to get an unambigous solution are presented:

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 simpler problem. Van der Tol et al. [16], [18] show that in this case the remote stations can be calibrated, provided the number of independent calibration directions is less then the number of core stations.

In the second approach, use is made of the fact that the effect of the ionosphere has a predictable frequency dependence [18]. 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 promising in this respect [19].

In the third approach, multiple samples in 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. Here use can be made from prior knowledge that not all parameters vary on the same time and frequency scales. In [16] it is reported however that this approach needs good initial estimates, since for instance the continuous phase polynomial is ambiguous to integer multiples of 2π.

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 have 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 the contributions from the strongest sources in the FoV are removed from the visibility data. The remaining residual visibility data is then corrected and imaged.

One visibility sample contains the combined contribution from all sources in the sky. Since LOFAR has a large 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 [17]. However, the non-coplanar baseline problem is better solved by the w-projection algorithm [20]. 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 source extraction algorithms will be used to estimate source parameters. This would then lead to an updated source model and a new cycle of the Major Cycle is entered.

V. CURRENT STATE AND FUTURE WORK

A. Core Station 1

Currently, four partially-built stations, also referred to as Core Station 1 (CS1), are functional [21]. One of those stations contain the hardware to observe with 48 LBAs or with 6 HBA tiles and 30 HBA elements. The other 3 stations are equipped with 16 LBAs and 4 HBA elements each. To create more baselines and achieve better spatial coverage, all stations 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.

The WAN infrastructure from the field towards the central location is realized for all four stations. At the central location, part of the input cluster and offline cluster are built, to match with the data streams coming from the field. These, clusters will be scaled up when more stations become available. Six BG/L racks are available for the main part of the online processing. The BG/L is capable of handling all foreseen future data rates.

Many LBA commissioning observations have been done so far. The development of the HBA was finished last year and the commissioning of the HBA tiles are still ongoing. Currently all the equipment in the field is operational.

Figure 8 shows a series of images that were made from data using CS1. In total 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").

Calibration is performed in two steps. In the first step, a point source model is used for both CasA and CygA, both at 20000 Jy flux and no polarization. An analytical beam shape is used and we solve for a single complex gain per station for the whole sky. In this way an estimate for the instrumental complex gains (e.g. due to clock drifts) and ionospheric phase differences is obtained. After correcting the data, a second step is performed where we estimate the complex gains 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 dominate 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 some 100 other sources become visible. This is shown in the right panel ("residual").
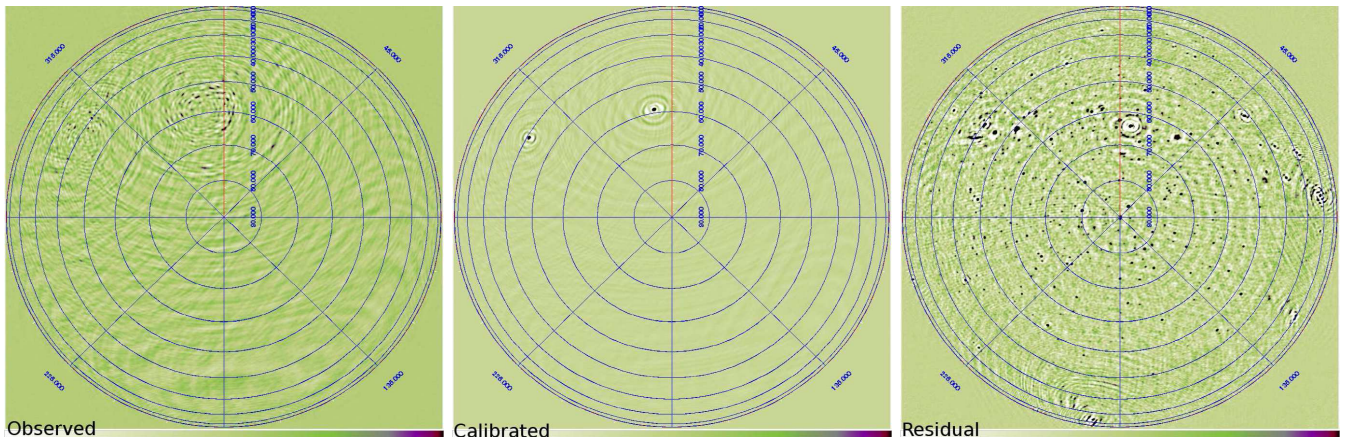
Fig. 8. All sky images from the LOFAR CS1 configuration using 48 hours and about 20 subbands of data. Observed: an image of the flagged, non-calibrated data. Calibrated: an image of the flagged, calibrated data showing CasA and CygA. Residual: an image of the flagged, calibrated data where CasA and CygA are removed from the data. Images courtesy of S.B. Yatawatta.

B. International stations

niet in een zin There is a growing international interest in building stations as well in order to achieve longer baselines. In this context the first LBA station, consisting of 96 antennas is currently operational in Effelsberg (Germany) and will be connected to the rest of the LOFAR system. Two additional German stations will be installed in Garching and Tautenburg. Further stations are planned in Germany (Potsdam), Sweden, France, and the UK and interest is shown by Poland and Italy.

For the international stations also stand-alone modes will be supported. In this way those stations can participate in the full LOFAR configuration or operate in a local mode.

C. Scaling LOFAR

In the course of this year 18 full stations will be manufactured and installed in the field. Also the WAN infrastructure and central processor facility will be scaled up to handle the data of the 18 full stations and the international stations.

The remainder of the stations will be built in 2009. Currently the software development is focused on the imaging mode. In the course of this year significant effort will be put in realizing the tied-array beamforming mode.

VI.

ACKNOWLEDGMENT

REFERENCES

[1] H. Butcher, "LOFAR: First of a New Generation of Radio Telescopes," *Proceedings of the SPIE*, vol. 5489, pp. 537–544, October 2004.

[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] 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.

[4] 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.

[5] 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.

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

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

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

[9] 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.

[10] 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.

[11] ——, "Array signal processing for radio astronomy," *Experimental Astronomy*, vol. 17, June 2004.

[12] 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.

[13] 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.

[14] 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.

[15] J. E. Noordam, "Lofar calibration framework," ASTRON, Tech. Rep. LOFAR-ASTRON-ADD-015, 2006.

[16] 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.

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

[18] 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.

[19] 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.

[20] 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.

[21] 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 is a digital system engineer at ASTRON and currently responsible for the development of the LOFAR system. He obtained a degree in Electrical Engineering from Twente University, Enschede, the Netherlands in 1999. Since, 2004 he worked on the development of the station systems in LOFAR. In 2006 he became responsible for the development of the overall LOFAR system for the astronomical applications. His research interests include (digital) system design and digital signal processing.

Ronald Nijboer Blah blah blah.

John W. Romein is a senior system engineer/researcher high-performance computing at AS-TRON, 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.