Astronomy & Astrophysics manuscript no. aartfaac_sys_design
July 19, 2016
©ESO 2016

Design and Commissioning of the AARTFAAC all-sky monitor

Peeyush Prasad[1,2], Folkert Huizinga[2], John Romein[1], Daniel van der Schuur[1], and Ralph Wijers[2]

[1] Universiteit van Amsterdam
[2] ASTRON, The Netherlands Foundation for Radio Astronomy

Received <date> / Accepted <date>

ABSTRACT

The Amsterdam-ASTRON Radio Transients Facility And Analysis Center (AARTFAAC) array is a sensitive, all-sky radio imager based on the Low Frequency Array (LOFAR). It generates images of the low frequency radio sky in near real-time with spatial resolution of 10s of arcmin, MHz bandwidths and a time cadence of a few seconds. The image timeseries are then monitored for short and bright radio transients. On detection of a transient, low latency triggers will be generated for LOFAR, which can carry out follow-up observations. In this paper, we describe the implementation of the instrumentation, and its capabilities.

Key words. Radio Interferometry - Imaging - Radio Transients - Correlators

1. Introduction

- What are celestial transients
  Transient astronomy deals with the detection and characterization of celestial transients, sources in the sky whose detectable properties can change on short timescales. These explosive events provide insight into a variety of astrophysics, ranging from emission mechanisms of jets to properties of the intervening medium [ref]. There is a rich history of the detection of such objects across wavelength ranges, with each wavelength regime probing a different parameter space [ref].
- Current state of radio instrumentation
  Studies of short duration radio emission from such objects has been restricted to either very short timescales (milliseconds to seconds, e.g. Pulsars), or to comparitively longer timescales (months to years) primarily due to instrumental or observational time constraints, the latter due to the narrow fields of view. Radio instrumentation is available in broadly two classes; single dish or phased array beam formed timeseries characterized by high time and frequency resolution, fields of view of a few degrees but poor spatial resolution. Aperture synthesis imaging observations address the latter to provide high spatial resolution, but have poor time resolution, typically needing several hours of observation time to build up adequate coverage in the UV plane via earth rotation aperture synthesis. The above classes roughly translate to coherent (short timescales) and incoherent (longer timescales) sources of emission.
- Current state of the science
  The serendipitous discovery of a new class of radio transient termed Fast Radio Bursts (FRBs) has galvanized interest in the field. The detected FRBs are characterized by high associated dispersion measures, high brightness and short timescales. They are non-repeating for the most part. Their unknown origin requires not only their discovery, but also rapid followup over a large wavelength regime to establish emission phenomena and associated parameters.
  The last requirement has led to the development of large field of view radio sky monitors, with an aim of continuously surveying large parts of the visible sky with shallow sensitivity and at high time resolution. A trigger is generated on the reliable detection of a transient in close to real-time, allowing other telescopes to carry out follow-up observations.
- Suitability of low frequency observations
  <TODO: Paragraph about transient sources, some stuff about spectral indices of coherent emission and expected class of sources, at what brightness levels can we expect to see things (take from Lazio LWA paper).>
  <TODO: Paragraph summarizing the current state of knowledge of low frequency transients: Stewart NCP transient, other searches at low freq. Conclusion: Need for more monitoring.>
- What is the AARTFAAC
  The AARTFAAC radio transient monitor is an All-sky radio telescope based on LOFAR. Its goal is to continuously scan the skies for bright transients, and on reliably detecting one, to generate a trigger to other telescopes for sensitive, broad band monitoring. It is a leading effort among a group of new radio telescopes aiming for detection of bursts of radio emission by continuous monitoring of the low radio frequency sky. Such telescopes are characterized by having moderate resolution and sensitivity as compared to contemporary telescopes, but with extremely wide fields of view (typically all sky), high availabilities and autonomous calibration and imaging in near real-time.
  The latter requirements make their implementations challenging. The antenna elements used to achieve the wide fields of view are typically dipoles, however, their low individual sensitivities requires an order of magnitude larger number of elements in the array. Bringing the resulting large number of data streams to a central location, as well as their correlation for carrying out aperture synthesis imaging thus poses a significant I/O and compute challenge. Further, the wide fields of view at the sensitivities of operation also result in direction dependent effects on the incoming signals, mostly due to the ionosphere. These pose a challenge to calibration, especially when carried out in an autonomous manner.

Apart from its primary goal of trigger generation on the detection of transients, the telescope products find use in a variety of science cases. These include wide field ionospheric monitoring via apparent flux and position variations of calibrator sources, Solar monitoring, RFI surveying, LOFAR beam model validation etc.

– AARTFAAC as a data transport, reorganization and computing problem.

The wide field of views necessary for an instrument like AARTFAAC can be achieved by sampling the sky with wide field dipoles. This, however comes at the cost of lowered sensitivity per receiving element. An instantaneously well sampled UV plane is needed to generate a PSF with low sidelobes. Both requirements can be met by spatially spreading a large number of dipoles. The highest sensitivities can also be achieved by the coherent correlation of the incoming signal, requiring access to the nyquist sampled signal at full resolution. Such an arrangement then requires the aggregation of high bandwidth data from the receiver elements, necessitating a high speed data network.
The incoming sampled voltages pass through various signal processing blocks, resulting in the generation of light curves for sources in the image. The estimation of spatial coherences requires the reordering of data to make optimum usage of compute resources. Thus, the functioning of the telescope depends on the optimization of the data transport, data reorganization and computing using available resources. An advantage of having an operating model consisting of signal processing blocks operating on high resolution data is the configurability of the telescope into different observing modes, as well as the tapping off of data from an upstream location. The latter ability makes a piggy-back instrument like the AARTFAAC possible. An important resource to optimize is the development time for each data routing or processing block, and this has been taken into consideration in the AARTFAAC.
In this paper, we describe the implementation of the instrumentation for the AARTFAAC array, and the commissioning of its various subsystems. Section 2 describes the array and the receiving antenna elements, its relationship with LOFAR, and introduces the full architecture of the instrument. Section 4 describes the hardware implementation in the field which allows creating a data path in parallel to LOFAR. This makes AARTFAAC processing independent of LOFAR to a large extent. In Section 5, we describe the implementation of a real-time, GPU based correlator for AARTFAAC, while Section 6 details the real-time, autonomous calibration and imaging implementation. Section 7 describes our control system for the full instrument, which also interfaces with LOFAR. In Section 8 we present performance metrics of the instrument as a whole.

2. The AARTFAAC array

We begin by summarizing the subsystems of the LOFAR telescope relevant for AARTFAAC processing in Section 2.1, and then elaborating on the scheme for creating a coupled data path for independent processing by AARTFAAC.

2.1. LOFAR telescope architecture

The LOFAR telescope **??** is a new generation radio interferometer covering the frequency range from 10-90 MHz using inverted V-dipoles known as Low Band Antenna (LBA), and from 110-240 MHz using Bowtie dipoles, also known as High Band Antenna (HBA). The antenna are linearly polarized, being made up of orthogonally placed dipoles in the E and H plane. The LBA dipole has a sensitivity pattern with a 6dB field of view of about 120^o, while the HBA dipoles first undergo an analog phasing within a 4x4 tile, which results in a field of view of about TODO. Due to this restriction, the AARTFAAC array utilizes only the LBA part of the telescope.

The telescope itself consists of a large collection of antennas, spatially organized into several 'stations', each spread over 60. The stations are laid out in a dense core: 24 (TODO: Check) stations within a 2km radius, while the long baselines of stations of upto a 1000km are also present. At the station level, the received and conditioned analog signals from a dipole are baseband sampled with a 200MHz clock and 10-bit quantized.(TODO: Check). The signal from each polarization is then split into spectral subbands of 200kHz via a polyphase filterbank implementation. In the regular LOFAR station level processing, the dipole subbands are then digitally phased in hardware towards the direction of an astronomical source to form a station beam, which is then transmitted over optical fiber for further interferometric processing with other stations.

A schematic representation of the LOFAR level processing is shown in Figure. TODO

2.2. The AARTFAAC system

The AARTFAAC array consists of 12-stations from within the core of the LOFAR telescope, with interdipole distances ranging from (TODO) within a station, and a maximum of TODO across stations. Due to the requirement of dipole level data in order to achieve all-sky imaging, the AARTFAAC creates a coupled data path to an independent processing architecture, prior to the phasing up of the dipoles in the LOFAR processing flow. This allows simultaneous observing with LOFAR, leading to high availability of the AARTFAAC system. A subset of the available subbands are correlated in a dedicated GPU based correlator in real-time. The estimated visibilities are sent over TCP/IP to dedicated servers for carrying out the autonomous and real-time calibration and imaging. The generated images are further sent to a software pipeline for the actual detection of transients, based on comparison of the image timeseries. A (planned) trigger generation subsystem will publish reliable triggers in the form of VOEvents [refTODO], which can be claimed by other telescopes.

2.2.1. Array configuration

The choice of stations is dictated primarily by imaging quality and sensitivity, but also due to constraints on the latency of calibration and imaging. The central six stations of the LOFAR telescope (called the superterp) form a densely sampled UV plane,and are ideal for wide field imaging due to their being coplanar to high accuracy (centimeter level). The outer six stations provide higher sensitivity and resolution. The salient features of the LBA_OUTER station configuration for the chosen stations are shown in table TODO. Figure 1 shows the LOFAR stations that are part of the AARTFAAC system.

TODO: Add 12-station beam characteristics, expected confusion noise contribution.

The station consititute the first component of the radio sky monitor, and are the only components shared with LOFAR. The AARTFAAC monitor consists of further subsystems which are

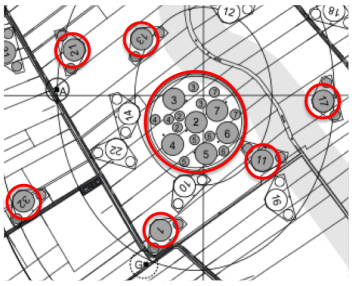

Fig. 1: The spatial distribution of AARTFAAC-12 stations within the core of LOFAR stations.

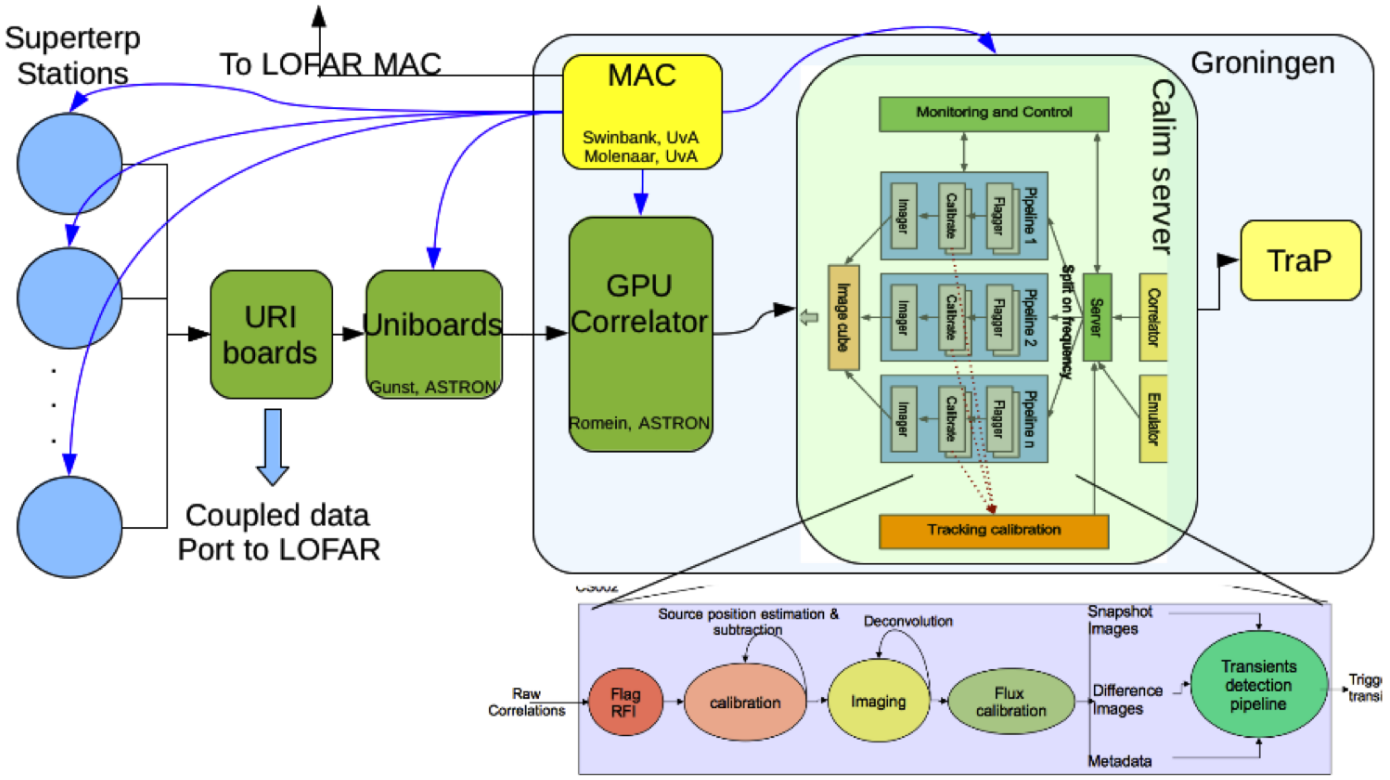

Fig. 2: Overall architecture of the AARTFAAC all-sky monitor depicting each processing subblock.

independent of LOFAR processing. Its overall architecture is shown schematically in Figure 2, and illustrates the main processing subblocks of the instrument. A user selectable subset of subbands from every dipole is transferred as UDP packets over a dedicated 10Gbit fiber connection to the central processing systems. These are received by a streaming software correlator implementation which aligns the data and estimates the spatial covariance matrix between every pair of dipoles. The generated visibilities are streamed over TCP/IP to a calibration and imaging pipeline component which carries out autonomous imaging. The images are then analyzed by a software tool (The Transients Pipeline, TraP), which extracts the light curves of sources within the image, and analyses them for variability using a number of parameters. The TraP is described in more detail in **??**.

The specifications of the AARTFAAC monitor are listed in Table TODO. We describe the various subsystems making up the AARTFAAC All-sky monitor in the following sections.

3. Remote Station Processing:

The Remote Station Processing (RSP) board handles the reception of sampled voltages from the dipole antennas. It carries out the first stage polyphase filter bank implementation common to the LOFAR and AARTFAAC telescopes. The above processing is carried out in the Remote Station Processor (RSP), an FPGA board which can handle the beamforming of TODO: X subbands for 4 dipoles. The further stages of beamforming are carried out by distributing the 4-dipole beamformed data over a station level ring network.

Available bit modes: The system offers the ability to trade off dynamic range in the polyphase filter bank outputs with the number of subbands available for further processing. The choice between the various bit modes depends on the RFI environment of the observation. An 8-bit complex representation of the filterbank outputs is found to be adequate for almost all observing conditions except during severe RFI.

Ring network sharing for dipole level data products:

Figure 4 depicts the station level ring network, whose bandwidth is shared between the beamformed subbands as well as the

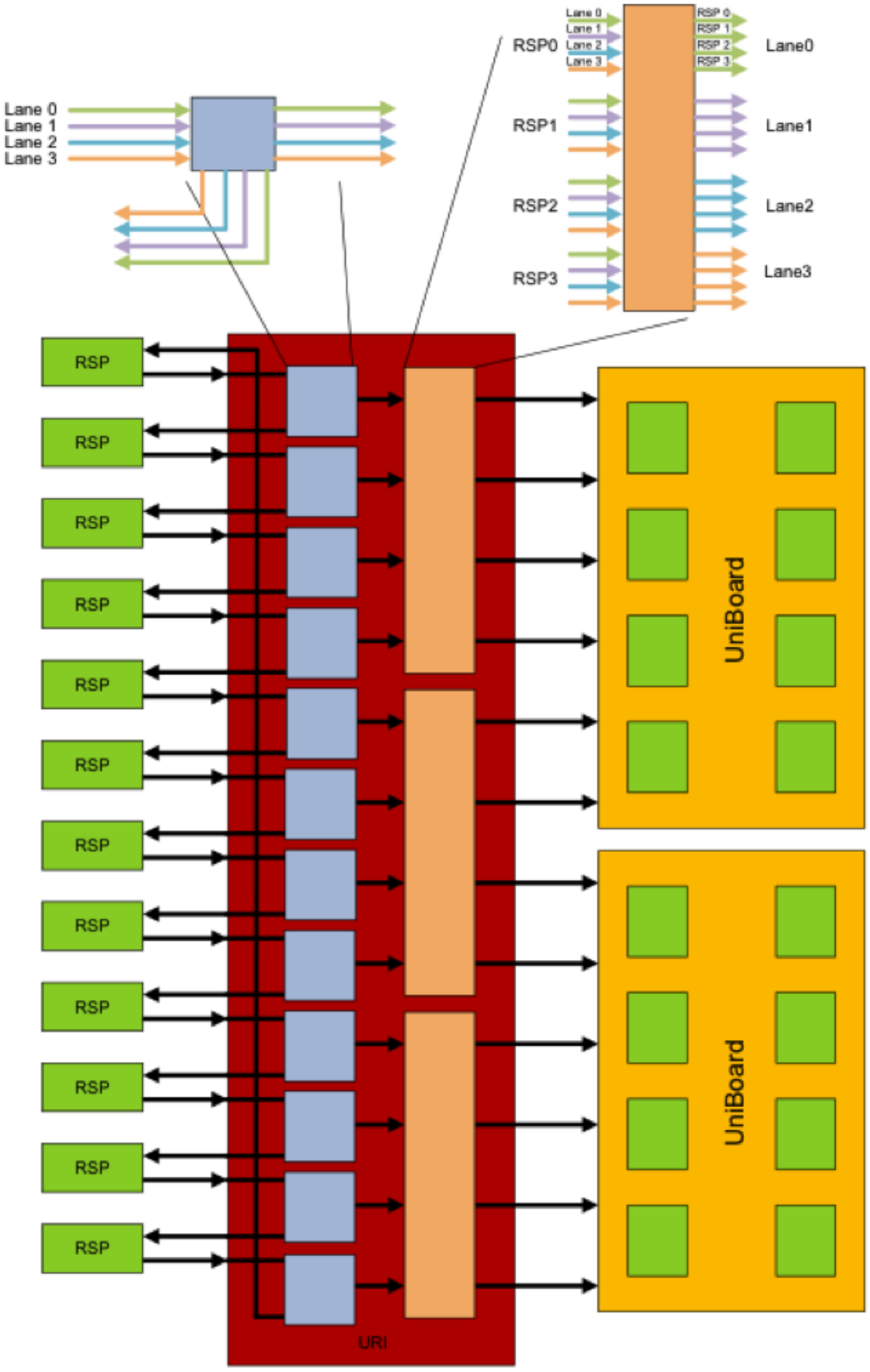

Fig. 3: The station level hardware changes allowing creation of a coupled data path for AARTFAAC data flow.

dipole level subbands. The ring network bandwidth constrains the processed AARTFAAC bandwidth to a fundamental maximum of 64 4-bit subbands, or about 12MHz. The URI boards in combination with the uniboards carry out a first level of the incoming data transposition.

3.1. Impact of sharing dipoles with LOFAR on AARTFAAC

LOFAR operates using either the LBA or the HBA antenna at a time. Further, due to limited station level electronics for stations within the core, only a subset of the available station dipoles can be utilized. This implies that the AARTFAAC telescope is de-

pendent on LOFAR for the choice of antenna and station configuration, reducing the availability for all-sky monitoring. Within the station, only the LBA_OUTER station configuration is currently deemded suitable for real-time imaging. This mode of LOFAR operation favorable to AARTFAAC depends on the observing schedule and the proposed observations. Table TODO shows some statistics from previous cycles on the fraction of observing modes favorable to AARTFAAC. Based on this, it may be resonable to expect AARTFAAC to operate TODO fraction of time, typically.

4. Station level hardware for piggy-back operation

- URI board description, data coupling scheme, constraints on achievable bandwidth
- The uniboard data reformatting (transpose), uniboard data transfer, output data format
- Available diagnostics, performance, commissioning tests

5. The AARTFAAC real-time correlator

The correlator subsystem estimates the spectral coherence between all pairs of the 1156 polarizations of the drooped dipoles, per subband. This is done by computing the average cross correlation between antenna polarization per subband. Its input is formed of the subbanded complex voltage timeseries from each polarization of every antenna. The output consists of a stream of visibilities with a chosen time and frequency averaging.

Correlators have a high input bandwidth and compute requirement, with a low compute per byte footprint. In the recent past,

- Enormous I/O requirements. Is a streaming, real-time application with large - I/O requirements. This can be a problem for implementation in many core - hardware. - Computational demands grow quadratically. Challenging computing step, since computation grows quadratically with number of inputs. - Recent trend is to correlate in software instead of hardware, due to flexibility and - to reduce development effort. - I/O requirements: 1156*8*16*195312.5 bits/sec. - Output resolution: Limited by time and frequency smearing effects, described more in a forward section. - Justification for correlating on a GPU as against special purpose hardware, or super computers (take from a GPU paper). - Logical blocks of the correlator: The correlator consists of the following logical blocks: - The input section: Job is to receive the incoming data. - The ring buffer for alignment of input streams: Forms a time aligned set of input data. All incoming data packets are copied into a slot in memory based on the timestamp from the data. - A second level polyphase filterbank is applied to the incoming subbanded data. This provides the necessary input frequency resolution, and also allows to carrout the subband amplitude modulation caused as a side effect of the first stage poly-phase filter bank. The correction is based on a theoretical response to the PFB. - The input is then passed on to the GPU via a host-device transfer, to carry out the actual correlation. The host-device transfer is on the critical path from latency and throughput perspective. - The combination of two stations is called a baseline, total number of baselines is N*(N+1)/2. This includes autocorrelations. - The correlator can operate on different input data modes. We describe the results in terms of 16-bit complex inputs. We address the question of scaling up the system in section [forward ref]. - Description of the operation of correlation. - Description of the output products. - Choice of doing floating point operations (get info from e.g. D'Addario) - How many FLOPS/byte of incoming data? (arithmetic intensity). - Memory organization in host memory. - Memory organization in device memory. Any optimizations for reduction of memory loads. - Tuning of tile size to implementation architecture. Make the tilesize as large as possible while fitting in registerspace for maximum utilization of loaded data. - Description of tile selection of data and resulting arithmetic intensity. - Table of properties of the chosen architecture (like table 3 of van niewpoort and romein, many core correlator architecture.) - Ratio between FLOPS and Bytes/sec of memory bandwidth: indicator of performance of memory system. - Performance is bound both by theoretical peak performance, and the product of memory bandwidth and arithmetic intensity. - Mention theoretical peak performance, get teh actual value from John Romein (See Sec. 3.6 of Nieuwpoort paper), what fraction of peak performance is achieved? - Kernel performance? - Achieving GPU theoretical performance depends on setting up an adequate tile size, this in turn depends on the nnumber of available registers, and the host-device bandwidth (?).

– Code related summary points – - Overall program architecture: What does the CPU do, what gets offloaded to the GPU? - What tile size is used? - Are the XY and YX hands calculated? What does -m9 do? - Buffer sizes, memory footprint?

- Correlation for transit mode observations: logical blocks.
- Description of processing flow.
- Motivation of chosen architecture for implementation.
- Supported time and frequency binning, motivation of choice.
- Required compute and memory bandwidth.
- Synchronization of incoming data (input buffer), output data format.
- Commissioning tests, performance.

6. Real-time calibration and imaging

- Architecture, implementation choices, performance
- Visibility and image buffering strategy for followup analysis of detected transients
- Quicklook images data path
- Unit test architecture
- Interface to TraP

7. The AARTFAAC control interface

Figure 5 shows the functional blocks of the AARTFAAC control system, and their interface to the LOFAR scheduling system.

- Control system description
- Interface with LOFAR
- Monitoring interface: AARTFAAC webpage

8. Commissioning results

- Overall latency of the system
- Long term performance of the entire system based on logs.
- Performance in various bit-modes, with different number of subbands, expected sensitivity.
- Imaging quality Vs. latency: 6 station to 12 station.

9. Discussion

- Long term operations of the instrument.
- Triggering mechanism: VOEvents?
- False alarms (?): Known sources of transients
- RFI situation

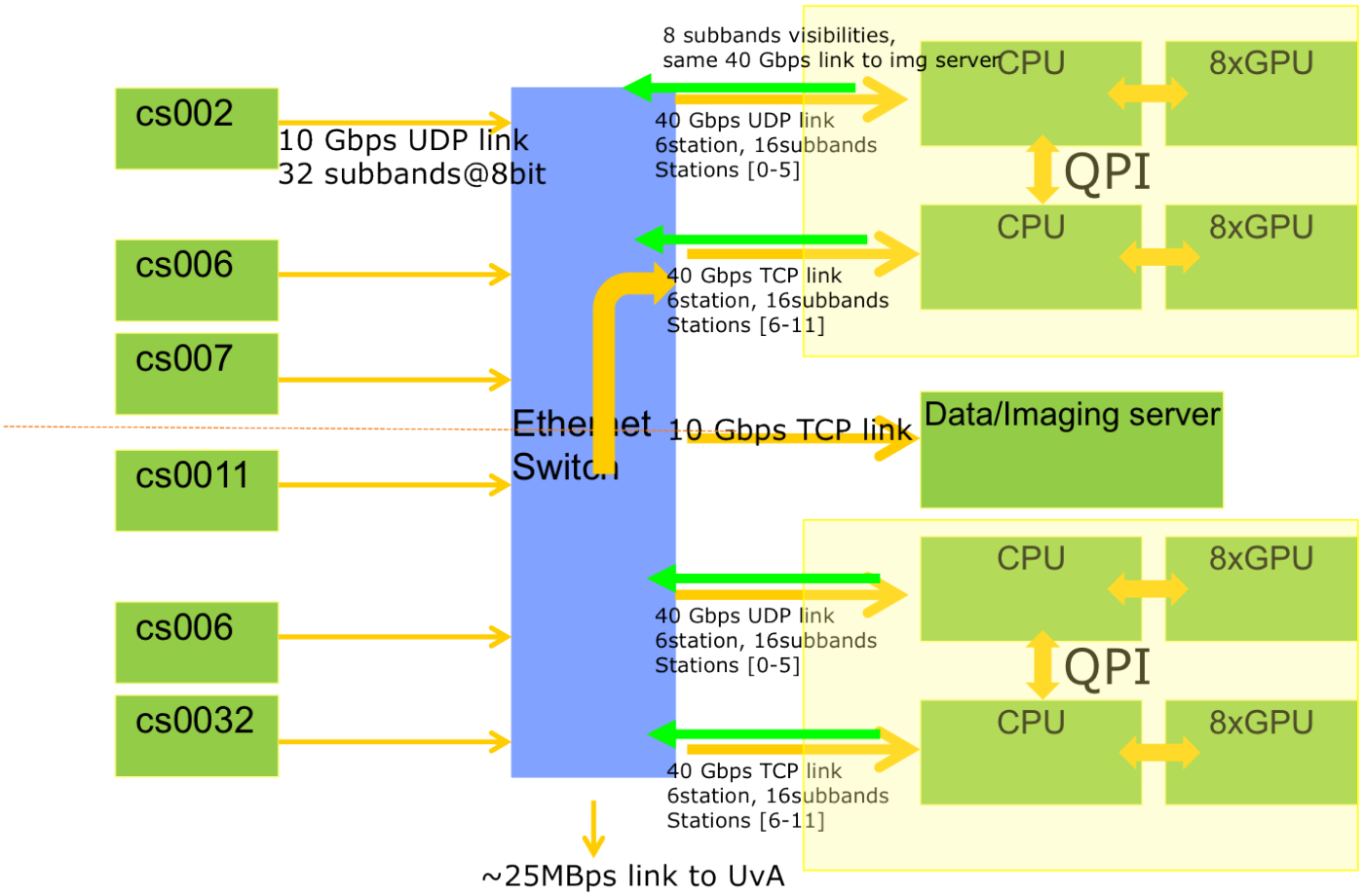

Fig. 4: The GPU correlator implementation using a pair of AMD(?) CPUs hosting AMD GPUs

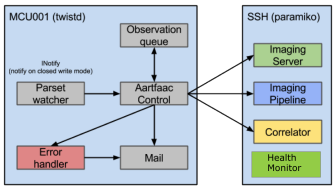

Fig. 5: The control system architecture which interfaces with the LOFAR observation scheduling system and triggers AARTFAAC observations.

10. Conclusions

Acknowledgements. This work was funded by the ERC grant <num> awarded to Prof. Ralph Wijers, Universitiet Van Amsterdam. We thank The Netherlands Foundation for Radio Astronomy (ASTRON) for support provided in carrying out the commissioning observations.