Large-N Correlator Systems for Low Frequency Radio Astronomy

Griffin Samuel Foster
St. Peter's College
Universitas Oxoniensis
Advisor: Steve Rawlings and Mike Jones

A thesis submitted for the degree of
Doctor Philosophiae
Hilary 2013

To my father, Don, and Steve who have left us during these few short years
but whose presence will be eternally felt.

Acknowledgements

The past few years have been incredibly enjoyable because of the people around me. Jack, Charles, and Danny (in order of relative love for 'Country Roads'), I would have given up along ago if you hadn't been around. Jack, for your cheerfully misanthropy. Charles, who may have been raised on the highveld but has shown me what it is to be an Oxford man [*jowl noise*]. Danny, your unrequested grammar, spelling, and typographical guidance is but one of your many traits we 'enjoy' (← you see that Oxford comma?).

I'd like to thank the post-docs who have infested this place. Kris, for his constant and steady study of methods in raising my blood pressure. Ian, the gentleman who crashed through Victor Maitland's window, who disabled an unmarked unit with a banana, who lured Taggart and Rosewood into a gross dereliction of duty at a strip tease establishment, who ruined the buffet at the Harrow Club this morning. The pride of our department. Aris, I am sure you do something, you are just very good at playing it cool. And that jerk Sascha.

To level 3, gender changers and all, you shabby old level. Kimon, do not fear, you will be able to turn left one day. Richard, I imagine you are somewhere reading an issue of the monocle and causing someone great ire. Thanks to Adam, David, Hiroyuki, and anyone else I conned into showing up to RadioLab. It was always just a little bit dodgy. And Zaira and her gripper, grip on you crazy gripper. gripper gripper gripper. Phil the particle physicist, I've got our ramshackle cover of 'Touch Me, I'm Sick' stuck in my head still. Come on Andrew, what is *really* going on?

I'd like to thank those brave few who made these past few years a pleasure. To those brave potatoes and Maria, of course Maria, the mistress of the tapas. Lindall, let us hope that one day there is an actual burning building so that we may finally know the truth. Zoë and that (diet) purple drank. On an somber note, let us never forget that fateful July morn of the Keyhaven Incident, all that we lost, and how our days would never be the same again. As it was so stoically put, 'abject lolz'.

Thank you to Thompson, Moran, and Swenson ; Taylor, Carilli, and Perley ; and Richard Lyons for the tomes this thesis is built upon.

And, thanks to all the help I received from beyond the DWB. Derek, I know you never sleep but I also think you have figured out how to slow down time too. The legend that is Captain Chris Williams, need we say more? Ben and Fred, like Sisyphus, eternally keeping the Stef at bay. The ASTRON crew: Andre, Eric, Gjis, Oleg, and Jan, who helped me along this path. It was there I learned, the devil always shits on the biggest pile. Whatever that means. Everyone at Medicina Observatory, living the dream. Ah, to be in Italy. Allora, Doza...

And finally, to my family - Lauren, Susan, Bobby, and Sterling - I do not spend nearly as much time with you as I would like.

Oxford, you beautiful town, I will miss you.

This is thesis.ThesisFinal

Large-N Correlator Systems for Low Frequency Radio Astronomy

Griffin Samuel Foster

St. Peter's College, Universitas Oxoniensis

A thesis submitted for the degree of Doctor Philosophiae in Hilary 2013

Abstract

Low frequency radio astronomy has entered a second golden age driven by the development of a new class of large-N interferometric arrays. The low frequency array (LOFAR) and a number of redshifted HI Epoch of Reionization (EoR) arrays are currently undergoing commission and regularly observing. Future arrays of unprecedented sensitivity and resolutions at low frequencies, such as the square kilometer array (SKA) and the hydrogen epoch of reionization array (HERA), are in development. The combination of advancements in specialized field programmable gate array (FPGA) hardware for signal processing, computing and graphics processing unit (GPU) resources, and new imaging and calibration algorithms has opened up the oft underused radio band below 300 MHz. These interferometric arrays require efficient implementation of digital signal processing (DSP) hardware to compute the baseline correlations. FPGA technology provides an optimal platform to develop new correlators. The significant growth in data rates from these systems requires automated software to reduce the correlations in real time before storing the data products to disk. Low frequency, widefield observations introduce a number of unique calibration and imaging challenges. The efficient implementation of FX correlators using FPGA hardware is presented. Two correlators have been developed, one for the 32 element BEST-2 array at Medicina Observatory and the other for the 96 element LOFAR station at Chilbolton Observatory. In addition, calibration and imaging software has been developed for each system which makes use of the radio interferometry measurement equation (RIME) to derive calibrations. A process for generating sky maps from widefield LOFAR station observations is presented. Shapelets, a method of modelling extended structures such as resolved sources and beam patterns has been adapted for radio astronomy use to further improve system calibration. Scaling of computing technology allows for the development of larger correlator systems, which in turn allows for improvements in sensitivity and resolution. This requires new calibration techniques which account for a broad range of systematic effects. And, a deep integration between DSP hardware and software data reduction into a single backend.

Contents

Glossary

σ Standard Deviation. 41

3GC Third Generation Calibration. 134

AARTFAAC Amsterdam ASTRON Radio Transients Facility and Analysis Centre. 76, 77, 101

ADC Analogue to Digital Converter. 2, 9, 27, 31, 40, 60–63, 152

ALM Advanced Logic Module. 46

ALMA Atacama Large Millimeter Array. 9, 27, 40

ASIC Application-Specific Integrated Circuit. 25–27, 36, 40

BAOs Baryonic Acoustic Oscillations. 24, 51

BEST-2 Basic Element for SKA Training II. 56, 57, 60, 63, 68, 71, 72, 74, 82, 83, 92, 94, 96, 153

BRAM Block RAM. 30, 32, 33, 37, 65, 83, 90, 91, 94, 95, 97, 98, 100

CASPER the Collaboration for Astronomy Signal Processing and Electronics Research. 56, 60–63, 65, 74, 76, 82, 89, 96, 98, 153

CDF Cumulative Distribution Function. 42

CEP LOFAR Central Processor. 78

Class A radio sources brightest radio sources in the sky, includes: Cassiopeia A, Cygnus A, Taurus A, Virgo A, Centaurus A, Sagittarius A. 12, 68, 114, 154

CLB Configurable Logic Block. 30, 37, 46–48

CMAC Complex Multiply and Accumulate. 10, 28, 36–38, 40

CMB Cosmic Microwave Background. 13, 15, 115, 122

CPU Central Processing Unit. 27

DAC Digital Analogue Converter. 61

DDC Digital Down Converter. 32

DDE's Direction Dependent Effects. 107, 108, 114

demultiplexer a device used to parallelize a single stream(at a rate of f_0) into D streams which at operate at a rate of f_0/D. . 31

DFT Discrete Fourier Transform. 104

DIE's Direction Independent Effects. 107, 117

DIF Decimation-in-Frequency. 33

DIT Decimation-in-Time. 33

DSP Digital Signal Processing. 1, 8, 9, 25, 26, 28, 30, 36, 40, 41, 46, 47, 54, 56, 60, 61, 63, 82, 152, 153

EOF End of Frame. 65, 90

EoR Epoch of Reionization. 13–16, 20, 22, 24, 40, 51, 77, 131, 151

FFT Fast Fourier Transform. 9, 10, 28, 32, 33, 41, 49, 51–54, 56, 57, 60–63, 65, 72, 78, 100, 118, 153

FIFO First In/First Out. 82, 83, 90, 93, 95, 96, 153

FIR Finite Impulse Response. 32, 65

FITS Flexible Image Transport System. 67, 119, 145, 154

FPGA Field Programmable Gate Array. 1, 3, 25–28, 30, 31, 36, 37, 39–41, 43, 46–48, 54–56, 60–62, 65, 66, 76, 77, 82, 83, 87, 97, 101, 152, 153

FWHM Full Width Half Max. 70, 128

GMRT Giant Metrewave Radio Telescope. 13

GPS Global Positioning System. 115

GPU Graphics Processing Unit. 27, 40, 54, 55, 145, 152

GRBs Gamma Ray Bursts. 17

HA Hour Angle. 116

HBA High Band Array. 21, 23, 76, 103, 104, 111, 112, 114, 116, 119, 121, 131, 132, 154

HDF5 Hierarchical Data Format, Version 5. 67, 68, 70

HDL Hardware Description Language. 25, 26, 61, 98

HEALPix Hierarchical Equal Area isoLatitude Pixelization. 108, 118, 122, 125, 145, 154

HPBW Half Power Beam Width. 116

IGM Intergalactic Medium. 13, 19, 20

IIR Infinite Impulse Response. 32

ILT International LOFAR Telescope. 21–23

ISM Interstellar Medium. 19

LBA Low Band Array. 21, 23, 76, 103, 104, 108, 109, 111, 114, 116, 119, 121, 129, 130, 132, 149, 154, 156

LNA Low Noise Amplifier. 1, 57, 104, 111

LOFAR Low Frequency Array. 11, 13, 20–24, 28, 36, 39, 76–78, 82, 91, 100, 101, 103, 104, 108, 112, 114, 116, 117, 126, 130–132, 149, 151, 153, 154, 156

LSB Least Significant Bit. 48

LST Local Sidereal Time. 104, 119, 122

LUT Lookup Table. 30, 37, 43, 46, 47, 83

LVDS Low Voltage Differential Signal. 35, 40, 60

MSB Most Significant Bit. 31, 34, 48

multiplexer a logic device used to select a single stream from D streams. . 65, 87, 89

NCP North Celestial Pole. 121

NRE Non-recoverable Engineering. 36, 60

OeRC Oxford eResearch Centre. 126

OOB Out of Band. 65, 87, 90

PDF Probability Density Function. 41, 42

PFB Polyphase Filter Bank. 9, 11, 32, 60, 63, 65, 152

PHY Physical Layer. 83

PSF Point Spread Function. 20, 71, 72, 109, 112, 115, 117, 119, 121, 122, 128, 132–134, 141, 142, 144, 149

QDR Quad Data Rate. 34, 61, 62, 65, 66, 82, 83, 89, 90, 93, 95–97, 101

RAM Random Access Memory. 30

RCU Receiver Unit. 78, 104

RF Radio Frequency. 40, 63

RFI Radio Frequency Interference. 4, 9, 27, 28, 33, 34, 40, 41, 50–52, 57, 63, 65, 68, 100, 104, 121, 122, 152

RIME Radio Interferometry Measurement Equation. 106, 107, 115

RMS Root Mean Square. 2, 44

ROACH Reconfigurable Open Architecture Computing Hardware. 56, 60–63, 78, 80, 82, 89, 98, 153

RSP Remote Station Processing. 76–78, 80–83, 87, 89, 90, 92, 100, 101

SEFD System Equivalent Flux Density. 111, 114

SEPcam SEPNET Widefield Camera. 24, 28, 76, 78, 80–83, 89, 97, 98, 100, 101, 103, 131, 132, 153, 154

SerDes Serialize/Deserialize. 78, 80, 82, 87, 89–92, 94, 100

SETI Search for Extraterrestrial Intelligence. 17, 25

SKA the Square Kilometre Array. 21, 24, 25, 52, 54, 55, 57, 77, 101, 151, 153, 154

SNR Signal to Noise Ratio. 6, 40, 42, 43, 71, 125

SPEAD Streaming Protocol for Exchange of Astronomical Data. 66, 67, 89, 96, 97

TVG Test Vector Generator. 26, 89, 98

UDP User Datagram Protocol. 67, 89, 97

URI UniBoard Ring Interface. 76, 78, 80, 82, 83, 87, 89, 100, 101

VLA Very Large Array. 13

VLBI Very Long Baseline Interferometry. 20, 141

VLSI Very Large Scale Integration. 40

XAUI 10 Gigabit ‘X’ Attachment Unit Interface. 60, 61, 65, 66, 80, 87, 89

Chapter 1

Introduction

> *You can't get started, because you can't get started.*
>
> -Don Backer

Observations at radio frequencies are deeply entwined with the instruments used to make the observation. *Interferometric arrays* are a collection of N antenna elements combined via a *correlator* to provide higher angular resolution than is available with a single component antenna element. These arrays make measurements at certain spatial resolutions which allow formation of synthesis images. Radio astronomy, a relatively young field of astronomy, less than 80 years old as of this writing, has advanced in unison with the growth of digital technology. The correlator is the central hardware used to make the interferometric measurement. Digital correlators allow the interconnect of many antenna elements into a single unified instrument. Modern correlator development is driven by the growth in Field Programmable Gate Array (FPGA) technology which increases in bandwidth and Digital Signal Processing (DSP) resources with each new generation. This has allowed for the growth in number of antenna elements in an array, processing bandwidth, along with all the advances in digital filtering techniques. At low frequencies an antenna element can be as simple as a coat hanger with a Low Noise Amplifier (LNA). These elements are cheap, low sensitivity, nearly omni-directional, and require no mechanical components. By building a large enough array with a complete digital backend these cheap, wide field of view elements can be combined to the desired sensitivity of an experiment. The challenge is shifted to the signal processing, system calibration, and imaging techniques. These stages are traditionally treated independently. But as the scale of the data bandwidth grows and the requirements of the signal measurement are raised these stages must be treated as a single digital processing unit.

There is a story which is being played out during each observation which has long been in narration. We aim to gain insight into a slice of that story with the instruments we construct. The output of our digital instruments are abstract complex numbers which somehow relate to a physical process. Our challenge is to develop an understanding of the universe by decoupling the instrument effects from the observation.

Two complex numbers represented as a base two number on a digital computer are multiplied together. The multiplication is a result of a series of logical operations performed on individual digital bits which are just analogue voltages within the error limits of a threshold. The resulting complex number is a representation of the correlation between photons arriving at two different radio antennas. Around 600 Megayears ago, electrons within the Cygnus A radio galaxy were accelerated due to the strong magnetic fields generated from the active galactic nucleus. This radial acceleration causes the electrons to emit broadband synchrotron radiation. These photons propagate away from their origin isotropically. A portion of these photons get absorbed along their paths from interactions within the propagation medium, but for the most part these photons continue on. Over following 600 Megayears the photons form an ever expanding shell of energy with the flux per unit area dropping as an inverse square law. Part of that shell of photons crosses the orbit of Earth and passes through the ionosphere on its destined path. For most of these photons it is the end of the line; they are absorbed by the Earth, a few are scattered back out to continue on some other path. Even fewer interact with the radio antennas we have placed throughout the Earth.

Fewer photons, still, induce an electrical current within the antenna. This induced current relates the energy of the photon to an electron charge, the relation between photon energy and electron charge is quantitatively understood with system calibration. Within a unit length of time the total induced current is an accumulation of all current induced by each interacting photon. This group of photons, after travelling 600 million light years from Cygnus A, has been reduced to a single voltage measurement. This signal voltage is too small to be measured reliably by our detectors. Thus at the cost of introducing additional noise the signal is amplified. Once the voltage Root Mean Square (RMS) is sufficiently high for an Analogue to Digital Converter (ADC) to sample the continuous voltage signal a discrete digital representation is produced. A sample in this discrete representation is related to the continuous stream as the average voltage measured during the characteristic sample period of the ADC. This digitization has the effect of quantizing the time and amplitude of the continuous voltage signal. That is, each digital sample is a time and amplitude average of the continuous voltage signal. Thus, each sample represents a group of electrons which in turn are related to the group of photons which interacted with the antenna which originated at Cygnus A.

We can then ask what is the relationship of signals from a pair of antennas. That is, what is the signal correlation? Each digital sample is a complex number made up of amplitude and phase. If the signals from a pair of antennas have the same phase, or are ‘in phase’, then the signals are said to be correlated. They are anti-correlated if they have opposite phases, or ‘out of phase’. Now, even though we have amplified the signals they are still weak compared to the system noise. This requires an accumulation of correlation samples to bring the correlated sky signal up above the noise.

We now have a measurement of the correlation between the photons which arrived at a pair of antennas, though the measurement is time averaged and frequency band limited. Each antenna pair, or *baseline*, is sensitive to a specific spatial component based on their orientation to the source’s position in the celestial sphere and the distance in which they are separated. The longer the baseline separation the faster the phase difference between the two measurements will change or *fringe* and the more sensitive the baseline is to the higher frequency spatial components. By computing the correlation between the antenna pairs in an array we are decomposing the voltage signals into spatial frequency components. The measured spatial frequency components can then be transformed as a whole into a spatial image. This image is composed of the various spatial frequency measurements, that is the correlation of digital signals between antenna pairs. Where each digital signal is related to the current which was induced by the interacting photons from the synchrotron emission generated in Cygnus A over 600 Megayears in the past. The insight into the photon energy density in turn gives us insight into the original physical processes which created this perceived emission, and that is our main goal.

Untangling the effect of each stage that leads from the original photons to the final image is the prime challenge of any measurement. Our ability to separate and mitigate each stage leads to a better understanding of the original processes. This chapter will lay the ground work of the correlation mathematics and their implementation in digital hardware. We will then discuss the science goals and advantages of using radio interferometric arrays at low frequencies. An overview of large-N arrays in which the context of this thesis can be placed will be presented. And finally, the advantages of FPGA-based hardware for digital correlation will be discussed. These topics will lay the ground work for the topics presented in the body of this thesis.

1.1 Cross Correlator Systems

Cross correlator systems are the heart of an interferometric array. They are used to compute the *correlation* between antenna pairs which is a measure of their interferometric response to sky signal. A radio interferometer measures an angular scale on the sky based on antenna spacing and orientation, and the observing frequency. Each correlation is thus a measure of a ***baseline*** signal spatial coherency. This may appear a vague statement, but is key to understanding how we can make measurements with interferometric arrays. We need to delve into the questions of what is a signal's spatial coherency and why is it an important observable?

1.1.1 Signal Coherency

The coherency of two signals quantifies how well one signal can be predicted based on the value of the other signal. Independent Gaussian noise signals will be incoherent as one cannot be used to predict the other. Two monochromatic sine waves at the same frequency will be coherent, if one of the signals changes the other will change at the same rate. Signals made of a superposition of multiple sine waves, noise signals, et cetera will be at some intermediate degree of coherence.

Spatial coherency is the coherency of a signal based on the relative location of two detectors, in the case of radio astronomy the detectors are antennas and the signal is a complex electromagnetic wave. Let us start with the simple case of noise free sky with only one monochromatic source in one dimension. Any astronomical object can be considered to be in the far field and not spatially coherent, i.e. we can model any resolved source as a collection of discrete point sources. This is not always the case for terrestrial Radio Frequency Interference (RFI) as local RFI can be in the near field. Figure 1.1 shows our single point source sky and the geometric delay between two antennas separated by a distance D based on the location of the source in celestial sphere (θ). The response of a monochromatic point source in the far field is approximated as a plane wave.

The geometric delay between antennas is $\tau_g(\theta) = \frac{D\sin(\theta)}{c}$, where c is the speed of light. Using the geometric delay and the observing wavelength, the relative phase offset is

$$\phi_g(\theta, \lambda_0) = 2\pi \left(\frac{\tau_g(\theta)c}{\lambda_0} - \left\lfloor \frac{\tau_g(\theta)c}{\lambda_0} \right\rfloor \right) \tag{1.1}$$

Equation 1.1 is cyclical since phase ranges $0 \leq \phi \leq 2\pi$. For the same physical distance the phase offset varies based on observing wavelength. Determining the phase offset provides the location of a source on the celestial sphere. But, the phase will generally wrap multiple times across the sky, thus the source spatial coherency is degenerate, Figure 1.2. More information is required in the form of using multiple antenna pairs or taking into account temporal coherency. As an astronomical source crosses the celestial sphere, really as the Earth rotates, the phase between antenna pairs will

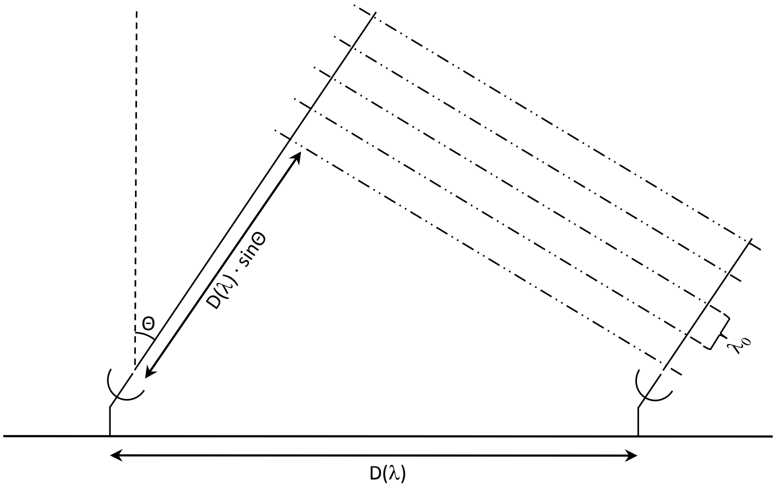


Figure 1.1: The response of a point source in the far field can be approximated as a plane wave. Based on the source position in the sky (θ) and observing wavelength (λ_0) there is time delay ($\tau_g(\theta, \lambda_0)$) between receiving elements based on their projected positions ($D(\lambda)$).

vary according to Figure 1.2 to produce a ***fringe pattern***. By measuring the spatial coherency using multiple antenna pairs a source can be localized to a single position on the celestial sphere.

In this model the source position can be easily determined with a few antenna spaces of different spacings. But, this model is not very realistic. There are multiple sources in the sky with different amplitudes and noise due to the thermal background and the receiver system. Nonetheless, the method behind measuring spatial coherence with antenna pairs remains unchanged. In a sky with multiple sources and extended sources the relative flux and positions on the celestial sphere can be determined down to the maximum antenna spacing which determines the array resolution.

The spatial coherency between a pair of antennas can be determined by computing their correlation. The correlation of two functions, $f(x)$ and $g(x)$ of real variable x, is defined as

$$(f \star g)(x) \equiv \int_{-\infty}^{\infty} f^*(w) g(x+w) dw \tag{1.2}$$

The $\star$ operator denotes a correlation. This function is very similar to the convolution function, and shares a number of properties. The correlation $f \star g$ is equivalent to the convolution $f^*(-x) * g(x)$. If $f(x)$ and $g(x)$ are complex voltages $V_i(t)$ and $V_j(t)$ received at antenna i and j, which are separated by a geometric delay τ_{ij}, then their correlation is

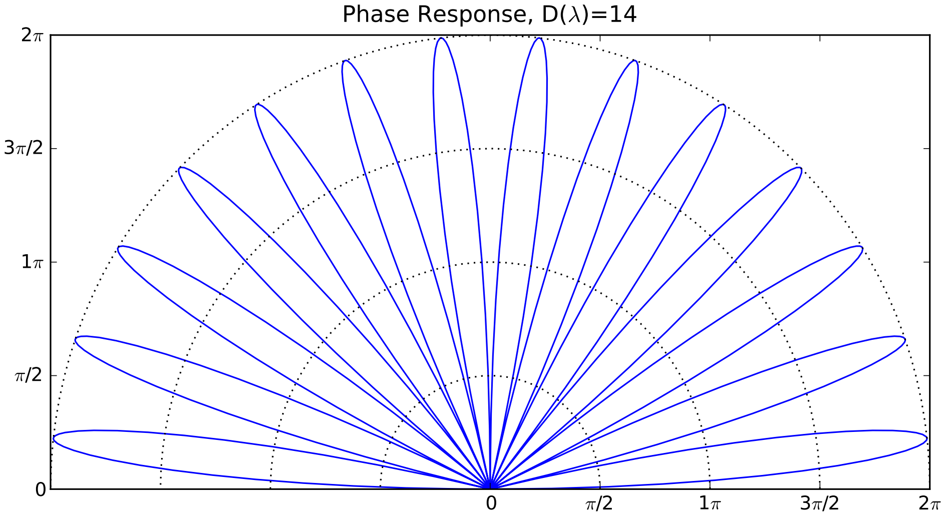


Figure 1.2: Geometric phase response (1-D) of a source on the celestial sphere for a pair of antennas separated by 14λ wavelengths. The degenerate source position can not be found without more information.

$$(V_i \star V_j)(\tau_{ij}) = \int_{-\infty}^{\infty} V_i^*(t) V_j(\tau_{ij} + t) dt \tag{1.3}$$

The voltages $V_i(t)$ and $V_j(t)$ are proportional to the complex electric field vectors $\mathbf{E}_i(t)$ and $\mathbf{E}_j(t)$ at each antenna. In practice the correlation can only be computed over a finite time range or ***integration period***, T_{int}. A finite time range is valid for computing the correlation when $T_{int} \gg \Delta\nu^{-1}$ where $\Delta\nu$ is the channel bandwidth. This is almost always the case in any correlator system as we desire to maximize the Signal to Noise Ratio (SNR) of the astronomical signal in a correlation. Though, the lower limit for integration periods is approached in systems designed to produce correlations on subsecond time scales to search for transients. The maximum integration period is set by the observing frequency, distance between antenna pairs and acceptable phase error (usually $\sim 1 - 10\%$). Equation 1.3 can be approximated as

$$(V_i \star V_j)(\tau_{ij}) \approx \int_{T_0 - T_{int}/2}^{T_0 + T_{int}/2} V_i^*(t) V_j(\tau_{ij} + t) dt \tag{1.4}$$

A constant time T_0 has been added to the range of the integral to account for the human-defined time over which the correlation occurs. In a digital system $V_i(t)$ and $V_j(t)$ are discrete functions sampled at a rate of $\Delta\nu^{-1}$. For the discrete case the correlation becomes

$$(V_i \star V_j)(\tau_{ij}) \approx \sum_{t=T_0-T_{int}/2}^{T_0+T_{int}/2} V_i^*(t)V_j(\tau_{ij}+t) = \langle V_i^*(t)V_j(\tau_{ij}+t)\rangle \tag{1.5}$$

Equation 1.5 states that the correlation of signals $V_i(t)$ and $V_j(t)$ is just their time averaged product. This is the key to digital correlator design. The right side of the equation can be implemented in a digital system as a complex multiplier and an integrator. The correlation will be maximized when the functions are matched, if the functions are independent the correlation will approach zero. This is an important observation about the correlation. If two signals are made up of covariant functions and independent noise components, then in the correlation the noise will average out to zero and only the product of the covariant functions will remain, thus eliminating a significant portion of system noise.

1.1.2 Development of Correlators

Interferometric arrays are made up of $N \geq 2$ receiving elements. Each possible pair of receiving elements, or baseline, in the array can be used as an independent interferometer. There are $\frac{N(N-1)}{2}$ unique pairs, in correlator terms these are called the cross-correlations. In synthesis imaging, a collection of baselines can be combined to form a synthetic aperture. Comprehensive coverage of synthesis imaging is presented in [Taylor et al., 1999] and [Thompson et al., 2001]. A correlator is digital or analogue hardware which computes the correlation between antenna pairs in an interferometric array. Two types of correlations are usually computed. The first is an *auto-correlation* which is the correlation of an antenna with itself and has a 'zero-spacing' baseline, which is related to a spectrometer power measurement through the Wiener-Khinchin relation [Wiener, 1930]. The second is the *cross-correlation*, this is the correlation between two unique receiving elements described in Equation 1.5. These are in the form of FX and XF architectures, explained further in section 1.1.3. Their naming is related to the order of operation for frequency channelization (F) and cross multiplication (X). Polarization has not been discussed but the conceptual framework can be extended to measure the correlation between polarizations to compute Stokes parameters. Maximum information from an array of N dual-polarization elements requires $4(N + \frac{N(N-1)}{2})$ correlations.

The first interferometer systems in radio astronomy did not compute correlations, rather the squared sum of the voltages was computed and low-pass filtered to select out only the slowly varying fringe patterns of astronomical sources. This method was based on the assumption that the voltage signal at two antennas is nearly the same, which is true for only the brightest astronomical sources. Weaker sources could not be

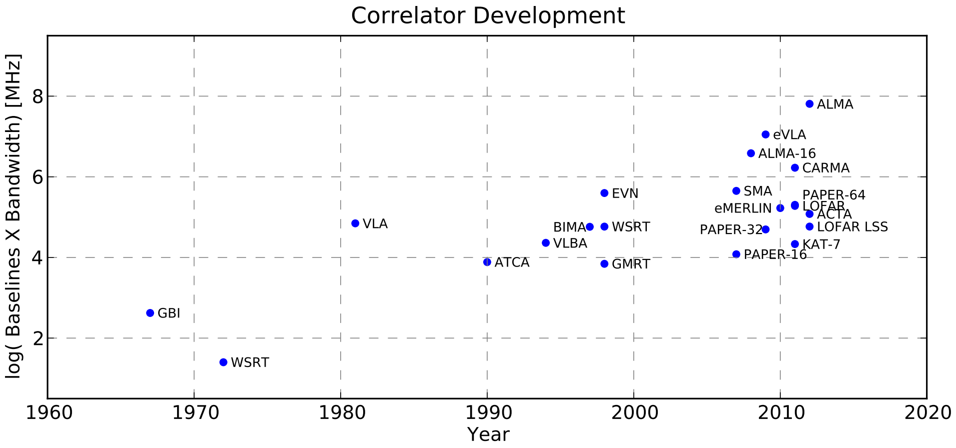


Figure 1.3: A selection of correlators developed for interferometric arrays. There has been a general growth in number of baselines and bandwidth as the technology for digital correlators has improved.

detected as they were below the electronic noise introduced during signal amplification. To counter this, *phase-switching* was introduced which would periodically flip the phase of one of the antennas by 180° [Ryle, 1952]. The interferometer hardware would then compute $(V_0 + V_1)^2$ and $(V_0 - V_1)^2$, by taking the difference of these two measurements a result proportional to $V_0 V_1$ would be produced. Phase switching became an essential tool in early interferometric work until the development of multiplying circuits. Modern systems use phase switching to mitigate antenna and electronic crosstalk effects.

By the early 1960s analogue and digital multipliers where introduced with the development of integrated circuits, this allowed direct computation of the correlation, supplanting the voltage squared sum hardware. The first use of a digital multiplier for correlation was presented in [Weinreb, 1963] which was used to compute the autocorrelation. Since then there has been a continued growth in correlator system bandwidth coverage and baseline computations which has allowed for the growth in number of antenna elements in interferometric arrays. Figure 1.3 shows the development of various correlators for radio astronomy. A metric, which is the product of the number of baselines and bandwidth coverage, shows a general increase in both over the decades. Though, this is not the perfect metric as each correlator is optimized for specific science cases which make it difficult to make direct comparisons. We have not included other considerations such as number of bits, channelization, or power usage per correlation.

1.1.3 Architectures

Correlators have been implemented using various architectures based on the available technology and science goal at the time of development. Analogue correlators consist of multiplier circuits, similar to operational amplifiers, which have a non-linear response using the Hall effect [Lofgren, 1958]. Digital correlators discretely sample the signal and use DSP logic components to perform the correlation. The digital architecture can further be broken down into DSP implementation as FX and XF designs. Most correlator designs can be separated into these categories, though there are some novel correlator-like designs such as the direct imaging correlator, an overview is presented in [Hickish, 2013].

Analogue correlators are used for large bandwidth arrays especially at high frequencies, and are implemented using a *lag* (or XF in digital) design [Holler et al., 2007]. The lag correlator design uses M time delays, or lags, for each baseline correlation. To compute the correlation in Equation 1.5, one of the antenna signals is held stationary while the other signal is delayed by M equally spaced delays and the product of the stationary and delayed signal is computed. This requires M multipliers per baseline. The product is then integrated over the nominal integration time T_{int} and written to an analogue recorder or sampled digitally and recorded to disk. The number of lags required is based on how fine of frequency channelization is required. To channelize a band of size ν_0 into channels of width $\Delta\nu$, $M = \nu_0/\Delta\nu$ lags are required. If no frequency channelization is required then $M = 1$ and the ‘zero lag’ is used. A review of lag correlator design is presented in [Taylor et al., 1999].

Analogue components can be made which cover tens of Gigahertz of bandwidth which is necessary for even small fractional bandwidths at high frequencies. Though, they are inherently variable in time, temperature, and frequency which require these systems to be constantly monitored and housed in stable environments. Generally, analogue correlators are only used for $\lesssim 10$ element arrays with coarse frequency channelization because of the cabling complexity and associated signal loss in the cables and splitters. But, the benefit of large bandwidths has made this cost acceptable. In recent years the sampling bandwidth of ADCs has increased and the cost of DSP has decreased allowing digital correlators to take over many traditionally analogue niches, e.g. the Atacama Large Millimeter Array (ALMA) correlator.

Digital correlators require voltage signals to be discretely sampled but allows correlators to scale with the Moore’s law growth of semiconductor technology. A limiting bandwidth design factor is the ADC sampling rate. The first digital correlators used ADC’s which only sampled a few Megahertz of bandwidth with low bit resolution. Currently, a typical ADC can sample at ~ 1 Gsps with 8-bit resolution [Semiconductors, 2009]. Though, there is a broad range of commercial ADC's available with sampling rates up to 10's of Gsps and bit resolutions up to ~ 16 bits. Digital correlators provide correlation on bandwidths of $O(100\text{MHz})$. This is sufficient to cover large fractional bandwidths of meter and centimeter wavelength arrays. Sampling rates are approaching those comparable to analogue correlator bandwidths and digital correlator can begin to cover high frequencies effectively. The real advantage of digital hardware over analogue hardware is the analytical known response of the hardware which is independent of operation time, environment, and frequency. Digital correlators are not plagued by many of the issues that arise in analogue correlators, though they do introduce their own unique issues. The continuous voltage samples must be discretely sampled which introduces sampling error, and at points in the DSP signals are quantized down to lower bit widths to make efficient use of the limited logic resources. Both of these issues must be taken into account during hardware design and accounted for during data analysis. The actual correlator implementation in digital hardware is based on FX and XF architectures, which as the names suggest are closely connected.

The FX correlator architecture is ideal for arrays with $N \sim O(10 - 100)$ antenna elements which require fine frequency channelization, Figure 1.4. In most radio astronomy applications the observed band needs to be channelized into narrow frequency channels. This is required to mitigate interband RFI, to reduce chromatic aberration, or observe spectral lines and a myriad of other scientific goals. The channelization is efficiently performed with a Fast Fourier Transform (FFT) which for n channels requires $O(n \log n)$ operations. The F component of the FX design represents the frequency channelization using a FFT. A Polyphase Filter Bank (PFB) is often used as the F component which uses a windowing function to suppress channel sidelobes, for an overview see [Price, 2012]. Each frequency channel can be thought of as an independent, narrow-band measurement. A correlation is then computed between every i^{th} frequency channel of each antenna pair, this is the X component. That is, in an array with N antennas and M frequency channels there are $M\frac{N(N-1)}{2}$ correlations. Though, this only requires at most $\frac{N(N-1)}{2}$ complex multipliers since each frequency channel is decimated by M time samples.

An individual correlation is implemented with a Complex Multiply and Accumulate (CMAC). The computational cost of the FX design scales as $O(M \log M + N^2)$. For $N \gtrsim 10$ the computational cost is dominated by the N^2 factor. With the FFT the F component computational cost is minimal and allows for high resolution channelization at a small additional cost. As will be noted below, the FX is a more efficient design than the XF design for large-N arrays.

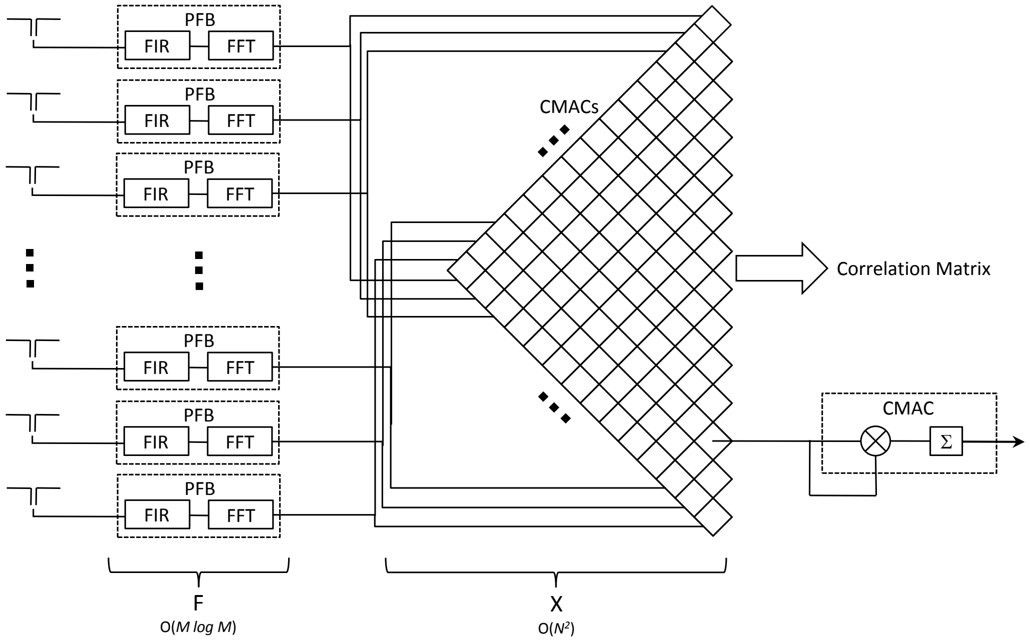


Figure 1.4: FX correlator architecture. For an N antenna, M frequency channel correlator the computational cost is $O(M \log M + N^2)$ operations. Adapted from Figure 4-5 [Taylor et al., 1999].

The XF correlator was retroactively named after the introduction of the FX architecture in [Chikada et al., 1986], this architecture is also known as a lag correlator as noted previous, Figure 1.5. The architecture originally consisted of the correlation, X, component and produced no frequency information. By introducing a range of time lags a temporal FFT can be applied after integration to produce a channelized spectrum. For an array of N antennas and M time lags the computation cost is $O(MN^2)$. The computational costs of the F component can be ignored since the signals are integrated in the X component and the FFT operation is called at a low rate. For systems with small N, i.e. $N \lesssim 10$, and coarse frequency channelization the XF and FX architectures are comparable in computational cost.

The FX and XF architectures can be shown to be mathematically equal. By the convolution theorem (Equation 1.6) the FX and XF architectures are equivalent, Figure 1.6. The Fourier transform of a function f is represented by $\mathcal{F}\{f\}$. The left side of Equation 1.6 is analogous to the XF architecture (convolution is directly related to correlation) and the right side the FX.

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\} \tag{1.6}$$

Even though the architectures produce equivalent correlations, FX tends to be the preferred architecture in modern correlator design. The computational advantage when

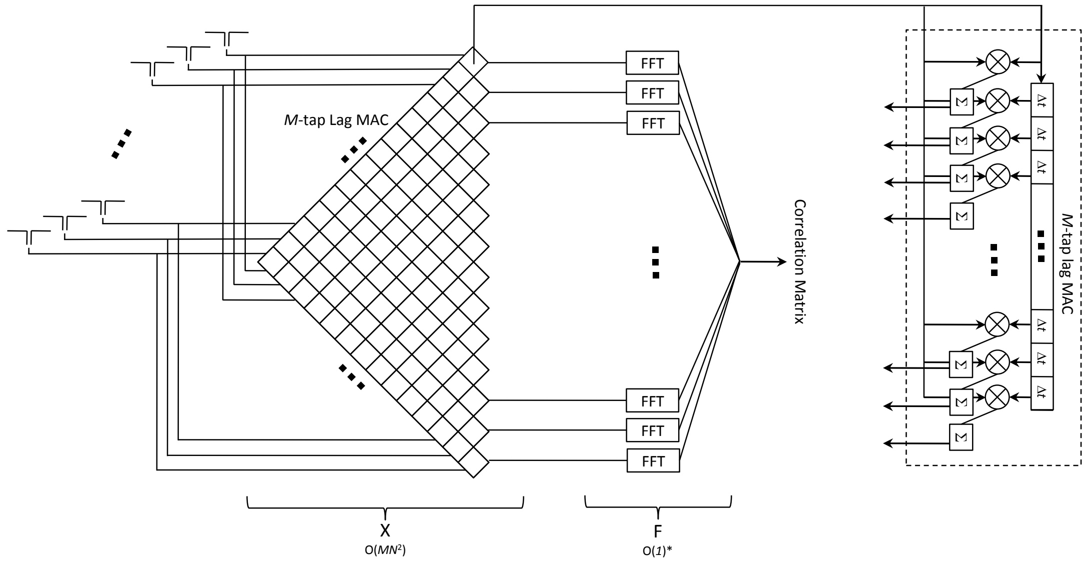


Figure 1.5: XF or lag correlator architecture. For an N antenna, M-tap lag correlator the computational cost is $O(MN^2)$ operations. The F stage computational cost is $O(1)$ since each baseline is integrated before the FFT.

the number of antenna elements grows is clear and there are additional system response benefits over the XF. Fractional sample delays can be added to antenna signals after frequency channelization by introducing a phase offset, this improves sensitivity and phase tracking. Also, FX designs have the choice of windowing functions in the PFB which allows for a better spectral response.

Based on these simple architectures hybrid correlators have been introduced such as the FXF WIDAR correlator [Carlson & Dewdney, 2000] for the eVLA. For low-frequency arrays, such as Low Frequency Array (LOFAR), aperture arrays are beamformed together into station elements and the stations are then correlated together in a BFX architecture [Romein et al., 2010].

Now that a basis for correlation and interferometry has been presented we will discuss the motivation of using interferometric arrays for observations in the low frequency band.

1.2 Science Goals at Low Frequencies

Detection of the first celestial radio sources with interferometric techniques came about in the post-war era of the late 1940s and early 1950s in a large part due to advances in radar technology. In the 1930s the first instruments of radio astronomy were single element antennas operating at long wavelength (meter and decimeter), with very large sky coverage. Karl Jansky's famous detection of the galactic center at 20.5 MHz was the first verification that celestial sources emit at radio frequencies [Jansky, 1933].

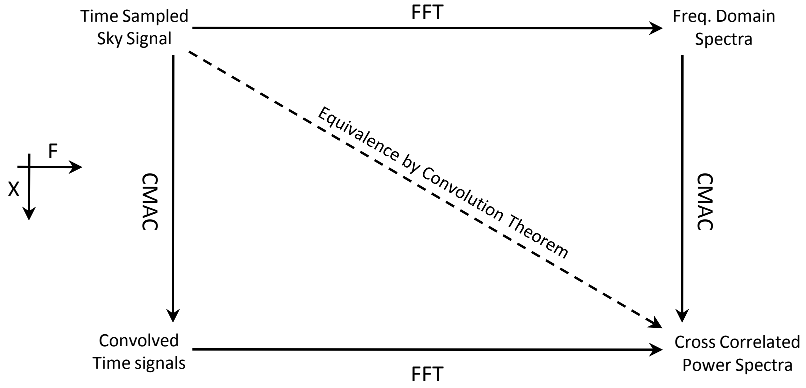


Figure 1.6: The FX and XF correlator architectures are equivalent under the convolution theorem. Thought, the order in which operations are performed incurs different computational costs and results in different systematic responses. Adapted from Figure 4-13 [Taylor et al., 1999].

Throughout the late 1930s and early 1940s Grote Reber built the first parabolic radio antenna and used it to construct a sky map at 160 MHz [Reber, 1940]. A multi-element radio interferometry, based on optical Michelson Interferometer techniques, was first used in 1946 to detect radio emission from the galaxy and measure the angular scale of the Sun [Ryle & Vonberg, 1947]. In the preceding years the first radio sky surveys were undertaken, and the Class A radio sources were observed at radio wavelengths. This so Class A radio sources of sources is made up of the brightest radio sources which includes Cassiopeia A, Cygnus A, Taurus A, Virgo A, Hercules A, and Centaurus A. With the desire to better localize weak sources new interferometric arrays shifted towards designs with tracking antennas operating at centimeter wavelengths and shorter. Operating at higher frequencies improved calibration, avoided the detrimental effects of the ionosphere, and reduced the noise component from the galactic synchrotron background. In many ways this put an end to the first era of low frequency radio astronomy.

What frequencies are considered low frequency? Somewhere around 300 MHz is usually chosen as the transition point, though it is a flexible number which may vary in the context of the science case. This point is chosen partially because 300 MHz neatly corresponds to a one meter wavelength but mainly it is around this point that the technology and methods we use to observe change. This will be discussed further in this chapter, but let us make the distinction now that low frequency radio astronomy covers frequencies between 10 MHz and 300 MHz.

Though solar physics and ionosphere studies continued to use low frequency receivers and arrays, radio astronomy for the most part left the meter wavelength band for greener pastures. In the early 1990s there was a return to low frequency astronomy with the installation of the 74 MHz receivers at the Very Large Array (VLA) [Kassim et al., 2007] and the construction of the Giant Metrewave Radio Telescope (GMRT) [Ananthakrishnan, 1995] with operating bands between 38 MHz and 1420 MHz. With significant advancements in technology and calibration techniques the low frequency band has become more manageable and provides a new window into the radio universe.

Throughout the 2000s and to present LOFAR and Epoch of Reionization (EoR) arrays have continued this trend of low frequency arrays. These new arrays are made of simple elements with wide fields of view connected via large digital processing backends. Beyond EoR studies these arrays are used to study the millisecond timescale transient universe, cosmic magnetism, and extragalactic sources, among other topics.

1.2.1 Epoch of Reionization

With the initial expansion of the Universe after the Big Bang, the ionized hydrogen and helium cooled to a neutral gas. The Cosmic Microwave Background (CMB) is the result of this phase transition at the surface of last scattering at $z \sim 1000$. The period between $z \sim 1000$ and $z \sim 30$, when the first luminous objects were formed, is referred to as the cosmic 'dark ages'. Little is known about this period of evolution. In the later part of the 'dark ages' formation of the first stars and quasars reionized the neutral hydrogen and is hence known as the Epoch of Reionization. A number of reviews [Furlanetto et al., 2006], [Zaroubi, 2012] among others, have made a significant effort to cover all aspects of the EoR along with the current state of research at the time of their writing. During this epoch bubbles of ionized hydrogen form around the first stars and quasars until later in reionization the entire Intergalactic Medium (IGM) is fully ionized, Figure 1.7.

It is clear the Universe went from a neutral state in the post-CMB era to the ionized state of the present Universe. But, we cannot say how ionization progressed or at which redshift. The redshift range of the EoR is uncertain, on the low end it has been constrained by Lyman-α quasar spectra to $z > 6$. Ionization began with the first luminous objects, Population III stars and quasars, which are expected to form around $z \sim 30$. Though, the exact range, which is expected to be short, is unknown. One major question during reionization is, what was the main source of the ultraviolet energy required to ionize hydrogen? A quasar dominant model leads to a small set of quickly growing ionization bubbles while a Population III star model is made up of many, slowing ionizing bubbles. Or is it the case that both these sources played an important part across different redshifts?

The EoR is one of the most significant stages of cosmic evolution, understanding the transition will lead to a greater understanding of galaxy evolution, star formation, and cosmological structure. How ionization proceeded during the epoch leads to how

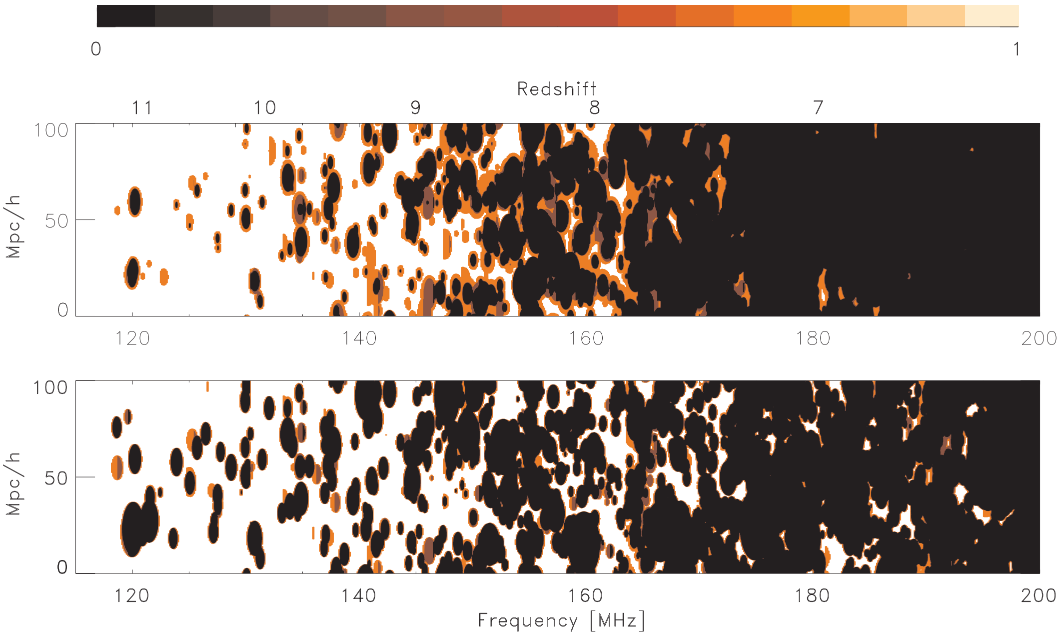


Figure 1.7: Simulation results showing the fraction of neutral hydrogen through a slice of redshift during the Epoch of Reionization. One simulation uses a quasar dominant model (top panel) and the other a star dominate model (bottom panel) to drive ionization. X-axis is observing frequency which is a measure of the age of the universe based on the redshift in the hydrogen 21 cm line. Figure from [Thomas et al., 2009].

clustering and galactic structure moulded the subsequent Universe. One of the most exciting avenues of study is in the detection and mapping out the redshifted hydrogen 21 cm spectral line. This will lead to understanding of cosmic evolution during the EoR along with the complex, low frequency radio foregrounds.

1.2.1.1 Hydrogen Hyperfine Transition

Spectral lines are a result of energy transitions in atoms and molecules which result in absorption or emission of photons within a narrow frequency band. The vast abundance of hydrogen in the Universe and the broad range of energy transitions: Lyman ($n \rightarrow 1$ transition), Balmer ($n \rightarrow 2$ transition), along with the simplicity of the hydrogen model makes it an excellent astronomical spectral line resource. Many of these lines tend to be most prominent in the optical and ultraviolet bands. In the radio band the hyperfine transition line of the ground state is used to detect neutral hydrogen. Hyperfine splitting is a higher order perturbation due to magnetic interaction between the proton and electron dipole moments, Figure 1.8.

When the electron undergoes a spin flip from being oriented parallel to anti-parallel there is a reduction in energy state, ΔE. Where ΔE, [Griffiths, 2005] reviews the derivation using perturbation theory, is

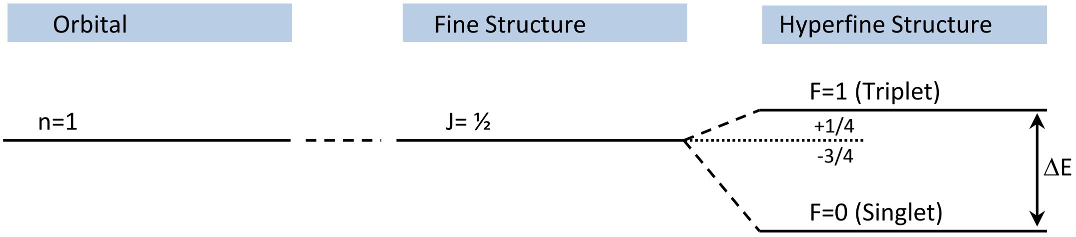


Figure 1.8: Hydrogen ground state fine (relativistic correction) and hyperfine (proton-electron magnetic dipole interactions) structure. The 'forbidden' hyperfine transition is the source of the 21 cm spectral line.

$$\Delta E = \frac{4g\hbar^4}{3m_p m_e^2 c^2 a^4} = 5.88 \times 10^{-6} \text{ eV} \tag{1.7}$$

where $g = 5.59$ is the Landé g-factor, a is the Bohr radius, m_p is the proton mass and m_e is the electron mass. This energy, in the rest frame, corresponds to a photon with frequency $\nu = 1420.41$ MHz or 21.1 cm in wavelength.

Hyperfine transition is very rare as it is considered a 'forbidden' transition but can occur spontaneously with a decay rate on the order of 10 million years [van de Hulst, 1982]. Galactic hydrogen was first detected with the 21 cm line by [Ewen & Purcell, 1951] and [Muller & Oort, 1951]. The long lifetime makes the ground state the only state in which the hyperfine transition occurs in any abundance. The rarity of the transition is balanced by the total number of hydrogen atoms to produce a strong, narrow spectral line. The most famous radio spectral line, study of the '21 cm' or HI line has been a major component of radio astronomy since its discovery.

1.2.1.2 Redshifted '21 cm' Tomography

The redshift of the HI line has been used to determine the distances to nearby galaxies and, is used with known redshifts from optical surveys for line stacking experiments. For the EoR the HI line will be spread across a frequency range corresponding to redshift. For a redshift of $z \sim [6 - 30]$ the corresponding redshifted HI line frequency is between approximately $50 - 200$ MHz. The signal power spectrum is expected to be on the order of milli-Kelvins. It is important to note that this is a 3D power spectrum which is a measure of the evolution from neutral to ionized hydrogen across a redshift range. This differs from the CMB which is a 2D surface at a single redshift.

The challenge of detecting the HI line during the EoR comes down to removal of foregrounds which contaminate the signal. These are up to six orders of magnitude brighter than the EoR power spectrum. Figure 1.9 shows the expected foregrounds in temperature for given l values (which is related to angular scale). Galactic synchrotron radiation ranges from 10000 K to 10 K across the EoR band even in patches of the sky

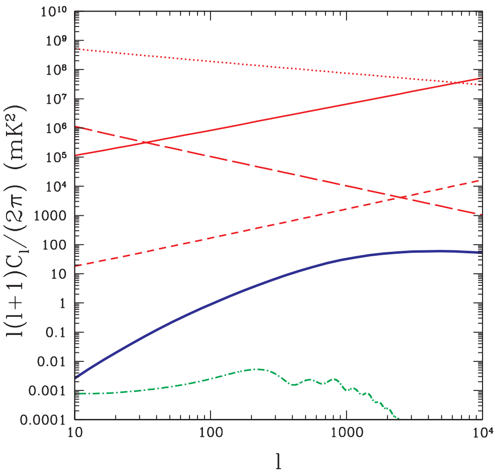


Figure 1.9: Expected 21 cm power spectrum signature at $z = 9.2$ (blue) and global foregrounds over a range of l modes. Galactic synchrotron (red, dotted) is the dominating source, followed by point sources (red, solid), galactic free-free interactions (red, short dash) and extragalactic free-free interactions (red, long dash). The CMB power spectrum at $\nu = 140$ MHz (green) is shown for reference. Figure from [Santos et al., 2005].

outside the galactic plane. Synchrotron radiation from nearby extragalactic sources contributes up to 1 K to the foregrounds. In addition to their brightness, these sources are polarized. Other foreground sources include diffuse free-free emission and radio galaxy point sources.

All of these foreground removal challenges are independent of the issues which arise during interferometric array calibration. Issues of element distribution, beam shape response, cross talk, polarization purity and projection effects, the ionosphere and sky modelling must be dealt with on the path to detection. Experiments leading the way to detect the EoR signal are required to develop novel techniques to deal with the unique issues of low frequency arrays. Though daunting the task may be, the rewards are great, and many groups (LOFAR [de Vos et al., 2009], PAPER [Parsons et al., 2010], MWA [Mitchell et al., 2010], GMRT [Paciga et al., 2011], LWA/LEDA [Ellingson et al., 2009]) are taking up the challenge.

1.2.2 Radio Transients

With the discovery, by Jocelyn Bell, of the first pulsar in November 1967 [Hewish et al., 1968] a view of the radio transient Universe was opened. Understanding the dynamic Universe is a challenging task due to the inherent unpredictability, and singular nature of transient events. Now what are transient events? Most objects in the Universe are stable, in some equilibrium, at least on the timescale on an observation or an average human lifetime. A transient event will change on some distinguishable timescale. They range from compact objects which produce emission on sub-second and millisecond timescales up to supernova afterglows which can last years. Transients can be further categorized into fast and slow transients, where fast transients occur on timescales at or below the rate in which the sky is sampled. For example, the Crab pulsar produces giant pulses, unresolved in time down to nanosecond timescales, which would be a *fast* transient. Of course, this is just a concept we have created, all objects are transient on some timescale. The delineation between *transient* and *stable* events is that we can observe a change in characteristics within a patch of sky across some length of observation.

Transients which occur on $O(\text{minute})$ timescales and longer can be observed across much of the electromagnetic spectrum using transient event detection networks. These networks use detection of a transient event to trigger fast follow up observations by other telescopes across the electromagnetic spectrum. This has proven very effective in observing Gamma Ray Bursts (GRBs) soon after an X-ray detection trigger is sent out. This kind of transient observation can use directed observatories with small fields of view if the event has been localized by the initial detection. Radio observations, especially low frequency radio, really become a useful tool in the sub-second regime of transient parameter space.

Radio transient events cover a wide class of observed and theoretical sources which each offer unique insight into the underlying source physics. [Lazio, 2008] provides a literature review of a broad range of radio transients. Many of the transients are from compact objects which emit short duration coherent burst such as pulsars. Every pulsar is unique which leads to a broad range of characteristics and emission such as millisecond rotation rates, rapidly rotating transients (RRATs), and giant nanosecond pulses. Pulsars are stable but periodic, their regularity is well defined even if they sometimes appear to turn 'off' or disappear during an observation. Both Supernovae and GRBs are predicted to have short duration prompt emission in the radio, though no such emission has yet been observed. Local to our solar system, the Sun undergoes dynamic changes across a broad range of timescales, Jupiter is known to have dynamic emission at low frequency (< 30 MHz), along with short duration emission from planetary weather. The more exotic transients that have been predicted include Search for Extraterrestrial Intelligence (SETI) transmissions, cosmic superconducting strings, and evaporating black holes. An informative, and densely packed, map of the transient parameter space from [Cordes, 2009] shows the variety in dynamic radio sources, Figure 1.10.

Detection of transients requires high time resolution along with enough sensitivity to detect a signal within the characteristic time length of the transient. If we have no foreknowledge of a transient's location then we can assume an isotropic distribution. Thus, an increase in the observing field of view provides a proportional increase in detection rate. This assumes that the transient sources follow some power-law distribution and a portion of the events are above the flux detection threshold. To achieve

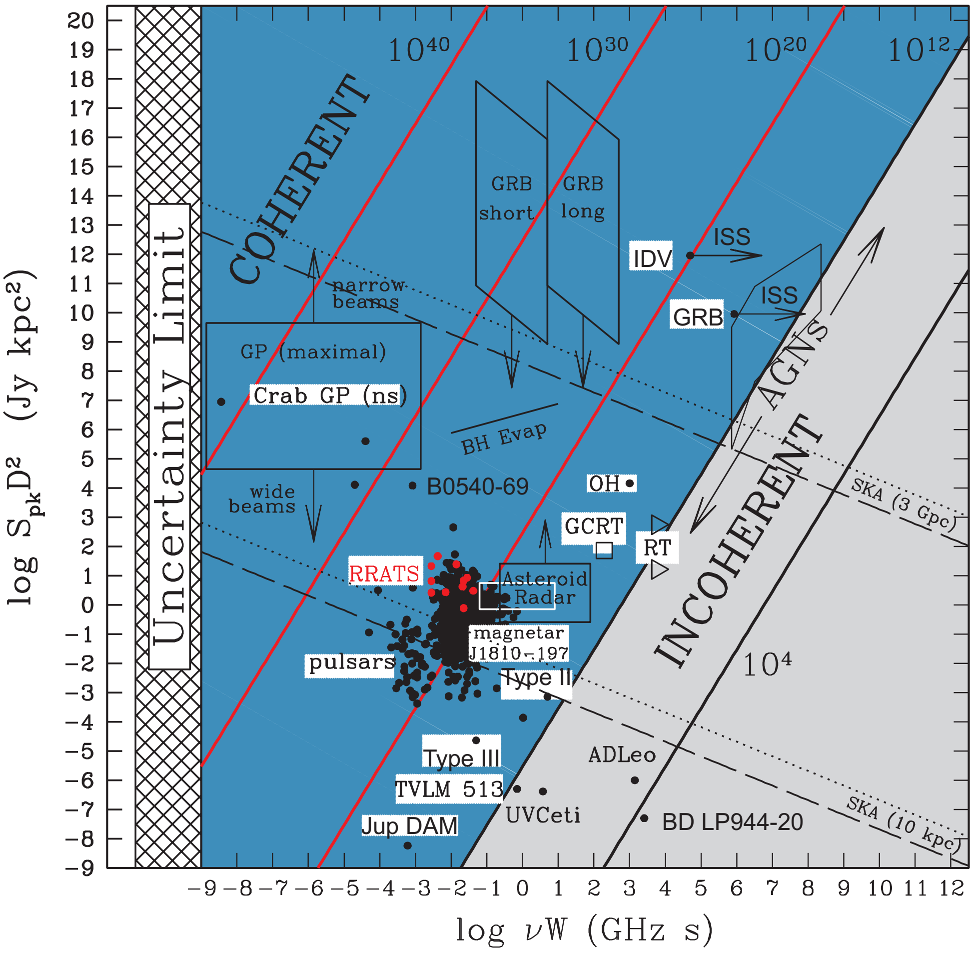


Figure 1.10: Time-luminosity distribution of observed and theoretical radio transients. The luminosity axis is the log of the product of peak flux, S_pk, and the nominal object distance squared, D^2, in units $\log(\mathrm{Jy}\cdot\mathrm{kpc}^2)$. The time axis is the log of the observing frequency, ν, and the transient width, W, in units $\log(\mathrm{GHz}\cdot\mathrm{s})$ Lines of constant brightness temperature are shown (solid red and black), along with the sensitivity of full SKA (dashed) and $\sim 10\%$ SKA (dotted) for given distances. Individual points are observed transients, regions show theoretical transient regions. Figure from [Cordes, 2009], where a full discussion is presented.

a large sky coverage a system must use wide field of view receiving elements and either be correlated together as an interferometric array or used as a *beamformer* with many beams on the sky. A beamformer is an instrument which coherently adds together multiple antenna signals into a single beam, for an overview of see [Van Veen & Buckley, 1988]. In addition to a sensitive array with a large field of view a powerful processing backend is required for signal detection. Fortunately, this is an active area of research which is constantly developing better algorithms and taking advantage of the ever increasing computing power available.

Though it is advantageous to search for transients across multiple frequency bands, observing at low frequencies has many advantages in terms of station design and digital processing over higher frequencies. At low frequencies transients have a higher spectral brightness. Using aperture arrays multi-beam observations can be made across the entire sky with high sensitivity at the cost of additional digital processing. Many octaves of bandwidth over the low frequency spectrum can be processed cheaply on current digital signal processing devices.

1.2.3 Deep Extragalactic Surveys

Large scale, high resolution surveys across the low frequency sky are desired to enable future science goals. Since the low frequencies have been avoided in the past, the characterization of sources in this portion of the electromagnetic spectrum is poor. Our understanding of the sky at low frequencies must be comparable in resolution and sensitivity to other regions of the spectrum. This will lead to better understanding of spectral source structure and can be used in conjunction with other surveys to provide a wealth of new science. Additionally, when we open up more of the observable parameter space this leads to discovery of new sources and a deeper understanding of the physics at work.

1.2.4 Cosmic Magnetism

Magnetic fields pervade every size scale in astronomy. From spinning dust to the web structure of galaxy superclusters, magnetism plays an important role in the physics of the environment. Low frequency radio observations provide a unique insight into a range of magnetic field effects. Synchrotron radiation, which has a steep spectral index and is polarized, is the result of charged electrons being accelerated in magnetic fields. The emission traces out galactic magnetic field structure and can be used to detect extragalactic sources.

Faraday rotation is the effect of phase rotation, or polarization rotation, on radiation from a source. The amount of rotation, the rotation measure, is proportional to the magnetic field along the line of the sight. This can be used to measure the magnetic fields within the Interstellar Medium (ISM) and IGM depending on the source. In order to measure the rotation measure of a source, that source must be at least partially polarized. Pulsars are excellent sources to be used as beacons in measuring Faraday rotation across the intervening ISM. The effects of Faraday rotation scale as λ^2. At low frequencies a large fractional bandwidth can be covered which provides broad access to the effects of Faraday rotation. On larger scales, measuring magnetic fields of nearby galaxies gives insight into the formation and structure during galaxy evolution. And on a cosmic scale, detection of magnetic fields in the IGM will lead to a better understanding of the large scale structure of the cosmic web.

1.3 Large-N Widefield Arrays

The terms 'large-N' and 'widefield' are flexible enough to cover a large range of definitions depending on their context. Here N is the number of elements in an array and widefield refers to the field of view of these elements. Especially 'large-N' which is likely to continue growing as technology allows us to build larger arrays. So, it is useful to define these terms now, as they are commonly discussed in the radio astronomy community today. What is considered a wide field of view is directly connected to the observing frequency and baseline length of an array. A wide field of view can be arcminutes in scale at short wavelengths (submillimeter) or during Very Long Baseline Interferometry (VLBI) observations. For long wavelengths (decimeter and centimeter) the field of view can effectively be the entire visible hemisphere of sky above an array. Since our discussion will be focused on low frequency array ($<$ 300MHz) let us define a wide field of view as that covering $\frac{1}{10}$ steradians ($10^\circ \times 10^\circ$ patch) up to the 2π steradians (entire hemisphere) of the sky. This range covers single dipole elements up to beamformed aperture array stations. At these scales widefield imaging distortion effect become prominent issues and array calibration has its own unique set of challenges.

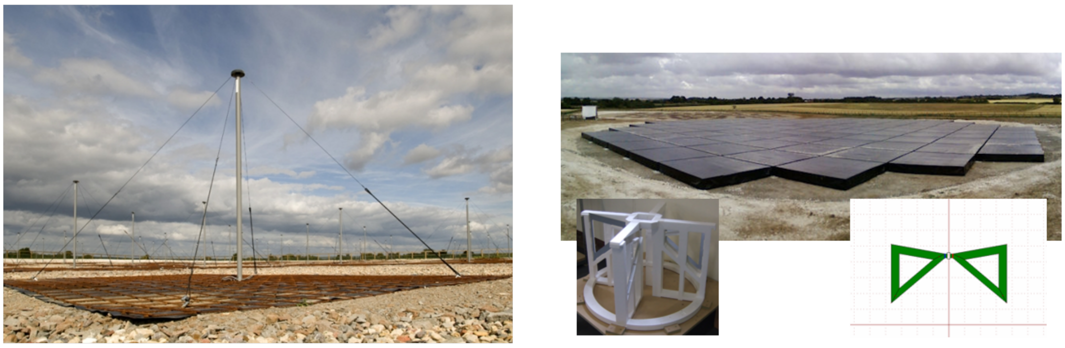
Currently, 'large-N' is generally considered to be any array with $N \sim O(100)$ correlation elements. Arrays pushing the 'large-N' threshold include the various EoR experiments: PAPER, MWA, LOFAR along with the ever looming future Square Kilometer Array. Not far on the horizon is talk of $N = O(1000)$ arrays, and it is not unreasonable to expect N to continue growing as the price of antenna elements, analogue hardware and correlation decreases.

A valid question is why do we need large-N, widefield arrays? An array with N correlated elements produces $O(N^2)$ baselines. The baseline length and projection distribution defines the synthesis imaging UV coverage. By correlating more elements we have more control over the baseline distribution to produce a smoother Point Spread Function (PSF) or to target certain modes in UV space. Sensitivity also increases with the number of elements. By building arrays made up a large number of elements which are sensitive to a widefield of view the survey speed of the array increases. This is a key factor in array design. Our ability to process widefield images is usually limited by the computation and memory availability which is a key factor in something like VLBI. As computing power and memory have decreased in price per unit operation and the calibration techniques have improved we gain access to wider field of view. This in turn opens up new possible science cases, e.g. power spectrum measurements and transient detection, and can help reduce the noise floor by reducing aliasing effects. Not all radio science cases require large-N or widefield arrays but these do provide access to a new sections of the parameter space.

The first radio interferometry arrays were made of meter wavelength, widefield elements. As the first radio sources where discovered and synthesis imaging techniques improved, arrays moved towards shorter wavelengths with directed dish elements in an effort to study individual sources. At the same time, baselines were extended to increase the resolving power of arrays at the cost of reducing the aperture filling factor. In the past few years we have returned to the meter and decimeter wavelengths with a few decades worth of improvement in technology and techniques. Instead of arrays made up of a few elements, covering narrow frequency bands these new arrays are made up of hundreds of elements with digital systems that cover huge fractional bandwidths. LOFAR is one of these ambitious arrays which, as of this writing, is finishing commissioning and undergoing an initial sky source survey. The goal behind the low frequency component of the the Square Kilometre Array (SKA) is to build a LOFAR like array which is 10 - 100 times more sensitive.

1.3.1 The Low Frequency Array

LOFAR is a distributed telescope network consisting of aperture array stations covering the 10 MHz to 240 MHz band [de Vos et al., 2009]. Stations are spread throughout Europe with the central core located in the Netherlands, Figure 1.12. It is one of the largest interferometric radio arrays in the world. Development and organization of LOFAR is led by the Netherlands Foundation for Research in Astronomy (ASTRON). As of this writing there are 41 verified stations which are currently in operation. Each station consists of two arrays, the Low Band Array (LBA) which is sensitive to the 10 MHz to 90 MHz band with a steep cutoff above 90 MHz to suppress the terrestrial FM radio band and, the High Band Array (HBA) which covers the 110 MHz to 240 MHz band. The LBA elements are ‘droopy’ dipoles which are half-wavelength dipoles at 52



(a) LBA ‘droopy’ dipole design sensitive to dual, linear polarizations. The LNA is located at the top of the pole with cabling fed underground to the local station receiver unit.

(b) Layout of HBA tiles at Chilbolton Observatory. Each tile consists of a 4×4 grid of bow-tie antennas and an analogue beamformer. The bow-tie design(insets) is sensitive to frequencies between 110 MHz and 240 MHz.

Figure 1.11: Antenna element designs for the two LOFAR arrays covering the 10 MHz to 240 MHz band. Photos from LOFAR-UK[1]and ASTRON[2].

MHz with a wire mesh ground plane, Figure 1.11a. The HBA tiles are made up of a 4×4 grid of bow-tie antennas which are analogue beamformed together, Figure 1.11b.

During a given observation only a single array is used at the station as a common digital backend is used for both arrays. To accommodate multiple observing modes using a common digital backend, sky signals are mixed down to baseband using a set of selectable local oscillators. Sky signals are digitized at 200 or 160 Msps. Band selection is set by selecting the appropriate bandpass filters for the desired Nyquist zone. After digitization the band is channelized into 512 subbands (195.3 kHz or 156.25 kHz resolution) of which up to 244 can be used in the station beamformer mode and International LOFAR Telescope (ILT) mode. Each station can be used as part of the ILT where signals are transported from each station to a central computer in Groningen, the Netherlands for beamforming and correlation using multiple stations. Or, a station can be used in stand-alone mode as a beamformer or a correlator. The data product produced at the station digital backend is a *beamlet*, up to 244 beamlets are available for processing. A beamlet is a single subband with a set of beamforming coefficients applied for pointing. On one end of the beamlet configuration 244 different subbands are used with the same set of beamforming coefficients as to give the most bandwidth coverage and sensitivity for a single pointing. On the other end the 244 beamlets can be pointing in all different directions, the beamlets can be set to the same subband or distinct subbands, to provide the most sky coverage in a single observation. An overview of the station architecture is presented in [Gunst & Schoonderbeek, 2007].

---

[2] http://blog.lofar-uk.org

[2] http://www.astron.nl

Figure 1.12: Map of LOFAR stations in operation as of July 2012; 24 core stations, 9 remote stations and 8 international stations. ©ASTRON 2012

In ILT mode, signals from each station are transported to the central processor in Groningen via a direct 10 Gbps fiber connection. Signals from each station are then further channelized (up to 1 kHz resolution) and correlated with an FX correlator implemented in real-time on an IBM BlueGene/P supercomputer. The calibration, imaging, and various other science-oriented processing pipelines are applied to the observations in post-processing. LOFAR has been designed to be expanded over time by adding of new stations and expanding the current central processing system.

The development of LOFAR is based on a series of science projects which can be uniquely served by low frequency observations:

Epoch of Reionization Detection of redshifted hyperfine transitions in hydrogen as a probe of cosmic evolution during EoR.

Extragalactic Surveys The high resolution and sensitivity of the ILT will allow deep surveys at frequencies which are under-observed for use in multi-frequency projects.

Transient Radio Events and Pulsars With the large field of view and beamforming flexibility the array is unique positioned as a pulsar timing array and a tool for detection of new pulsars and radio transients.

Cosmic Rays Cosmic rays generate particle showers in the upper atmosphere which produces low frequency radio emission.

Cosmic Magnetism The steep spectral index of synchrotron emission can be used to detect cosmic magnetism.

Solar Physics and Space Weather the Sun is a dynamic radio source, throughout the LOFAR bands numerous emission processes occur at various layers of the solar atmosphere.

At Chilbolton Observatory in the United Kingdom an international LOFAR station was completed in September 2010. The international station consists of 96 LBA dipole elements and 96 HBA tiles. This station creates some of the longest baselines in the ILT and will be essential in high resolution observations. In stand-alone mode the station is used for pulsar monitoring, transient searching, riometry, and imaging large scale galactic structure (see Chapter 5). The station backend is also being extended with a station element correlator, SEPNET Widefield Camera (SEPcam), discussed in Chapter 4.

1.3.2 The Square Kilometre Array

If the development of the SKA is successful it will become one of the great scientific wonders of the world. It is difficult to define the SKA specification as of this writing. A purposed system specification for phase 1 of the SKA is presented in [Dewdney et al., 2010]. The winds of time are constantly reshaping and moulding the sands of the SKA into new forms. Though we can say the SKA will cover large portions of bandwidth in between ~ 50 MHz to ~ 10 GHz and as the name implies consist of one square kilometer of collecting area. Broadly speaking, three independent arrays will be used to observe the band: SKA-low phased aperture array ($\sim$ 50MHz – 400MHz), SKA-mid phased tile element array ($\sim$ 400MHz – 800MHz), and SKA-mid dishes ($\sim$ 800MHz – 10GHz). For the low frequencies, roughly 50 to 400 MHz, sparse and dense aperture array stations will be used. Immediately it is clear that the knowledge gained from LOFAR will factor into design considerations. The science results and design considerations of LOFAR will guide the development of the low frequency band of the SKA. The middle frequencies, between 400 MHz and 10 GHz, will be covered by a mix of offset Gregorian dishes and phases array tiles. The engineering requirements of this array make it one the great scientific programs of the 21^{st} century and will require a long term investment in time, funding, and vision to make it possible.

Though the system design specification has not been set the science goals have been well defined and, unsurprisingly, are similar to that of LOFAR. I refer interested readers to [Carilli & Rawlings, 2004a] which covers in depth a broad range of science topics the SKA will make possible. The five main scientific drivers behind the SKA are presented in [Carilli & Rawlings, 2004b] as the following key science projects

General Relativity Tests of Gravity pulsars and black holes are invaluable tools in the study of general relativity. With increased sensitivity over current instruments the SKA will be able to detect thousands of new pulsars and put new limits on millisecond pulsar timing. The multi-beam mode of the aperture array will allow many pulsars to be observed concurrently and help lead to direct detection of gravitational waves.

Galaxy Evolution the original motivation for the SKA was for an instrument to map out the evolution of neutral hydrogen at high redshifts. This will give insight into star formation and galaxy structure along with the detection of Baryonic Acoustic Oscillations (BAOs) which will be key is constraining cosmological models.

Epoch of Reionization the main focus of the SKA low band will be the power spectrum measurement and mapping of highly redshifted hydrogen in the detection of the EoR signal to map out the cosmic 'dark ages'.

Cosmic Magnetism study of the ever elusive galactic and extragalactic magnetic fields will be made possible by the large bandwidth, sensitivity and resolution of the SKA.

The Cradle of Life detection, follow up, and possible imaging of extrasolar planets through spectral line measurements and high resolution baseline observations. Also, radio frequencies is an efficient method to transmit information which has led to great emphasis in SETI research in the radio bands. The sensitivity of the SKA greatly increase our ability to detect weak, narrowband signals.

Beyond the key science projects, the SKA will be a powerful instrument for unique experiments that will continue to provide unparalleled access to the radio universe for decades to come.

One of the great challenges of building these arrays is the interconnect, data bandwidth, and digital signal processing. In recent years, FPGAs have played a significant role in coping with these issues. These chips are made up of logic, multipliers, memory, and IO interfaces which are connected via a generic grid array.

1.4 FPGA-based Digital Instrumentation Design

FPGAs have proven to be an excellent technology for use in radio astronomy instrumentation design. The technology allows digital processing of large, streaming data rates ($O(10)$ Gbps) at a low operational power cost ($O(10)$ W). The reconfigurable nature of the technology allows modular instrumentation design using tools which have similarities to object-oriented programming.

An FPGA, as the name suggests, is an integrated circuit made up of an array of logic gates which can be programmed in the ‘field’, that is after the chip has been manufactured. Each chip is made up of an interconnected array of DSP in the form of logic gate, multipliers, memory space, and I/O interfaces. In this blank form the FPGA has no functionality. The chip will operate according to the firmware design which is loaded. The functionality of an FPGA is similar to that of an Application-Specific Integrated Circuit (ASIC), but an FPGA can be reprogrammed many times over the course of its lifetime where as the ASIC firmware is set during manufacturing. Firmware design involves constructing modular logic blocks in a Hardware Description Language (HDL) or with tools which reduce down to HDL code.

The focus of the FPGA market is on high speed, high bandwidth digital signal processing applications which cover a wide range of fields. Specialized boards for radio astronomy signal processing are developed for use as samplers, spectrometers, and correlators, Figure 1.13.

As FPGA firmware design is in a gray area between hardware layout and software programming, a number of unique challenges are introduced. Firmware is primarily written using HDLs such as Verilog and VHDL. Graphical design environments such as Simulink have been adapted by FPGA companies to be used as complete firmware design packages. Simulation is a key component to design. The FPGA operates on discrete time units and the output of each component can be analytically computed. Thus, the exact operation of a design can be determined on every clock. These tools come from ASIC design where the system must be verified before being sent to the foundry for production. Test Vector Generator (TVG)s are used to check for correct output during simulation.

DSP blocks are a limited resource on the FPGA. When designing firmware this limitation has to be taken into account. The first stage of the FPGA place and route process is *mapping* which determines the type and number of resources which are needed and attempts to lay them out on the logic grid. The second stage is *timing*. In an FPGA, at every clock edge every logical operation is executed in parallel. Signals are shifted from logic cells between clocks. This is a different paradigm to programming where operations are executed in serial and not tied to clocks. For example, if two

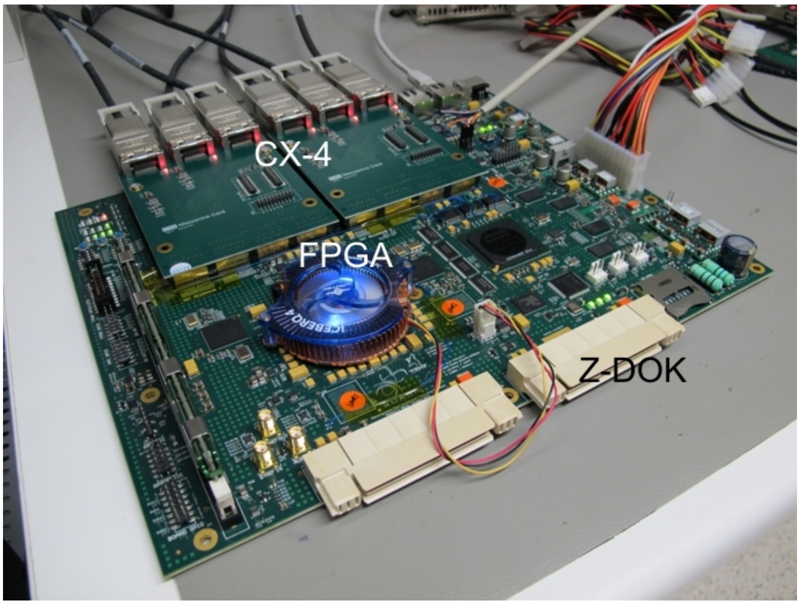


Figure 1.13: Xilinx Virtex 6 based ROACH2 board developed by CASPER. ADCs and DAC daughter boards are connected to the Z-DOK interface which has up to 40 Gbps I/O. Eight 10 Gbps high speed interfaces are used for interboard communication. Virtex 6 contains 400,000+ logic cells and 2,000+ specialized DSP slices. Photo via CASPER website.

signals are multiplied together in one region of the chip then those signals need to physically transit across the chip in the available time between clocks using the grid interconnect. If one of those signals cannot make arrive in time then the design will not meet timing. This requires either a reduction in the clock rate or an improvement in the place and route. One of the great challenges of FPGA firmware design is finding ways to improve the timing by adjusting the mapping of logic cells. Improving FPGA layout is an ongoing focus of research. In addition, as FPGAs grow in logic resources, efficiency becomes more challenging.

There are a number of competing technologies which are used for radio astronomy digital instrumentation. Though, competing technology is not right term as there are great advantages to be had by using different technologies symbiotically.

General computer hardware is incredibly flexible, easy to program and made of commodity components. The flexibility comes at a price though. A computing unit required 100's of Watts to operate and have limited I/O bandwidth. Computers can run complex algorithms which can be recompiled on second timescales. So, they are reserved for post-processing data after the rates have been reduced.

On the other end of the efficiency - flexibility spectrum are ASICs. Since these are custom designed for a single purpose they are the most efficient in terms of power per computational operation. They can also run at higher clock speeds because they can use a custom interconnect and not just fixed to a grid. The production cost and non-recoverable engineering costs are large. The time between iterations can be on month time scales and each iteration is a significant cost. ASICs are a good solution for a large system which will not be changing specification. ALMA has a very large correlator which is implemented using ASICs.

Somewhere in between general computing and FPGAs are Graphics Processing Unit (GPU)-based processors. These are a relatively new technology for scientific data processing, though the technology has been around for a few decades in graphics processing for general purpose computers. They are made up of many processing cores which operate in parallel asynchronously. This allows for a large amount of processing at a lower power cost then a Central Processing Unit (CPU). One of the current limitations to using GPUs is that they must be tied with a CPU server which limits I/O bandwidth. Many large-N arrays are moving towards hybrid technology arrays where FPGAs are used for the F component and the correlation, X component, is computed on GPUs.

1.4.1 Design Considerations for Low Frequency Array Correlators

Based on the low-frequency science goals and correlator architectures we can determine the optimal, in terms of computational cost, correlator architecture for low-frequency arrays. A digital FX correlator should be used for low-frequency arrays. First, the bandwidth requirements is at most $\sim$ 300 MHz which is a reasonable sampling rate for high bit resolution ADCs. Though, this bandwidth is the upper limit, many science cases require smaller bands because of the large fractional bandwidth coverage. RFI is ever present in the low frequency bands which will require digitization with $\sim$ 8-12 bit resolution. Once the time stream has been channelized the signal can be quantized down to a lower bit width to save resources as correlation can occur at low bit widths, see Section 2.2. Thus, performing the F stage first saves resources in the X stage. Fine frequency channelization is required to localize narrowband RFI to small frequency bands, which increases the fraction of usable band. Typically a width of $\sim$ 10 kHz is used. For 10 kHz resolution across a 300 MHz band a 2^{15} point FFT is required. An XF design with this level of channelization would require very large lag taps with 2^{15} delays and CMACs. Present and near future arrays will use 100-1000 antenna elements for correlation. This range fits in the large-N category which FX out performs XF designs. Even though large FFTs are required for channelization the $O(N^2)$ scaling of the correlation dominates computational costs. In Chapter 2 we will go into the details and challenges of implementing an FX correlator in digital hardware.

1.5 Progression of Thesis

Correlator design, calibration and simulation, and imaging all play important roles in reducing sky signals to scientifically useful data quantities. Understanding the design and interplay of these topics is necessary in this new regime of low frequency, large-N radio astronomy. In the next few chapters we will further develop the design of digital FX correlators for large-N, low frequency arrays. Implementation of correlators on multi-board FPGAs will be covered in Chapter 2. This will focus on interconnect methods and efficient DSP resource usage. The second half of the chapter will cover optimization techniques to reduce bandwidth and DSP usage while minimizing error.

Chapter 3 covers the development of a 32 input FX correlator for the BEST-2 system at Medicina Observatory. This system was implemented as a design study for the larger LOFAR SEPcam correlator presented in Chapter 4. A review of the digital design is presented along with the software wrappers used to control the system and record data. Verification of the system hardware along with imaging and calibration of observed sources is presented.

A 96 dual-polarization element correlator for the LOFAR station at Chilbolton Observatory has been developed based on a similar design as the one presented in Chapter 3. The design is modular and will be interfaced with the existing LOFAR station backend to be used commensally during normal observations. A review of the design is presented along with verification of the DSP subsystems. The chapter ends on the further work required to interface the correlator design with the LOFAR digital backend.

Chapter 5 introduces calibration and imaging techniques for LOFAR single station observations. Though commissioning of the SEPcam correlator is not complete we can use the LOFAR station calibration correlator to record correlation data. Using these observations a pipeline for calibration of the array is presented which is implemented in the form of the measurement equation. Snapshot imaging and generation of 'all-sky' maps is presented.

Chapter 6 continues on with the calibration challenges of widefield, low-frequency interferometry presented in Chapter 5. The chapter covers the use of shapelet basis sets to model extended sources and beams for use in interferometric calibration and simulation.

This thesis concludes with Chapter 7 on topics of future large-N correlator growth, extension of correlator systems to include imaging and calibration stages, and the challenges of improving sky models for low-frequency arrays.

Chapter 2

Topics on FX Correlator Design

> *I still think the solution is to just upmix the signals to optical and then use a lens to form the image.*
>
> -Jack Hickish, 2010

FPGA-based instrumentation design is about efficiency. Efficiency in resource usage, power consumption, and signal transport bandwidth. Design constraints are set by limited DSP resources, operational power budget, available I/O bandwidth, and available funding. To meet these limits, the resources used in a design are reduced while minimizing the loss to signal quality. This can be done by reducing the signal bit width representation, time or frequency averaging, and optimal algorithm implementations.

The DSP resources available on FPGAs are Configurable Logic Block (CLB)s, Block RAM (BRAM) memory, I/O transceivers, and specialized DSP circuits. The CLBs are arrays of logic 'slices' and Random Access Memory (RAM) which are routed to create the desired logic circuit. A logic 'slice' is the smallest logic unit of the FPGA. They are constructs made up of Lookup Table (LUT)s and flip-flop circuits. BRAM is localized memory spaces used for delays, reorders, and storing coefficients, they are on the order of tens of Kilobits in size. Specialized DSP circuits, such as Xilinx's DSP48e slices [XILINX, 2012b], are optimized to perform common operations such as large multiplication, adding, and accumulation. High speed I/O transceivers along with the CLBs are used to implement streaming and packetized inter-chip communication.

Correlator implementation on FPGAs is broken up into various subsystems: signal input/digitization, pre-channelization corrections, channelization, phase corrections, equalization and quantization, the cornerturn, cross correlation, and data output. The appropriate solution to each one of these subsystem depends on the required instrument specifications. The first half of this chapter will touch on solutions to each one of these topics. The remainder of the chapter will cover techniques which can be used to improve quantization effects, make efficient use of resource for multiplication, and reduce bandwidth.

2.1 Design Architectures for Large-N Correlators

FX Correlator design is split into two modules. The F module performs operations on single antenna streams, each stream is independent. This covers from the signal input to channelization to the cornerturn. Each independent module is called an *f-engine*. For an N element system the number of f-engines is proportional to N, the exact number depends on the ratio of the ADC clock to the FPGA clock. At the cornerturn, data units are blocked into windows for correlation, the stream become independent in terms of frequency subbands. Each one of these streams is then processed through an *x-engine* which performs the correlation. The number of x-engines in the system is proportional to N^2, the exact number depends on the clock speed and resource implementation.

2.1.1 Digitization

The ADC is the last analogue device before the signal enters into the digital domain. Once in the digital domain the signals are tied to the sampling clock used in the ADC. The time streams will be real or complex values depending on the sampling technique. The FPGA used in the F stage of the correlation is usually connected to this clock so that the FPGA and ADC are synchronous. A typical FPGA clock will be limited to around $\sim$ 250 MHz due to timing constraints during place and route. Though, with manual placement the clock can be further increased. If the ADC is sampled at 250 Msps then on each clock a new n-bit word is available. Where n is the bit width resolution of the ADC, $n = 8$ is a common value. When using high speed ADC such as 1 Gsps samplers the FPGA cannot be clocked at 1 GHz. The solution is to use a demultiplexer to convert a single high speed stream into multiple lower speed streams which can be processed in parallel. This allows for an increase in bandwidth at the cost of an increase in resource usage while the FPGA runs at a lower clock rate to the ADC. By demuxing a 1 Gsps stream by four an FPGA running at 250 MHz can process the entire bandwidth. The reverse process can be used in the case of multiple low speed streams, which can be multiplexed together so that fewer resources are required on an FPGA running a clock with an integer factor of the ADC clock, see section 3.3.

Signals are represented as values between 1 and -1, this simplifies operations further along by using the Most Significant Bit (MSB) as reference. A common notation to represent digital signals is a_b, where a is the total number of bits and b is the position of the decimal point, e.g. 1.00101 is a 6_5 number.

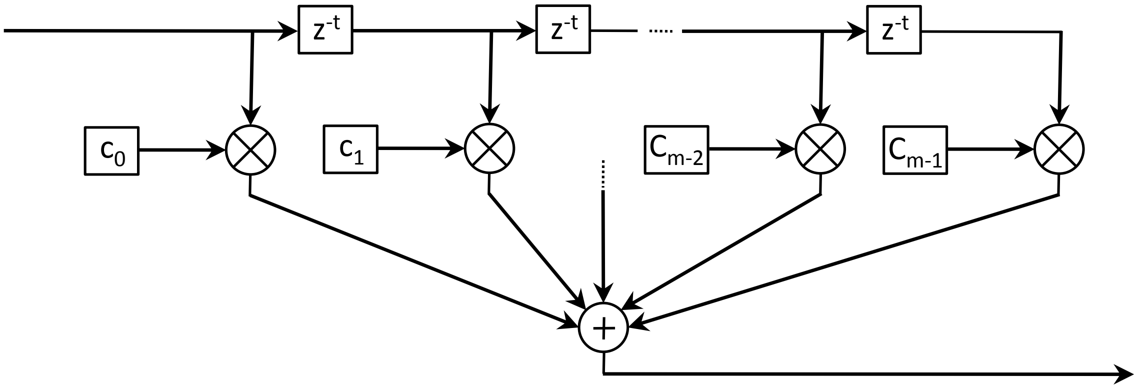


Figure 2.1: Diagram of an m-tap FIR filter used for digital band filtering and applying windowing functions in the polyphase filter bank. Each tap is the product of a windowing coefficients (C_i) and the streaming signal.

2.1.2 Time Domain Filtering

Before transforming to the frequency domain, filters and logic can be applied to the time domain signals. In the time domain delays can be added to align signals which improves correlation sensitivity for long baselines. This ‘coarse delay’ allows for signal alignment up to the resolution of the sampling rate. Once in the frequency space ‘fine delay’ can be added by introducing phase offsets. Coarse delay is implemented using BRAM as a circular queue.

Digital high/low/bandpass filters can be applied to the streams using a Digital Down Converter (DDC). The DDC converts a real signal to complex using the Hilbert transform. A phase is added to the complex signal and then a decimating Finite Impulse Response (FIR) filter is applied to select out a portion of the band [Lyons, 2004]. An FIR filter is a non-recursive, m-tap filter which is the weighted average of m regularly spaced samples, Figure 2.1. Infinite Impulse Response (IIR) filters are recursive, as they use a feedback loop. The coefficients in the FIR depend on the desired windowing function. A b-point rolling average is an example of an FIR with the coefficients in the b-taps set to $1/b$. The number of taps and bit width of coefficients determine the quality of the filter. Since the output of the DDC is the product of the coefficients and signal there can be significant bit growth. Quantization is often necessary to reduce the output bit width. As will be noted in the next section, an FIR is often used in frequency channelization.

2.1.3 Frequency Channelization

The transformation to frequency domain from time domain is performed with the Fourier Transform. In hardware a streaming FFT is used. But, the windowing function, a box, results in a poor frequency channel response. Applying a different windowing function, such as a Hann or Hamming window, to the signal can result in lower sidelobes and a flatter response across the frequency channel, see [Price, 2012] for a review. This windowing function is implemented with an m-tap FIR filter. The windowing FIR filter along with the FFT is known as a PFB. An m-tap FIR filter for a n-point FFT window uses $n(m-1)$ delays, nm coefficients, and m multipliers.

Transformation to the frequency domain is performed using an n-point FFT using $O(n \log n)$ operations. The Cooley-Tukey FFT algorithm [Cooley & Tukey, 1965] is commonly used. This algorithm recursively decomposes an n-point FFT into two smaller FFTs of size n_0 and n_1, where $n = n_0 n_1$. Commonly, $n = 2^d$ so that each recursive stage will use the Cooley-Tukey algorithm, this is known as a *radix-2* implementation. A further reduction in BRAM usage can be made by using the *radix-4*, $n = 4^d$, implementation. Decimation-in-Time (DIT) and Decimation-in-Frequency (DIF) methods are used to recursively split up the samples, these methods have similar implementations and produce equivalent results [Lyons, 2004].

If multiple time streams are undergoing transformation the FFT can be used in ***biplex*** mode to perform two FFT operations with one FFT by using the free clocks [Emerson, 1976]. This results is a reduction of logic by 50% and BRAM by 25%.

From Fourier theory, the transform of any real signal will result in a mirrored, complex spectrum. An n-point FFT of a real signal will result in $\frac{n}{2}$ unique complex values. The signal bit representation will have grown throughout the stages of the FFT. Quantization, as noted below, is used to reduce the output bandwidth before the cornerturn and correlation.

2.1.4 Phase Corrections

Once in frequency space the signal is complex. Phase offsets can be added to the channels which provide sub-timesample ‘fine delay’ control. This is useful for geometrical phase offsets for each baseline which is dependent on the pointing ‘phase center’ on the sky or complex gain calibration. The ‘fine delay’ can be updated on second or sub-second timescale through a software interface to set the phase offsets such that the phase at the center of the field center is constant. This is *fringe stopping* as it will stop any source at the center of the field from producing a fringe pattern. For long baselines, fringe stopping is required because the baseline fringe rates are faster than the accumulation length and will be washed out. Though, fringe stopping is not always required, baselines in short baseline arrays fringe slowly and the correlations can be accumulated with a minimum amount of phase error introduced in the accumulation. This allows phase offsets to be added post-correlations to select different phase centers, this is often done in large field of view arrays.

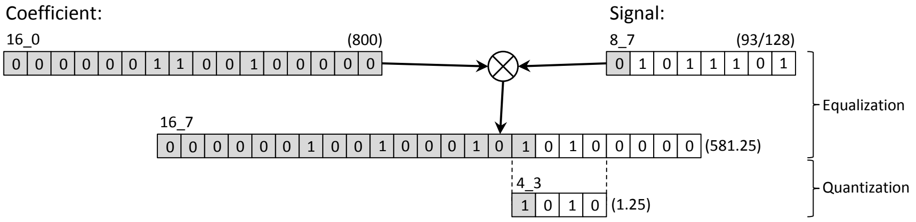


Figure 2.2: Example of equalization and quantization where an 8_7 value is equalized using a 16_0 coefficient and quantizing down to 4_3.

2.1.5 Equalization and Quantization

During frequency channelization and phase correction, the signal will undergo bit growth. A complex signal will commonly be represented by a 36-bit number (18-bit real and imaginary). Signals only need to be represented by a few bits for correlation, see Section 2.2. The original time signal was sampled with a large bit resolution to account for the dynamic range in strong RFI environments. Once the signal has been channelized the narrowband RFI is localized and the large dynamic range is no longer needed. Typically between 1 and 4 bits are used to represent the signal during correlation. We will mainly discuss 4-bit correlations as the following chapters cover 4-bit correlators. The channels with no discernible RFI can be approximated as a Gaussian signal. To quantize from 18 bits to 4 bits the signal is first equalized, Figure 2.2. The standard deviation of the noise signal in channel i is σ_i. For the quantized signal the standard deviation needs to be σ'_i, which for a 4 bit system is usually around $1.5 - 2$ bits [Backer, 2007]. The channel is scaled by a coefficient such that the signal standard deviation goes $\sigma_i \rightarrow \sigma'_i$ and is thus 'equalized'. Quantization then becomes an operation of selecting the 4 MSBs. This can be done by simply selecting those 4 bits which introduces a bias or by including rounding logic. Setting the coefficients is done using a software interface to account for fluctuations in power during an observation.

Equalization can also be used to apply complex calibration coefficients in real time. This is the goal of real time imaging systems, additionally it offers an improvement in sensitivity.

2.1.6 The Cornerturn

Up to this point the data is independent streams of antenna signals. By channelizing a window of M time samples from an antenna we now have M independent frequency subbands. The cross correlation operation only applies to single subbands but across all N antennas. Each antenna spectra is a frame of data with M words, one for each subband. A correlation frame is N words, one subband from each of the antennas.

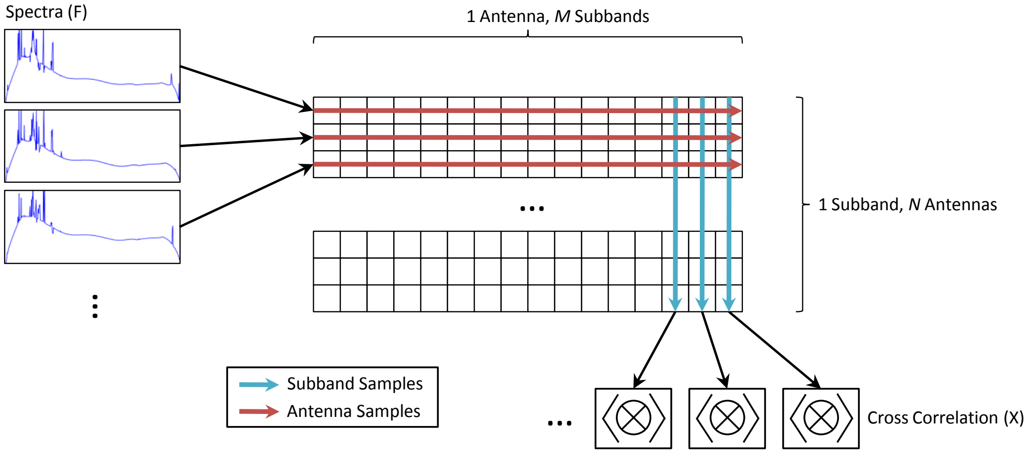


Figure 2.3: A cornerturn operation is used to group together subband sets from independent antennas for correlation. N frames of M antenna signal words are turned into M frames of N independent subband words.

The challenge is to efficiently reorder antenna frames into subband frames, this is the 'cornerturn problem', Figure 2.3.

A simple solution is to have an $N \times M$ block of memory which writes spectra into the memory space in one coordinate direction and reads subbands from the orthogonal direction. Similar to the arrows shown in Figure 2.3. When using the 'pipelined' cross correlator design discussed in the next section the memory space needs to be $N \times (ML)$, where L is the first stage integration size. This is a solution used for small or moderate N arrays or narrow bandwidth receivers. For example, the BEST-2 correlator presented in Chapter 3 uses Quad Data Rate (QDR) memory for the cornerturn operation. But there is a limit to the size of available memory which can be used, also there is a limit to the bandwidth into and out of the memory space. Further, if the system is distributed over many f-engine boards then all the signal need to be brought together at a single point. As N grows this becomes an unreasonable limitation. A hierarchical, distributed cornerturn is needed which uses custom backplanes, network switches, or a mixture of the two to perform the data reorders and allocations.

A custom backplane is a physical implementation of Figure 2.3 using a 1-to-all network topology, Figure 2.4. This has the advantage of a direct connection from each f-engine board to all x-engine boards. Since the backplane is customized for a specific layout then there will be a lower power rate compared to network switches with ancillary logic. The direct connection allows for simple data transfer and communications, usually over standard Low Voltage Differential Signal (LVDS) lines. But, inversely, it has the disadvantage of being fixed to a specific layout and is inflexible to changes in the network and scalability.

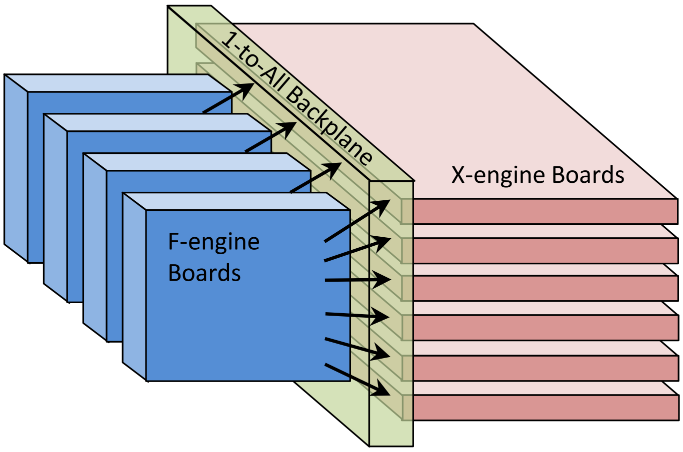


Figure 2.4: Custom backplane for a distributed FX correlator, a 1-to-all network topology is used to send subband sets from the f-engine boards to the x-engine boards.

The exponential growth in internet data rates has led to a boom in networking technology. 10 Gigabit Ethernet switches are common with 40 and 100 Gigabit Ethernet switches in development. The number of ports on these switches scale from a ~ 10 to thousands. Which means these switches can handle Terabit data rates across all nodes on a switch. This allows for significant flexibility is a correlator system layout. And, by using commercial switches the cost of development is avoided. To make use of these switches data must be packetized into common communication protocols such as Ethernet which introduces metadata overhead which takes up resources on the FPGA. The power usage of switches is higher compared to custom backplanes as they are more generic.

An analogy of the differences between custom backplanes and network switches can be made to the difference between ASICs and FPGAs. Network switches allow flexibility at the cost of power consumption and reconfiguration issues. While custom backplanes are low power and efficient but require initial Non-recoverable Engineering (NRE) costs and the design becomes fixed early in system design. It is advantageous to use both in a design. For multi-FPGA broads there should be a fixed topology to communicate between chips but using network switches between broads allows for design flexibility. In the end, a custom backplane maybe the better solution for a fixed system which will be used over many years where power consumption becomes a dominating cost.

2.1.7 Cross Correlation

The CMAC operation is simple, it consists of four multipliers, two adders, and two accumulators. For correlation of 4 bits and fewer, entire operation can be wrapped into large, custom DSP circuits available in most modern FPGAs, see [Hickish, 2013] for a review. The difficulty is that the number of correlation circuits scales as $O(N^2)$. The correlation of all antenna elements is an $N \times N$ matrix multiplication. Since the correlation matrix is Hermitian only half of the entries need to be computed.

N antenna signals sampled at a clock f requires $\frac{N^2}{2}f$ operations for correlation. For example, correlation of the full 100 MHz (200 Msps) for a LOFAR international station ($N = 192$) equates to 3.686 Tops/s. It should be noted that the number of operations for the cross correlation is independent of the number of frequency channels. In this example the number of operations exceeds the logic resources on a single FPGA. A Xilinx Virtex 6 SX475 contains 2016 DSP48e slices [XILINX, 2012c]. A typical clock speed for a Virtex 6 is 250 MHz. Each DSP48e slice can be used as a CMAC and assuming 100% utilization of DSP48e slices at this clock the maximum number of operations is 0.504 Tops/s. This is an upper limit on operations, a fully utilized design at 250 MHz would be difficult to achieve. Thus, the correlation is spread across multiple boards. The correlation is distributed across B x-engine boards, where each board processes $\frac{M}{B}$ subbands using $\frac{N^2}{2B}f$ operations.

The implementation of the full array correlation uses either a ‘pipelined’ or ‘matrix’ structure. A pipelined structure makes efficient use of the multipliers at the cost of delays implemented in memory. While the matrix structure minimizes memory usage over multiplier efficiency. The structures provide different advantages depending on which resources we choose to optimize. And, hybrid structures can be implemented which use both pipelined and matrix structures at different hierarchical levels.

2.1.7.1 Pipelined Structure

The pipeline design, presented in [Urry et al., 2007], is constructed out of an auto-correlation and $N/2$ cross-correlation ‘taps’ where the i^{th} tap computes the correlation between antennas A_j and A_{j+i} for every antenna A_j of N total antennas. Figure 2.5 shows a schematic layout of the pipelined x-engine. Using a loopback, the i^{th} tap also computes the correlation of antennas A_j and $A_{N/2+j+i}$ during the clocks when the redundant entires in the correlation matrix would have been computed. This reduces the number of taps to $N/2$, but still requires N delays of length L. Where L is the integration length of the accumulator in the CMAC.

The resource requirement for the pipelined structure is $N/2 + 1$ CMACs, N delays of length L, and some additional glue logic for the loopback. When using Xilinx Virtex FPGAs, a CMAC is usually implemented in a single DSP48e. Though it can be made in the CLBs using LUTs. Delays usually are implemented using BRAM, a typical integration length is $L = 128$. This value is dependent on the number of taps, and the output data rate and bitwidth. Again using the Virtex FPGA as an example, the

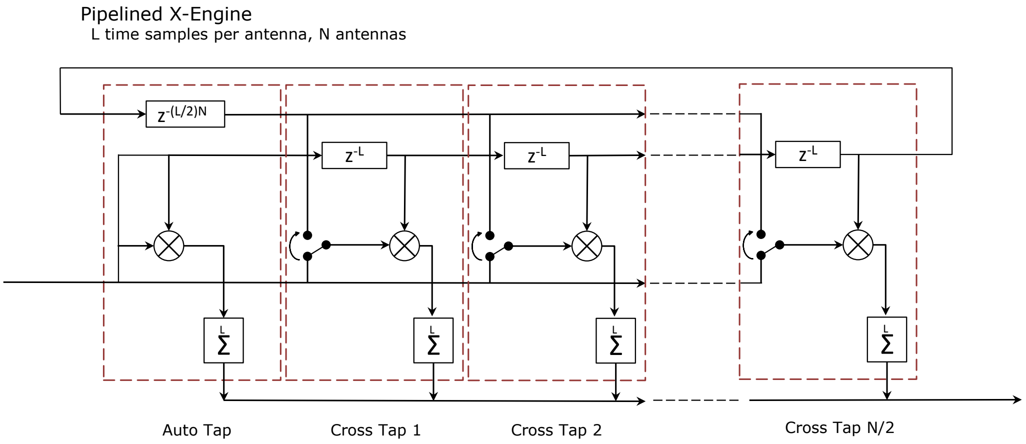


Figure 2.5: Pipelined X-engine design schematic. The input is ordered as L time samples per antenna per frequency channel. An accumulation stage after the complex multiply reduces the data rate of each tap. Outputs are multiplexed onto the same output using a valid signal.

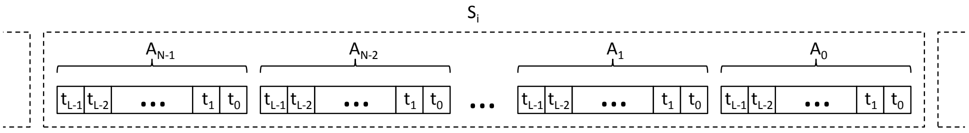


Figure 2.6: Input data ordering for the pipelined x-engine. Windows consist of L subband samples for N antennas. A cornerturn of size $N \times (ML)$ is used for this ordering.

minimum BRAM depth is 1024. This means the memory resources are underutilized until $L = 1024$. Distributed memory can be used but this would use up a large section of the CLBs and make timing difficult. Half of the delays are part of the loopback which can be implemented as a single delay of length $NL/2$, this is usually much larger than 1024 and would more efficiently use BRAM resources.

The input data structure of the pipelined x-engine is in subband windows of $N \times L$ words which are the values of all antennas (N) over an integration length (L) for subband S_i, Figure 2.6. As noted in the cornerturn section, this ordering requires a cornerturn of $N \times (ML)$.

The main advantage of the pipelined x-engine is efficient use of the multiplier on every clock. A pipelined x-engine (a single chip can have multiple x-engines depending on resources) can process a bandwidth of $\Delta f = \frac{f_x}{N}$ where f_x is the clock rate of the FPGA. In the past, multiplier resources on FPGAs have been limited. But, in modern FPGAs the multiplier to BRAM ratio is moving towards designs that will become BRAM limited before they are multiplier limited.

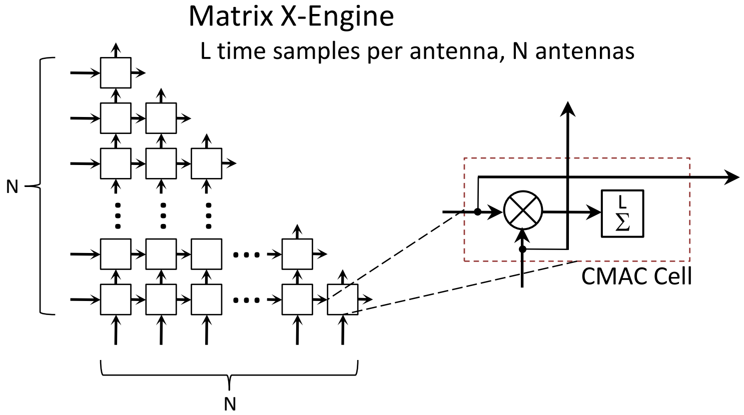


Figure 2.7: Correlation using a systolic array of CMAC nodes. This requires $N^2/2$ CMACs, but requires no memory for delays.

2.1.7.2 Matrix Structure

Matrix structure correlation is based on systolic array networks. These networks are laid out in grids (multi-dimensional) with each node performing interconnected parallel operations. A common use of systolic arrays is for matrix multiplication where $\mathbf{AB}$ is computed by passing $\mathbf{A}$ from left to right and $\mathbf{B}$ from bottom to top. The correlation operation follows the same method, Figure 2.7. The left to right input, $\mathbf{A}$, is an $N \times 1$ vector containing a single time sample, single frequency subband for all antennas, N times. The bottom to top input is $\mathbf{B} = \mathbf{A}^T$. Due to the Hermitian properties of the correlation matrix only half of the entries need to be computed. Each node contains a CMAC which accumulates L samples, this minimum length is set by the desired output data rate.

Using the matrix structure requires $N^2/2$ CMACs but requires no memory for implementing delays. The cornerturn operation only requires a $N \times M$ operation because there is a CMAC devoted to every baseline. The matrix structure has the advantages of reducing memory usage in delays and the cornerturn, and processing the entire bandwidth on an x-engine board which is clocked at the f-engine sampling rate. But, the number of CMACs scale from $N/2$ in the pipelined structure to $N^2/2$. A compromise can be made to reduce the number of CMACs by using memory for either input buffers or in the accumulators [D'Addario, 2011]. The bimodal distinction between the pipelined and matrix designs is formalistic. Really, they are just different ends of an efficiency spectrum. Usually correlator implementations use hybrid designs which are balanced between logic and memory based on the FPGA resources.

2.1.8 High Bandwidth Output

The baseline accumulation length is dependent on the output bandwidth and the science requirement of an observation. An x-engine will typically have a first stage accumulation length $L \approx N$ so that the output data rate is approximately equal to the input data rate. This sets the minimum accumulation length, a second stage *vector accumulator* will be used to further accumulate the baselines. The vector accumulator is a large memory space with a large enough bit depth to hold all the correlations and a bit width large enough for bit growth during accumulation. This second stage accumulator is software controllable to tune the accumulation length to the science case.

Output data rate from the correlator depends on N and the final accumulation length. Going back to the LOFAR correlator example of $N = 192$ and $M = 512$, each accumulation will be ~ 76 MB (using 32 bits for the real and imaginary components). At one second integrations the output data rate is 607 Mbps (~ 6.55 TBs/day). This can be comfortably handled by a 1 Gb Ethernet interface. But, for subsecond integrations required for the transient science case this data rate begins to grow. With 10 ms integrations the data rate goes up to ~ 60 Gbps. Processing, let alone storage, becomes a challenge. But that is left to another time. Current high speed network interfaces use 10 Gbps Ethernet going up to 40 and 100 Gbps. Using these interfaces, as is done in the cornerturn, facilitates transport of large-N, fine channelization, and short integration time correlations to processing hardware.

2.1.9 Additional Considerations

In the previous discussion I have used FPGA chips and boards interchangeably. But, multiple FPGAs can be placed on a single board such as the Uniboard developed by ASTRON [UniBoard, 2012]. On the board, interchip communication is simplified by using direct LVDS connections. Multi-chip boards will have a reduced power usage per chip because peripheral logic can be reused by all chips. Also, the physical dimensions per chip will be reduced. But, single FPGA boards offer the most flexibility in system design.

Though this section has focused on using FPGAs we should note the use of ASICs and GPUs for correlation. The systolic array leads to using ASICs for correlation, in fact it is a network structure often implemented in Very Large Scale Integration (VLSI). Each node is an ASIC made with an array of CMACs which performs a subset of the correlations. The correlator system is laid out using these chips to the required size. The ALMA correlator uses this method.

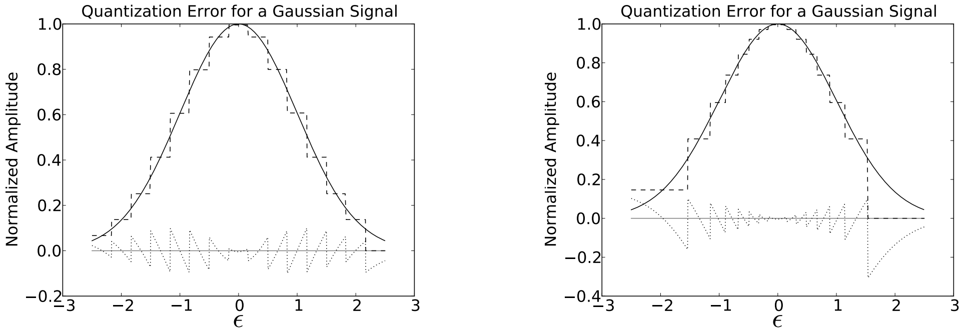
Large-N EoR arrays cover at most 100 MHz of bandwidth, but need fine channelization for RFI mitigation. The relatively small bandwidth and large correlation required for these arrays fit well into the GPU processing regime. This leads to using GPUs for the X stage of the correlation [Clark et al., 2011].

By taking into account the unique specifications of the end science goals additional optimizations can be made in the system design to reduce resource and data transportation costs with limited loss in information. In the following sections novel implementations of quantization, multiplication, and channelization are considered. A method of quantizing a signal based on an a prior model is shown to be useful in reducing data transport or increasing signal dynamic range. The low-bit multiplication used in the correlation operation can be implemented in a various DSP resources, including a 'multiplier-less' multiplication by transferring to logarithmic space. In large fractional bandwidth systems the overall resource utilization can be reduced by varying the channelization width based on the science aims.

2.2 Quantization Effects on Correlation

A digital signal is discrete in time and amplitude. Digital sampling is the conversion of a continuous analogue signal to a discrete digital representation. Quantization is a limit on SNR and dynamic range enforced on digital signals due to resource limitations. We would like to limit the amount of quantization performed in a digital design. With careful understanding of when quantization should occur and what effects it has on the data we can greatly reduce the resource cost in digital design at a minimal cost to the signal integrity.

The initial digitization of an Radio Frequency (RF) signal requires a broadband and well understood ADC with sufficient sampling levels to accommodate the required dynamic range of the radio environment. Quantization is usually implemented at later stages in a digital design. Typically in an FX correlator system after the frequency channelization, samples are quantized down to a lower bit width. The reasoning for this is multi-fold. After channelization in a typical correlator implementation the samples are represented as 18-bit complex. This large bit width is mainly to mitigate overflows in the FFT caused by narrowband RFI signals. Although the channels with RFI may have a dynamic range that covers the full 18 bit range, those channels are not useful for the end science observation. The clean channels which are free of RFI and contain an astronomical sky signal will have a much lower dynamic range, thus 18 bits are unnecessary to represent the sample. A typical FPGA-based correlator system will quantize the samples down to 4-bit complex. This reduces the overall internal and external bandwidth by a factor of 4.5, reducing the amount of interconnect required and improves FPGA timing constraints. Quantization also reduces the total resources



(a) Standard linear quantization: levels are equally spaced in standard deviation space.

(b) Gaussian quantization: levels are equally spaced in probability space.

Figure 2.8: Quantization error introduced when quantizing an 18-bit Gaussian distribution to 4 bits using different quantization schemes. The continuous distribution (solid) is quantized to 16 levels (dashed) with the resulting quantization noise (dotted).

required. The FPGA block RAMs and DSP multipliers can be utilized to pack in smaller bit width samples thus reducing total memory usage and compute multiple multiplications in a single DSP multiplier. This reduction in bandwidth and resource utilization comes at a cost of introducing quantization noise. Figure 2.8a shows the quantization error (dotted) introduced when an 18-bit Gaussian distribution (solid) is linearly quantized down to 4 bits (dashed). The distribution is defined in ϵ which is defined in [Thompson et al., 2007] as the quantization interval in units of Standard Deviation (σ), σ, of the distribution. The effect of this error has been encapsulated in the quantization efficiency η_Q, the signal to noise ratio of a digitally quantized system as a fraction of an ideal continuous Gaussian signal distribution [Thompson et al., 2007]. By quantizing down to 4 bits the quantization efficiency is reduced to 0.988457 as stated in Table 1 of [Thompson et al., 2007], the original 18-bit signal can be approximated to have a quantization efficiency of 1. A loss of 1.1% in quantization efficiency greatly reduces the system bandwidth and resource utilization costs. A balance between system resources and the introducing additional error must be struck in firmware design. For a Gaussian signal the marginal benefit of using a finer quantization scheme to improve quantization efficiency over the cost of resource utilization drops off around 4 bits.

2.2.1 Distribution in Sampling Space

If we have a priori knowledge of a signal distribution can we develop a better quantization scheme? Where ‘better’ means the scheme reduces the cumulative quantization noise of a signal. For a uniform distribution the optimal quantization scheme is a linear one since every value within the sample space is equally likely. Astronomical signals are made up of processes which follow a Gaussian amplitude distribution [Thompson et al., 2001], Section 1.2. A quantization scheme which takes into account this Gaussian distribution will reduce the cumulative quantization noise by redistributing the noise based on the likelihood of a value appearing. The Probability Density Function (PDF) for a Gaussian distribution, Eq. 2.1a, is the likelihood of a random variable to take a given value x. Where μ is the mean, and σ is the standard deviation of the distribution. Also useful is the Cumulative Distribution Function (CDF) of a Gaussian distribution, Eq. 2.1b, which is the probability of a random variable being less than or equal to x. The probability of a random variable X being between two points x_0 and x_1 is then $\mathrm{CDF}(x_1) - \mathrm{CDF}(x_0)$. When quantizing a signal which follows a Gaussian distribution to L levels an ideal encoding is one where each interval represents an equal area under the Gaussian PDF. This assures that each interval statistically will have an equal number of samples. It should be noted that each sample is made up of a real and imaginary component, which for this discussion are treated as two independent values. Transforming each sample into amplitude and phase may prove to be an interesting study for quantization, but is not discussed here.

$$\mathrm{PDF}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.1a}$$

$$\mathrm{CDF}(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right] \tag{2.1b}$$

This decomposition of the signal can be thought of as a one-dimensional Voronoi tessellation where the distance metric is the area under the Gaussian curve between two points instead of the Euclidean distance between points(in the case of linear quantization). Quantization using $L = 16$ of a Gaussian distribution is shown in Figure 2.8b. The interval spacing is narrow towards the center of the distribution and widens further out. The more likely samples near the center of the distribution will have reduced quantization noise at the cost of the more rare, further out samples being poorly quantized. This non-linear quantization technique is similar to the μ-law and A-law companding algorithms used in telephony to encode human speech at low bitrates with improved SNR at the cost of dynamic range.

The optimal choice of ϵ is dependent on the number of quantization intervals, both [Thompson et al., 2007] and [Jenet & Anderson, 1998] derived similar values for various levels. The values of ϵ derived from [Thompson et al., 2007] are used in this discussion.

The resulting quantization noise distribution from using a linear and Gaussian quantization scheme can be seen in Figure 2.9. Using $L = 16$, $\sigma = 1$, and $\epsilon = 0.3125$, N samples from a normal distribution are quantized using a linear and Gaussian

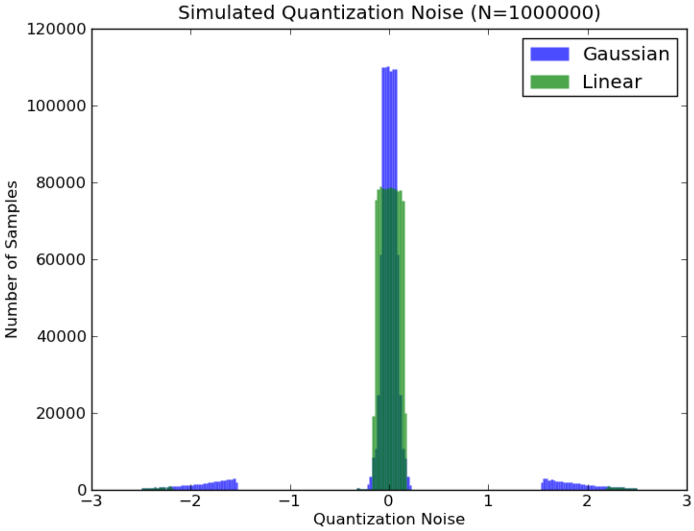


Figure 2.9: Quantization noise Monte Carlo simulation ($N = 1e6$) when using a linear (green) and Gaussian (blue) encoding scheme. The error 'wings' of the Gaussian encoding are the cost of improving the quantization noise for samples close to the center of the distribution.

scheme. The Gaussian scheme has a tighter noise distribution around the center but at the cost of higher 'wings' further from the center compared to that of the linear scheme. In an implementation of the Gaussian encoding scheme the noisy wings can be removed by flagging those samples which has an effect of decreasing the dynamic range but improving the quantization SNR.

2.2.2 Quantization for Signal Transport

The implementation of a companding algorithm for quantization of a Gaussian signal can be used to reduce the interboard bandwidth of a digital system and increase the SNR by reducing the quantization noise. To take advantage of digital logic resources such as multipliers signals must be in a linear encoding scheme, thus this quantization scheme's real advantage is in signal transport. Multipliers, adders and accumulators within a FPGA rely on a linear base-2 number representation to perform fast operations. Implementation of a multiplier which takes into account the scaling of a non-linear system would use a significant amount of resources over a standard multiplier core, though a LUT implementation could be done. An encoder/decoder system must be used to return the transported signal to a linear representation. Further, the quantization of the signal into a Gaussian encoding is limited to step sizes which are multiples of the final decoded signal step size. That is if we want to represent, after transport, an 8-bit number which ranges from -1 to 1 and has a quantization interval $\Delta l = \frac{1}{128}$ with a 4-bit Gaussian quantization scheme then each interval is required to be $\delta_{g,i} = a_i \Delta l$ where a_i is an integer. Figure 2.10 shows such an encoding where a 4-bit Gaussian encoding is used and is decoded to 5-bit linear after signal transport, each interval in the Gaussian encoding is a multiple of the linear 5-bit quantization interval.

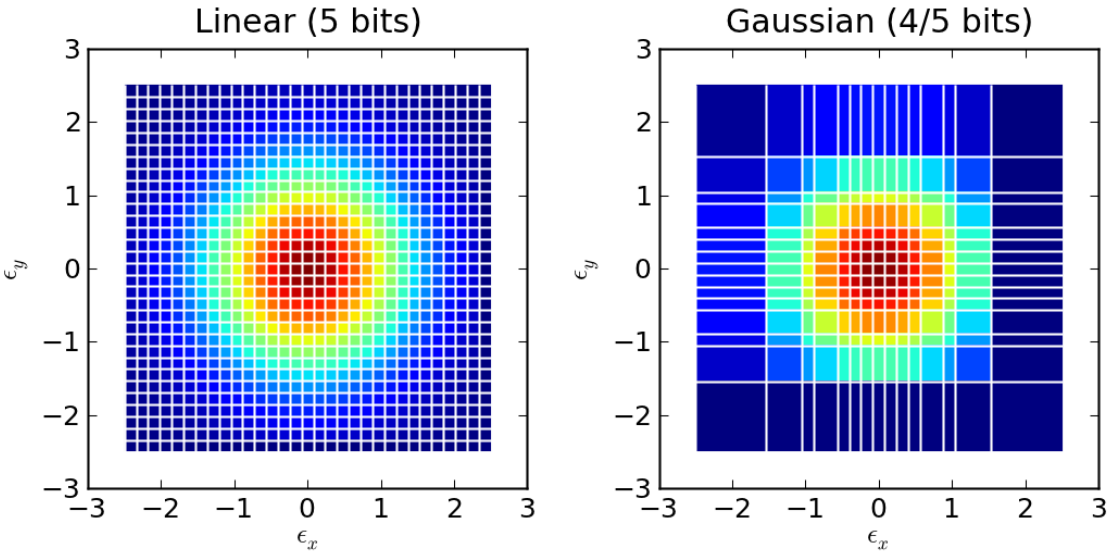


Figure 2.10: Linear 5-bit encoding of a Gaussian distribution(left) and Gaussian 4-bit encoding with intervals based on linear 5-bit multiples(right).

The immediate effect of Gaussian encoding before signal transport is a reduction in bandwidth. Transport of signals across broads and over networks involves a non-trivial energy cost to transmit and receive, and a well designed network topology. Reduction in bandwidth utilization have clear advantages in system architecture design which can reduce the overall interconnect and operational costs. But, of course, this savings in bandwidth comes at a cost. There is an improvement in the quantization noise using the Gaussian encoding over a linear encoding of the same number of encoding bits, n_{en}. The number of decoding bits, n_{de}, is always greater than n_{en}. In decoding back to linear from Gaussian there is a loss in dynamic range compared to if the signal was linearly encoded with n_{de} bits before transport. Thus the Gaussian encoding is an improvement over a linear encoding with n_{en} bits that after transport was re-encoded at n_{de} bits (there is no gain in dynamic range in this situation), but not as good as a signal originally encoded with n_{de} bits.

The quantization noise resulting from a particular encoding scheme (linear, Gaussian), S, using n bits can be defined as Eq. 2.2. Where X is a set of N random samples taken from a normal distribution. The function $Q_{S,n}$ is the quantization function using scheme S and n bits applied to X. The resulting RMS error between the quantized signal and the original is a measure of the quantization noise distribution.

$$\xi_{S,n} = \sigma(Q_{S,n}(X) - X) \tag{2.2}$$

In Table 2.1 various encoding and decoding schemes are listed along with their resulting quantization noise values based on Monte Carlo simulations using $N = 1e7$

Encode (bits)	Decode (bits)		Quantization Noise		
n_{en}	n_{de}	BW_{en}/BW_{de}	$\xi_{lin,en}$	$\xi_{gauss,en}$	$\xi_{lin,de}$
2	2	1	0.4990	–	0.4990
3	3	1	0.1921	–	0.1921
2	4	.5	0.4990	0.4990	0.0945
3	4	.75	0.1921	0.1221	0.0945
4	4	1	0.0945	–	0.0945
2	5	.4	0.4990	0.4990	0.0463
3	5	.6	0.1921	0.1202	0.0463
4	5	.8	0.0945	0.0609	0.0463
5	5	1	0.0463	–	0.0463
3	6	.5	0.1921	0.1161	0.0228
4	6	.67	0.0945	0.0584	0.0228
5	6	.83	0.0463	0.0296	0.0228
6	6	1	0.0228	–	0.0228
4	7	.57	0.0945	0.0567	0.0114
5	7	.71	0.0463	0.0291	0.0114
6	7	.86	0.0228	0.0152	0.0114
7	7	1	0.0114	–	0.0114
4	8	.5	0.0945	0.0558	0.0057
5	8	.625	0.0463	0.0278	0.0057
6	8	.75	0.0228	0.0141	0.0057
7	8	.875	0.0114	0.0075	0.0057
8	8	1	0.0057	–	0.0057

Table 2.1: Resulting quantization noise from various bit encodings using linear and Gaussian encoding schemes. Column 1 and 2 list the number of bits for signal encoding and decoding. Column 3 is the fractional bandwidth resulting from using a Gaussian encoding scheme for transport. The rows where $n_{en} = n_{de}$ are shown for reference, no Gaussian encoding is necessary in these cases.

samples. The quantization noise columns $\xi_{lin,en}$ and $\xi_{lin,de}$ list the resulting noise when using n_{en} and n_{de} bits respectively. As expected the $\xi_{gauss,en}$ column shows that the Gaussian encoding scheme always does as good or better than $\xi_{lin,en}$. As the difference between n_{en} and n_{de} shrinks the Gaussian encoding begins to approach $\xi_{lin,de}$ noise. The third column shows the bandwidth savings by using the Gaussian encoding with n_{en} bits before signal transport against using the linear encoding with n_{de} bits. By using a Gaussian encoding scheme for transport the increase in quantization noise can be reduced against that of a linear encoding while still maintaining a reduction in bandwidth.

Hardware implementation of this Gaussian companding method requires additional resources on top of the quantization logic. To encode a n_{de} bit signal using n_{en} bits the original signal should be quantized down to n_{de} bits initially. This quantization logic is dependent on the DSP implementation which can be as simple as a bit truncation or include rounding operations. A $n_{de} \times n_{en}$ conversion table will convert the linear n_{de} signal to a Gaussian n_{en} bit encoding. This can be implemented in block RAM using either shared memory, which allows run time configuration of the encoding, or as a ROM with fixed values. For small values of n_{en} and n_{de} the table can also be implemented using LUTs. After signal transport the same table can be used to convert the n_{en} bit Gaussian signal back to n_{de} bit linear.

Signal quantization is always a balance between signal integrity and resource allowances. Ideally we have can afford as many bits as necessary in the computation and data transport. But there is a cost point both in the science and engineering at which quantization becomes necessary. By having knowledge of a prior signal distribution the computation and transport cost can be reduced or the signal dynamic range can be improved.

A Gaussian signal distribution has been used in this section as it best represents an astronomical signal in a correlator system. But, any quantization scheme can be created to optimize a given distribution. For example, in an environment with a strong source such as a satellite transmitter the distribution of bits can be separated between the astronomical Gaussian signal and the strong Satellite transmitter.

2.3 Optimal Multiplier Implementation

Binary multiplication can be performed with a variety of algorithms. They vary in resource efficiency, number of clocks cycles, and simplicity of design. On FPGAs multiplication is either done with specific multiplier circuits, such as the DSP48 slices on Xilinx chips, or in logic. In logic a multiplication is implemented using partial products with LUTs. On Xilinx Virtex-5 and higher the LUTs are 6 input, and 1 or 2 output [XILINX, 2012a]. Each CLB contains four of these LUT6's. Altera's basic logic building block, Advanced Logic Module (ALM), has a single 8-input 'fracturable' LUT which can be programmed into smaller LUTs [ALTERA, 2011]. As the size of the multiplication grows the amount of logic requires grows roughly as the square of the bitwidth. To alleviate this resource drain modern FPGAs include custom multiplier circuits. A single Xilinx DSP48e can perform up to a 18×25 multiplication, or be chained together for larger multiplications. The Altera Stratix series include 18×18 multipliers. For large multiplications these custom DSP slices are the optimal solution [XILINX, 2011a] [ALTERA, 2004]. But, for correlation only a few bits are used. The optimal solution for these multiplications vary depending on bitwidth and FPGA architectures.

Using a 18×18 or 18×25 DSP slice for a 4×4 multiplication is a waste of resources. Multiple small bit multiplications can be performed with a single DSP slice in parallel

[XILINX, 2003]. Fitting multiple signed multiplications in a DSP slice requires an n-bit number to be sign extended to $2n$ bits. But, an unsigned multiplication only requires a single bit of padding. For this savings it can be advantageous to convert signed to unsigned numbers and keep track of the sign with additional logic, then convert back to signed after the multiplication.

Before the widespread use of reconfigurable logic and multipliers in DSP, various logarithmic approximation methods were used. These methods performed a logarithmic encoding, added the signals, and then decoded back to linear. The first method, called Mitchell's algorithm [Mitchell, 1962], used a linear fit to the logarithm curve. Implementation of this algorithm for FPGAs along with the resulting error from the approximation is discussed below.

Table 2.2 shows the resource usage of multipliers using 'soft' logic using CLBs, embedded DSP48e's, and Mitchell's algorithm. The table only shows resources for unsigned values. Signed values can be converted to unsigned and the sign can be tracked during multiplication with a small amount of logic overhead. Signed multiplication has the effect of increasing the number of LUTs in logic-based multiplication by $n - 3$ (n is the number of bits of the input signals). In embedded multiplication there is a larger effect of using signed values, the number of multiplications possible is reduced due to sign extension.

The error rate of Mitchell's algorithm is below 1 bit for 4×4 and smaller multiplications. With larger multiplications the linear approximation to the logarithmic curve breaks down and begins to introduce significant error. Though, this error rate might still be considered reasonable with larger multiplications if the desire is to reduce resource usage. More advanced logarithmic approximation algorithms exists which better approximate a linear response for larger bit depths [Lee, 2012].

For multiplication of signals where $n \leq 3$ logic implementation is the most resource efficient based on the current configuration of Xilinx CLBs. At $n = 4$ all three methods are roughly equal, thought Mitchell's algorithm has a slight improvement of resources over the standard 'soft' multiplier. This is good news for 4-bit correlators, as this shows a design would make efficient use of resources by using DSP48e slices and logic for multiplication. In the $n = 5$ and $n = 6$ region Mitchell's algorithm provides the best use of resources but as noted there is an included error. Beyond $n = 6$ embedded multipliers should be used as the interconnect between logic in the 'soft' multipliers becomes the dominating cost.

Configuration	Type	Slices	Flip-Flops	LUT6's	DSP48e's	# Multipliers
2×2	Logic	6	12	3	0	1
2×2	Embedded	32	64	32	1	4
2×2	Mitchell's	6	8	6	0	1
3×3	Logic	11	21	13	0	1
3×3	Embedded	36	72	36	1	3
3×3	Mitchell's	9	12	9	0	1
4×4	Logic	14	28	17	0	1
4×4	Embedded	24	48	24	1	2
4×4	Mitchell's	11	14	13	0	1
5×5	Logic	24	46	33	0	1
5×5	Embedded	30	60	30	1	2
5×5	Mitchell's†	15	18	19	0	1
6×6	Logic	28	55	39	0	1
6×6	Embedded	36	72	36	1	2
6×6	Mitchell's†	15	20	21	0	1
7×7	Logic	37	71	62	0	1
7×7	Embedded	14	28	14	1	1
8×8	Logic	41	81	71	0	1
8×8	Embedded	16	32	16	1	1

Table 2.2: Resource utilization on Xilinx FPGA's for unsigned multiplication using logic 'soft' multipliers, embedded DSP48e slices, and Mitchell's logarithm approximation algorithm. 1-bit multiplication is an AND operation and has not been included. Resource usage for the embedded multiplier entries are assuming full use of the DSP48e slice. † Above 4 bits the error rate using Mitchell's algorithm becomes larger than one bit for certain values.

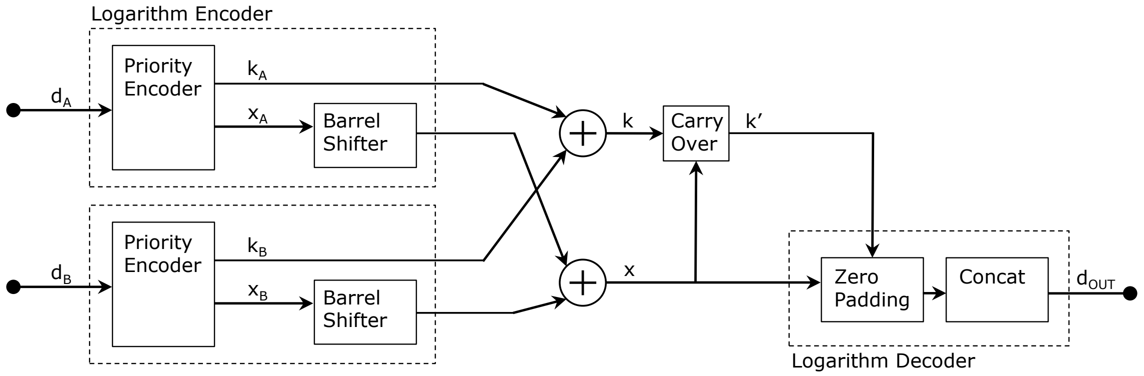


Figure 2.11: Circuit diagram of Mitchell's Algorithm for unsigned multiplication using logarithmic approximations. The encoder converts linear values to logarithmic based on an straight line approximation.

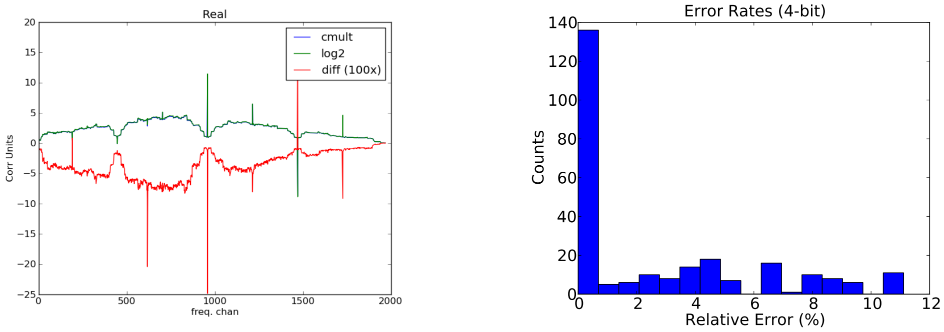
2.3.1 Logarithmic Approximation Algorithm for Low Bit Multiplication

Since the development of integrated circuits a number of algorithms to convert signal between linear and log space have been proposed [Lee, 2012]. These methods vary in circuit complexity, error response to the approximation and resource utilization. For low bit correlation, a straight line approximation is sufficient to not introduce error above a single bit. One of the earliest, and simplest, algorithms is Mitchell's Algorithm [Mitchell, 1962].

To implement Mitchell's Algorithm on an FPGA three circuit components are used. The encoder to convert a linear signal to a logarithmic signal. An adder, or subtracter (if performing a divide). And a decoder to return the summed signal to linear space. Figure 2.11 shows a block diagram of algorithm functions.

The encoder determines the characteristic (k) and the mantissa (x) of the logarithm. A priority encoder determines k based on the position of the most significant bit which is equal to one. Based on k the input signal is barrel shifted and zero padded to produce x. k and x are summed separately with carry over logic from x to k. The decoder zero pads the MSBs of k based on the value of k and concatenates k with x. The Least Significant Bit (LSB)s of the output are clipped based on the required output bitwidth.

Since this method uses an approximation we should note the resulting error made in the approximation. Mitchell's algorithm produces a negatively biased approximation to the logarithmic curve. The resulting error in multiplication is between 0 and -11.1% [Mitchell, 1962]. The distribution in error between a standard 4-bit product and the approximation is shown in Figure 2.12b. Over half the values are equivalent to the real values. The remaining relative error distribution is roughly flat. Figure 2.12a shows the



(a) Spectrum (real portion) comparison of Mitchell's algorithm (green) to a standard multiplier (blue) for 4-bit correlation of a white noise signal. The difference (red) is multiplied by a factor of 100.

(b) Relative error rates when performing a 4-bit multiplication using Mitchell's logarithmic approximation.

Figure 2.12

results from a 4-bit cross correlation using the logarithmic approximation algorithm. The resulting difference is on the order of a few percent.

2.4 Fractional Channelization of Large Bandwidths

When channelizing a time series into frequency subbands each resulting subband has the same fixed resolution in frequency. This is because channelization usually involves the FFT to transform equally spaced, discrete time samples into equally spaced frequency subbands. Given the intended science goals of an observation this can result in an unnecessary over-sampling of the desired frequency channelization. By reducing this over sampling we can reduce the cost of correlation, data transport, and cornerturn memory at no loss to the data quality.

Let us consider a few common science observations using correlator systems. When observing a continuum source the sensitivity is related to the available bandwidth. The limitation on maximum channel width is in the case of a high RFI environment where narrow channels are required to excise narrow band signals while preserving the rest of the band. And when bandwidth smearing creates distortions within an image. When tracking a phase center during observations fringe stopping is used, this is a phase applied to each antenna signal to account for the changing baseline projections. This phase value is updated on a timescale necessary to keep the phase at the phase center flat to within some nominal phase error level. The time scale is usually related to the longest baseline which will have the fastest fringe rate. The fringe rate is dependent on the baseline length, which is in turn based on the observing frequency. Each channel will have a different central frequency in which the desired fringe stopping phase is based. But a channel has a finite bandwidth Δf, the phase only accounts for the center frequency of that bandwidth. On the edges of the channel an additional error is introduced which leads to radial smearing in the image that increases further away from the phase tracking center. The amount of smearing in an image is proportional to the fractional bandwidth $\frac{\Delta f}{f}$ and the distance from the phase tracking center $\sqrt{l^2 + m^2}$, [Taylor et al., 1999] Chapter 17 and 18. Since bandwidth smearing scales in relation to the distance from the image phase center distortion increases in images covering larger fields. For a constant smearing factor across the band the frequency channel widths can thus scale as $\frac{\Delta f}{f}$.

On the other side from continuum observations is spectral lines observations which require sufficient resolution to detect narrow spectral lines and possibly resolve them. Spectral lines are narrow absorption or emission features in a continuum source causes by quantum transitions. Broadening of spectral lines is due to three processes which define the line profile [Carroll & Ostlie, 1996].

Natural broadening due to the natural uncertainty of the energy measurement.

Doppler broadening thermal motion of the atoms induce a broadening of the line due to the Doppler effect.

Pressure broadening due to interactions with the surrounding medium.

Natural and pressure broadening produce a similar line profile where $\Delta f \propto f^2$. The two processes produce comparable profiles, though pressure broadening can at times be much wider. Thermal broadening has the relation of $\Delta f \propto f$ from the Doppler effect. Channel width during a spectral line observation need at least a $\frac{\Delta f}{f}$ relation to account for the Doppler broadening.

Further, experiments to make measurements in spacings that are constant in redshift, such as in EoR or BAOs power spectrum measurements, follow a frequency spacing which scales as $\frac{\Delta f}{f}$.

From [Thompson et al., 2001], Section 13.6 the measurement of Faraday rotation due to galactic magnetic fields and the ionosphere has a frequency dependence of $\frac{\Delta f}{f^2}$.

Lastly, though RFI comes in many forms, including broadband and intermittent sources, strong narrowband transmitters tend to have intrinsic widths proportional to $\frac{\Delta f}{f}$ as set by radio spectrum allocation agencies.

Any arbitrary exponent factor d can be chosen to accommodate the science goals of an observation and the radio environment of the array such that channel widths are $\Delta f \propto f^d$.

For a band from f_s to f_f which is channelized into M channels the width of a channel k is Δf_k (Eq. 2.3). Where Δf_{min} is the minimum channel width required for any part of the band, this is usually the channel width of the first channel, i.e. $\Delta f_{min} = \Delta f_0$. This minimum channel width will be set by the size of the FFT window and sampling rate. The central frequency of channel k is f_k and d is the exponent scaling factor of the channel width. For the general case of constant Δf_k across the band $d = 0$ and f_{min} is the width of the FFT channels. The $d = 1$ case is a special case and a transform called the constant Q transform is often employed in audio and music signal processing since it is flat in logarithmic space which is similar to the human aural response [Brown, 1991]. In order to approximate a constant fractional bandwidth response a step function is employed since the channel widths will have an intrinsic quantization width based on the size of the FFT window and the sampling rate, e.g. a channel can not be of size $2.5\Delta f_{min}$.

$$\Delta f_k \equiv \Delta f_{min} \left(\frac{f_k}{f_s} \right)^d \tag{2.3}$$

To choose the set of central frequencies for the M channels across a band with a scaling factor d we set the restriction that the k^{th} channel is an integer multiple of Δf_{min}, (Eq. 2.4). As M increases, Δf_{min} decreases and the channel width better approximate a continuous f^d curve, in the $d = 0$ case any number of M perfectly represents the curve, as it is flat. Figure 2.13 shows the scalings for $d = 0, 1, 2$ across four octaves of bandwidth. The rate in channel width over center frequency has been normalized such that $\Delta f_{min} = 1$. The step functions in the figure are at the integer multiples of Δf_{min} for $d = 2$ (blue) and $d = 1$ (red), such that channel widths are that multiple of Δf_{min}.

$$\left(\frac{f_k}{f_s} \right)^d = \chi_k, \ \chi_k \in \mathbb{Z} \tag{2.4}$$

For small bandwidth systems, that is systems with bandwidth that have fractional octave bandwidths ($\leq$ 1octave), using $d = 0$ channelization is sufficient as $\frac{\Delta f_0}{f_0} \approx \frac{\Delta f_{M-1}}{f_{M-1}}$. The bandwidth in octaves for a given system which covers a band from f_s to f_f is $(f_f - f_s)/f_s$. But for systems which cover multiple octaves of bandwidth fractional channelization can reduce the cost of the digital processing system. By reducing the number of channels from the number of FFT channels, M_{FFT}, to M_{frac} the correlator bandwidth is decreased by a factor of $\frac{M_{FFT}}{M_{frac}}$. Further, the number of correlations required scales linearly with the number of frequency channels, by reducing the number of channels a significant savings can be made in resource utilization of a large bandwidth correlator.

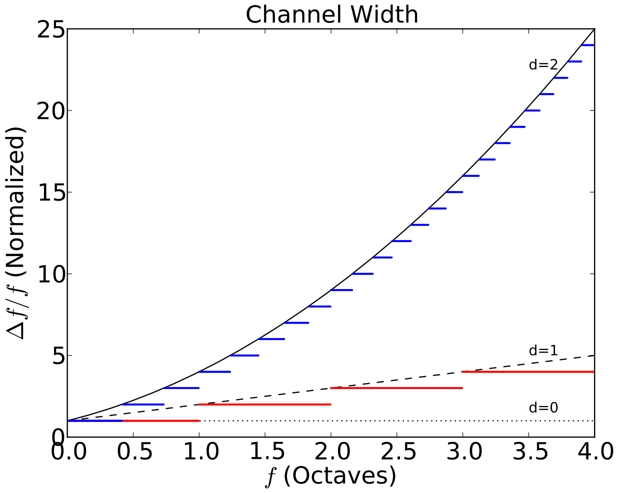


Figure 2.13: Channel scaling relative to a Δf_{min} channel width across a system with four octaves of bandwidth based on d scaling. The black lines show the ideal channelization scaling. The red ($d = 1$) and blue ($d = 2$) lines are the channelization scalings as an integer multiple of Δf_{min} A typical channelization scheme uses $d = 0$ (dotted line).

Using the latest system description for the SKA phase 1 low and mid-frequency designs [Dewdney et al., 2010] as an example wideband array we can look at the potential correlator design savings by using fractional bandwidth compression. The relevant design specifications are listed in Table 2.3 along with the fractional savings in resources and bandwidth. Since the resource requirements and bandwidth scale linearly with the number of channels for a correlator we can define the saving of using fractional bandwidth compression as $S_d = M_d/M_{FFT}$. Channel compression is dependent on the d scaling factor and the bandwidth coverage in octaves. The SKA-lo band covers over five octaves, this allows for a significant reduction in channels. For observing modes where $d = 1$ the total correlator resources can be reduced to 43% of the original resources while still maintaining the same observing bandwidth coverage. This is further improved if making rotation measure observations, $d = 2$, where the total resources are reduced by a factor of five. With the SKA-mid band the frequency coverage is only an octave and so there is no improvement in resource utilization for the $d = 1$ case, in the $d = 2$ case resources are reduced to 66% of original. These compression rates are lower limits on the resource savings as fine channelization of portions of the band will still be required for narrow band RFI and science requirements.

Now let us consider the noise characteristic of a compressed spectrum. For constant channel bandwidth the noise for any channel is $\sigma \propto \frac{1}{\sqrt{\Delta f_{min} \tau}}$ where τ is the integration time which is inversely proportional to the sampling rate. The noise for a channel k in a compressed spectrum is $\sigma \propto \frac{1}{\sqrt{N_k \Delta f_{min} \tau}}$ where N_k is the bandwidth of channel k in units of Δf_{min}. This channel dependent noise characteristic should be noted when imaging and determining noise levels. The number of channel compression steps, as shown in Figure 2.13, can be controlled by the size of Δf_{min}. By decreasing Δf_{min}

	SKA-low Phase 1		SKA-Mid Phase 1
	(coarse)	(fine)	
f_s (MHz)	70	70	1500*
f_f (MHz)	450	450	3000*
BW (MHz)	380	380	1500
Δf_{min} (kHz)	125	1	22
$M_{chan}(d=0)$	3060	$3.8e5$	$6.7e4$
$M_{chan}(d=1)$	1328	164834	$6.7e4$
S_1	0.434	0.434	1.0
$M_{chan}(d=2)$	580	71945	44385
S_2	0.190	0.189	0.662

Table 2.3: Frequency bandwidth and channel resolution specifications for SKA-lo and SKA-mid phase 1[Dewdney et al., 2010]. *The SKA-mid will cover multiple bands, all of which cover an octave of bandwidth. S_d is the resource and bandwidth compression rate relative to a $d = 0$ constant channel bandwidth implementation.

more steps are used and better approximate a continuous $(\Delta f/f)^d$ curve at the cost of implementing a larger FFT.

2.4.1 Implementation of Correlator Channelizer

A frequency compression channelizer can be implemented in hardware in two ways: by modifying the standard FFT or by adding additional memory logic post-FFT. The modified FFT design only computes a subset of the complete N-point transform. This can be done by using a subset of the FFT twiddle factors and setting the remaining factors to zero or by 'pruning' the FFT logic of unnecessary stages. Using a subset of twiddle factors imposes no additional resources on an N-point FFT, 'pruning' can even reduce the overall cost. The downside to this implementation is the inflexibility. The amount of channel compression must be decided during the firmware design, changes to the compression require recompiling the design.

Flexibility in channel compression is possible during runtime with the second method of implementation, a block diagram is presented in Figure 2.14. First, the full N-point FFT is performed. Two memory spaces of depth N are used, the first is a buffer $C_N(x)$ where accumulated values are stored and the second is $A_N(i)$ which contains a list of memory addresses to the $C_N(x)$ buffer. At the beginning of each FFT window the C_N buffer is initialized to zero. On the i^{th} clock of the FFT window the address $x = A_N(i)$ is read out and used to read out the value of the buffer at $C_N(x)$, this value is added to the i^{th} sample of the FFT output and written back into the $C_N(x)$ memory space along with a valid flag. The buffer is then read out at the end of the window and filtered for valid channels. This implementation requires additional resources

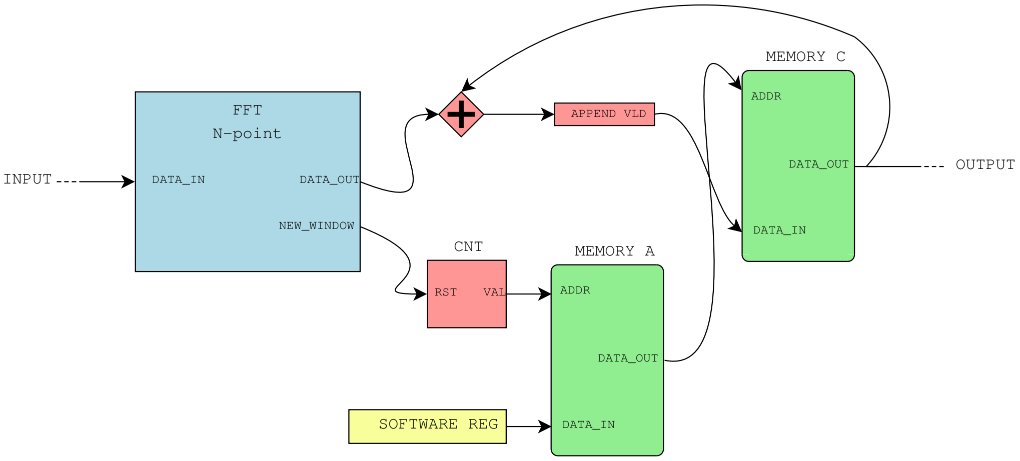


Figure 2.14: Layout of a flexible, post-FFT frequency compression design. Memory A contains the memory addresses to memory C which designate how the channels are collapsed and reordered, these addresses can be rewritten during run time. Memory C is an accumulation buffer which can hold up to N valid channels, the values are reset at the beginning of a new FFT window. Data out of Memory C is filtered for valid data only.

of two memory spaces N words deep and an adder but allows the memory addresses, and thus the channel compression, to be rewritten during runtime. The standard FFT output can be passed through this buffer by setting all the addresses to N unique values. A buffering setup also allows on the fly reordering of the FFT channels. Since the main DSP resource cost of a large-N FX correlator is the X-Engine, these additional resources will have a minimal impact on the total design while allowing for flexibility in observing modes.

With the development of multi-octave bandwidth systems a fractional channelization frontend can greatly reduce the computation and data transport cost with very little loss to the science. Thus far this has not been implemented on any radio astronomy system, but should be considered in the design of large bandwidth systems such as the SKA.

2.5 Implications for Future Correlators

The correlation operation is an easy to understand and simple to implement digital processing operation. The challenge is scaling up the design to account for the continued increase in bandwidth and number of antennas in new arrays. This requires consideration of power consumption, network topology, and computing architecture. FPGAs have a clear advantage in operations per Watt over general computing and GPUs. But, FPGAs are also difficult to program and have limited flexibility. For large N arrays the best solution with the current technology is to implement the F-Engine on

FPGAs and the X-Engine on GPUs. A few years ago the only X-Engine solution would have been on FPGAs but the growth in GPU technology has lead to a change. Though GPUs still use a more energy per operation than FPGAs, their ease in programming have proven to be a strong advantage in quickly building a system.

For the SKA correlator design the goal is not to merely build a correlator but to efficiently build a correlator. Power consumption, energy use per operation, is the key factor in the design of the SKA correlator. The technology to produce a correlator on the scale of SKA Phase 1 exists in the form of FPGAs and GPUs. Though GPUs, in their current level of development, and even perhaps FPGAs are not low power enough to sufficiently meet the power specification of the SKA digital system. Continued growth, in the classic form of Moore's Law, will bring further reductions in power per operation in both FPGAs and GPUs. A heterogeneous digital backend made up of various technologies to balance power usage, ease of programming, and acceleration will used for the SKA correlator.

Chapter 3

Instrumentation for the BEST-2 Array at Medicina Observatory

If it compiles, ship it.

-Kristian Zarb Adami, 2010

The Basic Element for SKA Training II (BEST-2) array is a subset of the Northern Cross cylindrical array, at the Radiotelescopi di Medicina in Italy [Montebugnoli et al., 2009b]. In this chapter I present a new digital backend designed for this array, implemented on FPGA-based hardware from the Collaboration for Astronomy Signal Processing and Electronics Research (CASPER)[1] [Parsons et al., 2006]. This flexible array has provided an excellent test bed for developing new correlator implementations and verification using astronomical observations. Development of an FX correlator for the BEST-2 array had a two fold purpose. In order to use the spatial FFT imager complex gain calibration coefficients need to be derived before an observation and applied in situ. Secondly, most of the DSP used in this design has been written to be configurable and is scaled up to be used in the large LOFAR-UK station correlator. The BEST-2 system is an invaluable tool for verifying digital instrumentation beyond simulations.

The digital backend developed for BEST-2 comprises a 32 element digitizer and channelizer, an FX correlator, spatial Fast Fourier Transform imager, and beamformer, implemented on Reconfigurable Open Architecture Computing Hardware (ROACH) FPGA boards. The FX correlator computes all possible baseline pairs, the number of which scales as $O(n^2)$ with the number, n, of antennas. The spatial FFT imager, which represents an alternative route to sky maps without the need of an FX correlator, takes advantage of the regularly gridded nature of the BEST-2 array to perform a spatial FFT

---

[1] https://casper.berkeley.edu/

to compute the correlation along with beamforming capabilities. For an overview on the spatial FFT imager see [Hickish, 2013]. Digital firmware development was designed and tested at Oxford and the designs transferred with minimal changes to the available hardware at Medicina.

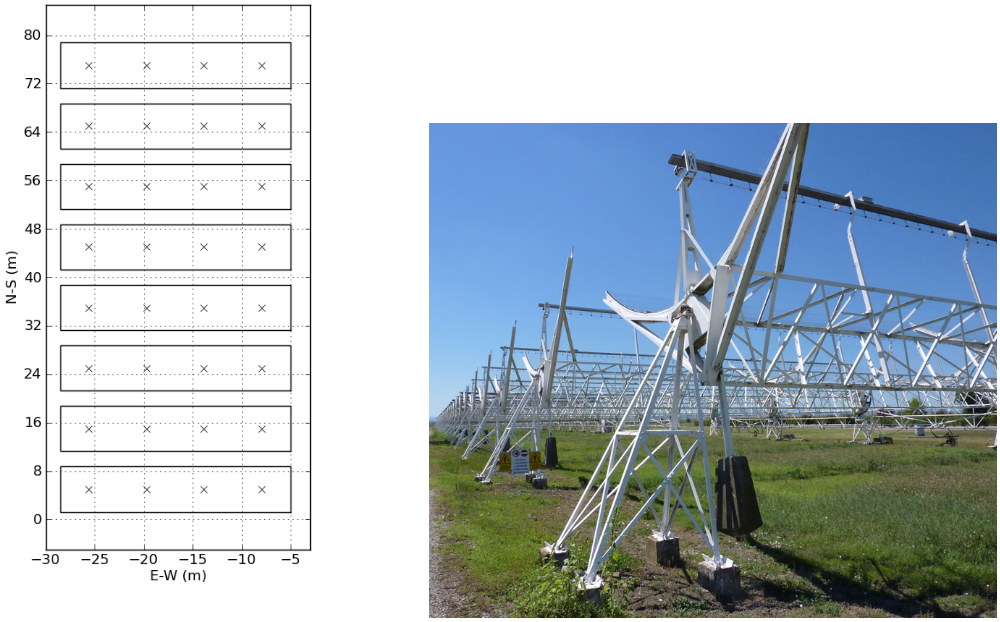
Since deployment, a number of sources have been successfully observed with the correlator and imager. Observations have been reduced and calibrated using a combination of custom software and existing, popular radio synthesis imaging packages.

This project is the work of myself and Jack Hickish. My focus of the instrument is on the FX correlator and his is on the spatial FFT imager. As will be noted in this chapter they are very much integrated as a single instrument. My focus for this chapter will be on the FX correlator as a test design for larger N arrays. I recommend viewing his thesis chapter [Hickish, 2013] on our Medicina instrumentation to complete the treatise on this work. A short instrumentation paper, [Foster et al., 2012a], provides an overview of the complete digital backend. Further, we have a results paper, [Foster et al., 2012b], which reports on the comparison of data quality between the FX correlator and spatial FFT imager.

3.1 BEST-2 Array

The BEST-2 testbed consists of 8 East-West oriented cylindrical concentrators, each with 64 dipole receivers critically sampling a focal line at 408MHz. Signals from these 64 dipoles are summed in groups of 16, resulting in 4 channels per cylinder, and a total of 32 effective receiving elements laid out on a *4-by-8* grid, shown in Figure 3.1a. The longest north-south baseline is 70 meters and east-west 17 meters. BEST-2 is an upgraded block of eight cylindrical reflectors of the 64 cylinders in the north-south arm of the Northern Cross Telescope, originally built in the 1960s.

BEST-2 was developed as a reliable, low cost frontend to be used in SKA development, with a core design requirement of simplicity in interfacing with different digital backends [Montebugnoli et al., 2009b]. Extensive documentation of the development of the analogue chain developed for BEST-2 can be found in a number of papers ([Perini, 2009]; [Perini et al., 2009]). The top level specifications of the array are listed in Table 3.1. The major improvements over the previous analogue system was a reduction in the system temperature by a factor of 10 and an increase in the analogue bandwidth by a factor of 6.4. Before the upgrade to BEST-2 the analogue system covered a 2.5 MHz band centered at 408 MHz with a system temperature $\sim$ 800 K. System temperature was reduced by installing upgraded LNAs, converting portions of the coaxial copper signal chain to optical cabling and improving the band filter design.



(a) The 32 effective receiving elements of BEST-2, indicated by crosses, lie on a regular 4 × 8 grid. Each receiver is the analogue sum of 16 dipoles, critically spaced at 408 MHz in the east-west direction.

(b) Each cylinder has 64 single polarization receiving elements. Eight cylinders of the Northern Cross make up the BEST-2 system.

Figure 3.1: BEST-2 layout and receiving system.

The bandwidth of BEST-2 is 16 MHz. This limit is chosen to match the usable, RFI-free portion of the band around 408 MHz in northern Italy. Correlation between the BEST-2 sub-array and other sections of the Northern Cross is possible as both analogue chains are still deployed. Though, the correlation is only valid over the common 2.5 MHz bandwidth between BEST-2 and the Northern Cross.

In 2008, the initial digital correlator backend of the BEST-2 array was based on previous generation FPGA boards from the CASPER group [Montebugnoli et al., 2009a]. The upgraded digital backend reduces the size of the instrument and increases the overall stability, Figure 3.2. The FX correlator design presented in this chapter provides finer frequency channelization, improved PFB window response, and higher output data rates to allow for O(millisecond) integrations. The spatial FFT imager portion of the design is a completely new addition to the backend which provides direct imaging and beamforming capabilities.

BEST-2 Array Specifications

Array Properties		
Number of Cylinders	8	
Total Number of receivers	32	
Total Collecting Area	1411.2	m^2
Antenna Temperature	35	K
Receiver Temperature	51	K
System Temperature	86	K
A_{eff}/T_{sys}	11.65	m^2/K
Longest Baseline		
E-W	17.04	m
N-S	70.00	m
Bandpass		
Central Frequency	408	MHz
Analogue Bandwidth	16	MHz
Primary Beam		
Primary Beam Size	37.62	deg^2
Declination	5.7	deg
Right Ascension	6.6	deg
PSF		
PSF FWHM	0.9	deg^2
Declination FWHM	0.52	deg
Right ascension FWHM	1.73	deg

Table 3.1: The top level specifications of the BEST-2 array, a subset of the collecting area of the Northern Cross, located in Medicina, Italy.

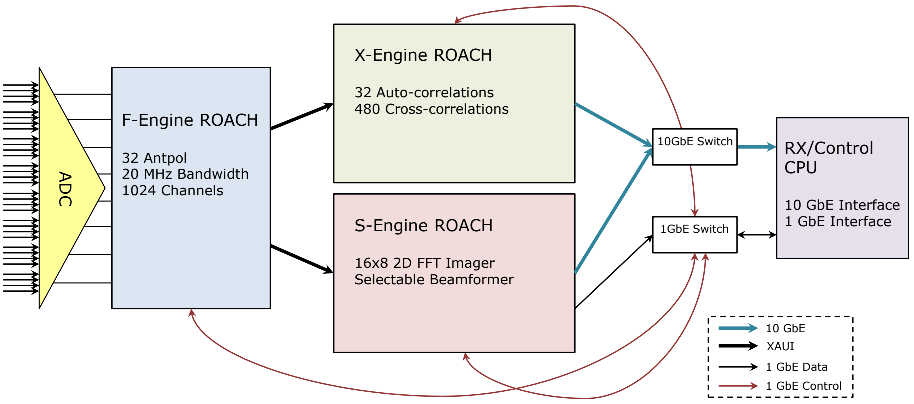


Figure 3.2: Digital backend for BEST-2 based on a three ROACH design where the F-Engine board distributes duplicate signals to a correlator and spatial FFT imager. Correlations are written to disk over a 10 Gb Ethernet interface. The spatial FFT imager outputs bother beamformed signals and integrated images.

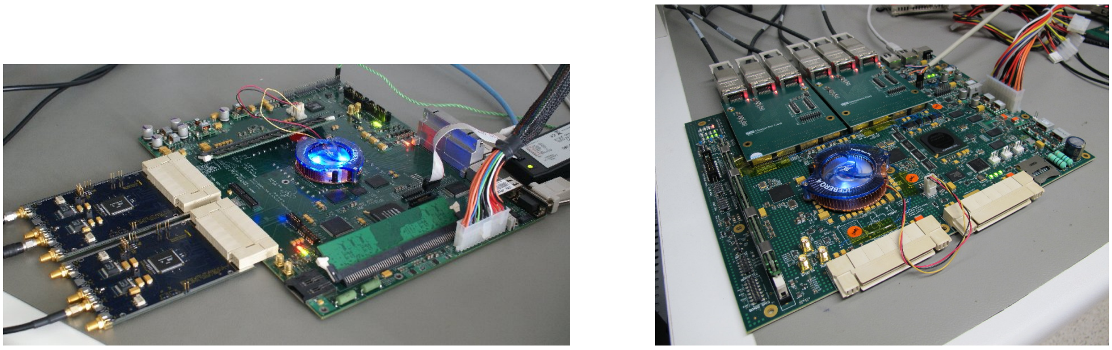
3.2 CASPER Open Source Instrumentation

The Collaboration for Astronomy Signal Processing and Electronics Research (CASPER), is a collaboration between many research groups and radio astronomy facilities throughout the world with the goal of providing a common set of digital signal processing boards and libraries for instrumentation development. By using common hardware and firmware the time and cost of digital instrumentation development is greatly reduced as much of the NRE costs associated with the development of new instrumentation are reduced. Board design also incorporates industry standard communication interfaces such as 10 Gigabit 'X' Attachment Unit Interface (XAUI) and 10 GbE to allow standard network switches to be used in building large scale systems such as correlators without requiring a custom backplane.

The CASPER DSP boards are designed to be generic with an emphasis on optimal use with radio astronomy signal processing. These boards feature high I/O data rates, adapters for attaching ADCs and other daughter cards, and FPGAs with a high density of multipliers. The current generation of CASPER boards, the ROACH series, are based around a single Xilinx FPGA. ROACH is a single Xilinx Virtex 5 SX95T FPGA based board, Figure 3.3a. The board interfaces include two Z-DOK+ 40 differential pair LVDS connectors to connect daughter cards such as ADC and Digital Analogue Converter (DAC) boards. Interboard and network communication may be performed with the four CX-4 high-speed serial interfaces. Firmware allows each CX-4 interface to be used as streaming XAUI, 10Gb Ethernet or a custom protocol interface.

The successor to the ROACH is the ROACH2 board, Figure 3.3b, which uses a Xilinx Virtex 6 SX495T FPGA. This FPGA contains about four times more logic, multipliers, memory and I/O interfaces as compared to the Virtex 5. The ROACH2 board follows the same design as the ROACH board, as it includes two Z-DOK+ connectors for attachment of daughter cards. The increase in the number of high speed serial transceivers on the Virtex 6 allows for an increase in bit-width for the off chip QDR RAM and up to eight 10 GbE connectors. Since production of the original ROACH design a new standard connector, SFP+, for 10 GbE has been introduced. To maintain backward compatibility and allow for possible future changes the high speed serial connectors the ROACH2 uses add-on mezzanine cards for flexibility in choosing which connectors to use. This also allows direct access to the GTX transceivers for use in custom interfaces.

Beyond the CASPER hardware is a set of generic DSP blocks developed to be used for firmware design. This is one of the core achievements of the CASPER community, and makes the hardware a powerful, yet easy to use, system. Much of radio astronomy signal processing relies on the same set of functional cores such as streaming FFTs,



(a) ROACH board.

(b) ROACH2 board.

Figure 3.3: CASPER developed FPGA-based signal processing boards the ROACH and ROACH2.

polyphase filter banks, and correlation. The CASPER libraries are designed and tested for optimal FPGA resource utilization and reusability.

The main firmware design process is centered around the Matlab Simulink environment which provides a graphical interface to system level design, Figure 3.4. Additionally firmware implemented with HDL such as Verilog and VHDL can be incorporated into a system design by a 'black box'. We have extended the base CASPER libraries here at Oxford to include a number of additional DSP blocks which have been added back into the main libraries for use by groups in the CASPER collaboration. The current Oxford based CASPER library is hosted on Github[2].

3.3 FX Correlator Design

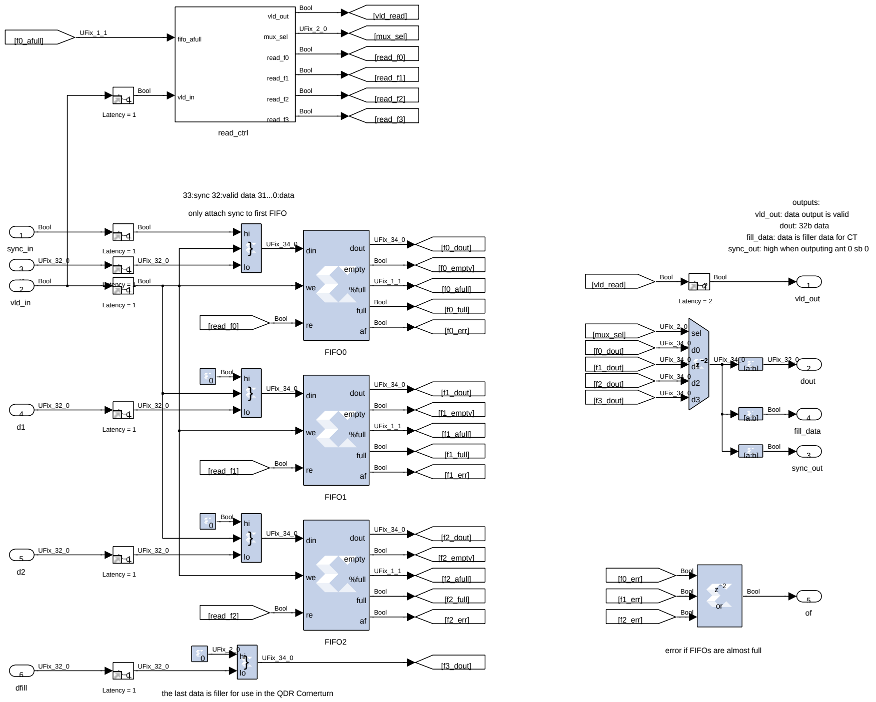
The hybrid design of the digital backend make it possible to use the FX correlator and spatial FFT imager concurrently. Both instruments use the same digitization and channelization frontend. This allows a streamlined process for calibrating the spatial FFT imager, reduces the amount of hardware required for digitization and channelization, and allows for simultaneous observation with both instruments.

The ROACH platform is based on a XILINX Virtex 5 SX95T[3] FPGA with interfaces to DRAM and QDR memory, high speed CX-4 connectors and a generic Z-DOK interface for connecting ADCs and various daughter boards, Figure 3.5. Additionally, the board has a PowerPC running the BORPH [So & Brodersen, 2008] operating system, a variant of Debian Linux, which allows access to software registers and shared memory on the FPGA. Firmware for the ROACH is designed using the MATLAB Simulink environment, Xilinx System Generator DSP blocks, and CASPER's open source DSP

---

[2] https://github.com/oxfork

[3] http://www.xilinx.com/support/documentation/virtex-5.htm


33:sync 32:valid data 31...0:data

only attach sync to first FIFO

the last data is filler for use in the QDR Cornerturn

outputs:
vld_out: data output is valid
dout: 32b data
fill_data: data is filler data for CT
sync_out: high when outputing ant 0 sb 0

error if FIFOs are almost full

Figure 3.4: Matlab Simulink is a graphical tool for development of FPGA firmware. DSP blocks are representations of HDL logic which is implemented of FPGAs using logic slices and BRAM memory or Xilinx specific logic cores such as the FIFOs in this figure.

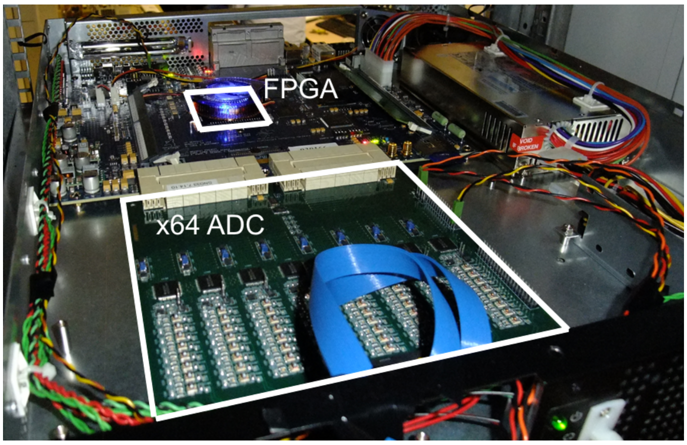


Figure 3.5: The 'F-engine' ROACH board, a Virtex 5 SX95T FPGA board, with the 64-input ADC board connected via two Z-DOK connectors (front). Two of the four CX-4 interfaces are utilized to send frequency subband data to the correlator 'X-engine' and spatial imager 'S-engine' boards.

blocks[4]. Design specific DSP blocks and hardware interfaces have also been created. Design models and control software are available from the project repository[5].

Instrument design specifications are presented in Table 3.2. The digitizer samples a 20 MHz band at 12 bits which covers the 16 MHz analogue band. Each antenna stream is digitized to $\sim$ 20 kHz frequency resolution with a 4-tap PFB which is sufficient to localize narrowband RFI. A full correlation matrix is formed using the 4-bit quantized signals in parallel with the spatial FFT imager.

3.3.1 Digitization and Channelization

Signal digitization is performed using the Texas Instruments ADS5272 8 channel, 12-bit ADC. The ADC board, developed by Rick Raffanti[6], uses eight ADC chips to channelize 64 streams at up to 65 Msps connected via the two Z-DOK+ connectors, seen in Figure 3.5. In the BEST-2 design, 32 signal streams are digitized at 40 Msps which covers the 16 MHz analogue band of the BEST-2 system. The ADC is clocked with a 160 MHz clock which is locked to a local maser source. During the analogue stage the RF signal, centered at 408 MHz, is mixed down to baseband. Prior to digitization the last amplifier stage of the analogue chain has per signal adjustable gain useful for setting levels for optimum ADC quantization. This ADC is connected via a dual Z-DOK+ interface to an 'F-Engine' ROACH which performs the frequency channelization. A block diagram of the design layout is shown in Figure 3.6.

The ROACH board is clocked at four times the sample rate such that four signals are time division multiplexed onto a single stream. Frequency channelization is performed with a four-tap Hann filter, 2048-point PFB to produce 1024 channels per real antenna stream. The standard CASPER PFB has been specifically modified to account for the

---

[4] https://casper.berkeley.edu/

[5] https://github.com/griffinfoster/medicina

[6] https://casper.berkeley.edu/wiki/64ADCx64-12

Digital Backend Specifications		
Digitizer/Channelizer (F-Engine)		Notes
ADC Sampling Rate	40	Msps
ADC Sampling Precision	12	bit
Antenna-polarizations	32	single pol
PFB-FIR	4 tap	Hann window
PFB-FFT	2048 point	radix-2 biplex real FFT
Freq. Channel Width	19.53	kHz
Equalization Coeff.	512	32 bit, complex
Quantization	4	bit
FX Correlator (X-Engine)		Notes
Auto Correlations	32	
Cross Correlations	496	
Minimum Integration Length	6.55	ms
Output Bitwidth	2 × 32	bit, complex
Output	10 GbE	SPEAD protocol
Spatial FFT Imager (S-Engine)		Notes
2D FFT	8 x 16	
Beams	128	
Minimum Integration Length	1	s
Imager Output	1 GbE	SPEAD protocol
Imager Bitwidth	2 × 32	bit, complex
Beamformer Output	10 GbE	up to 8 beams

Table 3.2: BEST-2 digital backend, a three ROACH design where the correlator and spatial FFT imager use the same digitizer/channelizer frontend.

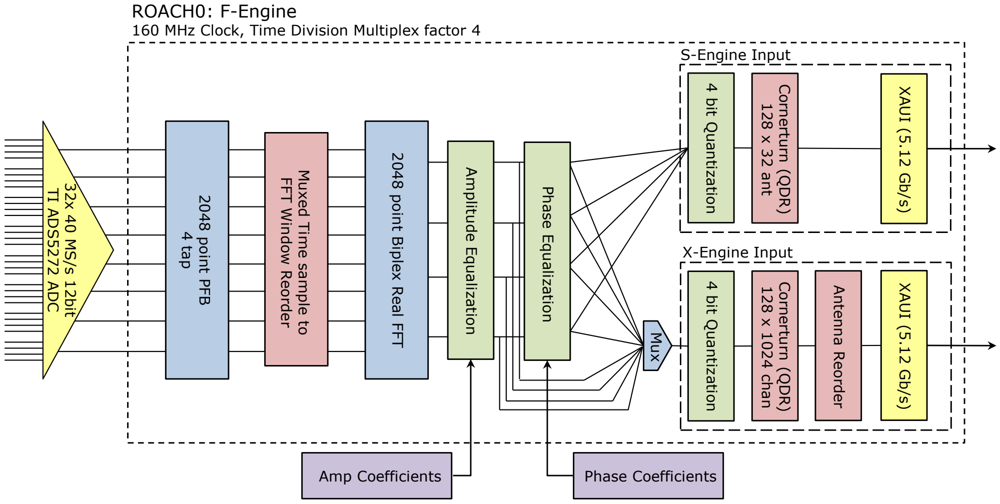


Figure 3.6: Block diagram of the ROACH F-Engine. During observations amplitude and phase coefficients are applied to scale the power for the 4 bit correlation and apply phase corrections.

signal multiplexing. The PFB is a series of FIR filters which require a continuous signal stream. Each signal stream is made up of four antenna signals, by increasing the delay line by four and holding on each PFB coefficient for four clocks the windowing function can be properly applied to the time division multiplexed stream. This introduces an additional cost of four times the BRAM usage required for signal delays but the PFB coefficients are reused and requires no additional storage cost.

After applying the windowing function of the PFB the signal streams are reordered into FFT windows of length 2048, an antenna is then repeated every four FFT windows. Each channel has a width of 19.5 kHz and the output of the FFT stage is 18-bit complex. This bit growth from 12-bit real to 18-bit complex is due to scaling from the FFT twiddle factors. Xilinx's Virtex series of FPGAs use contain a set of 18-by-18 (Virtex 2) or 18-by-25 (Virtex 5 and 6) bit multipliers, there is no additional cost of increasing the bit width to 18. The narrow channel widths and PFB windowing are required for good frequency separation in the high RFI environment at the observatory.

After channelization the samples are quantized down to 4-bit complex. An adjustable, per channel complex gain equalizer is used for amplitude and phase corrections before quantization. A selectable multiplexer is available to skip the phase coefficients on the FX correlator data stream.

Post equalization, the data stream is duplicated and split for specific reordering for the correlator and imager, Figure 3.7. The correlator data stream is blocked into 128 samples of a single antenna, and single frequency channel. This requires the use of a large memory block for performing the corner turn operation, this has been implemented using the on board QDR RAM.

After data reordering, the stream is sent over high speed XAUI at a rate of 5.12 Gbps to the 'X-Engine' board. XAUI operates at 156.25 MHz and streams up to 64 bits each clock which is equivalent to a full 10 Gbps. Additionally on each clock 8 bits of Out of Band (OOB) data can be sent to be used to signal frame type, headers and End of Frame (EOF) signals.

There are sufficient resources to include additional features in the 'F-Engine', including finer channelization, an increase in antennas or an increase in bandwidth should a future expansion of the design be warranted. This is discussed in Section 3.5.

3.3.2 Pipelined Correlation

An asynchronous architecture is used between the F-Engine and X-Engine boards. The X-Engine board has been clocked at 200 MHz, well above the 160 MHz F-engine board clock, this assures the X-Engine board will never have input buffer overflows during the

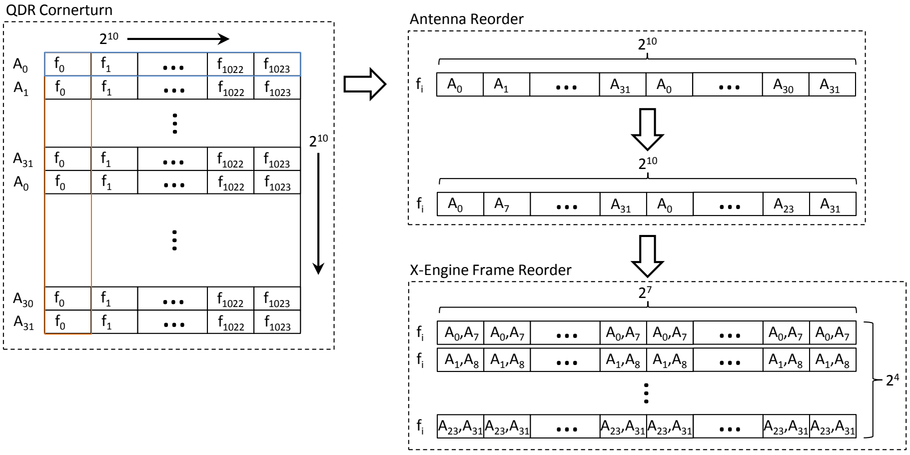


Figure 3.7: The X-Engine cornerturn operation converts independent antenna streams into frequency blocks of 128 time samples from each of the 32 antennas. Output order is 128 times samples for frequency subband j from antenna i followed by antenna $i + 1$ up to antenna 31, then the $j + 1$ subband is output up to subband 1023. The QDR input is shown in blue, output in orange.

X-Engine windowing stage. The XAUI interboard connection is a streaming interface which guarantees the same output order as input order but with variable latency. A buffer is required to capture a complete frame of the XAUI stream, which is made up of 32 64-bit data words and a 64-bit header. In rare cases the XAUI interface can drop a 64-bit word during the streaming, this requires an initial error correction stage to track the number of words received within a XAUI frame. In case of missing words the entire payload is dropped and counters reset for the next header.

A correlation is performed on a per channel basis. The channelized band can be split up into portions and processed in parallel across multiple X-Engines. This allows a larger bandwidth to be processed at the cost of increased logic and multiplier resource utilization. The design currently under utilizes the available resources on the FPGA. For this design, two pipelined X-Engines are used which each processes half of the band, Figure 3.8.

The X-Engine design requires a continuous stream of data for 128 samples of all antennas for a single frequency channel. Prior to the X-Engine, samples are buffered up into windows to guarantee valid data during a valid window of the X-Engine. During the X-Engine stage an initial accumulation of 128 sample is performed after the complex multiply to reduce the output rate to roughly the input rate. This limits the minimum integration time to 6.55 ms.

A vector accumulator using the on board QDR RAM is used for longer integration lengths. This second accumulator is software controlled with integration lengths ranging from milliseconds to minutes. A completed integration is sent to a receive

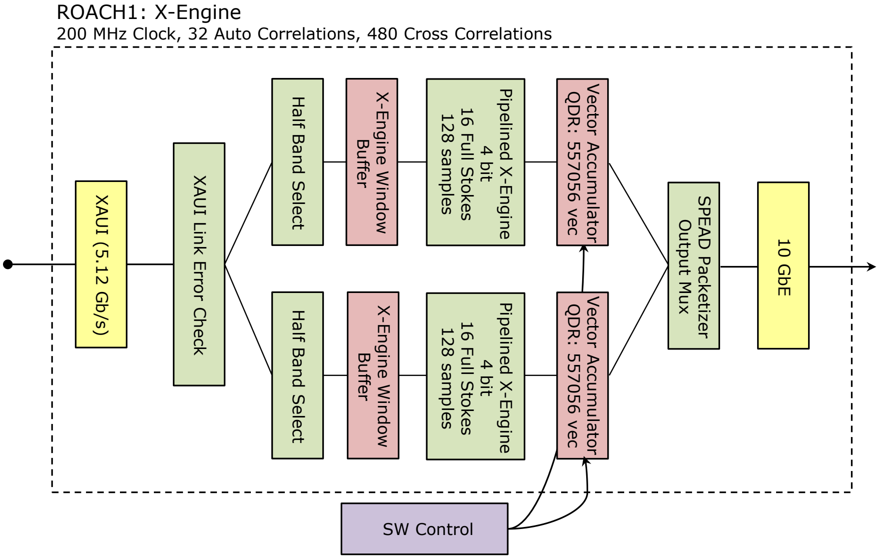


Figure 3.8: ROACH X-engine block diagram, two parallel pipelined x-engines are used, each processes half of the band.

computer over a 10 GbE connection. Integrations are packetized using the Streaming Protocol for Exchange of Astronomical Data (SPEAD) protocol[7] and transmitted over User Datagram Protocol (UDP).

3.4 Results from Deployment Observations

Instrumentation was installed and tested during March 2012, during which time various bright radio sources were observed. Since the Northern Cross is a transiting array there is a limited period of time each day in which a source is in the primary beam. Bright sources such as Cygnus A, Cassiopeia A and Taurus A along with a number of 3C sources were observed along with multiple constant declination 24 hour cycles.

Raw data from the correlator is recorded to Hierarchical Data Format, Version 5 (HDF5) files using a SPEAD protocol receive script. A suite of python scripts have been written to interface and manipulate the data in this pre-calibration stage. A python FITS-IDI package has been written to convert HDF5 files into the standard Flexible Image Transport System (FITS) format which can be read by AIPS and CASA[8]. This allows for conversion to the Measurement Set format which standard interferometry and imaging packages can interface with.

---

[7] https://github.com/ska-sa/PySPEAD

[8] https://github.com/telegraphic/pyfitsidi

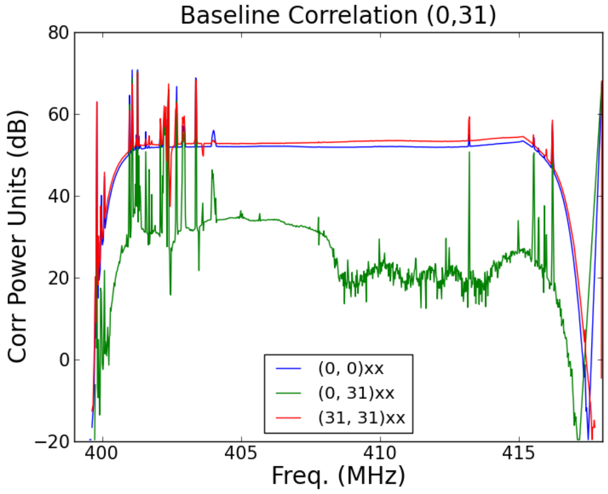


Figure 3.9: Auto (red, blue) and cross (green) correlations between antennas 0 and 31. The cross-correlation has been accumulated for 10 minutes. The high correlation below 408 MHz, and to a lesser extent above 415 MHz, is a known effect introduced in the analogue signal chain. Between $400 - 405$ MHz strong RFI from local communication towers dominate the band.

3.4.1 System Verification

Verification of the digital system began by checking the auto and cross correlation bandpass, Figure 3.9. The auto-correlations produce the characteristic bandpass of the BEST-2 analogue system which has a steep high-pass filter at 400 MHz and low-pass filter at 416 MHz. Between $400 - 405$ MHz the band is dominated by narrowband RFI. Visible in the cross-correlations is crosstalk effects introduced in the analogue paths. The usable portion of the band in the cross-correlation is 30 dB down from the auto-correlation.

The source Cygnus A was observed over a 100 minute window during its meridian transit. Bright fringe patterns were detected which follow a smooth linear cycle, Figure 3.10. East-west baselines show fringe patterns at a rate in ratio to the baseline east-west projected distance. North-south baselines only show a slow fringe pattern as the source transits across the sky. The large dynamic range of bright sources like Cygnus A require proper quantization to fully utilize the available 4 bits to produce a linear amplitude and phase response.

Strong detections of the Class A radio sources radio sources and a number of 3C sources were made during the initial deployment. These sources are used later in the calibration process to derive the complex gain corrections when imaging weaker sources.

3.4.2 Data Visualization and Editing

A suite of python-based software, *poxy*[9], has been developed for instrument control and monitoring, data capture, and editing. Correlation, along with metadata, are written to a custom HDF5 structure during observations. Post-observation additional metadata

---

[9] https://github.com/griffinfoster/medicina

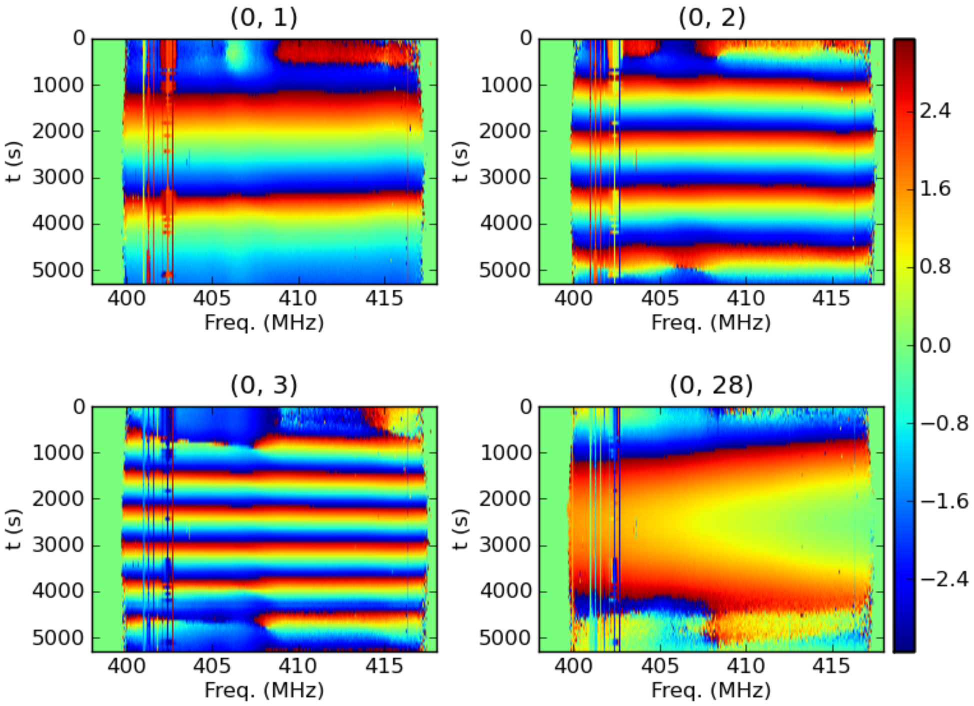


Figure 3.10: Phase plots of four baselines during a two hour transit of Cygnus A. Color scale is in radians. The $(0, 1)$, $(0, 2)$, and $(0, 3)$ antenna pairs are east-west baselines. These baselines are integer spacing $[1, 2, 3]$ of the $(0, 1)$ baseline, thus the fringe rates are in integer ratios. Baseline $(0, 28)$ is north-south aligned, the fringe rate varies slowly due to projection effects.

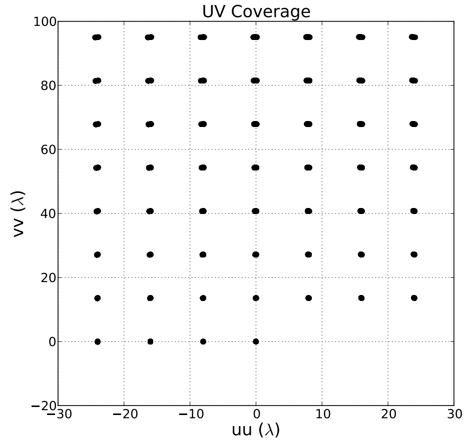


Figure 3.11: uv coverage during a 'snapshot' pointing, $\lambda = 73$ cm. Only one side of the coverage is shown, mirror sampling is also used in imaging due to the Hermitian properties of the correlation matrix. The highly redundant array produces the grid-like response in the uv plane which leads to strong grating lobes in the image plane.

is incorporated into the HDF5 file to include antenna positions and observation notes. No source information is written during an observation, only the pointing declination and timestamp.

During observations no phase tracking is applied to the antenna signals. The integration time is sufficiently short $O(1s)$ that this does not affect the phase. This allows the array phase center to be tracked to any position in post processing. Before conversion to measurement sets a phase correction is added to each integration. The uv coverage is essentially the same for any observation, with some variation due to pointing declination, Figure 3.11.

With corrected HDF5 data sets a number of scripts have been written to manipulate the correlation matrices. The afore mentioned phase tracking, time and frequency dimension manipulation, and a general purpose plotting script are all used extensively. Wrappers for converting the HDF5 files to FITS-IDI and Measurement Set formats have been developed. Once low level data editing has been performed in HDF5 the data is converted to one of these formats for use in common radio calibration and imaging packages.

3.4.3 Imaging with a Regularly Gridded Array

The east-west Full Width Half Max (FWHM) beam size of an individual element is $\approx 11.25°$ which translates to a 45 minute 'observation time' for a source, as seen in the measured primary beam, Figure 3.12. For the bright class A sources; Cassiopeia A, Cygnus A, Taurus A, and Virgo A, this time can be extended since they remain the dominating source well after crossing the FWHM. Observations of 80-90 minutes are possible which gives an small improvement in uv coverage at the cost of properly accounting for the amplitude modulation due to the primary beam. This warrants a small sensitivity benefit for a large calibration cost. Most images are created using data

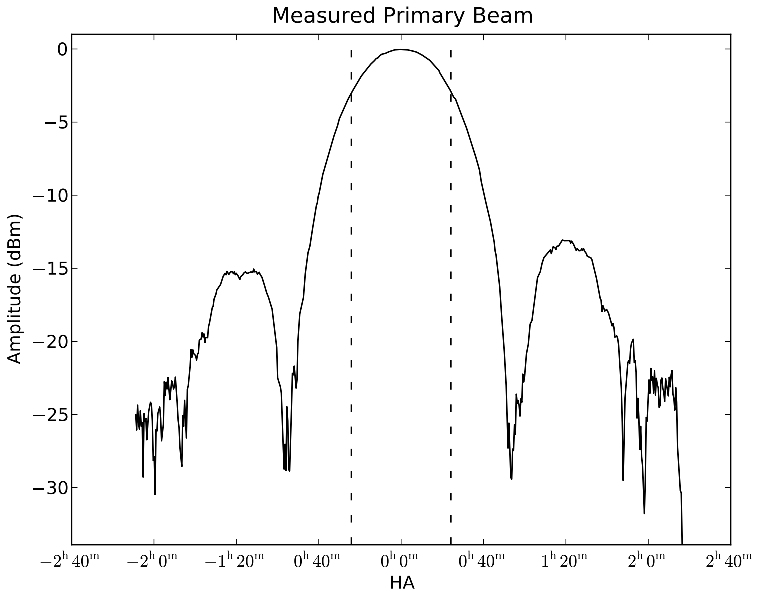


Figure 3.12: A measured east-west primary beam based on the correlation between antennas 1N-6-1 and 1N-1-3, which are representative of typical antennas, based on a Cassiopeia A transit. The dashed lines indicate the FWHM points. The antennas have a Sinc response to a point source. The first sidelobes are -15 dB down from the peak.

within a few minutes of the source transit time. The high sidelobes in the primary beam, Figure 3.12, cause the bright class A sources to dominate even when they are far from the field of view and makes it difficult to perform calibration and imaging near these sources.

A transiting array provides a unique challenge of gain calibration since a source's apparent gain will change as it transits the primary beam. To account for this variation, a two stage gain calibration method is used. Phase calibration and setting the flux scale is accomplished by using Cassiopeia A and Cygnus A observations as point source sky models set to their 3C flux levels [Bennett, 1962] and known spectral indices [Baars et al., 1977]. Since these sources are very bright, only a few seconds at transit are needed produce a high SNR dataset to use for calibration. Over this period the primary beam can be approximated as flat. A time independent complex gain is derived for an initial calibration. After applying the gain corrections an observation will be set to a flux scale relative to the flux of the calibration source. Each individual source is then self calibrated in MeqTrees [Noordam & Smirnov, 2010] based on a local sky model taken from the 3C catalog. This stage is calibrated on short time intervals to account for amplitude changes from the primary beam.

An effect of the density of the antenna layout is a low PSF to field of view ratio (spatial fidelity), thus images tend to contain at most a few spatially separated point

Source	3C Flux (Jy)	Peak Flux (Jy)	FX Noise Level (Jy)
3C10	134	67.2	1.1
3C48	47	27.3	0.9
3C123	175	91.1	2.1
3C144	1420	729.5	2.0
3C157	210	116.5	1.6
3C196	59	34.6	0.9

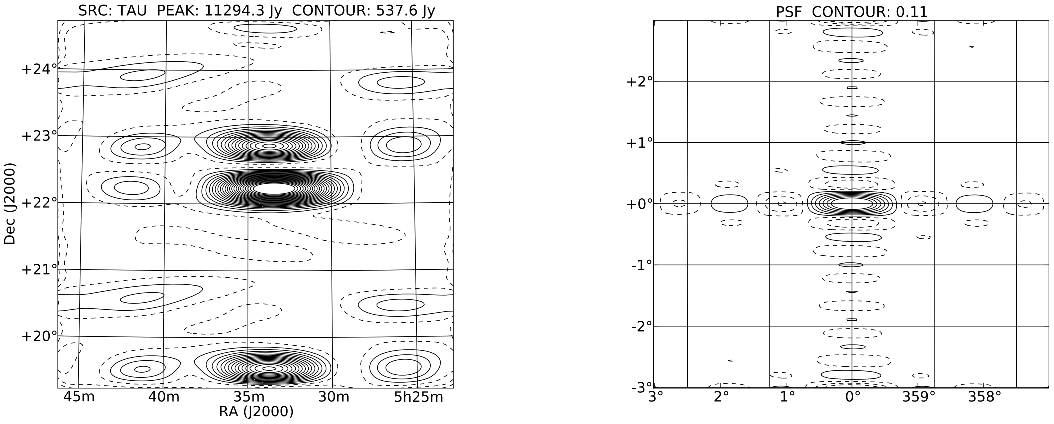
Table 3.3: Noise level for various 3C sources calculated for images formed with data from the FX correlator.

sources. The grid layout of the BEST-2 array produces strong sidelobes and grating lobes in the PSF as seen in Figure 3.13b. Observations of bright point sources produces calibrated dirty images in which the PSF is visible, Figure 3.13c. After cleaning the measured residual noise is around 2 Jy over a number of two minute snapshot observations, Table 3.3. Imaging using longer source transit times has shown the noise floor to decrease at the expected rate proportional to $\frac{1}{\sqrt{\tau}}$. The various stages of imaging and calibration for the source Taurus A is shown in Figure 3.13. Before phase calibration there appears to be a point source with an asymmetric sidelobe, this is characteristic of any uncalibrated field using BEST-2. Using a point source model the calibration fits a bright point source in the center of the field with the north-south and east-west PSF first sidelobes appearing. Performing a small number of clean iterations (~ 100) produces a single point source image.

Figure 3.15 shows a number of 3C fields which have been imaged with the array. These images show distinct point sources at the phase center of the field, 3C10, 3C48, and 3C157 also show weaker sources in the field. Observation of Cassiopeia A shows an equally bright source, slightly resolved. This second source is attributed to the Sun. The source has the same right ascension of the Sun on this day but the declination differs by over $50°$. The wide difference in declination suggests that the Sun is directly illuminating the receiver dipoles or appearing in a strong sidelobe.

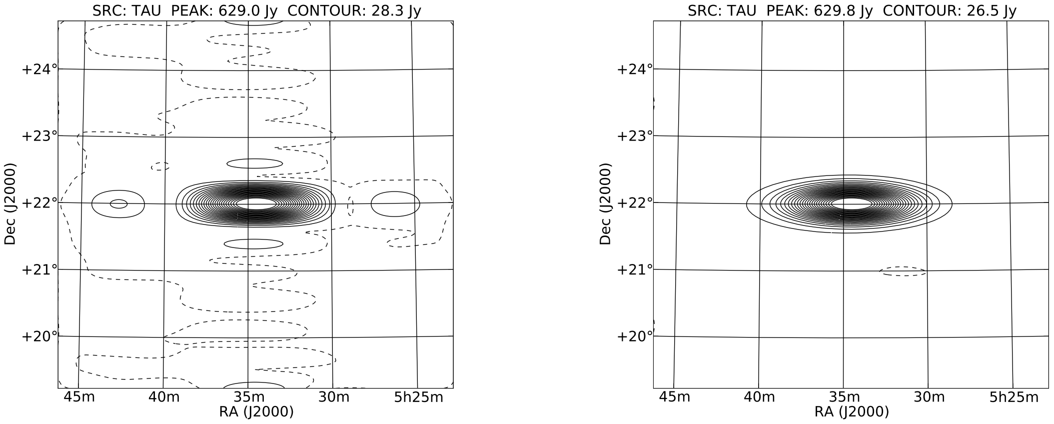
3.4.3.1 Weighting for a Regularly Gridded Array

As noted, the regular gridded array produces many highly redundant baselines, Figure 3.13b. When forming the dirty image the choice of weighting greatly affects the outcome of the image. In the highly redundant array case of BEST-2 the affects of using uniform versus natural weighting can be seen in Figure 3.14. Uniform weighting significantly reduces the sensitivity of the image for a small improvement in spatial resolution. Due to the high redundancy of the baselines and the strong PSF sidelobes a natural weighting scheme has been used throughout to produce images.



(a) Uncalibrated image of Taurus A.

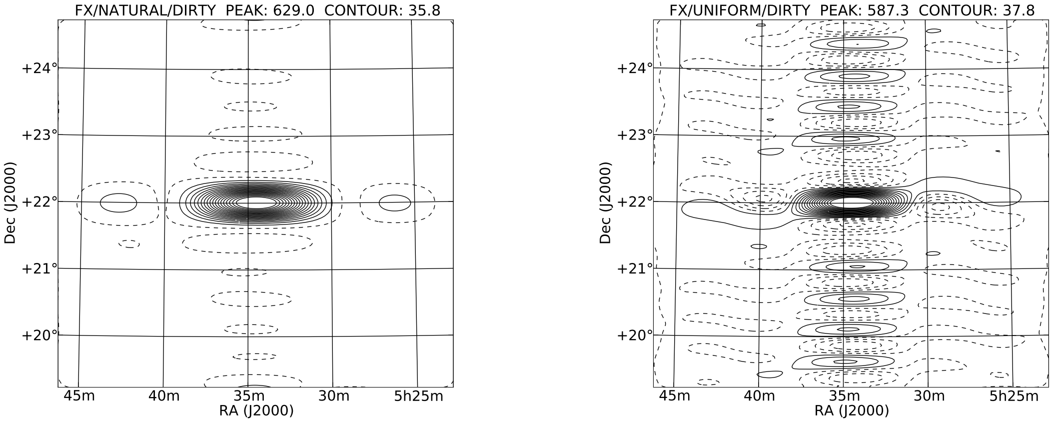
(b) Point spread function. The high, regular sidelobes are an effect of the regularly gridded array.



(c) Dirty image formed using natural weighting, after applying complex gain solutions. The structure from the PSF is clearly visible. The dynamic range of this image is ~ 150.

(d) Cleaned image of the field. Taurus A is the dominating point source with a peak of 730 Jy, the image has a dynamic range of ~ 350.

Figure 3.13: Images and PSF formed from an FX correlator two minute snapshot observation of Taurus A.



(a) Dirty image with natural weighting of the point source Taurus A.

(b) Dirty image with uniform weighting of the point source Taurus A.

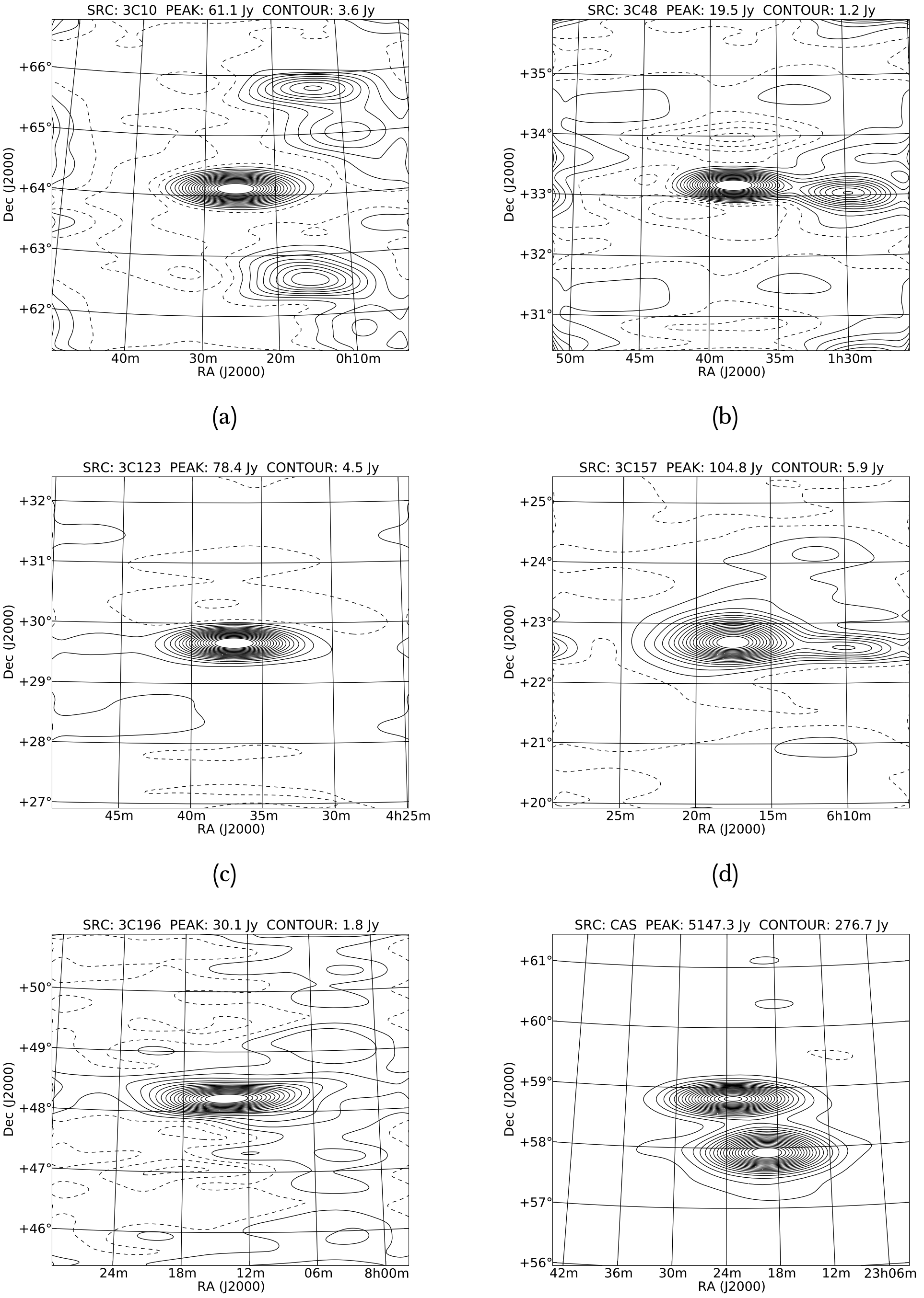
Figure 3.14: Natural versus uniform weighting of a point source.

3.5 Further Work

Deployment and observations with the new digital backend for BEST-2 has been shown to be successful. The digital hardware has improved on the previous generation of hardware and added the functionality of the spatial FFT imager. A pipeline has been developed to record observations to disk and produce calibrated images using the MeqTrees environment.

Further work can be done to turn the BEST-2 digital backend into an automated system. A real time system to derive calibration gains and generate images would greatly add to the capabilities of the system. Images could be formed on sub-minute timescales to image strips of declination through out the day. Gains have been shown to be stable over many hours, deriving new gains on this timescale should be a relatively easy process to implement. With a real time calibration process the system could be used as a pulsar and transient monitor.

The design and deployment of the new Medicina FX correlator on CASPER hardware leads directly in development of the hardware correlator for the LOFAR-UK station at Chilbolton Observatory, see Chapter 4. Many firmware components and design decisions developed for the Medicina system are reused with limited additional development. This work is continued on in the next chapter.



(a) (b)

(c) (d)

(e) (f)

Figure 3.15: Cleaned images formed from observation of various sources with the FX correlator. Sources are 3C10 (fig. 3.15a), 3C48 (fig. 3.15b), 3C123 (fig. 3.15c), 3C157 (fig. 3.15d), 3C196 (fig. 3.15e), Cassiopeia A with the Sun in a sidelobe of the array (fig. 3.15f).

Chapter 4

LOFAR-UK Station Correlator

I can still recall the twinkle in Steve's eye when we first formulated this plan.

-Aris Karastergiou, 2012

LOFAR, as noted in Section 1.3.1, is a low frequency interferometric array with stations spread throughout Europe. Each station is made up of antenna elements which, during international observations, are beamformed into digital beams for correlation. Instead of beamforming the elements into a single station they can be correlated to produce widefield, low spatial resolution images of the viewable sky. This is the motivation for SEPcam, a correlator which has been developed for the international LOFAR station at Chilbolton Observatory. The firmware is an FX correlator based on a ROACH2 board with a Virtex-6 FPGA developed by CASPER. See Figure 4.16 at the end of this chapter for a block diagram description of the board. The instrument correlates a full set of Stokes parameters for up to 36 subbands ($\sim$ 7 MHz total bandwidth) using all 96 dual-polarization antenna elements in the array. And, can output complete correlation matrices on up to $O(100\text{ms})$ timescales. The correlator is interfaced with the existing station backend using a UniBoard Ring Interface (URI) signal duplication board and modified Remote Station Processing (RSP) firmware. The URI board is under development by ASTRON for use with SEPcam and Amsterdam ASTRON Radio Transients Facility and Analysis Centre (AARTFAAC) correlators [Prasad & Wijnholds, 2012]. The instrument will operate commensally during normal observations when using the LBA but will require setting the tile elements to 'all-sky' mode with the HBA. An overall diagram of the SEPcam system is presented in Figure 4.1, each section of this diagram will be explored through the chapter. Once completed the SEPcam system could be readily duplicated for use on other international LOFAR stations.

Development of SEPcam has both engineering and scientific aims. Just as the BEST-2 correlator, discussed in Chapter 3, is a prototype for SEPcam, so is SEPcam a

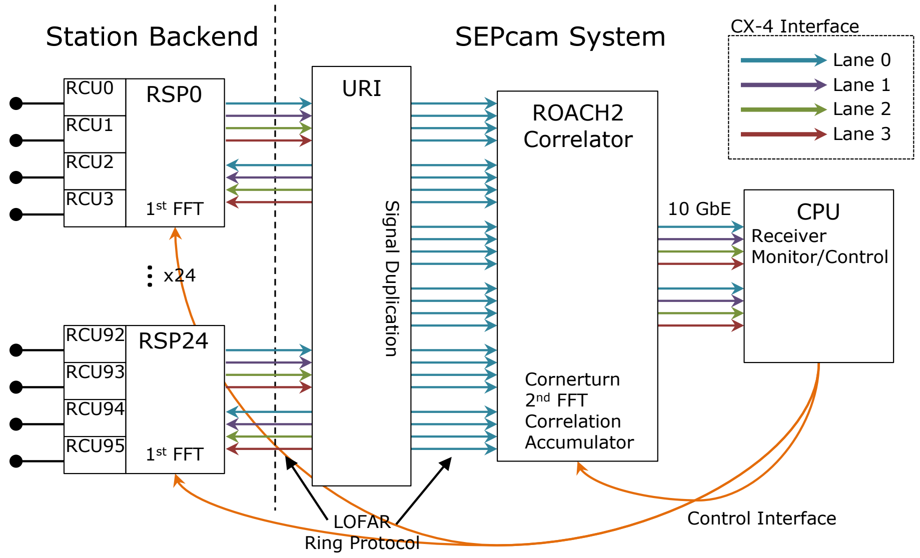


Figure 4.1: Block diagram of the SEPcam correlator system. The ROACH2-based x-engine is connected to the existing digital backend over a single CX-4 lane from each RSP board via the URI duplication board. Correlation matrices are dumped over 10 Gigabit Ethernet to a receive computer for real time imaging or archiving. Between the RSP boards and the ROACH2 input the custom LOFAR ring protocol is used to transfer data.

prototype for larger correlators. As will be noted in Section 4.2, SEPcam is modular in design allowing for scalability in bandwidth with minimal or no updates to the firmware. The large-N correlation of the station elements is performed on a single FPGA, this is made possible by optimized firmware design, an efficient implementation of the x-engine and the growth of available resources on the FPGA. Correlators for large-N arrays such as SKA and the various EoR projects can continue to be implemented on FPGAs.

A number of science case have been presented in Chapter 1 which are relevant to single station observations. Bright radio transients will be detected, and localized to within a few degrees (depending on observing subband). This is the main science justification for the purposed AARTFAAC correlator [Prasad & Wijnholds, 2012]. For LOFAR stations the transient detection threshold will be $\sim 1 - 10$ kJy depending on duration and frequency. Large scale synchrotron structure is a strong foreground which will need to be properly modelled for EoR projects. Chapter 5 presents early results from observations using the station calibration correlator. The international stations in independent operation mode, but in unison, will be a power array for observing rapid changes in the ionosphere And finally, there is an element of ‘if you build it, they will come’. New experiments come about with an extension of the current instruments. Old instruments that are repurposed for experiments they were never built for. Hopefully, everything we build will lead to a better understanding of our scientific goals, whether intended or not.

Briefly noted was the station calibration correlator. This is an already existing $N \times N$ correlator on the RSP boards which generates a complete correlation matrix for one of the 512 subbands per second. This is used to derive calibration solutions for station beamforming. It has proven an invaluable tool in developing the initial software for calibration and generating sky maps. See Chapter 5 for an in depth discussion.

There are multiple advantages the SEPcam system has over the current calibration correlator: bandwidth, time resolution, and flexibility. Only a single subband can be observed using the calibration correlator. SEPcam will be able to process up to 36 independent subbands, selectable during observation. With future expansion, SEPcam can include a second stage FFT for further channelization. The integration length for observations are software controllable ranging from 10's of milliseconds to 100's of seconds. The modular design and reprogrammability of SEPcam will allow for future expansion of bandwidth, channelization and adding additional modes as they are necessary.

At the stage of this writing the completed system is still in development. The ROACH2 firmware has been designed and simulated. Though, the high speed I/O interface firmware between the correlator and the station backend is still under development. The URI board and the station firmware is currently being developed at ASTRON. With further work SEPcam could make an excellent addition to the science capabilities of all LOFAR stations.

4.1 LOFAR Station Digital Backend

Before delving into the correlator design let us briefly discuss the station digital backend and the purposed interface for SEPcam. The main function of the station digital backend is to form *beamlets* which are sent to the LOFAR Central Processor (CEP) for correlation. A beamlet is a single subband with a set of beamforming coefficients applied to each antenna element. Antenna signals are digitized on the Receiver Unit (RCU) boards, Figure 4.3a. Subband channelization is performed on the RSP boards [Poiesz et al., 2007]. This is the F component of the correlator design. Future updates to the SEPcam firmware will allow a second stage FFT to further channelize the subbands in an FFX architecture. A ring network is used to connect the RSP boards to form beamlets which are then sent to the CEP [Lubberhuizen & Kooistra, 2007]. Figure 4.2 shows the general layout of the station backend.

Two data products are transmitted on the RSP ring network; beamlets and crosslets. Beamlets are the final product sent to be correlated at the CEP. Crosslets are cross

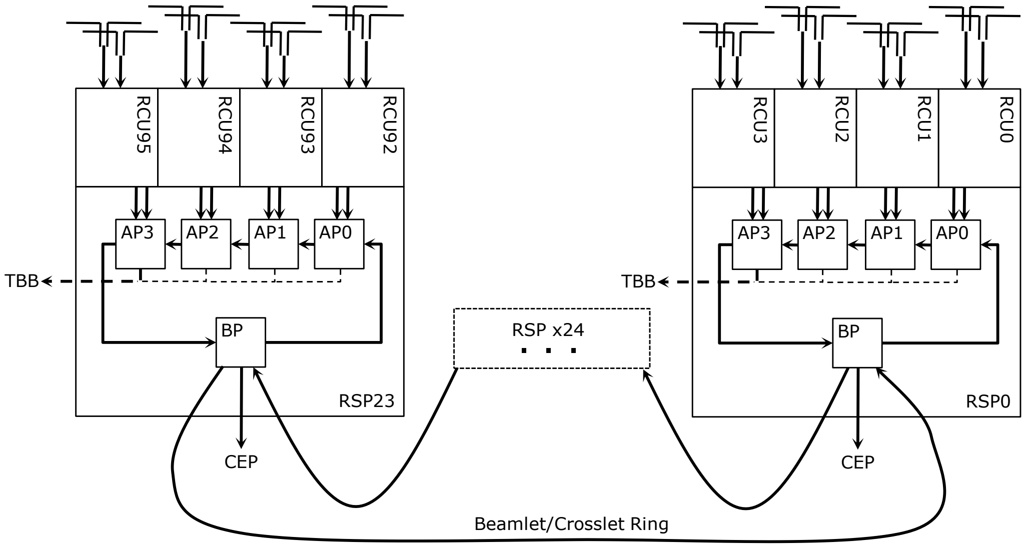
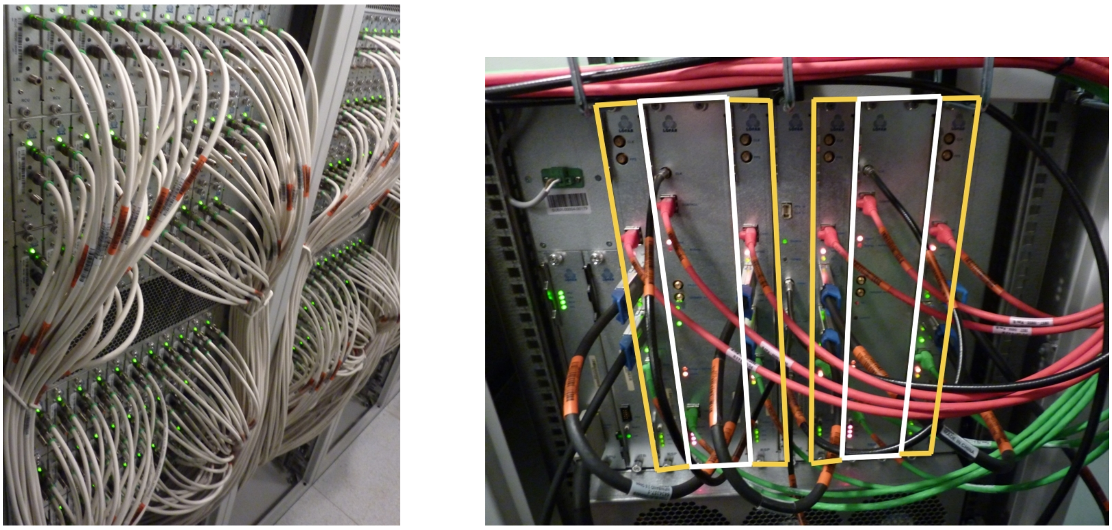


Figure 4.2: Digital backend of a international LOFAR station. The signal from each antenna elements (dual polarization) is digitized by the RCU unit. Four RCU units are connected to a RSP board which contains four Antenna Processor (AP) FPGAs and a Beam Processor (BP) FPGA. The APs channelize the signals into 512 subbands. Each AP is connected, via the backplane, to Transient Buffer Boards (TBBs) which can be used to capture sample snapshots. The BP forms beamlets out of the subbands from the AP boards and the other beamlets travelling through the ring network. Cross correlations (crosslets) for calibration are also formed on the BP. Completed beamlets are sent to the central processor (CEP) over the backplane and a high speed network connection.



(a) RCU racks digitize antenna element signals. 192 coaxial cables are connected to 96 RCU boards.

(b) A subrack containing four RSP boards (yellow boxes, connected to each other with black XAUI cables), two Transient Buffer Boards (white boxes), and a clock board.

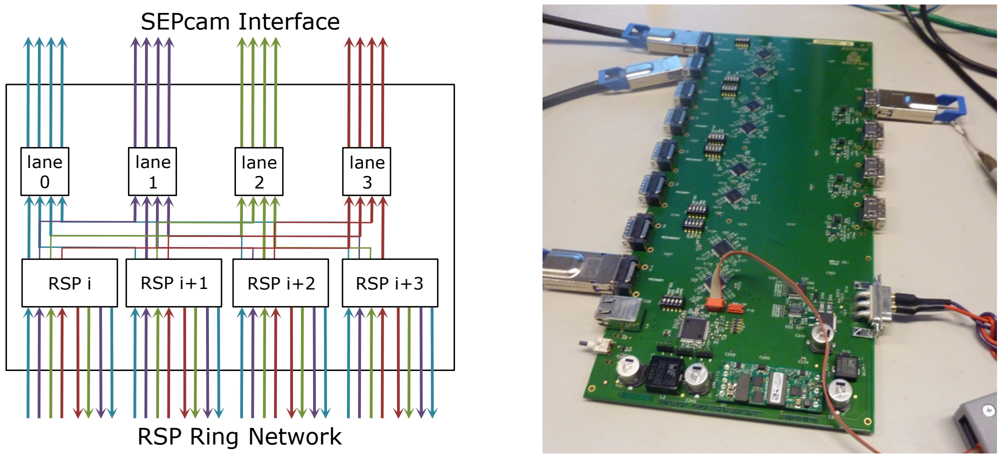
Figure 4.3: Components of the LOFAR station digital backend.

correlations of individual antenna elements. A distributed $N \times N$ correlator is implemented on the RSP boards for deriving complex gain corrections.

The ring network connecting the RSP boards use a 8/10b Serialize/Deserialize (SerDes) encoding over CX-4 cables, Figure 4.3b. Each cable has four lanes, which in SerDes mode, operate independently at up to 2.0 Gbps (125.0 MHz clock). Beamlet data takes up 4.65 Gbps and crosslets use 2.7 Gbps plus overhead. This load is distributed across all four lanes evenly, 1.84 Gbps is used on each lane. But with a firmware modification, three of the lanes can be used at full capacity freeing up the last lane to be used for SEPcam. The modification to the digital backend for SEPcam requires new hardware and firmware to along the station to function seamlessly with this new instrument. This portion of the project is being undertaken by ASTRON.

4.1.1 URI Interface to SEPCAM

Connecting SEPcam to the current backend requires additional hardware to select out signals from the ring network. This is done with a custom board developed by ASTRON called a URI board which duplicates a lane from each input CX-4, and outputs four of the extracted lanes to a single CX-4 interface, Figure 4.4. 24 input CX-4 cables from the RSP boards results in 6 CX-4 interfaces for each of the four lanes. Three sets of these interfaces will contain beamlet and crosslet data, and so can be ignored. Though, in future expansions, the RSP firmware could be written to allow for dedicated station observations which would allow all four sets to be used for



(a) URI board diagram. Each URI duplicates four RSP signals to go back into the ring network. The four lanes from each CX-4 is rewired to output the same lane set on each of the four outputs. SEPcam interfaces with one of these four outputs in commensal observation mode.

(b) URI board during testing. Four links on the left are duplicated and transmitted back out the remaining four links on the left. The links are interconnected for mixed output on the right.

Figure 4.4: URI board for interfacing SEPcam to the LOFAR station RSP ring network. The completed SEPcam system will use six URI boards.

correlation. SEPcam could then be replicated to increase the bandwidth by a factor of four. The current development path is to use a single set of lanes to be connected to six of the eight CX-4 interfaces on the ROACH2.

The ROACH2 board includes two breakout interfaces to the Virtex-6 GTX transceivers [XILINX, 2011b]. A breakout board has been produced which connects four CX-4 cables to 16 GTX pairs, enabling eight CX-4 interfaces which have bandwidths of up to 10 Gbps using 8/10b encoding, Figure 4.5. Typically these interfaces are used for XAUI or 10 Gb Ethernet communication. Both of these communication layers use the four lanes in conjunction. Since the ring network lanes are independent a custom interface is used to process each SerDes lane independently.

Communication between RSP boards uses a 125 MHz clock, by increasing the clock to 156.25 MHz (standard 10 GbE clock) the available bandwidth could be improved by 20%. This crystal change would not require any other modifications to the current backend.

4.1.2 RSP Firmware Modifications

The input bandwidth to SEPcam is limited by the available network bandwidth in the ring. Nominally there is over 2.0 Gbps free, but the total bandwidth is quantized by the lane bandwidth (2.0 Gbps). A total of 48 Gbps of bandwidth is input into SEPcam.

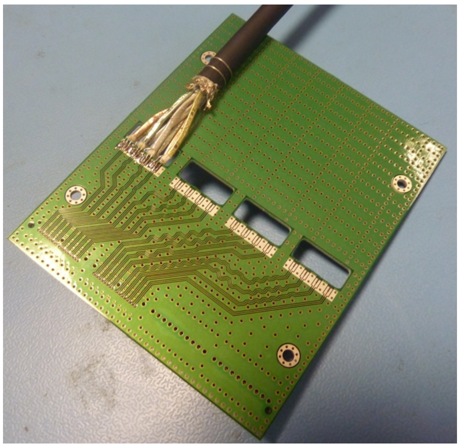


Figure 4.5: CX-4 × 4 breakout board for the ROACH2 with one of the interfaces populated with a cable. Two breakout boards are connected to the ROACH2 via the daughter card interfaces. Six CX-4 interfaces are used for unidirectional data streams from the RSP boards. The remaining two interfaces output the correlation products.

This allows up to 36 subbands. There are 96 dual-polarization antenna elements represented by two(real and imaginary) 16-bit numbers. Subbands are 195 kHz wide. This resulting number of subbands is $40 = 48$ Gbps/ $(96 \times 2 \times 2 \times 16\text{b} \times 195\text{kHz})$. But, the network overhead reduces the number down to 36.

The current RSP firmware is not setup for this mode. ASTRON will be undertaking the modifications required. First, the lane data must be re-balanced so that beamlet and crosslet data only use three of the lanes. A frame generator will be added before the beamlets are formed to select out subbands from each of the eight elements connected to the RSP. A software configurable interface will be added to select the 36 subbands from the 512 available. The subbands will not be required to be contiguous and can be changed without reloading the firmware. In addition to the first round of firmware modifications there is a possibility of increasing the number of subbands available for correlation by reducing signals from 16 bits to 8 or 4. Also, the URI board makes all lanes available outside the ring network. This would allow a dedicated mode which uses all the lanes for SEPcam, increasing the bandwidth by a factor of four.

4.2 Design and Implementation of DSP Firmware

The SEPcam FPGA firmware is made up of previously implemented DSP, new logic, and community developed blocks from the CASPER DSP libraries. Various components of the BEST-2 correlator, e.g. the quantizer, First In/First Out (FIFO) buffers, QDR vector accumulator, and 10 Gb Ethernet packetizer, have been reused or are the basis for the new logic developed for SEPcam. This is possible due to the generalized design of the logic blocks and similarities between the ROACH and ROACH2 boards. Only the hardware specific 'yellow blocks' used in the CASPER toolflow changed between boards. Specifically, for SEPcam, the BORPH software interface, QDR RAM, and high speed transceivers were needed. The BORPH and QDR interfaces are generic to the

CASPER collaboration and have been developed during the ROACH2 initial board bring up. The high speed transceivers interface use a application specific protocol and has been developed specifically for the interface between SEPcam and the URI boards, discussed further in this section. The DSP logic developed for ROACH, though Xilinx specific, works with little additional effort on ROACH2.

A block diagram of the firmware is presented in Figure 4.6 and 4.7. The design is separated into three data window frames with FIFO buffers between each frame. The SerDes frame captures blocks of data from each of the six CX-4 interfaces from the URI board. This frame assures data alignment, corrects for any word loss, and performs quantization. In the next frame a cornerturn is applied to each stream to transfer the streams from antenna streams to frequency streams. Typically this is done on the 'F-engine' or in the network switch, but the specific architecture of the LOFAR backend requires this be done at the 'X-engine'. The last frame, the X-Engine frame, performs the correlation and is essentially the same as the BEST-2 X-Engine frame. Data is buffered into windows for correlation and first stage accumulation. A second stage accumulation, in QDR, is runtime configurable and used to reduce the output data rate. Finally, correlations are output over 10 Gb Ethernet to the post-processing CPU.

The current SEPcam firmware fits well within the resource limitations of the Virtex 6. A system clock of 210 MHz has been used, which is more then sufficient to process the data streams without overflow issues. At 210 MHz the design can process up to 8.75 MHz of bandwidth. Pushing beyond this clock rate is unnecessary for the current specification. But, with a thorough place-and-route process this rate could be increased by a significant amount if more processing bandwidth was desired.

FPGA logic utilization is shown in Table 4.1. Currently the FPGA is under utilized in logic, the limiting factor is the input data bandwidth. The design implements four 'x-engines' which require 784 DSP48e logic cells for multiplication, the remaining 96 multipliers are used in the equalization stage of the quantizer. The majority of the BRAM is used in the FIFO buffers and cornerturn reorders. Since the cornerturn scales as $O(N^2)$ and the 'x-engine' is linear, future expansion in input bandwidth would likely be BRAM limited. Register and LUTs take up different resource in logic slices, many of the slices are uses for both logic types.

The default Xilinx ISE place-and-route routine takes a shotgun approach to resource placement, Figure 4.8. For the SEPcam design, which is not a full design and clocked at a reasonable rate, this approach is sufficient to meeting timing. As as design requires more resources or is clocked at higher rates timing issues arise which require manual placement of logic.

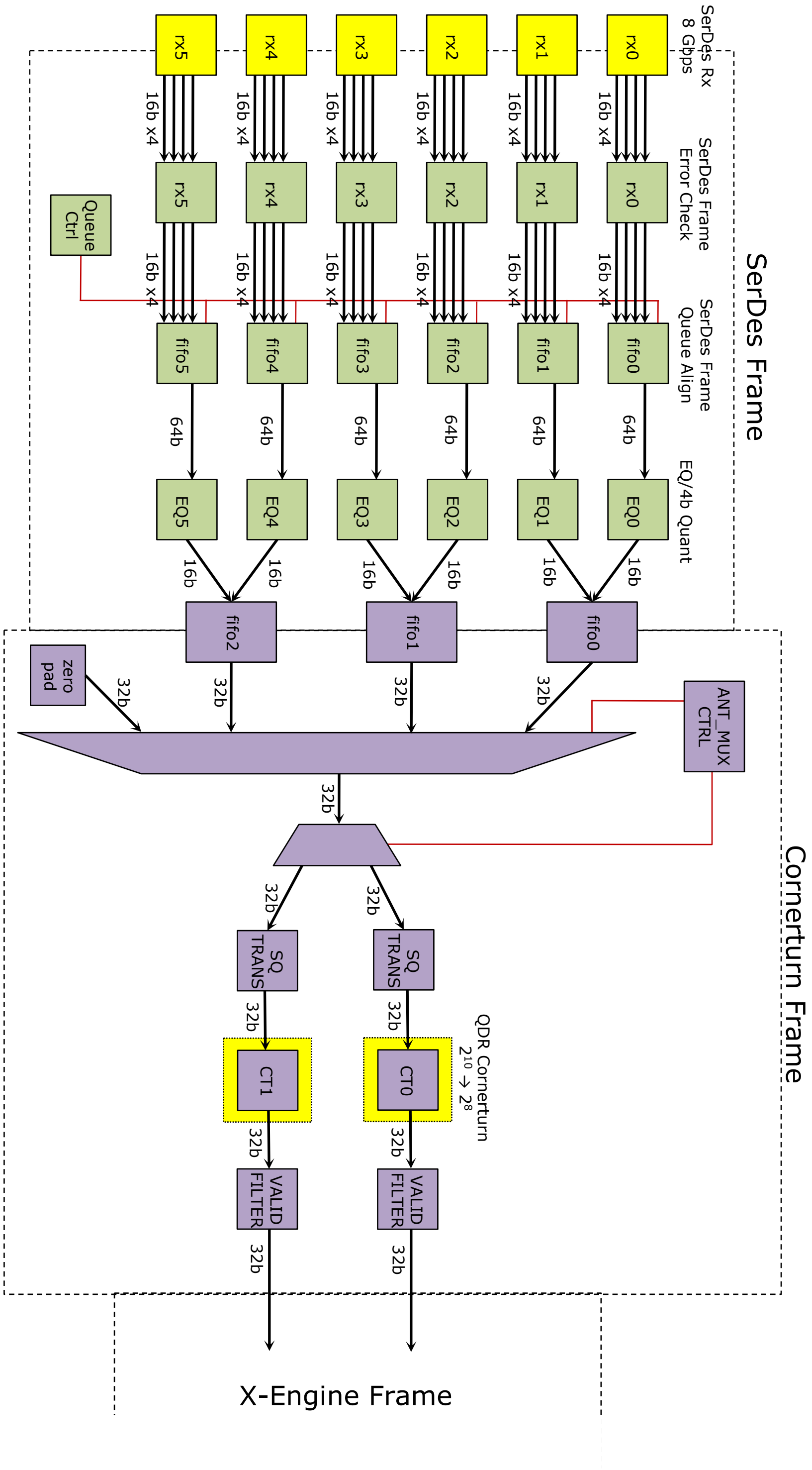


Figure 4.6: Firmware diagram of the SEPcam correlator. The design is separated into three data domains. A SerDes frame for receiving and aligning streaming data from the URI interface. A cornerturn frame for reordering signal streams from antenna data to frequency data for use in the pipelined correlation. The last frame, the X-Engine frame is shown in Figure 4.7. Yellow blocks represent off chip interfaces such as the SerDes interface, QDR memory and 10 Gb Ethernet. Black lines are data streams with their associated bitwidth, red lines are logic control interfaces.

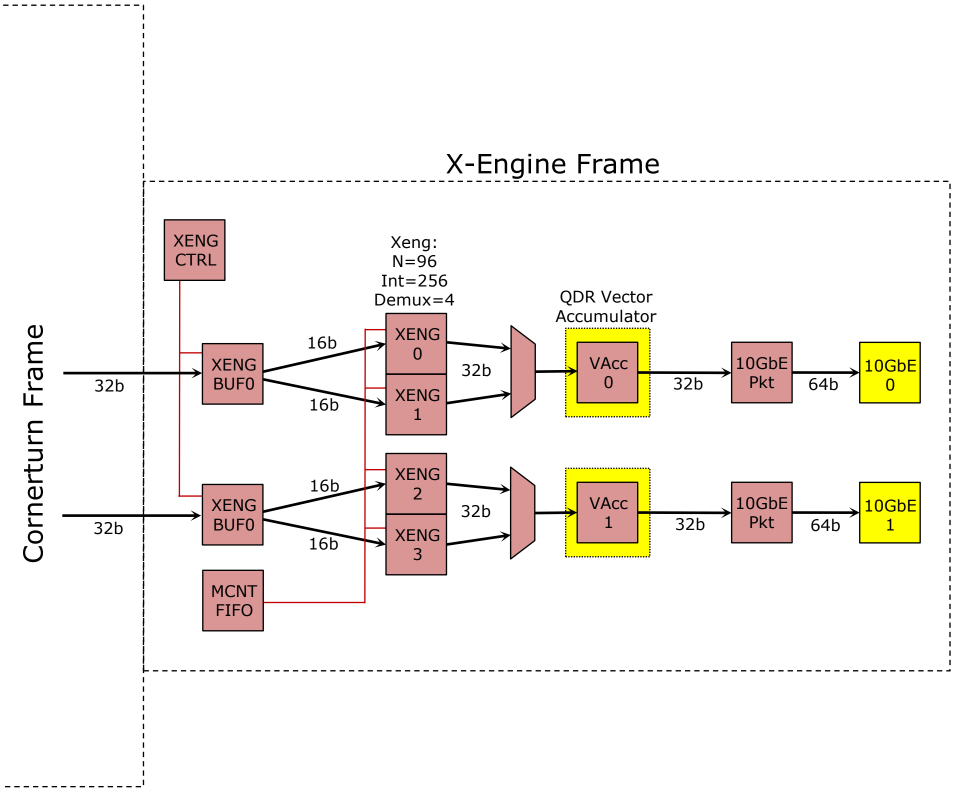


Figure 4.7: Last data domain of the firmware diagram, continued from Figure 4.6. The X-Engine frame performs the complex multiplication, accumulation, and packetizes data for transmit over 10 Gb Ethernet. Yellow blocks represent off chip interfaces such as the SerDes interface, QDR memory and 10 Gb Ethernet. Black lines are data streams with their associated bitwidth, red lines are logic control interfaces.

Resource Utilization (Virtex 6 SX475T)		
System Clock	210 MHz	
Slice Registers	89722 / 297600	15%
Look Up Tables	55820 / 297600	19%
BRAM (36 × 1k)	312 / 1064	29%
DSP48e (Multipliers)	880 / 2016	44%
CX-4 Interface	8 / 8	6:Lane-mode SerDes, 2:10 Gb Ethernet
QDR Memory	4 / 4	2:Cornerturn, 2:Accumulator

Table 4.1: DSP implementation for current SEPcam firmware.

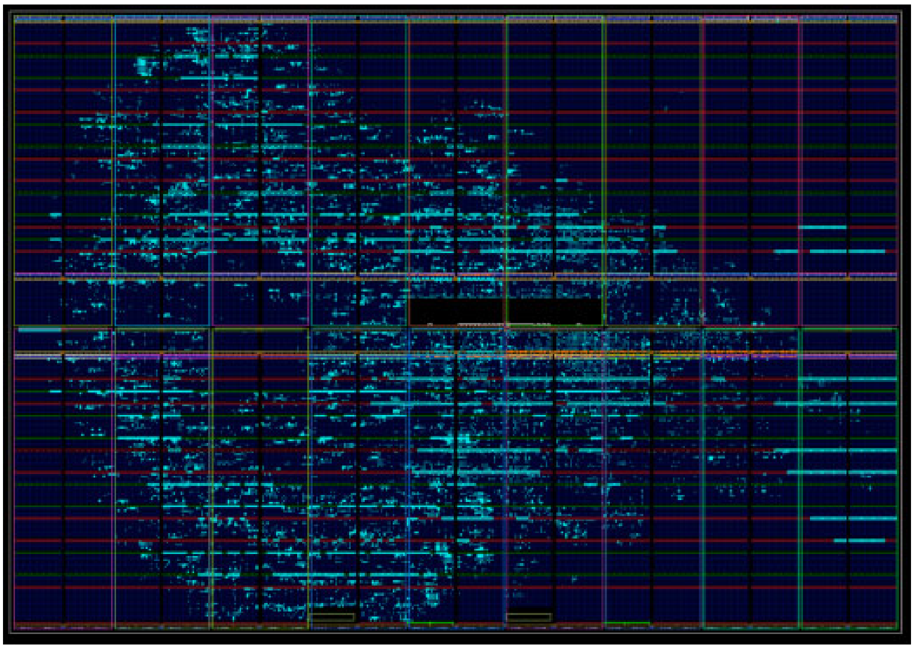


Figure 4.8: Resource placement of SEPcam firmware onto a Virtex 6 SX475T. Cyan regions are resource being used in the design. The DSP48e (red horizontal lines) and BRAM (green horizontal lines) resources are lined across the chip. Between these resources are the logic slices made up of four registers and eight flip-flops. The edges are the LVDS interfaces to external hardware. The current firmware under-utilizes the Virtex 6 FPGA. Future expansion of features in the firmware is possible. The under-utilization doe not require custom place-and-route techniques.

4.2.1 High Speed I/O Interfaces

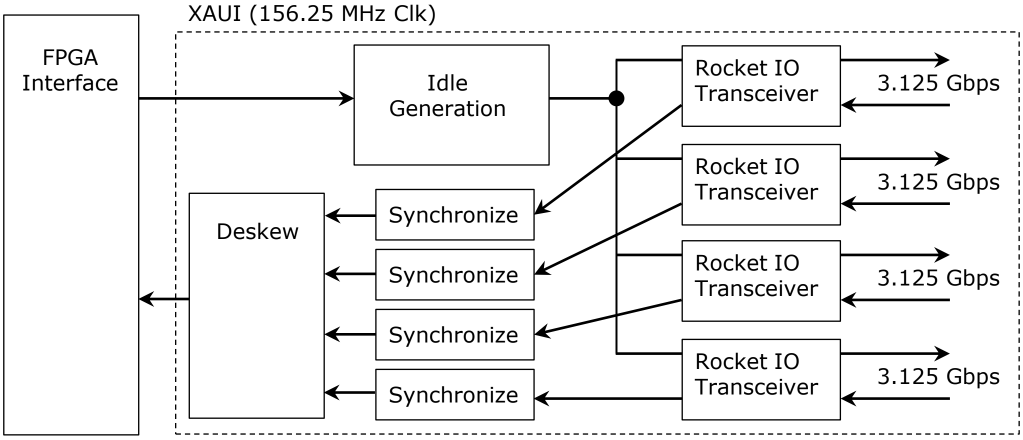
Of the eight CX-4 ports, six are used as a uni-directional interface to the URI board. The RSP boards use an external Physical Layer (PHY) chip [PMC-Sierra, 2004] to transmit and receive data in a 'lane-mode' SerDes protocol. Each lane (there are four lanes per connector) are treated as independent streams. This is not a standard protocol used in XAUI communication. In XAUI communication the four lanes are aligned as a single stream. To interface with the RSP data streams a custom interface must be written.

4.2.1.1 Lane-mode SerDes

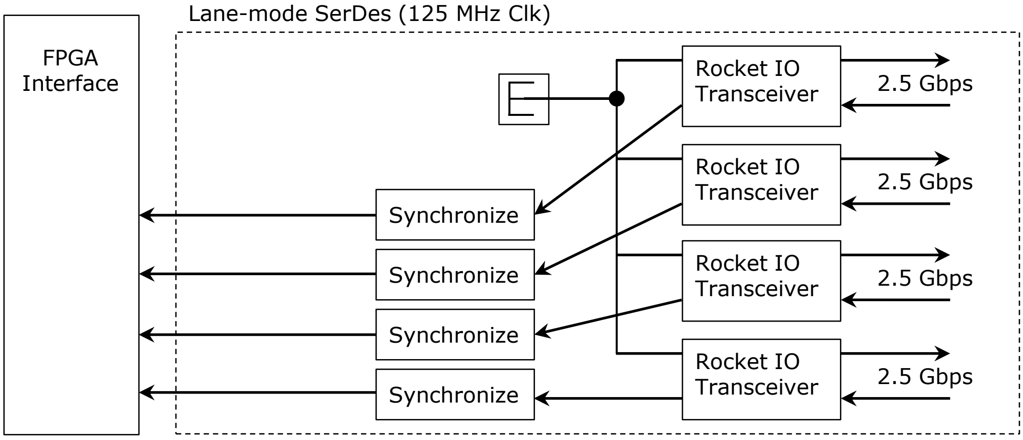
Xilinx provides a low-level framework called Aurora[XILINX, 2010] to allow custom protocols to run on the GTX high speed transceivers. Using Aurora the GTX transceivers can be setup to receive 'lane-mode' SerDes signals. The Virtex 6 SX475T has 36 GTX transceivers, of which 32 are used to make eight XAUI interface. The last four transceivers are left unconnected. The XAUI layer channel bonds together the four lanes, Figure 4.9a. The lanes are synchronized and aligned before being sent to the FPGA interface. Data is presented as 64-bit on the positive edge of a 156.25 MHz clock, along with 8 bits of OOB data. Since there is a clock interface between the FPGA and XAUI clocks an invalid data generator is used for output during idle clocks. In lane mode there is no channel bonding or alignment, Figure 4.9b. But clock synchronization is still required when crossing clock domains. To simplify the interface further, output over the transceivers can be hard coded to constants since the URI interface is uni-directional. The idle clock generator logic is dropped. Lane data synchronization will be required, but that is part of the SerDes frame logic, discussed below (Section 4.2.3).

The GTX transceiver is designed to be as flexible as possible by using a set of selectable parameters and multiplexer logic. For a uni-directional interface, a simplex RX Aurora module interfaced to the GTX transceivers is implemented. Table 4.2 lists the configuration parameters for the core used in the design. This bare bones core only receives, expects frame independent/streaming data, and uses no flow control(uni-directional interface). A core is generated for each of the 24 lanes.

In the current design the Aurora core has been integrated into the firmware, but has not been fully tested. Testing is ongoing to assure data streams from the RSP boards is correct. This will require the development of a TX/RX firmware design which transmits the expected data streams over one GTX transceiver and receives over another transceiver. The actual format of this signal is still under development. ASTRON will provide a data specification documentation when they begin work on the



(a) Block diagram of the XAUI layer. Four lanes are channel bonded together to be presented as a single data stream Each transceiver runs at a maximum of 3.125 Gbps, for 12 Gbps total. But, a 8/10 bit encoding is used for signal integrity which reduces the usable bandwidth to 10 Gbps.



(b) Lane-mode SerDes block diagram. No bonding and alignment logic is used, each lane is presented as an independent data stream. Synchronization is still necessary to align each stream to the FPGA clock edge. Since the URI interface is uni-directional, the GTX transceiver outputs will be tied to constants to reduce logic utilization. A 125 MHz clock is used which reduces the total usable bandwidth to 8 Gbps.

Figure 4.9: Reduced data interface for the uni-directional, lane-mode SerDes compared to standard XAUI. Diagram adapted from Xilinx documentation.

Aurora 8B10B Core Configuration (v7.1)	
Lanes (GTX)	1
Lane Width	2 Bytes
Line Rate	2.5 Gsps
Reference Clock	125.0 MHz
Dataflow Mode	RX-only Simplex
Interface	Streaming
Flow Control	None

Table 4.2: Aurora core parameters used to generate the lane-mode SerDes interface.

RSP firmware modification. Until then, a format which would logically be expected will be used. Though this may require future modification, the changes should be minor. Once this has been tested and we receive a URI board, the component will be wrapped into the loopback test design. For simulating the SEPcam firmware a TVG is used for each lane. A software controllable multiplexer is used to select between the lane-mode SerDes interfaces and TVGs.

4.2.1.2 10 Gigabit Ethernet

Post accumulation, the correlation matrix is output over UDP on standard 10 Gb Ethernet. The 10 Gb Ethernet logic is a layer wrapped around the lower level XAUI layer. Both are standard logic blocks in the CASPER library. A complete correlation matrix is 5.11 MB (4656 baselines, 36 subbands, 4 Stokes parameters, 32-bit for real and imaginary components). Neglecting the packet overhead, which is small compared to the payload, the shortest integration time possible over the 20 Gbps line (2 × 10 Gb links) is 2 ms. This is slightly longer then the absolute shortest integration length, 1.31 ms, which occurs when there is no second stage accumulation in the QDR. Thus, at least two correlations must be accumulated in QDR; making the minimum integration length 2.62 ms. This is an incredibly short integration time for correlation, and would difficult to process at present. Though, we plan for the future. UDP is a lightweight protocol used to maximize streaming data rates over packet integrity. SPEAD, a CASPER developed application layer protocol[Manley et al., 2010], is used for the output correlation matrices. The SPEAD transmission logic is intertwined with the QDR vector accumulator and will be discussed further in Section 4.2.6.

4.2.2 Domain Windowing

As shown in Figure 4.6 and 4.7, the SEPcam firmware is separated into three domain windows: SerDes, cornerturn, and x-engine. At each domain edge, buffers are used to transfer data from one domain to the next, similar to how clock domains work. The first portion of the SerDes domain can be thought of as a buffer between the RSP board and ROACH2. In this domain the 24 lanes are aligned such that each lane has the same subband output from the queue on a given clock.

After equalization and quantization, eight lanes are concatenated together and a four way time-division multiplex is used to transfer 32 lanes (24 real, 8 zero lines) into a single stream containing eight lanes. This orders the data streams into a convenient format to perform the cornerturn operation. This time-division multiplex by four is possible because each SerDes lane will have valid data on less than 25% of the clock cycles. If the input bandwidth is expanded by a factor of two in the future a duplicate window can be setup to process the other half of the subbands.

In the X-Engine domain the data window is 96 antennas $\times$ 256 integrations. Once a full window of data is available in the X-Engine input FIFOs the window is streamed out at the beginning of the next X-Engine sync pulse. An X-Engine window occurs every 96×256 clocks, if that window is valid the output of the pipelined x-engine block is added to the QDR vector accumulator. A switch mode vector accumulator which uses twice the memory required to store an entire correlation matrix. When a accumulation is finished and ready to be read out over 10 Gb Ethernet the write address is switched to the other half of memory space. That way an accumulation can be read out while the next accumulation is being written to memory.

4.2.3 SerDes Domain

Since the 'lane-mode' Aurora core is a bare bones implementation of the GTX interface, the alignment and error checking logic is incorporated into the SerDes domain. After alignment and error checking, equalization and quantization down to four bits is performed before transitioning to the cornerturn domain. The input data to each SerDes frame is in the form of a subband frame, Figure 4.10. Each subband frame contains 17 16-bit words, with 2 bits of OOB data. The header information determines the subband and spectrum ID which is used in the alignment and error checking logic. The four OOB values are 0: payload word, 1:header word, 2: EOF + payload word, and 3: header + sync. It should again be noted that the exact input frame format has not been defined, but will logically be of this form. Minor modifications maybe required when connecting to the completed system.

In the first stage of logic, each SerDes lane is processed through the error checker. This verifies that a complete subband frame is received. A counter keeps track of the number of words between the headers and that EOF signals are received. If a frame is missing words a `DISCARD` message is sent to the queue so that frame is known to be bad and the proper amount of memory should be allocated in the queue but should be considered invalid. Also, an error counter is incremented for monitoring of the system. In a stable system a frame should rarely be dropped. If many errors start arising this could be due to a bad cable connection, the signal integrity dropping off with cable length, or invalid data from the RSP board.

After error checking the queue aligns all 24 SerDes lanes to output the same subband data on the same clock. This is implemented with a dual-port RAM which can read and write to a block of memory on a single clock. The memory is 2048×18 bits, this comes from 36 subbands $\times$ 16 words $\times\ 2 = 1152$ and $\lceil \log_2(1152) \rceil = 11$ bits. Two

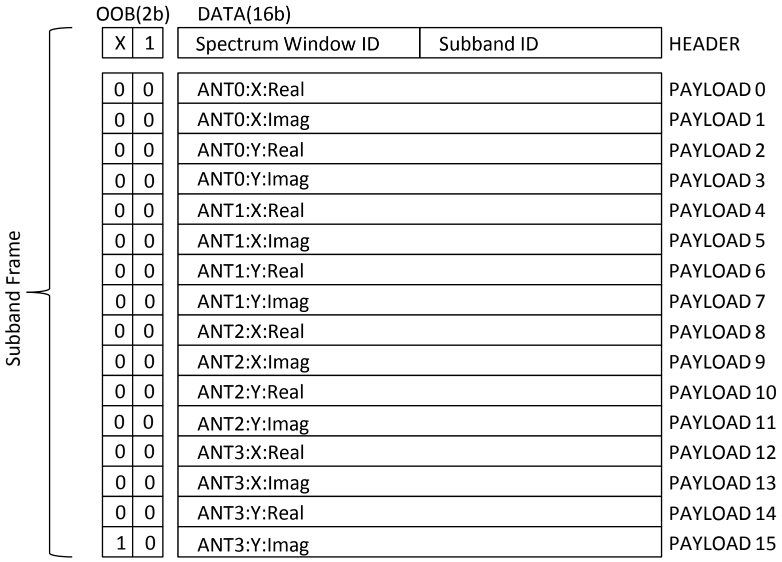


Figure 4.10: Format of the subband data frame for a single SerDes lane. Each frame represents a single subband. Each RSP board outputs data for four antennas, both polarization. A frame is made up of 17 words $16 + 2$ bits in length, 16 bits of data, 2 bits of out of band signal. The first word is the header, the 8 MSB are the spectrum window ID which increments every new FFT window. The 8 LSB identify the subband. The following 16 words are the real and imaginary components of the 8 antenna polarizations for that subband. The number of clocks between each $16 + 2$ bit word is variable.

dual-port RAMs are wrapped up into a single 36-bit wide BRAM. Of the 18 bits, 16 are used for data, one as a valid data signal, and the last is unused. The factor of two in the RAM depth is used to read out a fully filled spectrum window while in the other half of the memory a spectrum window is being written. This removes the possibility of any conflict between reading and writing to the same address. The address of a data word encodes the subband ID, that information does not need to be maintained.

The depth of the dual-port RAM and complexity of the logic has been reduced with the assumption that all the SerDes lanes will deliver, within a few clocks, the same subband data for the same spectrum window. This may not be a valid assumption. The spectrum windows arriving may differ widely between SerDes lanes. If this is the case there are two solutions. First, ignore the issue. If the spectrum window ID between SerDes lanes is stable and different spectrum windows are correlated this reduces the phase coherency. But, as is the case with the LOFAR station, on short baselines this is small effect and will have a minimal effect on the data quality. The second solution would be to add coarse delay block which could be configured to delay each SerDes lanes by a few spectrum windows at the cost of increased BRAM usage. At this stage it is unknown if this issue will arise. This logic will likely require further refinement when a test system is available.

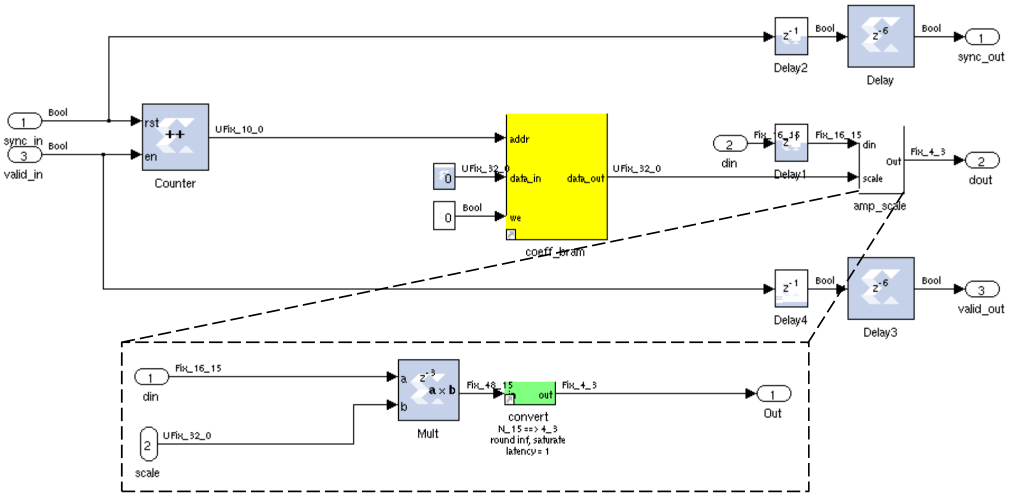


Figure 4.11: Equalization and quantization logic used to reduce 16_15 values to 4_3 for correlation. The `coeff_bram` (yellow block) is a software writeable memory space which stores 576 complex 32-bit values.

Since the dual-port RAM only reads out a filled spectrum at a time, from this point in the design the data is now blocked into a window of 1152 consecutive clocks of valid data, separated a variable length of clocks with no data. This will become useful in the next stage, quantization and equalization.

4.2.3.1 Equalization and Quantization

Quantization logic reduces the 16_15 values from the RSP boards to 4_3 values for correlation, Figure 4.11. First, the equalizer multiplies the subband data with real time configurable coefficients. Then, the signals are converted to 4_3 values. A unique 32_0 number can be uploaded for each of the 576 complex values in the spectrum window. This allows both amplitude and phase correction on a per subband, per antenna basis. This is essential for performing the quantization. Each 16_15 values is multiplied by a 32_0 value to create a 48_15 value which is then quantized down to 4_3. Since this logic has been adapted with minimal changes from the BEST-2 f-engine design, Section 3.3.1, the same software can be used to determine bit selection and write gains. A discussion of quantization and bit selection is presented in Section 2.2. Memory ‘snap’ blocks have been included before and after the quantization stage to be used in the equalization software and for visual inspection of the data. Each snap block captures a single spectrum window, it is triggered by the start of a new valid window. From here we enter the cornerturn domain which starts by time-division multiplexing the signals into a single stream.

4.2.4 Cornerturn Reorder

On the output of the SerDes domain there are six 16-bit wide streams, with frames 1152 clocks in length. For hierarchical convenience, four 4-bit streams are concatenated into a single 16-bit stream. The `ANT_MUX` logic time-division multiplexes six 16-bit streams into two 32-bit streams. One half of the subbands goes to each of the 32-bit streams. The `ANT_MUX` also adds in 'zero-pad' data into the streams which is necessary for the cornerturn operation. Since the output of the `ANT_MUX` is two 32-bit streams, the six 16-bit streams are concatenated into three streams and written to three FIFOs. The output of the `ANT_MUX` is then one of these FIFOs or 'zero-pad' data. Each FIFO has a depth of 2^{12} which can store almost four complete spectrum windows. One FIFO will always be reading out data if there is a spectrum available so no FIFO will ever need to store four full windows. More over, there are two output streams, so no one FIFO should ever need to store more than two windows. An error counter, attached to each FIFO, is readable from a software register. Though this should not pose a problem as long as the board clock (210 MHz) is above the input bandwidth (24 lanes $\times 7$ MHz $= 168$ MHz) plus the zero-padding overhead.

Zero padding the data streams is necessary to maintain a 2^N input window length for the cornerturn. A complete spectrum window is 192 ant-pols $\times$ 36 subbands. The spectrum window is split into two streams, 18 subbands each. Further, each 32-bit word represent 4 ant-pols. The input spectrum window to each cornerturn is $18 \times 48 = 864$ 32-bit words. This needs to be zero-padded to 1024 words. After each FIFO has been read out the window is zero-padded with 160 words before the next iteration. This adds an effective $\sim$ 10 MHz of input bandwidth, which still puts the input bandwidth well below the board clock. To differentiate the zero-pad data from the real data an extra bit, `ZERO_DATA`, is added to each 32-bit word.

Before the cornerturn a square transpose is used to convert four independent, serial streams into one stream where antenna data is in parallel, Figure 4.12. Every fourth clock an independent stream is repeated, but in parallel, 16-bit form. This is necessary because the real and imaginary, X and Y polarization, need to be formatted into a single 16-bit word into the cornerturn. Additional logic is added to correctly reassign the `ZERO_DATA` and `VALID` bits after the transpose.

The cornerturn operation presents data streams in the format required for the pipelined x-engine, Figure 4.13. Output of the cornerturn is in antenna windows which are M integration long windows, in the case of the x-engine implementation $M = 256$ time samples. Each word in the frame is 32 bits wide, two antennas with both polarizations. The `VALID` bit is used to only write in valid words to the QDR, the `ZERO_DATA` bit is used in the filter post-cornerturn. The filter simply sets the output

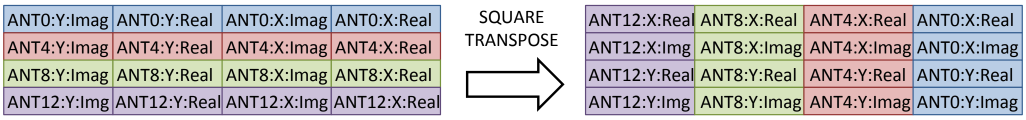


Figure 4.12: Square transpose operation reorders four inputs from serial input to parallel output. Each valid 16-bit word now represents the two polarizations of a single antenna, complex.

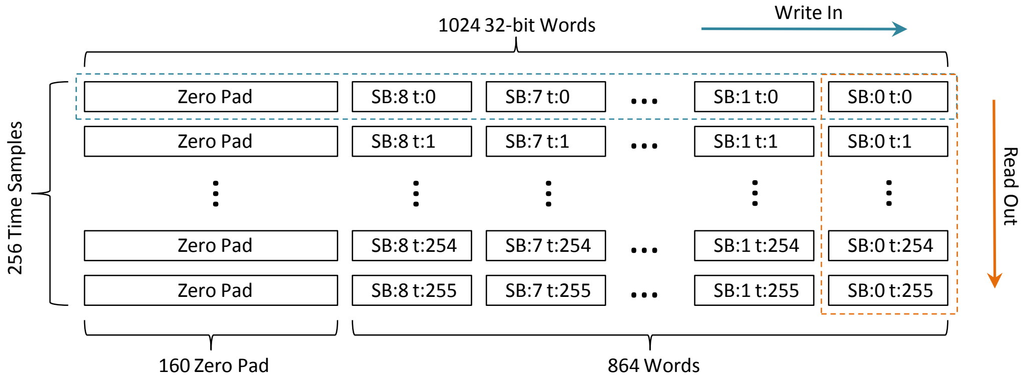


Figure 4.13: Cornerturn operation, 256 spectrum windows are written into QDR (with zero-padding to 1024 samples). 1024 antenna windows (with 256 time samples) are read out.

valid line to low if the `ZERO_DATA` bit is high. A `SYNC` is also inserted at the beginning of a new cornerturn window to be used in the x-engine domain.

4.2.5 Pipelined X-Engine

Correlation is performed with four 49-tap pipelined x-engines, each x-engine processes 9 subbands. An x-engine data window is 24576 (96 antennas × 256 time samples) 16-bit words in length. Before correlation an x-engine window needs to be created by buffering up sufficient data from the cornerturn output. A similar setup to the SerDes queue logic is used where a dual port RAM is used to write in an x-engine window length of data while reading out the previous window in the other half of the memory. This is the largest BRAM-based memory usage in the design, and requires careful layout to overcome timing issues. Because the pipelined x-engine is a streaming correlator, new data can only start to be input at the beginning of a window. Therefore the dual-port RAM read logic only outputs data when at least one half of the memory is filled and at the start of a new window.

The pipelined x-engine design is the same as the design used in the BEST-2 correlator, Section 2.1.7 and 3.3.2. Except the first stage integration is 256 time samples, and uses more cross correlation taps to perform the full correlation. Inputs to the pipelined x-engine is `SYNC` (bool), `DATA` (16-bit), `VALID` (bool), and `MCNT` (48-bit). This last input, `MCNT`, is the master counter which is used to associate a completed integration with a timestamp. On initialization of the correlator, `MCNT` $= 0$ is assigned to the current universal time. The `MCNT` is then incremented every new x-engine window. This value is written out with a completed integration.

The output of a correlation tap is 128-bit. The 4-bit multiplication is $4\_3 \times 4\_3 = 8\_6$, but can be reduced to 7_6 with no loss of information since both values are ≤ 1. Addition/subtraction results in a 8_6 number, of which 256 are accumulated to produce a 16_6 number. For a real and imaginary component, and full Stokes this results in $8 \times 16\_6 = 128$ bits.

This output would require 8 parallel vector accumulators. Each vector accumulator would integrate a vector of length 41904. That length is in an awkward region where it is too large for internal BRAM and too small for a dedicated QDR. Ideally, only one large QDR-based vector accumulator would be used. To accomplish this, the pipelined x-engine output is demux'd onto fewer lines. This is one of the primary reasons behind choosing an initial integration length of 256 sample. With this integration rate the 128-bit output can be demux'd down to 32-bits. As will be noted in the next section, this allows the output of all four pipelined x-engines to be accumulated using two QDRs.

The Virtex 6 is required for the design to accommodate the input data rate, as stated earlier, and for the resource requirements of the pipelined x-engines. The pipelined x-engine logic uses an optimized design which drastically reduces the number of DSP48e slices and amount of BRAM, [Hickish, 2013]. This has greatly improved the design flexibility by freeing up resources.

4.2.6 QDR Vector Accumulator

The second stage vector accumulator is performed using the QDR memory. This allows for large vectors ($>$ 100000 64-bit words) to be accumulated for a run-time configurable length. Each QDR module can write and read 72 bits of data per clock. Data is presented at a fix latency of 11 clocks after a read is issued. The 16-bit words output from the pipelined x-engine are accumulated using up to 32 bits. This allows two words to be written to a single address in QDR memory. Bits $31:0$ and $67:36$ are used to store the accumulated values and bits $35:32$ and $71:68$ are unused in this design but could be used for additional accumulation or as flag bits. There are 8 16-bit words that need to be written to QDR on a valid output of the pipelined x-engine (the four x-engines are in parallel). But only four 16-bit words can be written per clock to the two available QDRs. A post-correlation mux must be used. Each pipelined x-engine was mux'd by 4 already, but could not be further mux'd in line as the output window (37248 words) would be larger then the input window (24576 words). The final mux takes into account that not every x-engine window will be valid. Less than one in three windows will be valid, this means that if the valid windows are spaced at least every second available window then a post-correlation mux will not cause any overflow issues. Over two x-engine windows (49152 clocks) 37248 32-bit words are written into each QDR.

For two pipelined x-engines the mux is performed with a FIFO which writes in the values of one of the 32-bit outputs while the other output is written to QDR. The FIFO depth is 2^{14} which is sufficient to handle the maximum queue depth. While the pipelined x-engine output is not valid, values from the FIFO are read out and written to QDR. Since both pipelined x-engines use the same `SYNC` and have the same valid windows the outputs are predictable. This assures the ordering input to the QDR will always be the same, though a post-accumulation reorder is necessary to have sensible data. But, this can be done in software at a later time.

Each QDR is used to accumulate vectors of length 670464, stored in 335232 memory address. Similar to the dual-port RAM, twice the number of memory spaces for storage are used so that one half of the address can be used to for integration while the other half are read out as completed integrations. A block diagram of the accumulator logic is presented in Figure 4.14, this diagram is some what generalized to the actual logic. The accumulator has been adopted from the BEST-2 design, which in turn was based on the CASPER QDR-based vector accumulator. Words are burst written and read from the QDR in large blocks (~ 1000 words). This is done because of the fixed latency between read calls and output, and the irregularity of data from the pipelined x-engine. The reading and writing controller is linked to a software register to set the accumulation length, and the SPEAD controller to control output rates. The main changes from the BEST-2 accumulator is the inclusion of the x-engine FIFO to mux the outputs to a single line. And, increasing the QDR width to 72-bits. This required a modification to the adder logic to correctly slice and add the two values from the QDR.

The integration length of the first stage accumulator is $\frac{256}{195\text{kHz}} \approx 1.31\text{ms}$. Due to the logic of the QDR vector accumulator the minimum accumulation length is two correlation windows, which puts the absolute minimum accumulation length to 2.62ms. As noted earlier in the chapter, this results in a large data rate output, and just at the limit of two 10 Gb Ethernet interfaces. Typically, integration lengths of ~ 10 seconds can be used for non-phase tracking, short baseline arrays.

Since there are two available 10 Gb Ethernet ports, each vector accumulator is attached to an individual port. The UDP packet adds a minimal amount of overhead to the data to encapsulate each packet payload with the necessary SPEAD header

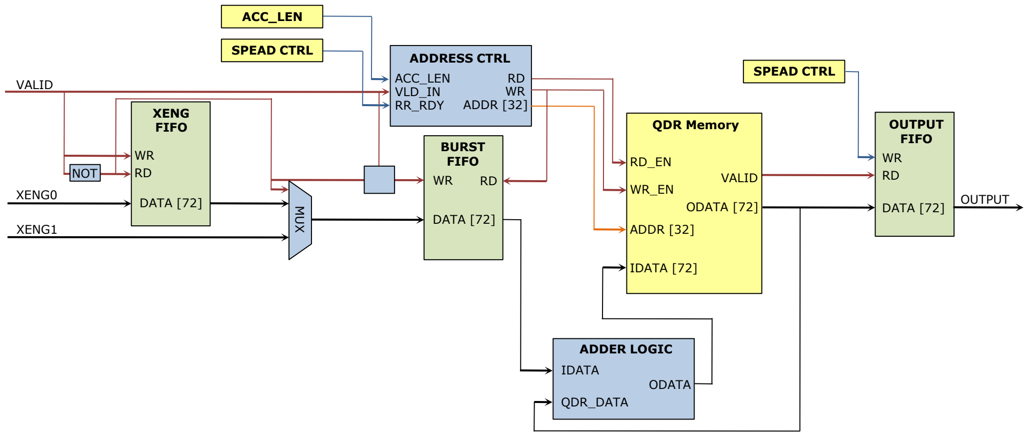


Figure 4.14: SEPcam QDR vector accumulator adapted from the CASPER QDR vector accumulator. Two pipelined x-engine outputs are demux'd onto a single line. The address controller starts burst read and write operations on the QDR. Addition and reset is done in the adder logic.

information. The header contains a SPEAD packet identifier (`SPEAD_ID`), heap ID (`HEAP`), offset within the heap (`OFFSET`), and packet payload size (`PLD_LEN`). A heap in terms of the correlator is a correlation matrix, which has a total size of 1340928 32-bit words. The maximum payload size per packet is 512 64-bit words. Thus, each interface will transmit 655 packets per heap. The position of an individual packet in the heap is set by the `OFFSET` value. All the packets will have the same payload length except the last packet which will be 128 64-bit words, which is registered in the `PLD_LEN`. To assure the 10 Gb Ethernet buffers do not overflow the SPEAD controller logic send a request data signal to the vector accumulator.

4.2.7 Software Interfaces

The software interface between the FPGA and PowerPC allow for runtime system configuration an monitoring. Software registers, 32-bit wide, act as status monitors and controllers. The subsystems described above include software register interfaces to monitor errors or overflows, reset and trigger sync lines, and set adjustable rates such as in the QDR vector accumulator. 10 Gb Ethernet interface parameters (IP address, port, MAC address) are set at run time through software registers. A sync pulse system is used to align data, link the master count to a timestamp, and reset overflows.

Shared BRAM, 36-bit by 2^n memory blocks, allow capture of continuous data either by manual or automated system triggers. This is useful to capture blocks of data at intermediate points in the firmware, such as at domain interfaces. These 'snap' blocks are essential system testing and troubleshooting. These memory spaces can also be used to write coefficients for subsystems to access such as the equalizers. Coefficients are cycled through in the equalizers based on antenna and subband.

4.2.8 Simulation and Digital Verification

When designing a large system testing must be done on a number of levels to verify the system. A large design is split up into smaller subsystems, which are in turn made up of smaller subsystems. A down-up approach to testing is taken such that the basic blocks are verified before the larger system is built. Many times the completed system can not be simulated. In the case of SEPcam this is true, and even if the entire system could be simulated it would not be a useful exercise. For example, in the cornerturn domain the data window is the size of the cornerturn, and simulating a few cornerturn windows is sufficient to verify that domain. While in the x-engine domain the pipelined x-engine can be simulated over an x-engine data window. Trying to simulate both the cornerturn and x-engine at once would lead to a massive simulation time ($\sim 2^{18} - 2^{20}$ clocks).

The individual data domains provide encapsulated regions which are the highest level of the hierarchy. Then, the interface between the domains is tested. Put together, in a compiled design the full system should work as expected. At the lowest level are the Xilinx provided core blocks and HDL code. These are expected to work as documented. The first level which needs verification is subsystems which include these basic components. Blocks from the CASPER libraries are expected to be throughly tested before hand. But, testbenches are used to verify the expected results of using these blocks. A standard test bench will made up of a TVG which inputs known vectors into a subsystem, and a verifier which compares the output of the subsystem to the expected output. The subsystem can then be treated as a black box. As long as the simulated output matches the expected output the subsystem is verified.

The down-up approach to design leads to parameterization of subsystems. The testbench must also be parameterized. But, this allows development of a hierarchical system where parameters can be propagated from the top levels down to the basic components. This is a common practice in the CASPER community.

Simulators such as ModelSim, Xilinx's iSim, and Simulink are excellent and indispensable tools for simulating complex systems over a large number of clocks. These tools use TVGs which are are verified at the subsystem level. A simple example TVG is presented in Figure 4.15. Here, two signed 4-bit multiplications are being performed in a single 18-bit multiplier. After settling the scope `msb-correct` and `lsb-correct` should be flat at 0. Which occurs for `msb-correct` but there appears to be a problem in `lsb-correct`.

Compiled TVGs are also used. These are inserted into a compiled design before an operation. A snap-block BRAM then capture the output and runs the verification in software. These are very useful to check if errors from one subsystem of the design is being propagated to another subsystem.

Simulation and testing from the basic component level all the way up to the completed system are essential to a completed system. The small issues that arise at the low levels amplify up to higher levels in odd and unintuitive ways. Frustrations from issues in a subsystem may arise from the lower level components which lurk below.

4.3 Conclusions

The SEPcam correlator firmware has been implemented on the ROACH2 board based on the known system specification with assumptions about the as yet undetermined specifications. Design of most subsystems uses reconfigurable parameters to allow minimal redesign with changes in system specifications. As the correlator is further tested and connected to the LOFAR backend this will likely be required.

4.3.1 Further Work

Further development and testing is required for the lane-mode SerDes interface. Verification of the Aurora core requires development of a loopback design to test the receiving system. Incorporating the URI board into the loopback design will allow checking of the signal integrity through the interface and time latency. Finally, connecting the interface to the RSP board will require testing with new RSP firmware.

A major challenge of the design will be integration of the correlator with the LOFAR backend. As noted, this requires significant changes to the RSP firmware which will be undertaken by ASTRON. A revision of the current firmware will likely be required once this is available.

Though much work remains to have a fully operational SEPcam system, the correlator firmware has been designed to be generic and parameterized. This will minimize the time and effort required to accommodate any changes to the system specifications.

Beyond the correlator is the development of an automated imaging and calibration system. Work towards this end is presented in Chapter 5. The imaging system will open up a new avenue of study to develop calibration techniques for low frequency, large-N arrays.

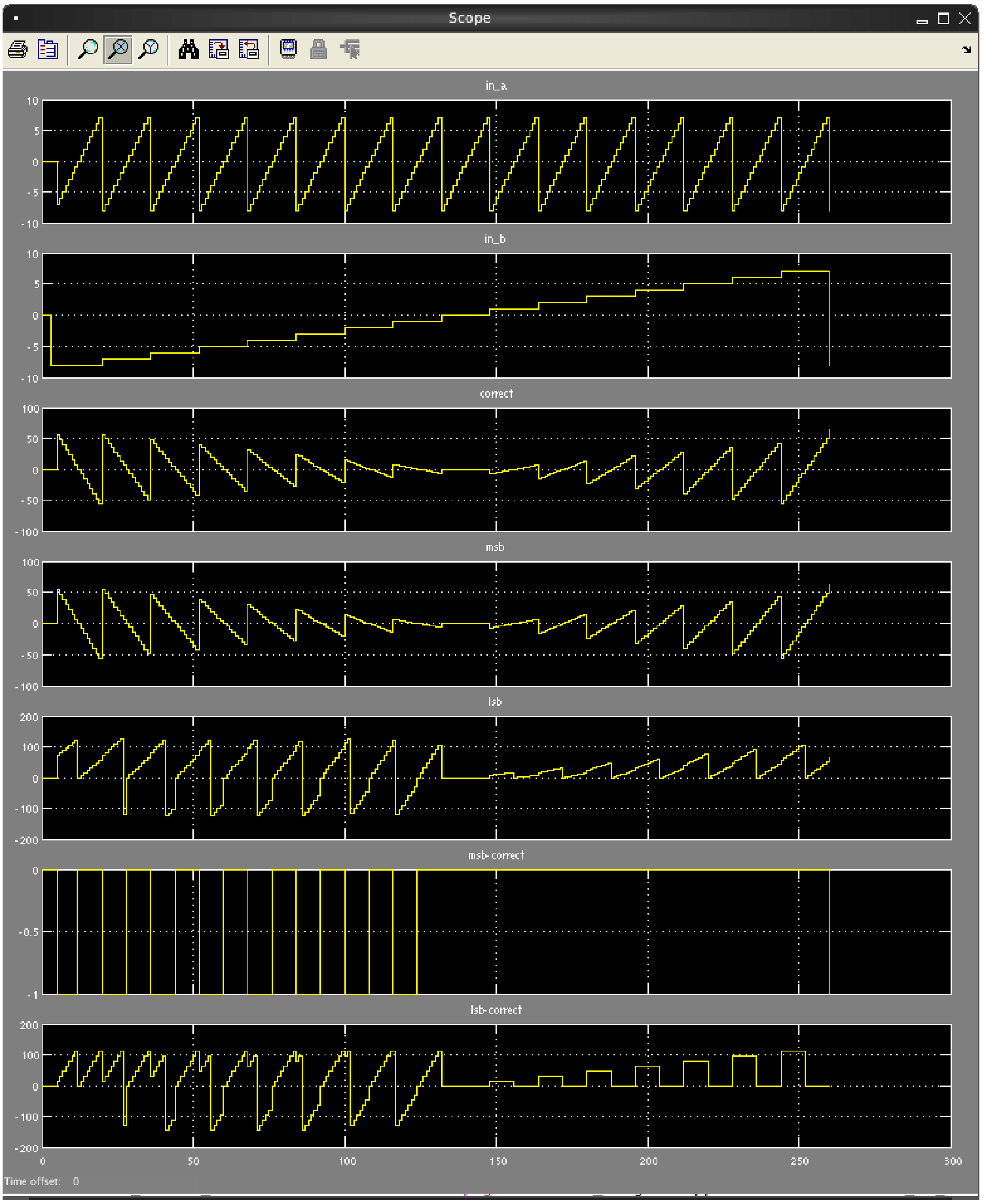


Figure 4.15: Testbench for two 4-bit multipliers in a single unsigned multiplier. The first and second scope shows the `A` and `B` inputs as it cycles through all signed 4-bit input pairs. The third scope, `correct`, is the expected output of a single signed multiplier. Scopes `msb` and `lsb` are the output of the two 4-bit multiplications. The final scopes, `msb-correct` and `lsb-correct`, shows the difference between scope `msb` and `correct`, and scope `lsb` and `correct`. The desired output of scope `lsb-correct` is a flat line at 0.

4.3.2 Additional Features

Once the 'version 1.0' firmware has been installed into a working SEPcam setup there are a number of changes and features that can be added the firmware. This includes second stage frequency channelization, coarse delay, and increased bandwidth processing.

Currently, frequency channelization is performed on the RSP board to $\sim$ 195 kHz resolution. A second stage FFT can be incorporated into the firmware. After the SerDes domain a small cornerturn and FFT domain can be added. An FFT size of 8, 16, or 32 would provide further channelization to mitigate narrowband RFI. The impact to the firmware would be the FFT logic, cornerturn memory, vector accumulator memory, and the x-engine cornerturn which will be required to be larger. The x-engine cornerturn memory usage will be the limiting factor of the FFT size. But, the x-engine will not require any further scaling.

A coarse delay may be required to accommodate the different travel latencies between RSP boards for spectrum window alignment. This is accomplished by including configurable BRAM delays before the SerDes queue. Fine delay is already included as part of the complex gain coefficients in the equalizer.

With a further modification of the RSP firmware the processing bandwidth could be increased by a factor of two or four. The subband data arriving currently uses 16 bits for the real and imaginary components. But, after queue alignment, each stream is quantized down to 4 bits. If this equalization and quantization stage was shifted to the RSP firmware then subband data could be output as 8 or 4 bit data, thus decreasing the data bandwidth per subband. The current SEPcam firmware could readily handle 8-bit data, with minor reconfiguration of the block parameters. Though, 4-bit data would be more of a challenge. Additional pipelined x-engines would be required, and the first stage integration would increase to 2^9 spectra. Both of which can be accommodated in the available FPGA resources. The main issue that would arise would be the cornerturn memory size. The QDR memory space is not large enough, and the cornerturn would need to be shifted to the DRAM. The freed QDR could then be used as vector accumulators. A major change to the firmware, but could be accomplished with the available FPGA resources.

4.3.3 Implications for Future Correlator Systems

Once a full setup has been built and tested at Chilbolton the system can be duplicated and installed at any other LOFAR international station. A number of international station operators have already expressed interest.

SEPcam has a number of similarities to the AARTFAAC system. In many ways SEPcam is little brother of AARTFAAC. The AARTFAAC system will correlate three times the number of antennas by connecting six LOFAR core stations. Both systems will use the URI board and new RSP firmware. Much of the design and issues that have arisen in SEPcam development will be useful for AARTFAAC. It would to the advantage of both projects to communicate on firmware development, imaging and calibration techniques, and shared resources when possible.

The specifications for the SKA low frequency and mid frequency systems are beginning to solidify. For phase 1 the low frequency array is expected to be made up of ~ 50 aperture array stations, covering $\sim 70 - 450$ MHz. The SEPcam correlator processes nearly four times the number of baselines for a ~ 7 MHz band. A simple scaling argument says that the phase 1 SKA low frequency correlator could be built with ~ 14 SEPcam like systems (per beam). This is likely an over estimate, as each board could process more than ~ 7 MHz of bandwidth.

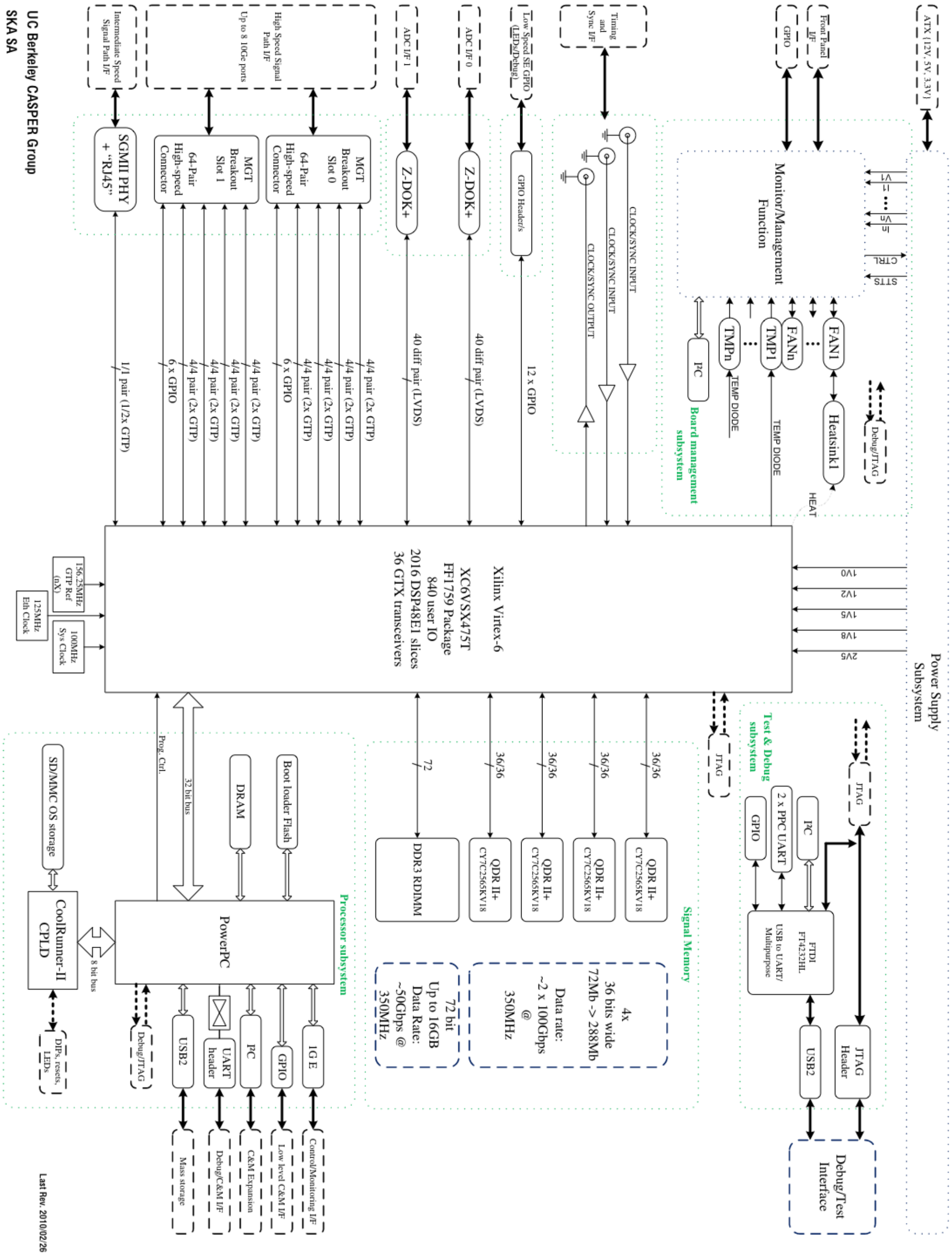


Figure 4.16: ROACH2 board diagram showing connections between the Virtex 6 and peripheral memory and logic. All eight high speed transceiver pairs are utilized for communication. Four QDR memory chips are used for the cornerturn operation and vector accumulation. Software interfaces are accessed through the on board PowerPC. From CASPER ROACH2 documentation page.

Chapter 5

LOFAR Single Station Observations

> *This would all make sense if you wrote it as a matrix.*
>
> -Daniel Price, 2011

In a parallel effort to the firmware design and deployment of the SEPcam correlator (see Chapter 4) initial interferometric observations were made with the current LOFAR-UK digital backend at Chilbolton Observatory. This has allowed for the development of a code base and pipeline to be used for transforming correlations into calibrated datasets which can be imaged and stacked into sky maps. These initial observations have been essential in understanding the array response, widefield imaging, and the low frequency sky.

Within the LOFAR observing bands there is steep evolution of the dominant synchrotron emission from the galaxy. A snapshot imaging method is used to generate maps of the total sky accessible from the LOFAR-UK station. Individual correlation snapshots over 24 hours are calibrated and imaged. Calibration requires a model for the element beam patterns and a sky model to derive complex gain coefficients. The resulting images are then gridded onto a sky map projection. This has been completed for several subbands to study the evolution of the galactic synchrotron radiation at low frequencies. Bright, class A point sources are also tracked to map out a generic beam pattern for the station elements.

5.1 LOFAR-UK Calibration Correlator

While deployment of the SEPcam correlator continues the current station calibration correlator can be used for interferometric observations. Part of every LOFAR station digital backend is a calibration correlator used to periodically update corrections to the complex antenna gains for coherent station beamforming. This correlator performs

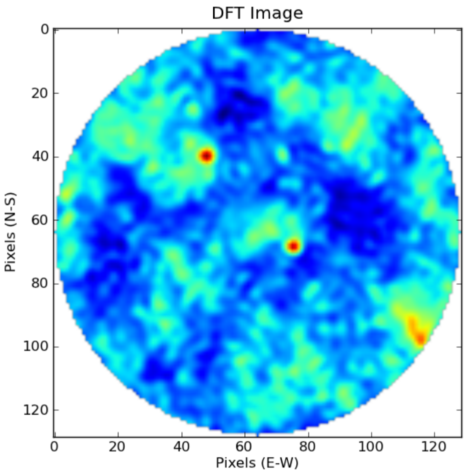


Figure 5.1: LOFAR-UK station snapshot image at 78 MHz. The field of view is horizon to horizon(π radians) and is projected onto a 2-D plane. The two bright point sources are Cygnus A and Cassiopeia A. In the lower right corner the galactic plane can be seen.

a complete $N \times N$, dual polarization cross correlation of the LBA or HBA station elements. In the case of the international stations N is 96 low band dipoles or 96 high band tiles. But, this correlator only computes the correlations of a single subband for a fixed length of one second and then shifts to the next subband. As there are 512 subbands the correlator cycles through all the subbands every 512 seconds. Correlations are written out to a single binary file of size $N^2 \text{ correlations} \times 512 \text{ subbands} \times 16 \text{ bytes}$. No phase tracking is applied to the antenna signals, each snapshot correlation has the zenith position as the phase center. The phase center declination is the latitude of the array and the right ascension is current Local Sidereal Time (LST) of the station. Applying the correlation matrix to the uv sampling based on the antenna layout and performing a Discrete Fourier Transform (DFT) produces a snapshot image where the bright sources Cygnus A, Cassiopeia A, and the galactic plane are visible, Figure 5.1.

Observation setup of a LOFAR station requires setting the RCU mode which sets the array type, system clock, and band filters. Each RCU board digitizes two analogue signals, usually the two orthogonal polarizations from an antenna element. Setting the sampling clock (160 MHz or 200 MHz) and setting filters for selecting Nyquist zones (1, 2, or 3) allows for a range of observing modes, Table 5.1. Since there is only one digital backend per station only one array is used at a time. Two modes can be used with the LBA. Below 30 MHz there is significant terrestrial communication RFI, a 30 MHz high-pass filter can be used in RCU mode 4. A 10 MHz high-pass filter is used With RCU mode 3, below 10 MHz the dipole response drops off, RFI increases in power, and the ionosphere turbulence makes observations impractical. The resonance knee at 57 MHz in the LBA band is due to the RLC circuit implementation of the bandpass filter before the LNA [van Cappellen et al., 2007]. Between 87.5 MHz and 110 MHz is the terrestrial FM radio band which is extremely strong and must be filtered prior to digitization in both the LBA and HBA bands, Figure 5.2. RCU modes 5, 6, and 7 cover

RCU Mode	Clock (MHz)	Array	Bandwidth (MHz)	Nyquist Zone	Usable Band (MHz)	Subband Width Δf (kHz)
3	200	LBA	100	1	10-90	195.3
4	200	LBA	100	1	30-90	195.3
5	200	HBA	100	2	110-190	195.3
6	160	HBA	80	3	170-230	156.2
7	200	HBA	100	3	210-250	195.3

Table 5.1: Available RCU configuration modes set the array type, the system clock and the band filter set. By setting the analogue filters and sampling clock different portions of the band can be digitized by using different Nyquist zones.

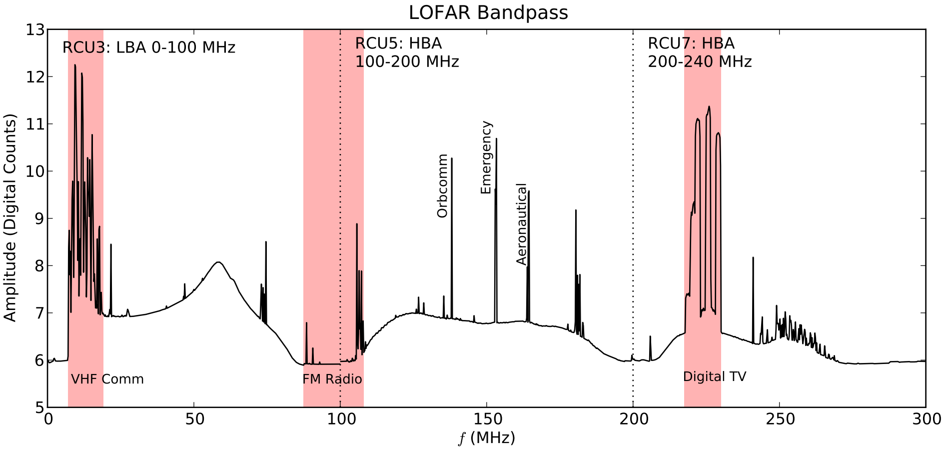


Figure 5.2: LOFAR band, $10 - 240$ MHz. RCU band 3 covers 10-90 MHz in the LBA band. RCU bands 5, 6(not shown), and 7 use the HBA and cover bands from 100-240 MHz. The FM radio band between 87.5-108 MHz is particularly strong and requires a set of low pass and high pass filters for the two arrays.

the HBA band from 110 MHz to 250 MHz by selecting different Nyquist zones. Within the HBA band there is an amalgamation of different RFI sources including digital TV, Orbcomm communication satellites, and aeronautical bands.

5.2 The Measurement Equation

The measurement equation is a mathematical model used to describe an electromagnetic signal as it propagates from a source through a medium and to a correlator system. The Radio Interferometry Measurement Equation (RIME) was first formalized in the mid-1990's in [Hamaker et al., 1996] and [Sault et al., 1996] using a Mueller matrix formalization. It is not a hyperbole to state that the RIME is the cornerstone to understanding and calibrating modern interferometric radio array. Recently, a re-formalization using Jones matrices has been introduced in a series of papers [Smirnov, 2011a], [Smirnov, 2011b], [Smirnov, 2011c]. The Jones formalization will be used in the context of this chapter. This section gives a brief overview of the RIME using the conventions presented in [Smirnov, 2011a].

As with any good radio astronomy model let us begin with a single, monochromatic point source which is described by the polarization vector $\mathbf{e}$. As the signal propagates from the source to the correlator system various transformations are applied to the signal, e.g. Faraday rotation, antenna beam gain, etc. These transforms are assumed to be linear and hence reversible during calibration. A transformation is described by $\mathbf{J}$, a 2×2 Jones matrix [Jones, 1941]. Multiple effects can be represented as a chain of transformations which are applied to the electromagnetic signal $\mathbf{e}$ before it is converted to a voltage $\mathbf{v}$ at the receiver, Eq. 5.1, where the order is dependent on the order in which the effects occur. The complex voltage is represented as a two polarization column vector, Eq. 5.2. If multiple effects are applied to the signal at the same time and are independent then the order of those effects does not matter. The goal of calibration is to return to $\mathbf{e}$ given $\mathbf{v}$ by inverting the $\mathbf{J}$ matrix.

$$\mathbf{v} = \mathbf{J_0 J_1 \dots J_{k-1} e} \tag{5.1}$$

$$\mathbf{v} = \begin{pmatrix} v_a \\ v_b \end{pmatrix} \tag{5.2}$$

Now in an interferometric array with N elements each baseline is a correlation between two spatially separated elements p and q. This correlation can be represented in the visibility matrix, Eq. 5.3, where v_{qa}^* is the complex conjugate of v_{qa} and the angle brackets are a time average. The factor of 2 in Eq. 5.3 is for maintaining historical convention.

$$V_{pq} = 2 \begin{pmatrix} \langle v_{pa} v_{qa}^* \rangle & \langle v_{pa} v_{qb}^* \rangle \\ \langle v_{pb} v_{qa}^* \rangle & \langle v_{pb} v_{qb}^* \rangle \end{pmatrix} \tag{5.3}$$

The visibility matrix, Eq. 5.3, can be rewritten by moving the time averaging and using Hermitian conjugate notation to put V_{pq} in terms of $\mathbf{e}$:

$$V_{pq} = 2\langle \mathbf{v_p v_q^H} \rangle = 2\langle \mathbf{J_p e} (\mathbf{J_q e})^H \rangle = 2\langle \mathbf{J_p (e e^H) J_q^H} \rangle \tag{5.4}$$

Here the component $\langle \mathbf{e e^H} \rangle$ is directly related to the Stokes parameter of a source, also known as $\mathbf{B}$, the brightness matrix. Which gets us to the general form of the RIME for a point source:

$$V_{pq} = \mathbf{J_p B J_q^H} \tag{5.5}$$

Now the $\mathbf{J}$ matrix is the general form of the translation matrix, let us consider some specific forms of this matrix which are used in practice. These translations or corruptions represent physical processes applied to the $\mathbf{e}$ vector. A phase delay applies a phase offset to both polarizations equally, this it is a scalar transform and be described by a scalar matrix, $\mathbf{K}$:

$$\mathbf{K} = e^{i\theta} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{5.6}$$

Complex antenna gain, the so called $\mathbf{G}$ term, is a translation applied to each $\mathbf{e}$ component independently, thus is represented by a diagonal matrix:

$$\mathbf{G} = \begin{pmatrix} g_x & 0 \\ 0 & g_y \end{pmatrix} \tag{5.7}$$

Intermixing the components, such as from crosstalk in the electronics, Faraday rotation in the ionosphere, or change of polarization basis are represented by off-diagonal matrices. All these matrices have an implicit time and frequency variability which is related to the physical process the matrix models. The characteristic time length and frequency shape for each translation can be used to separate out different effects when solving for a calibration. For a simple point source model being correlated between antennas p and q with a complex gain applied the RIME is:

$$V_{pq} = \mathbf{G_p K_p B K_q^H G_q^H} \tag{5.8}$$

The point source RIME, Eq. 5.8, can be used to create a more general RIME for S sources:

$$V_{pq} = \mathbf{G_p} \left(\sum_s \mathbf{E_{s,p} K_{s,p} B_s K_{s,q}^H E_{s,q}^H} \right) \mathbf{G_q^H} \tag{5.9}$$

The measurement equation is the sum over S sources in the field, each with a brightness $\mathbf{B_s}$ and phase offset relative to the antennas based on their location in the sky. The complex gain term is source independent and can be brought out of the sum, this fits into the Direction Independent Effects (DIE's) class. Additionally there is a source dependent term, $\mathbf{E}$, which represents the Direction Dependent Effects (DDE's). The $\mathbf{E}$ term is used to model beam patterns which have traditionally been approximated as flat or at least static over an observation. A LOFAR station, with an extremely large field of view, has a dipole beam pattern which must be taken into account to model the change in apparent brightness of a source as it crosses the celestial sphere. Using a formalism that takes into account DDE's makes it possible to calibrate interferometers which observe wide fields of view and have time and frequency variability. By increasing the number of sources the sky model can be improved and Eq. 5.9 can be extended further to model a continuous brightness distribution.

Inversion of the Jones matrices would be a simple task in a perfect system which does not take into account noise. The introduction of noise to the system requires the signals to be sufficiently high that minimization can be performed to solve for the solutions, i.e. the inverted Jones matrices. This requires a large enough number of samples in the time and frequency domain to raise the modelled source signal above that of the noise. However, the time and frequency intervals should be short enough as to not wash out the short period effects, especially when it comes to solving for DDE's. When performing a calibration on an observation we are free to choose which effects to include in a model. It is without saying, the better the model the better the calibration solution so we should strive to create a model which accounts for all systematic effects above the nominal noise level. The specific model used to calibrate the LOFAR single station observations is described in detail in Section 5.4 of this chapter.

5.3 Station Observations

An automated pipeline for generating sky maps from the station correlation data has been implemented using a suite of radio synthesis calibration tools, imagers, and custom code. Observations are converted to measurement sets and calibrated using MeqTrees [Noordam & Smirnov, 2010] to solve for $\mathbf{G}$. A common antenna gain solution is applied to a set of observations. The resulting corrected data are imaged and cleaned. Snapshot images are gridded onto a Hierarchical Equal Area isoLatitude Pixelization (HEALPix) surface and combined in the image domain to produce sky maps at multiple frequencies which cover the viewable sky from Chilbolton Observatory. This method can be used to build up low resolution maps to study the evolution of the galaxy at with in the LOFAR bands.

5.3.1 Low Band Array

During observations with the LBA the calibration correlator can continuously output correlation matrices. Over 24 hours the entire viewable sky from Chilbolton is observed covering all right ascensions and declinations $-10^\circ \leq \delta \leq 90^\circ$. The observable declination range is set by the response of the element beams and the latitude of the array (51.14°). Each dipole element is randomly distributed within a 60 meter diameter circle, Figure 5.3a. This position distribution minimizes the baseline redundancy and fills in the uv coverage, Figure 5.3b with a hole in the shorter baselines due to the

ν (MHz)	λ (m)	$A_{eff,dipole}$ (m^2)	$A_{eff,arr}$ (m^2)	ΔS_{ij} (kJy)	ΔS_{arr} (Jy)
15	20.0	51.92	4985.10	10.59	156.86
30	10.0	12.98	1246.27	7.23	107.14
45	6.67	5.77	553.90	5.79	85.72
60	5.0	3.25	311.57	4.94	73.18
75	4.0	2.08	199.40	4.37	64.72
90	3.33	1.44	138.47	3.95	58.55
110	2.73	1.56	150.00	4.38	64.80
130	2.30	1.56	150.00	2.86	42.32
150	2.0	1.33	127.82	2.33	34.48
170	1.76	1.04	99.52	2.17	32.19
190	1.58	0.83	79.67	2.04	30.28
210	1.43	0.68	65.22	1.94	28.66
240	1.25	0.52	49.93	1.80	26.63

Table 5.2: Effective collecting area and sensitivity of the LBA dipoles and HBA bow-tie elements. The $A_{eff,arr}$ and S_{arr} for the HBA only takes into account one of the 16 dipoles per tile. These values are upper limits, as they do not account for systematics and electronic passband gain. Sensitivity presented is for a single polarization, an I Stokes image will have a $\sqrt{2}$ improvement.

physical size of the elements. A smooth and filled uv coverage, Figure 5.3c, produces a smooth PSF with minimal sidelobes, Figure 5.3d. The dense distribution of elements and the long wavelengths (3.33-30 meters) of the LBA band assure the aperture is at least critically sampled for most observing frequencies and should be considered filled.

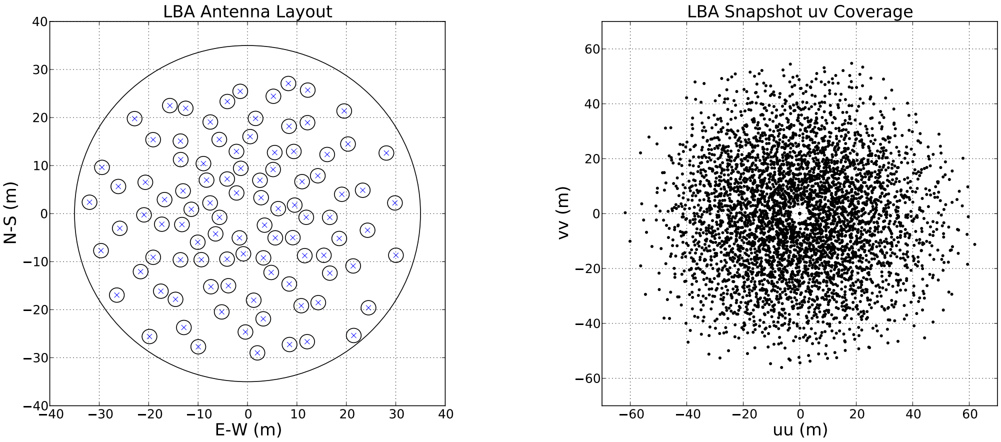
Over the course of the station build-out and the subsequent station validation (October 2010-June 2012) observations over multiple epochs were recorded with the station calibration correlator using the LBA.

Let us discuss the sensitivity and effective collecting area of the LBA. The LBA elements are half-wavelength dipoles where $\lambda_c = 1.4$ meters. From Eq. $4 - 92$ in [Balanis, 2005] the maximum effective collecting area of a half-wavelength dipole is:

$$A_{eff} = \frac{\lambda_c^2}{4\pi} D_0 \simeq 0.13\lambda_c^2 \tag{5.10}$$

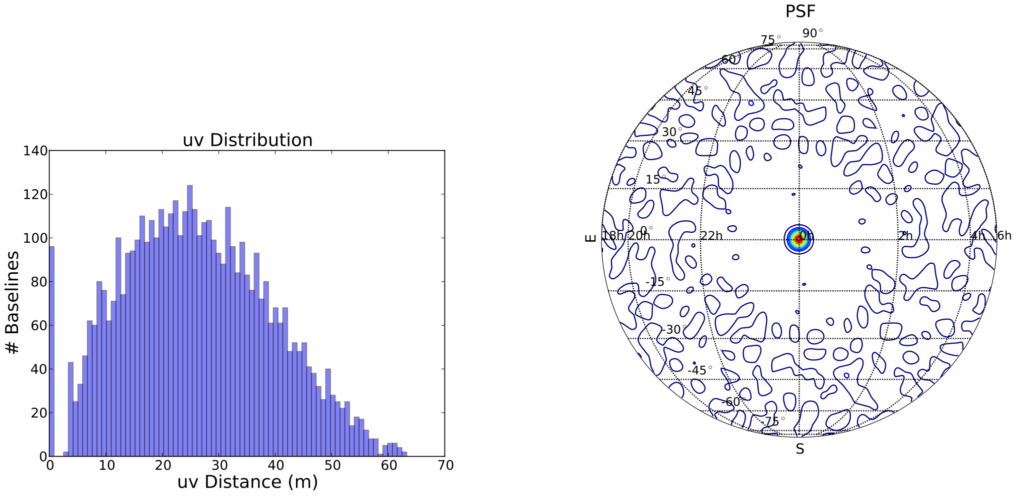
where D_0 is the directivity of an antenna, for a half-wavelength dipole $D_0 = 1.643$. We will use the approximation that the element is a half-wavelength dipole at all frequencies in the LBA band. At longer wavelength the total A_{eff} of the array will decrease as some sections of the array are above the critical spatial sampling rate. The element A_{eff} and the total A_{eff} of the array presented in Table 5.2 should be taken as an upper limit.

In low frequency radio astronomy we are in a unique situation where the sky temperature, in this case the galactic synchrotron emission, dominates the system temper-



(a) Position layout of the 96 dipoles of an international LOFAR LBA. Circle size represents the physical size of the elements.

(b) Instantaneous uv coverage during a zenith pointing. Due to the wide range of wavelengths, dimensions are in meters.



(c) Distribution of baselines in uv distance ($\sqrt{U^2 + V^2}$). The lower limit in uv spacing is due to the physical size of the dipoles.

(d) Instantaneous PSF response across the entire hemisphere field of view at 78 MHz, contours are in a linear scale.

Figure 5.3: Configuration and response of the international LOFAR LBA at Chilbolton Observatory. Dipoles are randomly distributed (5.3a) to fill the uv plane (5.3b). The uv coverage fills in all the spacing within the array (5.3c), to produce a smooth PSF response (5.3d).

ature [Taylor et al., 1999], Chapter 29. The sky temperature at 30 MHz ranges from $6.6e4$ K at the galactic center to $1.05e4$ K at the galactic pole with a spectral index of $T_{sky} \propto \nu^{2.55}$ [Cane, 1978]. The system temperature is a function of the LNA receiver noise, antenna efficiency, cable losses, and the sky. This can be broken up into two components, $T_{sys} = T_{ant} + T_{sky}$. For the LBA $4\text{dB} < T_{sky}/T_{ant} < 14.6\text{dB}$ [van Cappellen et al., 2007]. This ratio decreases as the observing frequency increases. The LBA dipole System Equivalent Flux Density (SEFD), a measure of the element performance is:

$$SEFD = \frac{2k_b T_{sys}}{\eta_a A_{eff,dipole}} \tag{5.11}$$

Here k_b is the Boltzmann constant, η_a is the antenna efficiency which is ~ 1 for an LBA dipole. Sensitivity is defined as the measure of the weakest radio source detectable, usually around 3σ where σ is the standard deviation of the noise level. The baseline sensitivity of antenna i and j, and correlated array sensitivity in Janskys is:

$$\Delta S_{ij} = \frac{SEFD}{\eta_s \sqrt{2\Delta\nu\tau_{acc}}} \tag{5.12a}$$

$$\Delta S_{arr} = \frac{SEFD}{\eta_s \sqrt{N(N-1)\Delta\nu\tau_{acc}}} \tag{5.12b}$$

Sensitivity values are presented in 5.2 where $T_{sys} \sim T_{sky}$ is used. The system efficiency, η_s, is taken to be ~ 1 thus the sensitivity is an upper limit. The SEFD for each element is taken to be the same. Bandwidth, $\Delta\nu = 195$ kHz, is the subband width and $\tau_{acc} = 1\text{s}$ which is the station correlator integration length. For Eq. 5.12b, $n = 96$ for the number of antennas in an international station. Based on these sensitivity estimates only a few point sources should be detected as well as the galactic structure.

5.3.2 High Band Array ‘all-sky’

When using the HBA for station imaging two complications are introduced: the regular spacing of the element tiles and the analogue beamformer in each tile. The regular layout of the HBA, Figure 5.5a, produces a set of highly redundant baselines and a poorly sampled uv plane, Figure 5.4a. Additionally, beamforming the 16 dipoles in each tile increases the beam directivity but reduces the field of view. With the regular tile layout this produces strong grating lobes, Figure 5.4b. Imaging in the regular tile beamformed mode is impractical, but both of these problems can be over come at the cost of sensitivity.

During international LOFAR observation mode a two stage beamformer is used, the first is the analogue tile beamformer, the second is the digital station beamformer. The station calibration correlator extracts signals before the digital beamformer. The analogue beamformer is configurable to select sections of the sky by setting delays between the dipole element in each tile. Each dipole element can also be disabled, when the HBA is not observing all the dipoles are disabled. But we can enable a single dipole in each tile to enable the 'all-sky' mode of the HBA. This significantly reduces the tile array sensitivity to $\frac{1}{16}$ of the original sensitivity, but allow the HBA to be used for station imaging.

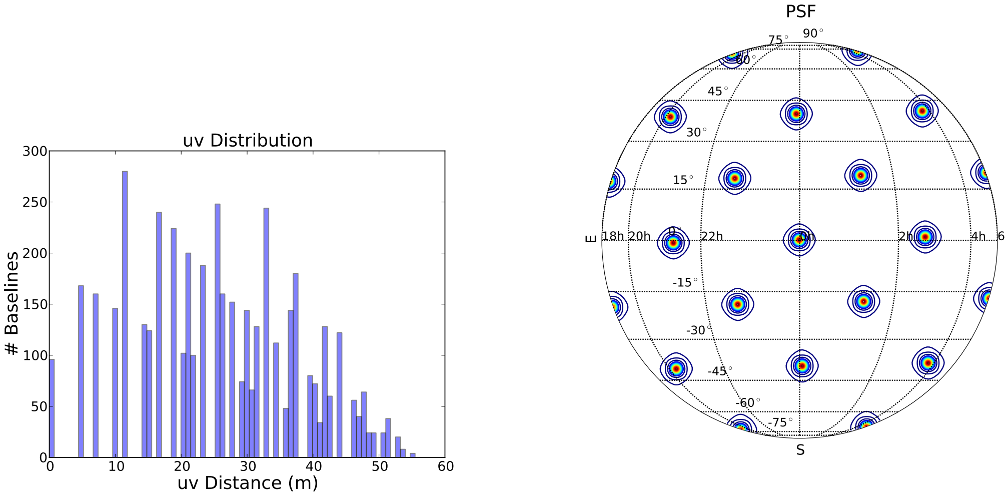
Enabling only a single dipole per HBA tile allows for a 16-fold increase in the number of possible baselines. By randomly selecting one of the 16 dipoles in each tile the uv coverage can be filled in, Figure 5.5c. Though the elements are still on a regularly spaced grid the baselines are better distributed, Figure 5.5b, which in turn leads to a smooth PSF with no grating lobes, Figure 5.5d.

Using the HBA in the 'all-sky' imaging mode requires direct control over the array and can not be done during commensal observations due to the unique settings in the tile beamformers. Because of this, only a few observations have been made using the HBA in this mode so far.

The effective area and sensitivity of the elements and HBA array are computed, see Table 5.2, using the same method as for the LBA presented earlier in the chapter. Each bow-tie element is modelled as a half-wavelength dipole and the total physical collecting area is $1.5625\ \mathrm{m}^2$. In the HBA band the $T_{sky} \sim T_{ant}$ [Kant et al., 2007], both are included in the SEFD calculation. The bright Class A radio sources will be detected in a single integration. The steep spectral index of the galactic synchrotron means only the very strongest structure will be detected in a single integration. To improve the sensitivity and sky coverage the individual integration images are added in the image plane.

5.4 A Measurement Equation for the LOFAR Station

In order to make use of common radio interferometry calibration packages such as AIPS, CASA, and MeqTrees the correlation matrices are converted into measurement set files. Each measurement set consists of a single time integration and one or a group of subbands. The large beam size of the array means that over a few seconds the sky will not have changed much, this allows multiple subbands which are close to each other to be grouped into a single time slot at the cost of a slight phase gradient across the subbands. In principle, multiple time integrations can be placed in the same



(a) Distribution of baselines in uv distance ($\sqrt{U^2 + V^2}$). The regularly spaced layout of the tiles produces a highly redundant array and creates gaps in the uv coverage.

(b) Instantaneous PSF response across the entire hemisphere field of view at 140 MHz with the tiles beamformed towards zenith, contours are in a linear scale.

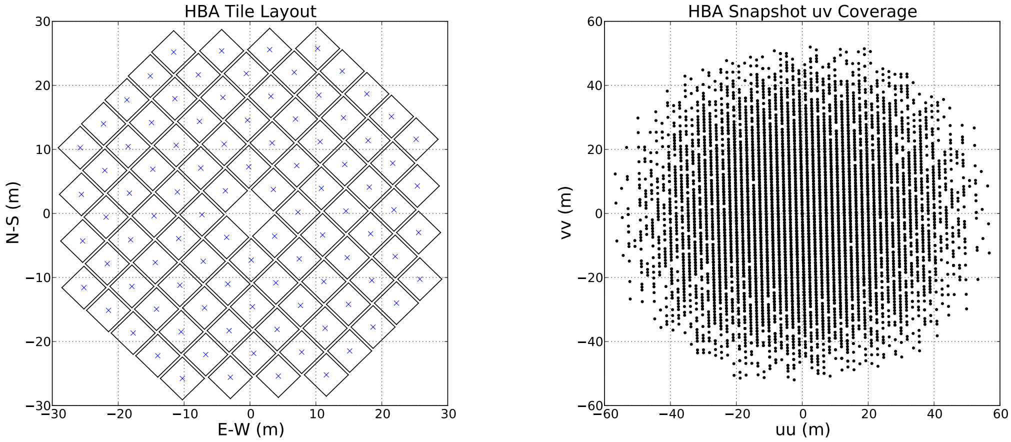
Figure 5.4: Configuration and response of the international LOFAR HBA at Chilbolton Observatory with beamformed tiles. The grid layout of the HBA tiles, Figure 5.5a, create a highly redundant array with gaps in the uv coverage (5.4a) which leads to a strong grating lobe response (5.4b). This makes imaging with beamformed HBA tiles a difficult, if not impossible, task.

measurement set but this requires adding the necessary phase correction to each time integration such that all the integrations have the same phase center. This introduces a number of problems related to aliasing and widefield effects as noted below. Calibration and imaging is performed on single time integrations and the observations are combined in the image plane.

A number of issues arise during calibration of the station data related to the wide field of view, beam shape, the dynamic sky among others. This requires use of a third generation calibration method with can correct for DDE's. The MeqTrees [Noordam & Smirnov, 2010] software package has been used to derive antenna gain solutions by inverting the Jones matrix based model of the array. Eq. 5.9 is the model used for the LOFAR station array.

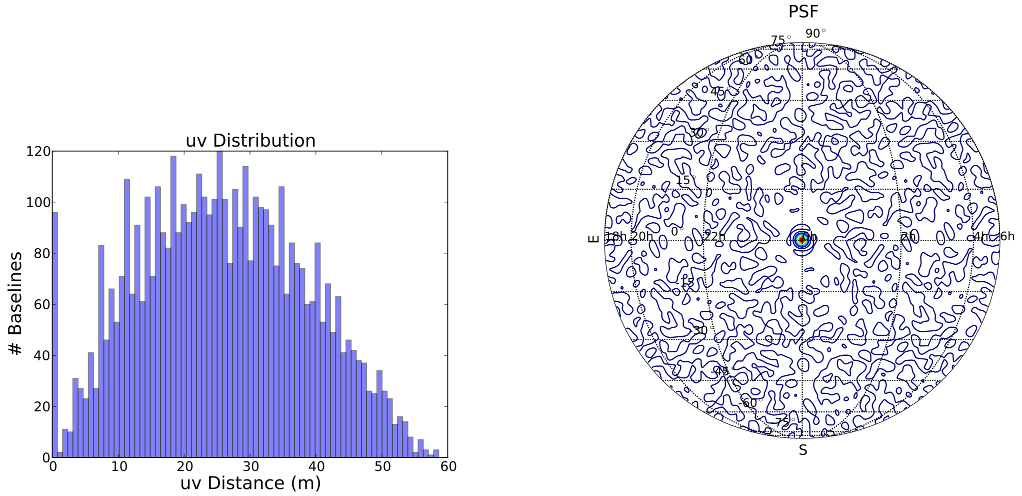
5.4.1 Local Sky Model ($\mathrm{B_s}$)

In the northern hemisphere, in the LOFAR bands the dominant sources are Cygnus A, Cassiopeia A, the Sun, and the galactic center. Jupiter has bright, variable radio emission (~kJy) below 40 MHz due to current interchanges between itself and Io, but



(a) Position layout of the 96 tiles of an international LOFAR HBA. Each tile is made up of 16 'bow-tie' dipoles which are analogue beamformed.

(b) Instantaneous uv coverage during a zenith pointing. Due to the wide range of wavelengths, dimensions are in meters.



(c) Distribution of baselines in uv distance ($\sqrt{U^2 + V^2}$). By selecting one of the 16 dipoles randomly the gaps in uv coverage can be filled in.

(d) Instantaneous PSF response across the entire hemisphere field of view at 140 MHz using a single dipole in each tile, contours are in a linear scale.

Figure 5.5: Configuration and response of the international LOFAR HBA at Chilbolton Observatory using an 'all-sky' setup. HBA tiles are on a grid (5.5a), to reduce the grating lobes and fill the uv plane (5.5b) a single, random dipole is used from each tile. The uv coverage fills in all the spacing within the array (5.5c), to produce a smooth PSF response (5.5d).

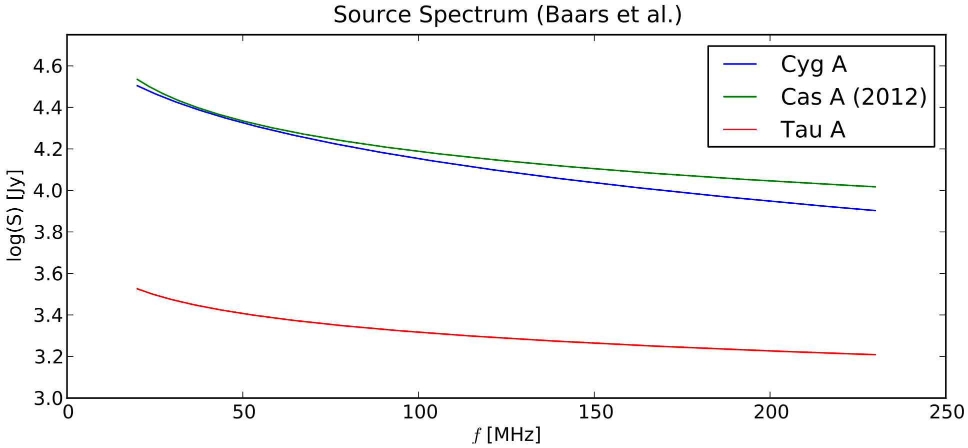


Figure 5.6: Radio spectra of Cygnus A, Cassiopeia A and Taurus A based on the [Baars et al., 1977] model. The spectrum of Cassiopeia A has been adjusted for 2012 based on the modelled flux evolution.

has a very steep spectral index and is only visible in the low end of the LBA band [Zarka, 1998]. The Sun's declination varies across the seasons, ranging from $+23°$ to $-23°$ through out the year, which is an altitude of $16°$ to $62°$ with respect to the array latitude. The sources Cygnus A and Cassiopeia A have been well studied over a broad range of the radio spectrum. Cygnus A is a stable source with known spectrum, Cassiopeia A has been shown to be decreasing in flux at a known frequency dependent rate [Baars et al., 1977]. Figure 5.6 shows the spectrum of the three brightest extrasolar radio sources: Cygnus A, Cassiopeia A, and Taurus A. Cygnus A and Cassiopeia A have comparable brightnesses within this frequency range, and are almost an order of magnitude brighter than the next brightest source, Taurus A.

Besides these bright sources the galaxy is also a very bright source at low frequencies due to synchrotron emission. Regions near the galactic center have comparable flux to that of the class A sources [Guzmán et al., 2011]. The galaxy can be modelled as a set of point sources which are unresolved by the array PSF. Such a model has been developed for CMB foreground removal [de Oliveira-Costa et al., 2008] which takes into account the spectral evolution of the galaxy based on an interpolation between previous low frequency sky surveys. A simple model consisting of the class A point sources has been used for calibration. This is due to the limitations of the beam model discussed below. Only sources near the beam center are used for calibration. Since the Sun and the galactic center just skim the lower altitudes of the beam and the beam model is not well known at low altitudes they are not included in the model. Cygnus A and Cassiopeia A both cross near the beam center and are extremely bright, unresolved sources which are ideal for calibration. The sky model flux scale is set using the values derived from [Baars et al., 1977]. These are interpolated scales, further work is need to set an absolute scale based on calibrator sources.

5.4.2 Phase Delay ($\mathrm{K_s}$)

The phase delays used in the RIME are based on the antenna element positions and the sky model source positions. Both of these positions have some inherent error in measurement. For the antenna positions this is mainly due to the error in the Global Positioning System (GPS) measurement, which is assumed to be below the phase noise detection limit. This error is much shorter than the observing wavelength and can be ignored. The sky model source positions will be based on telescope pointing error and resolution from when the astrometry was determined The LOFAR stations are very low resolution, between $.5^\circ$ and 3° depending on subband frequency, that this does not pose a problem. We win at low resolution and low frequency when it comes to using correct phase delays.

5.4.3 Dipole Beam Model ($\mathrm{E_s}$)

A main issue with using the station elements for correlation is the unknown beamshape of the elements. But we can make a model of the beam based on a theoretical model. This is just the first stage of model refinement though. Once an initial model is developed and used to calibrate we can use stable radio sources to map out the element beams and update the beam model. This method is explored in section 5.7.

The LBA antenna elements are a ‘droopy dipole’ design which is a half-wavelength, orthogonal polarization dipole with each of the quarter-wavelength feeds at an angle to the plane of the array. Each antenna element is also made up of a wire mesh ground screen which increases the sensitivity by $\sqrt{2}$. Each wire feed is 1.4 meters in length which means the nominal wavelength is $\lambda_c = 5.6$ meters and $\nu_c = 53.5$ MHz for the dipole response. For calibration it is useful to make the assumption that across the usable LBA band the antenna element can be approximated as a half-wavelength dipole for the central observing frequency. The Half Power Beam Width (HPBW) of a half-wavelength dipole is 78°, for $\lambda > \lambda_c$ the HPBW increases asymptotically towards 90°, as expected for $\lambda < \lambda_c$ the HPBW decreases, see [Balanis, 2005] Chapter 4 for a full treatment. ASTRON has cited the effective HPBW of an LBA element to be 80° [van Cappellen et al., 2007]. The normalized power pattern of a half-wavelength dipole is Eq. 5.13 from Eq. $4 - 87$ in [Balanis, 2005]. We use a beam power model of $\sin^3\theta$ and wrap the normalization factor into the $\mathbf{G}$. Since the antenna is approximated as a half-wavelength dipole for all frequencies the model will break down at lower altitudes and derived gains when sources are low in the sky may provide poor results. But near the center of the beam the approximation is good at all LBA frequencies, we will use observations which are at times when calibration sources are near Hour Angle (HA)$= 0$ for deriving gains.

$$U = (E_{\theta n})^2 = C \left[\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} \right]^2 \backsimeq C \sin^3\theta \tag{5.13}$$

The same first order approximation of a half-wavelength dipole beam pattern will be used for the HBA bow-tie (or fat dipole) elements. This ignores any effects of the densely packed HBA, possible cross coupling effects, and the actual element beam pattern. Further refinement of the beam pattern will require a better model and use of celestial sources to probe cuts in constant declination of the beam.

5.4.4 Complex Gains (G)

This is a catch-all term for slowly varying in time DIE's in the system. At the end of calibration we want to derive a G for all antennas which can then be applied to all observations during a period of time, in the case of the LOFAR station a new set of G values are derived every 24 hours of observation. These complex gains will set the antenna phase corrections, the flux level based on the sky and beam models to convert the digital counts into Janskys. During calibration only a subset of observations are used which have Cygnus A and Cassiopeia A high in the beam and the galactic plane below or near the horizon. Also, if possible the sun will be set below the horizon also. An average G$_{avg}$ based on the G values for each observations in the subset is then applied to all observations over a 24 hour period. Over the subset of observations used for calibration the phase is stable and the amplitude is flat with a slight slope most likely due to the changing galactic structure within the field of view, Figure 5.7.

The sky model, once fit, can be subtracted from the corrected data. This will remove Cygnus A and Cassiopeia A models which to reduce their PSF sidelobe levels in the images.

5.5 Sky Maps

Sky maps are generated by griding hemisphere snapshot images onto a HEALPix projection. This allows images produced over the full right ascension range to be combined into a single map.

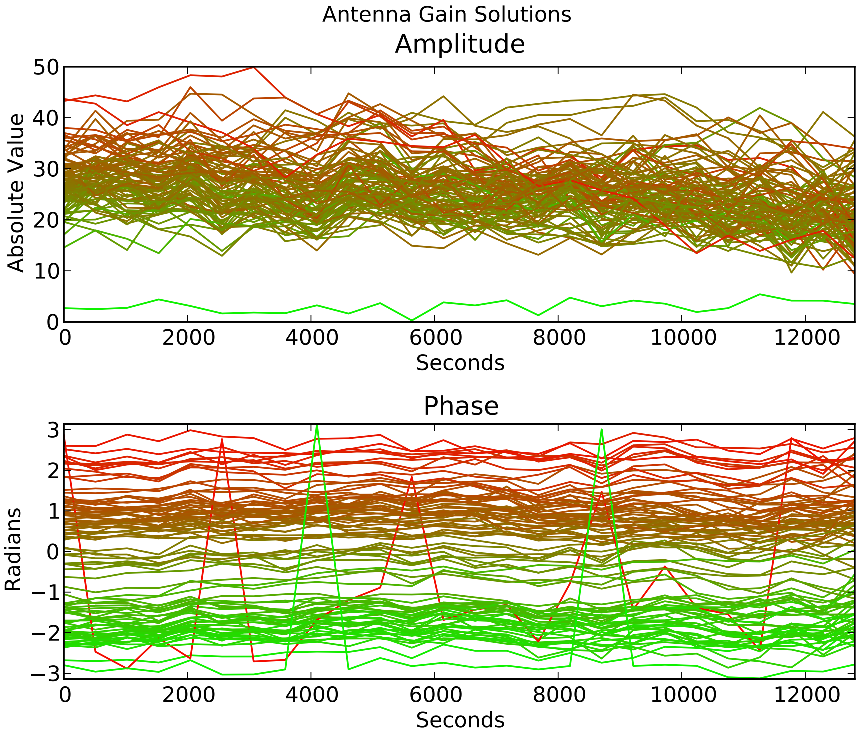


Figure 5.7: Gain solutions for the 96 LBA elements using observations covering 3.5 hours when the Cygnus A and Cassiopeia A are near the local meridian. The colors vary to differentiate between the 96 solutions. One of the elements (green line at low amplitude) is not functioning properly which is causing erratic phase changes(noisy red line). All baselines using that element are flagged during imaging.

5.5.1 Widefield Images

Imaging projects the celestial sphere within the field of view onto a 2D plane. For a small field of view the sphere can be approximated as flat and no projection distortions are introduced. In widefield imaging sources far from the phase center will become distorted, a full discussion is presented in Chapter 17 of [Taylor et al., 1999]. The visibility equation 5.14 in the general form is a measurement of the signal coherence between two elements p and q based on the sky brightness, $I(l, m)$.

$$V_{p,q}(u, v, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{I(l, m)}{\sqrt{1 - l^2 - m^2}} e^{-2\pi i \left[ul + vm + w\left(\sqrt{1-l^2-m^2}-1\right)\right]} dl\, dm \tag{5.14}$$

An approximation of Eq. 5.14 is made to reduce the form to a 2D Fourier transform. This is made by placing the measurement elements onto a common plane, thus making the array a co-planar array or by making the assumption that the field of view is small and thus that region of the celestial sphere can be approximated as a plane. Both of these scenarios has the effect of setting the w term to zero and the visibility equation is simplified to a Fourier form, Eq. 5.15.

$$V_{p,q}(u, v, w \equiv 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{I(l, m)}{\sqrt{1 - l^2 - m^2}} e^{-2\pi i [ul + vm]} dl\, dm \tag{5.15}$$

By not accounting for the w term a linear phase error is introduced in (u, v) plane proportional to the distance squared of a source from the phase center, the source will appear to move position during the image. This is a minimal effect when only imaging a small region of the sky. When the term $2\pi w \left(\sqrt{1 - l^2 - m^2} - 1\right)$ becomes large, greater than unity, the Fourier approximation to the visibility equation is no longer valid. Combining together multiple time integration, which cover different portions of the uv plane cause sources to become smeared out. This distortion can be corrected by multiple methods: faceting smaller fields within the entire field, a computationally intensive 3D FFT, applying a w-projection correction terms, or snapshot imaging [Cornwell et al., 2003]. The last method, snapshot imaging, fits well into the station correlation datasets.

For a single snapshot observation there always exists a common w-plane which can be used at the cost of a known coordinate distortion when projecting the celestial sphere to the imaging plane, [Bracewell, 1984]. This means that the Fourier approximation (Eq. 5.15) still holds [Cornwell et al., 2003]. If the phase center of an observation is chosen to be zenith the resulting coordinate distortion is relative to the observation LST and array latitude, Figure 5.8.

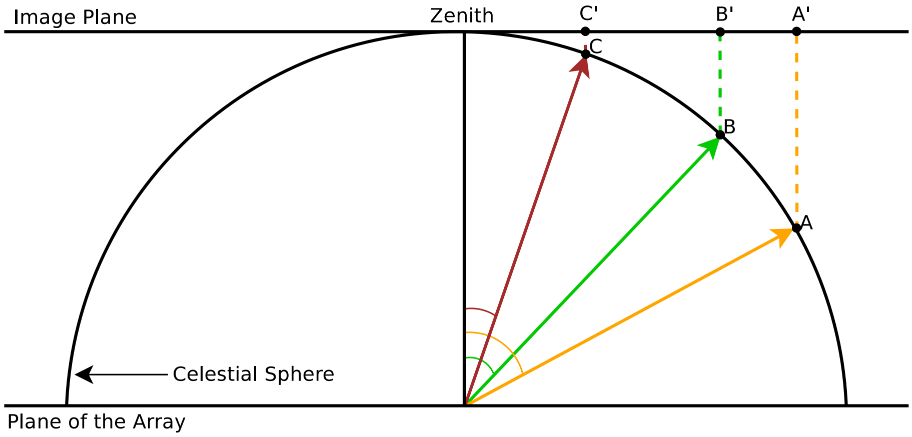


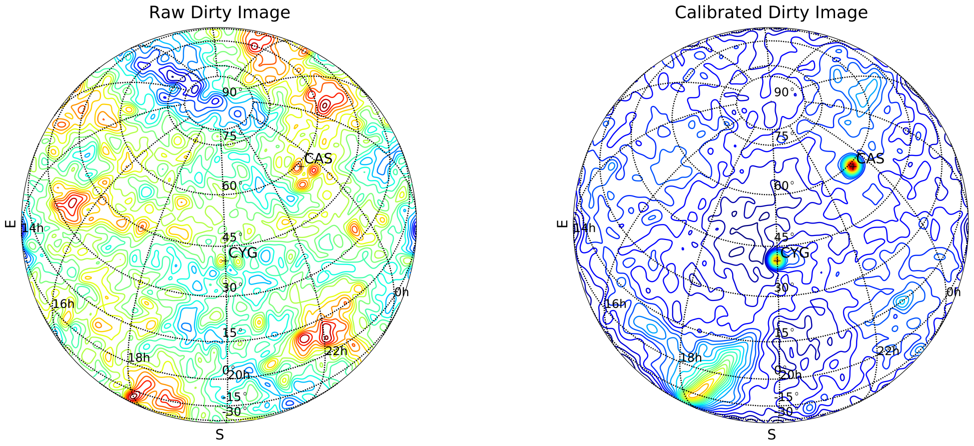
Figure 5.8: Snapshot projection of the celestial sphere onto the image plane when observing with a zenith pointing array. For a snapshot observation there always exists a plane which does not introduce w-term distortions in the image.

By imaging a single time integration, which is directed towards zenith, the w-term is eliminated. The resulting snapshot images are stored as FITS files using a non-linear, sine projection coordinate system [Greisen, 1994]. Since the phase center was chosen to be at zenith for all observations then all images will have a common w-plane coordinate distortion, this will come in useful when producing sky maps. This projection is taken into account when griding images onto a celestial sphere during the map making process.

An example snapshot LBA observation when Cygnus A is crossing the local meridian is shown in Figure 5.9. This figure shows imaging stages from the raw visibilities to the calibrated dirty image to the cleaned image with residuals. By imaging the entire hemisphere of sky over the array it is assured that no sources are aliased into the field. But, any source in the field has an extend PSF response which will alias, especially sources near the horizon. For weak sources the aliasing is below the noise. The extended galactic center, which is very bright at low frequencies, is always near or just below the horizon. The resolved, bright structure can introduce artifacts at the aliased PSF sidelobe positions. With so much extended structure a portion of the galactic flux is not properly modeled during cleaning. This requires development of a better galactic model.

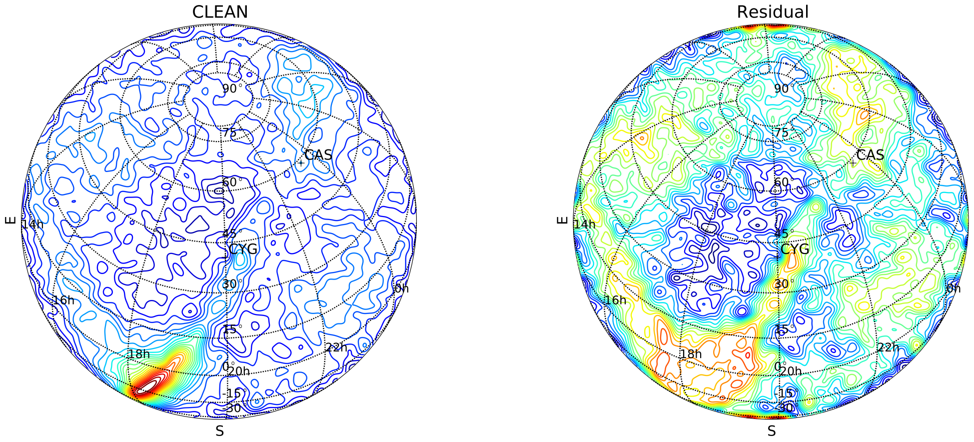
5.5.1.1 HBA 'all-sky' Images

The process of calibrating and imaging with the HBA is similar to that of the LBA. A common measurement equation is used for both arrays, but the HBA beam pattern has not been included at this point. Figure 5.10a shows a snapshot image at 115 MHz using the HBA. Cygnus A and Cassiopeia A are the dominating sources, as in the



(a) Image formed from the raw visibilities before applying complex gains. No sources are visible. Contour levels are 25 in digital counts.

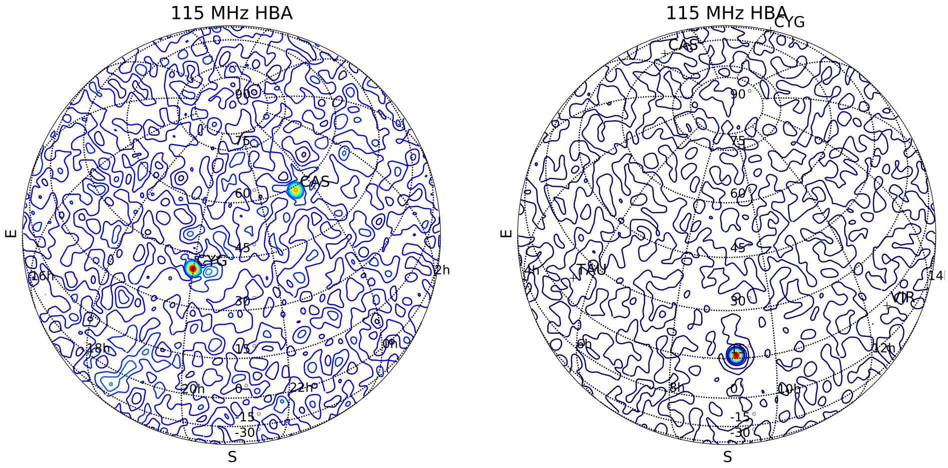
(b) After applying gain corrections the galactic plane, Cygnus A and Cassiopeia A are visible. Contour levels are 800 Jy.



(c) Cleaned image with the sky model removed, only the residual of Cygnus A and Cassiopeia A are present along with the galactic plane. Cleaning was performed using the Cotton-Schwab clean algorithm, 3000 iterations. Contour levels are 600 Jy.

(d) Residual flux remaining after clean iterations. Flux from the galactic plane remains in the residuals. Contour levels are 250 Jy.

Figure 5.9: Single subband, one second integration, widefield image at 78 MHz using the LOFAR LBA array at Chilbolton Observatory. A sky model based on the class A radio sources is used to derive complex gains from the raw visibilities (5.9a). Gain corrections are applied to the observation to produce a calibrated dirty image (5.9b). The sky model is removed and the resulting image is cleaned using the Cotton-Schwab clean algorithm (5.9c) with the resulting residuals (5.9d).



(a) Snapshot image at 115 MHz from the LOFAR HBA 'all-sky' mode. The class A radio sources dominate the field. Unlike the LBA band, the galactic synchrotron structure is barely visible due to its steep spectral index.

(b) Snapshot image at 115 MHz from the LOFAR HBA 'all-sky' mode. Image from an observation taken during midday in August 2012 where the sun is the only visible source. The noise floor is ~ 4 times higher then when the sun is set below the horizon.

Figure 5.10: Widefield images formed from HBA observations at 115 MHz.

LBA. But the galactic synchrotron structure has almost disappeared from the snapshot images. Integrated sky maps, such as Figure A.4 in Appendix A, shows visible galactic structure. At higher frequencies longer integration are required.

Calibration of the HBA is difficult, especially at the high end of the band, because the source flux decreases at high frequencies and the weak sensitivity of the single elements. Imaging also poses issues because of the small PSF dynamic range, RFI sources appearing at the North Celestial Pole (NCP), and strong sidelobes for the sun. Even though the PSF is much smoother in 'all-sky' mode (Figure 5.4b compared to Figure 5.5d), the dynamic range is small. The noise floor in Figure 5.10a is dominated by the sidelobes of the class A sources. Deconvolution is required to properly remove these sources. In the LBA band the sun has a comparable flux to the other bright sources, but in the HBA band the source fluxes have rolled off much faster then the sun. When the sun is visible from the array it is the dominant source which makes calibration difficult. Images are dominated by the sun's sidelobes, Figure 5.10b.

5.5.1.2 RFI Environment

The main sources of RFI at Chilbolton Observatory are local terrestrial transmitters, apparent as sources on the horizon, and intermittent transmitters from aerial sources. These intermittent transmitters tend to be much brighter then any astronomical source and vary with location. Aeronautical transmitters are narrow band (sub-195 kHz) and are localized to a unique allocated frequency spectrum. Thus, that portion of the spectrum is not used in imaging. Aeronautical radar is broadband but rare. In the case of broadband radar the entire integration is dropped from the map making process. Terrestrial transmitters remain active for long periods of time if not continuously. They appear as bright sources on the horizon, which do not change position with LST. Figure 5.11 shows plots of terrestrial RFI sources in some of the 'clean' subbands which are used for imaging. These sources vary in strength and location. The sources also appear larger at lower frequencies due to the PSF size. Most subbands have low level terrestrial RFI sources at the horizon. In the map making process the image is masked below a set altitude (dependent on frequency) before being combined in the image plane.

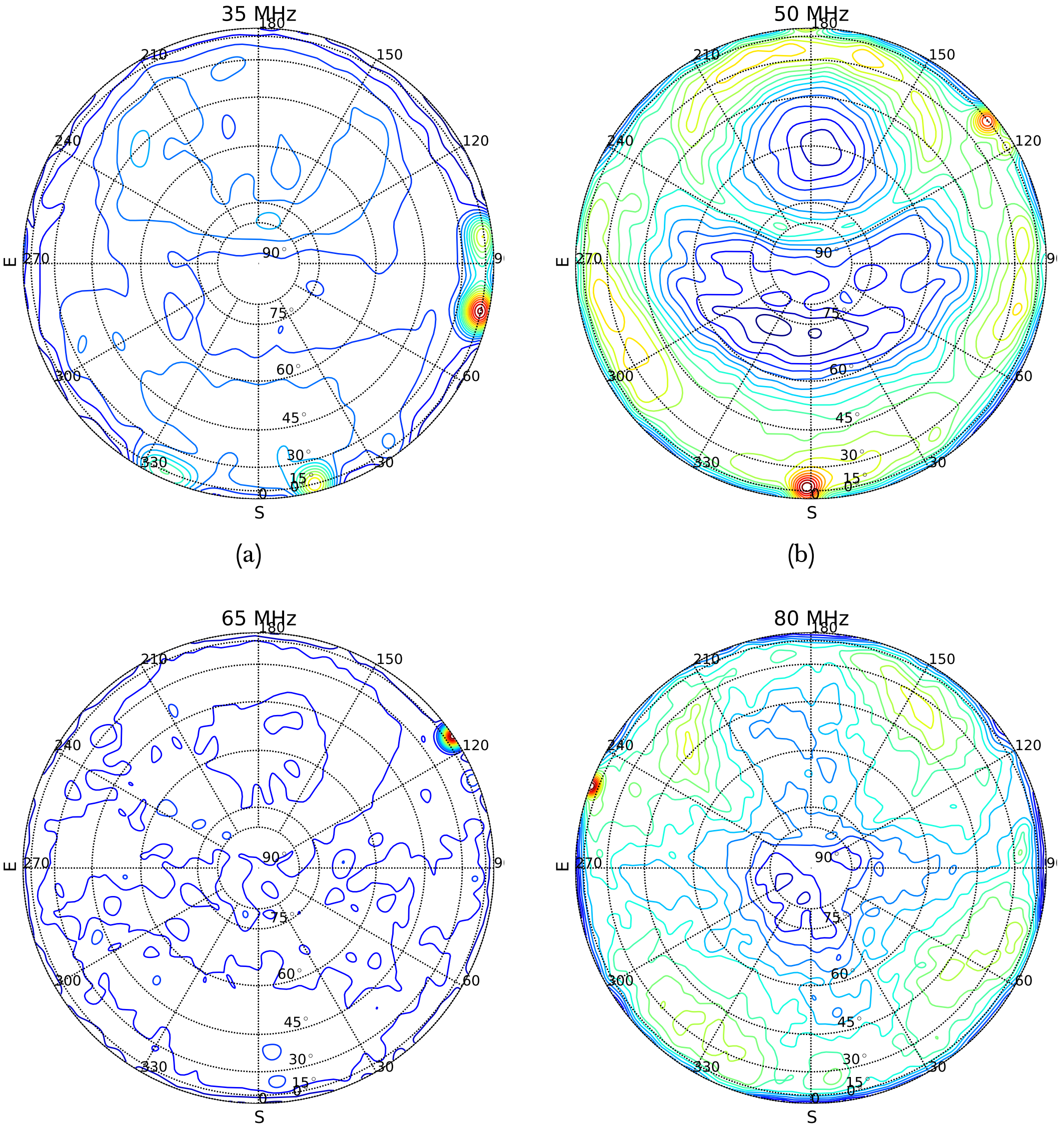
5.5.2 Combining Observations in the Image Plane

To create a complete sky map, 24 hours of observations are used, this amounts to 160 snapshot images which cover the complete LST range. As noted above, a constant complex antenna gain term is applied to each antenna in each observation. A set of images are produced for each observation: calibrated dirty, cleaned, residual, and model images. The calibration and imaging stages are run as automated scripts using the MeqTrees pipelining interface and Python scripts.

Due to the coordinate distortion, terrestrial RFI, and the beam response at low altitudes a mask is applied to images in regions below 20° altitude. The produced sky maps are limited to a minimum declination of $\delta = -20^\circ$.

Snapshot images are converted to HEALPix format maps. HEALPix is a mapping scheme which pixelizes data on a sphere such that all pixels cover equal area [Górski et al., 2005]. This format was original designed for CMB experiments where large portions of the celestial sphere are mapped. This in an ideal encoding for widefield images.

Once each snapshot has been converted to a HEALPix map it is trivial to combine them into a single map, Figure 5.12. In this map, Cygnus A and Cassiopeia A models have been removed though the large beam size makes it difficult to separate the point source from the galactic emission and residual flux remains. Also visible is Taurus A on edge on the map (both east and west) and Virgo A just above the North Galactic Spur. The dynamic range of the map is ~ 15, close to the simulated map developed in Section 5.5.3. Dynamic range is calculated as the ratio between the average flux of the galactic center to the average flux of a region away from the galaxy. Additional sky maps at a selection of subbands are presented in Appendix A. As we can see in these maps, the galaxy is the dominate feature at low frequencies. The galactic center never rises above 40° altitude in the beam yet is the brightest feature.



(a) (b)

(c) (d)

Figure 5.11: RFI environment at Chilbolton Observatory across the LBA band. Images formed from summing snapshot images across a 24-hour period. At 50 MHz, Figure 5.11b, the tracks of Cygnus A and Cassiopeia A are apparent.

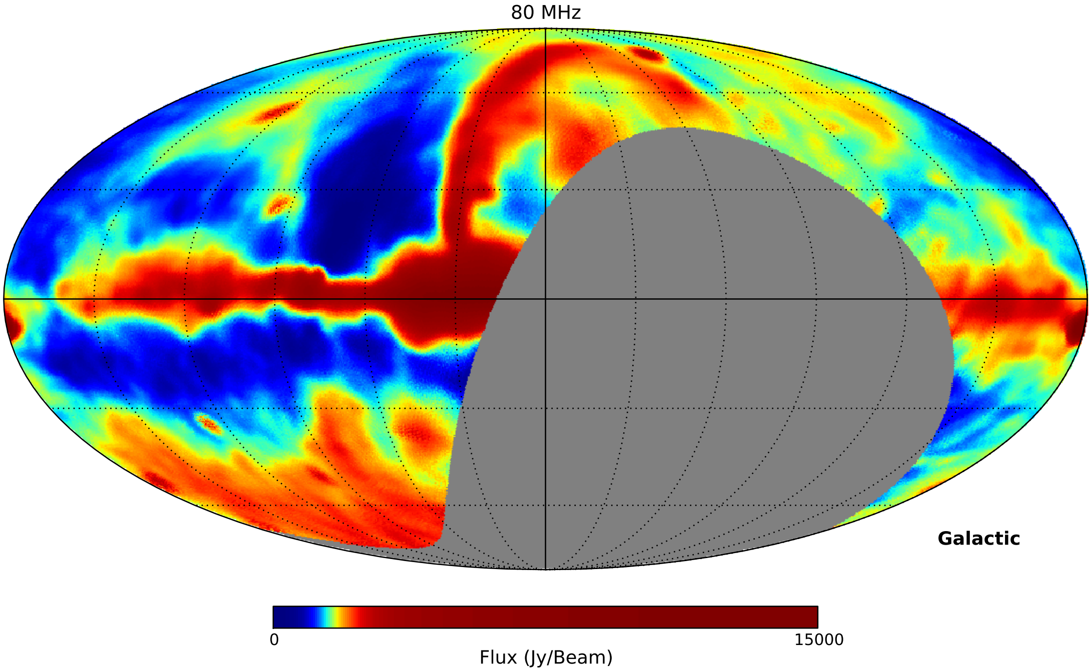


Figure 5.12: Sky map at 80 MHz produced from snapshot images taken over 24 hours using the LBA at Chilbolton Observatory. Models of Cygnus A and Cassiopeia A have been removed from the snapshot images. Taurus A, the next brightest source, is visible on the edge of the map at $l = 90^\circ$ The flux level has been clipped at 15 kJy, the galactic center and Taurus A are the only regions above this level. The map dynamic range is ~ 15. The flux values are based on the source models and should not be taken as an absolute flux.

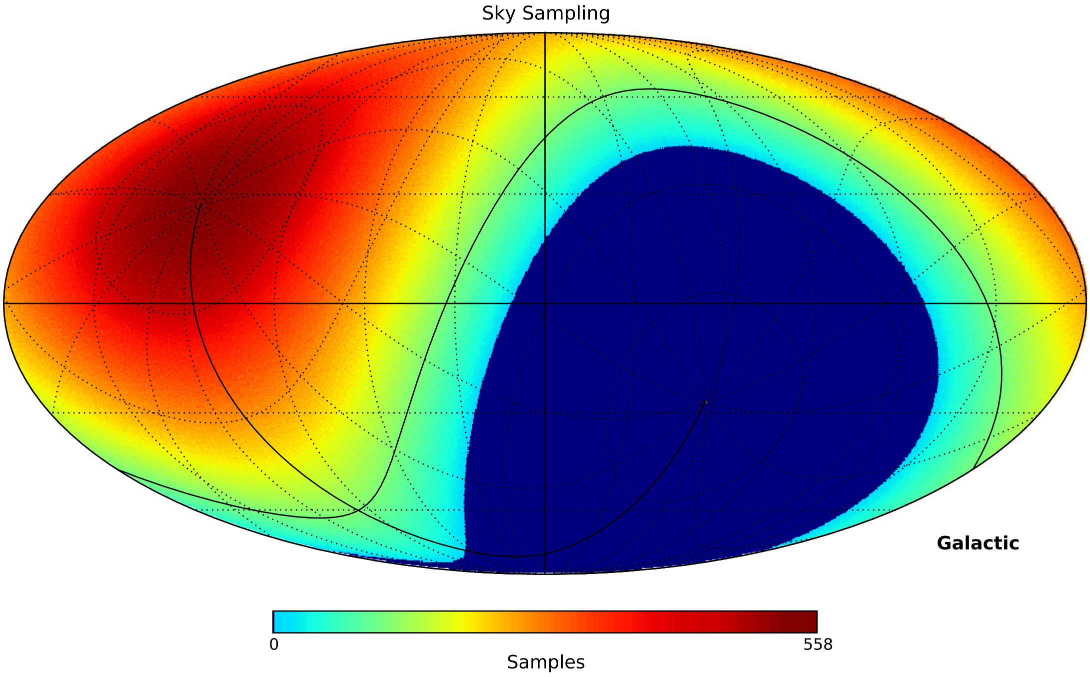
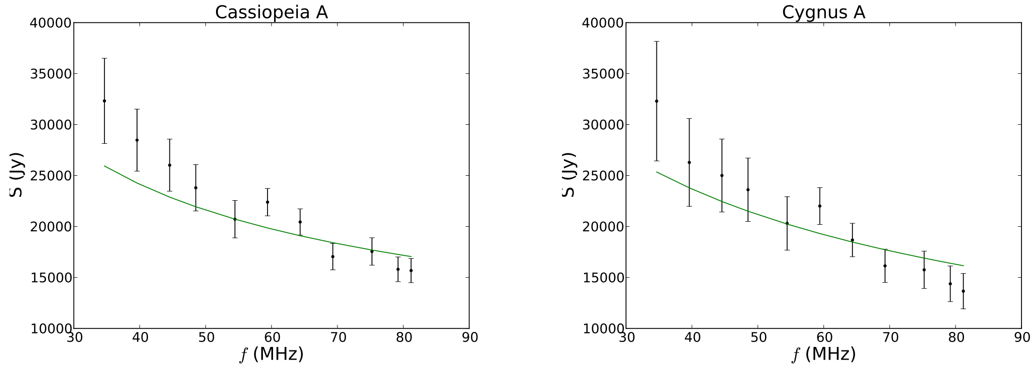


Figure 5.13: Number of samples per pixel based on the gridding of snapshot images to HEALPix maps. The signal to noise improves as declination increases. Galactic projection overlaid with equatorial coordinates.

From the latitude of Chilbolton Observatory declinations above $\delta = 51^\circ$ are always visible, and above $\delta = 70^\circ$ are always visible in the masked images. The number of samples per HEALPix pixel drops off as the declination decreases. When adding together individual HEALPix maps the sampling must be divided out of the final map to adjust each pixel to a normalized value. This average weighting means that the SNR increases as the declination increases, Figure 5.13.

The dynamic range of the sky maps can be improved by accounting for a beam model. When combining images on a HEALPix grid the beam is taken to be flat since a beam model has not been experimentally measured. Source near the edge of the beam will be attenuated, thus any flux scaling is an under estimate and varies with declination. An accurate beam model is essential to setting a stable fluxscale.

An absolute flux calibration is difficult to set as an accurate model of the specta of both Cassiopeia A and Cygnus A must be know. These are the dominate sources, Cassiopeia A never sets below the horizon and Cygnus A below the horizon for only a short period of time. The initial flux levels are set using those derived from [Baars et al., 1977]. From the sky model the flux can be extracted for different subbands. The resulting spectral evolution, Figure 5.14a and 5.14b, follow the predicted values, as



(a) Measured flux of Cassiopeia A (blue) extracted from sky maps across the LBA band. The expected flux (black solid line) derived for 2012 from [Baars et al., 1977].

(b) Measured flux of Cygnus A (blue) extracted from sky maps across the LBA band. The expected flux (black solid line) from [Baars et al., 1977].

Figure 5.14: Spectra of Cassiopeia A and Cygnus A.

they were used for the initial scaling. Though, both sources appear to have a slightly steeper spectral evolution. Constraining the source spectral indices will require further refinement of the calibration model and beam model.

Now that the process for producing sky maps has been constructed the next step will be to develop a galactic model for the LOFAR bands. This is an interactive process to refine the model to incorporate all the structure up to the map noise floor. A better model will in turn provide a better calibration solution. Let us first look at an existing global sky model and compare it to the observed maps.

5.5.3 Simulated Sky Map

Using a sky model of the galaxy and the array layout we can create a simulation of the expected sky to compare to our results. The OSKAR-2 simulator tool [Oxford Electronic Research Centre, 2012] developed by Oxford eResearch Centre (OeRC) was built to simulate beamforming and interferometry with large-N aperture arrays. A global sky model at 80 MHz was created using the *GSM* software presented in [de Oliveira-Costa et al., 2008]. The Haslam 408 MHz all-sky survey [Haslam et al., 1981], [Haslam et al., 1982] is the basis for the sky model. The observational metadata from the snapshot observations over 24 hours was used as the basis for the simulated measurements. Combining the model, the observational metadata and the simulator a set of simulated snapshot images are produced which can be converted to a sky map using the same process as presented above. No additional noise, phase errors or beam models were included in the simulation. The resulting map, Figure 5.15, is expectedly similar to the observed map, Figure 5.12. The simulated map dynamic range is ~ 17,

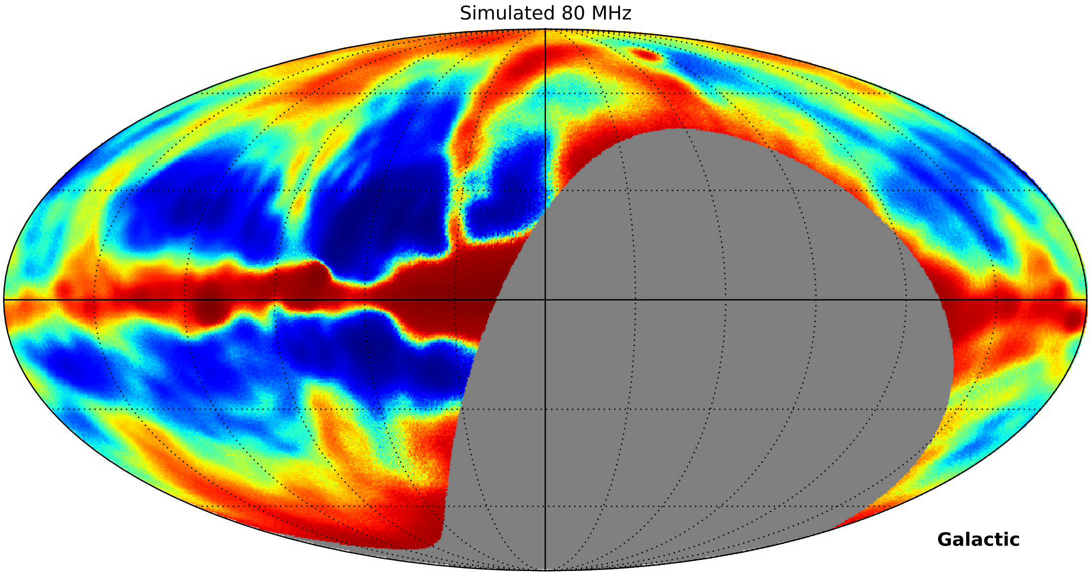
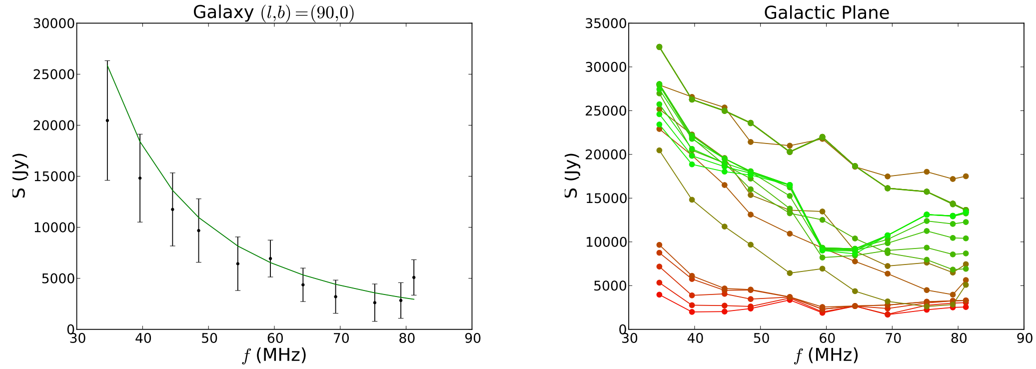


Figure 5.15: Simulated sky map at 80 MHz using the same observation metadata from the 24 hour observation used to generate Figure 5.12. The sky model is based on the Haslam 408 MHz map scaled down to 80 MHz using *GSM* [de Oliveira-Costa et al., 2008]. The resulting sky map has a dynamic range of ~ 17.

the observed dynamic range is ~ 15 which can be due to a number of effects including beam attenuation and signal noise. It is striking how close the maps are in dynamic range which suggests the observed sky map is very nearly confusion limited. Features in the observed map, not present in the simulated map, such as the North Galactic Spur is due to scaling error from high frequencies. Few sky surveys exist at low frequencies. A new sky model can be developed using these observed sky maps. This is discussed further in Chapter 6 in the context of modelling extended sources with shapelets.

5.6 Evolution of the Galactic Structure at Low Frequencies

Emission from the galaxy at low frequencies is dominated by synchrotron radiation which follows a power-law spectrum where the spectral index is $\beta = -2.55$ [Guzmán et al., 2011]. Synchrotron emission is generated when relativistic electrons interact with magnetic fields. The power-law spectrum is due to the distribution of electron velocities. At low frequencies there is a turnover in the power-law spectrum because of synchrotron self-absorption where the plasma becomes opaque to its own radiation.



(a) Measured flux from $l = 90°$ in the galactic plane (blue). The expected galactic synchrotron spectral evolution, $\propto \lambda^{2.55}$ is plotted in black.

(b) Spectral evolution along the galactic plane, $l = 180°$ to $l = 15°$ in 5° increments is colored red to green. The galactic positions with the highest flux correspond to regions near Cassiopeia A and Cygnus A.

Figure 5.16: Flux within the galactic plane across the LBA band.

The radiation produced from synchrotron interactions is highly polarized and thus is an excellent tracer of the magnetic field lines within the galaxy.

Synchrotron flux is proportional to the plasma density, flux should be higher within the galactic plane as opposed to outside the plane and should increase in the galactic plane as $l \rightarrow 0$. Figure 5.16a shows the measured flux level at $l = 90°$ in the galactic plane along with the expected $\lambda^{2.55}$ evolution curve. Sampled positions along the galactic plane show the evolution of the synchrotron emissions at different plasma densities, Figure 5.16. At the galactic anti-center(red lines) there is little flux and the spectral evolution flattens at higher frequencies when the flux is comparable to the noise floor. As $l \rightarrow 0$ the flux increases and a synchrotron evolution curve is present across the band. Two of the sampled galactic plane positions have a higher flux and are out of the expected order of the flux increasing as $l \rightarrow 0$ since these positions are picking up additional flux from nearby Cassiopeia A and Cygnus A.

Sky surveys which cover large extended structure are usually presented in terms of a spectral brightness temperature. Flux density, in units of Janskys, is used for discrete sources. Where as spectral brightness or specific intensity is a measure of surface brightness subtended over an area of the celestial sphere in units of temperature. Though the maps presented here are in units of Janskys it is more accurate to say the maps are in units Janskys per beam, since the main source of flux is the extended galactic emission. Here the beam size is taken to be the FWHM of the array PSF. The PSF size is frequency dependent and subtends a larger area at lower frequencies. To convert a measurement in Janskys per beam to brightness temperature we start with the flux density per beam equation

$$S_\nu = \int_{\Omega_b} I_\nu(\theta, \phi) \cos\theta d\Omega \tag{5.16}$$

Where Ω_b is the solid angle subtended by the beam and I_ν is the spectral brightness. For long wavelength radio, the approximation, $\frac{h\nu}{kT} \ll 1$, can be made such that Planck's law for black-body radiation reduces to the Rayleigh-Jeans law

$$I_\nu = B_\nu \approx \frac{2kT_b\nu^2}{c^2} \tag{5.17}$$

Since θ_b, the angular scale of the beam, is small the approximations $\sin\theta \approx \theta$ and $\cos\theta \approx 1$ can be made. In spherical coordinates Eq. 5.16 reduces to

$$S_\nu = \int_0^{2\pi} \int_0^{\theta_b} I_\nu \cos\theta \sin\theta d\theta d\phi \approx \int_0^{2\pi} \int_0^{\theta_b} I_\nu \theta d\theta d\phi = \pi I_\nu \theta_b^2 \tag{5.18}$$

Combining Eq. 5.18 with Eq. 5.17 the low frequency observations with beams on order of a few degrees in scale the conversion between flux density per beam and brightness temperature is

$$S_\nu = \frac{2\pi k T_b \nu^2}{c^2} \theta_b^2 \tag{5.19a}$$

$$T_b = \frac{S_\nu c^2}{2\pi k \nu^2 \theta_b^2} \tag{5.19b}$$

From these equations we can see where the spectral index relation $\beta = 2 + \alpha$ between temperature and intensity comes from, β is the temperature spectral index and α the flux density spectral index.

Now converting the sky maps into brightness temperature we can make an comparison between the observations and historical surveys. The 85 MHz survey by [Landecker & Wielebinski, 1970] covers declinations $\delta = 25°$ to $\delta = -25°$. Using the 82 MHz map and accounting for the power-law change between the two map frequencies, the measured temperature from the LBA observations is within 15% of the temperature values measured by [Landecker & Wielebinski, 1970] for galactic longitudes $60° > l > 15°$. On the low end of the band the temperature values of the galactic plane between $90° > l > 15°$ in the LBA 45 MHz sky map are within 10% compared to that of the 45 MHz sky survey from [Guzmán et al., 2011]. Both these surveys indicate that the calibration and fluxscale applied to the LBA observations are in agreement with [Landecker & Wielebinski, 1970] and [Guzmán et al., 2011].

5.7 Mapping Beam Patterns with Bright Sources

A generic model of the element beam patterns can be developed from observations of the bright class A radio sources as they transit the array. The class A radio sources are stable, far field point sources. Variation in their apparent flux can be attributed to amplitude modulation from the element beam patterns. The path of each source through the sky makes unique cuts through the element beams.

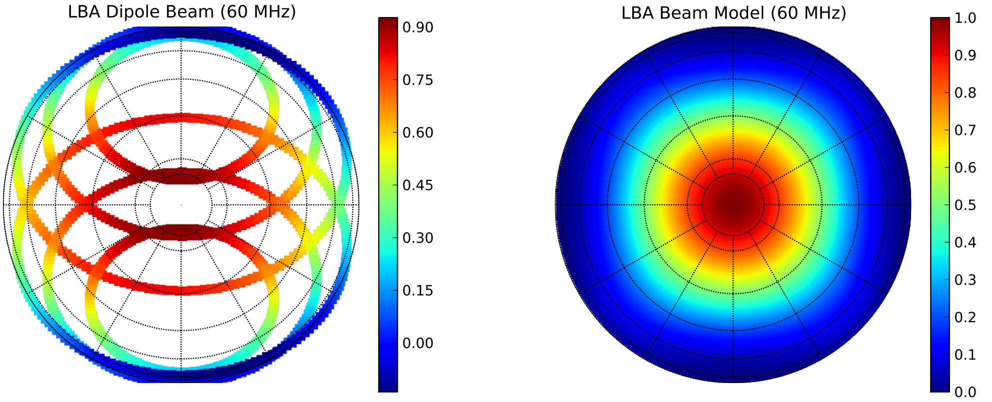
Source flux is extracted from each snapshot image in a beam-sized region around the source center. The flux values are normalized based on the expected source flux at the observing frequencies so that the fluxes become a measure of the beam attenuation. In this normalization a source at zenith will have unity flux. A polynomial is then fit to each track based on the assumption of a smooth beam.

We can also make an assumption about beam symmetry to introduce a further calibration. The beam should be symmetric about its east-west axis, so any source track can be reflected about that axis. This introduces points in the beam where two sources cross. At those points the beam gain is independent of source. Thus a correction factor can be applied to each flux track based in assuring that the fluxes at the point in the beam are equal. The resulting measurements, Figure 5.17a, shows a flat response across much of the beam and even has a strong response close to the horizon.

A $\cos^2\theta$-based analytical function is fit to the source tracks using a least-squares method, Figure 5.17b. The resulting beam model fits the measured flux samples with a standard deviation of $\sigma = 0.052$. Noticeably, the beam model looks square which is due to the source track sampling of the beam where the beam is well sampled in the east-west direction but has gaps in the north-south direction. The derived beam pattern is a generic one which is an average of all LBA dipole element patterns. This is sufficient since the beam pattern of each element is expected to be nearly the same.

5.8 Discussion

This chapter has introduced a process for converting LOFAR station correlations into calibrated maps covering the viewable sky from the station. The science topics discussed: galactic modelling, spectral index measurements, and beam modelling are just the beginning of what we can possibly learn from this widefield, low frequency array data. The spectral index map is an ideal experiment to be completed with the LOFAR stations with further observations. And, it has the added bonus of not requiring an absolute calibration since the spectral index is a relative measurement. An elephant in the room, as always, is polarization. I have avoided the topic by using only I Stokes images, though calibration is performed on each polarization. Understanding the sky



(a) Tracks of Cassiopeia A, Cygnus A and Taurus A across the element beam with fluxes normalized. A polynomial has been fit to each track and the resulting flux value plotted. Assuming a north-south beam symmetry allows flux corrections based on source crossing points.

(b) Beam model of a generic LBA element at 60 MHz based on least-squares fit to normalized source flux tracks using a $\cos^2 \theta$ function.

Figure 5.17: LBA beam model fitting based on class A radio source tracks.

polarization along with time dependent polarization induced by dipole projection effects is a topic left to future work. The HBA has only been briefly touched upon in this chapter, this array has its own calibration and imaging challenges along with its own rewards including EoR related science. Development of a galactic and extragalactic model, which includes polarization, across the LOFAR bands will be essential to foreground removal in high redshift hydrogen EoR projects. Additionally, when the SEPcam correlator becomes available for observation the increase in time resolution and bandwidth will further open up the scientific possibilities.

The generation of calibrated sky maps from correlation matrices has been completely automated using a suite of common radio astronomy software and custom scripts. This is not to say the process is finished, each stage can be further improved which will require continued iterations on the current process. Deriving calibration solutions currently requires human interaction to verify a good calibration has been determined. But, this could also be automated with a variety of methods: machine learning, statistical analysis, or simulated annealing. This is a relatively easy problem to solve as we gather more data and improve calibration.

Calibration is, as stated in the chapter, applied on 24 hour timescales. An open question currently is if this is a good timescale? Would the image quality be greatly improved if calibration solutions was derived on one hour timescales? Or, do we only need to update the solutions once a week? This question is best answered by further observations covering multiple epochs and looking at how the solutions change on hour, day, week, month timescales.

The measurement equation I have developed for the station works well to produce stable calibration solutions across multiple frequencies. As more observational data is made available the measurement equation can be improved. Though a half-wavelength dipole model is used across all frequency bands this is not the exact model. As more source tracks are observed across many epochs the actual beam model can be determined. The beam model problem is very connected to the sky model. A simplistic sky model using the class A sources has been used, their flux based on previously reported values. As has been shown in the sky maps presented, the galaxy is a complex structure with comparable flux at low frequencies. The galaxy should be included in the sky model for calibration. A beam and sky model are further tied together in setting an absolute fluxscale using physical units. An absolute fluxscale will come about from a number of different experiments to isolate time variability, measure system linearity response, and check hardware stability. These are still early days for LOFAR and setting a fluxscale will surely come about with time.

To generate the sky maps only dirty images have been used. There are a few reasons for this. First, the snapshot PSF response of the LBA and 'all-sky' HBA is relatively smooth with no significant sidelobes. The images are not dominated by sidelobes of the bright sources, in some cases the sources have been removed in the uv plane which will suppress the sidelobes. Secondly, broad, extended structure is difficult to clean and the residuals are dominated by extended structure which has not been incorporated into the clean model. Once a galactic model has been developed cleaning images should result in much lower residuals.

Adding correlations in the image plane has the advantage of not introducing smearing or w-projection distortions. This is the ideal method to combine LOFAR station observations. But, adding correlations in the image plane requires knowledge about the element beam patterns. The beam pattern introduces time dependent amplitude modulation on sources in the image plane. Before adding together images, each image should have an inverted beam model applied which has the effect of correcting source fluxes on the edge of the beam. Currently, this is not done and is evident in the reduced apparent flux of the galactic center compared to the expected flux.

With all these improvements possible in calibration, imaging, beam models, and sky models it would seem that none of this should actually work. But, it does and that shows that all the approximations and models being used are valid and good enough to get started with. The next step is to adjust and experiment with the process, and check for repeatability of the experiment. One of the end goals of developing this calibration and imaging process is to make the system fully automated to produce flux and phase calibrated sky maps. The SEPcam correlator is only the first stage of a system to produce images of the sky as it is observed and generate deep maps of the sky at low frequencies. We now have a foothold at the base of our Eiger, we only need to take the next step.

Chapter 6

Shapelets for Modelling Extended Sources and Beam Patterns

> *The two-step program to be a contented radio astronomer: (i) stick to X-band (ii) don't spatially resolve anything*
>
> -Ian Heywood, 2012

Images and two-dimensional datasets can be decomposed into a set of functions using orthogonal basis sets. A number of basis functions (wavelets, Fourier modes) have been used for image decomposition covering a number of applications. In [Refregier, 2003] a new orthogonal basis set using Hermite and Laguerre polynomials modulated with Gaussian functions which are termed *shapelets* is presented. The shapelet basis set is similar to the commonly used wavelet basis set. Wavelets can be thought of as decomposing images using different spatial frequency components. Where as shapelets decompose images into basic shapes. This proves to be a useful decomposition for astronomical objects which are compact, smooth and do not have sharp edges. Shapelets were originally used to model the extended structures of galaxies in optical images along with the imaging PSF. The shapelet basis sets is particularly useful for astronomical image reconstruction. They can be extended to source modelling in interferometric imaging [Chang & Refregier, 2002].

Image decomposition is useful for compression and modelling. For example, wavelets are often used in lossy compression algorithms such as the common JPEG image format. Indeed [Refregier, 2003] uses this method to reduce the database size in the Hubble Deep Field analysis. More importantly though, shapelets can be used to represent extended sources with a set of smooth functions. This allows for easy rescaling, rotation and projection. Beam pattern can also be represented using shapelets to avoid the issue of interpolation of a pixelized beam pattern. The shapelet basis set has the useful property that it is its own Fourier transform with a scale factor. This property is used to quickly transform a source from image space to uv space.

Improvements in calibration using Third Generation Calibration (3GC) techniques require extensive knowledge of element and array beam patterns. Analytical forms of the beam patterns, either derived from models or by observations will improve the computation time for evaluating the measurement equation. A shapelet-based analytical form can easily be derived for complex beam patterns.

As the resolution and sensitivity of large-N arrays grow more sources will become visible and many will be at least partially resolved. Current CLEAN-based methods rely on sources being point source like and under represent the flux of extend sources. Shapelets can be used to represent extended sources in simulation and calibration. A python-based implementation of the shapelet modelling and fitting code is currently hosted on Github[1]. This chapter presents a stand-alone software package and add-on module to the MeqTrees environment which adds shapelet functionality for interferometric calibration and imaging.

6.1 A Mathematical Model for Modelling Extended Objects

An orthogonal basis set is made up of normalized η functions which have the property

$$\int_{-\infty}^{\infty} \eta_i(x)\eta_j(x)dx = \delta_{ij} \tag{6.1}$$

Where δ_{ij} is the Kronecker delta. The product of two orthogonal basis sets is itself an orthogonal basis set. Shapelets are constructed out of Gaussian functions and a shape basis set. The Gaussian function it to taper off the source intensity and account for the ideal PSF. The shape basis set is usually Hermite polynomials in the linear form or Laguerre polynomials in the polar form. The choice of basis sets is to take advantage of their Fourier transform properties and the quantum harmonic oscillator framework which is ubiquitous in physics. But, any orthogonal basis set could be used in practice.

Each dimension of the shapelet, in this case x and y, are treated independently. Thus, multi-dimensional shapelets are products of the 1-D case. Any 1-D profile, $f(x)$, can be expressed using an infinite set of shapelet basis functions, Eq. 6.2. Where $B_n(x;\beta)$ is the n^{th} shapelet basis function and f_n is the coefficient.

---

[1] https://github.com/griffinfoster/shapelets

$$f(x) = \sum_{n=0}^{\infty} f_n B_n(x;\beta) \tag{6.2}$$

An approximation to $f(x)$ can be made to within a desired error using a subset of values from n. The maximum order of n, n_{max}, determines the overall compression and loss of higher order structure. To obtain a optimized solution to a shapelet decomposition of an image the character scale β, maximum order of decomposition n_{max} and the centroid of the image x_c parameters are optimized. The scale factor is related to the minimum, and maximum, spatial resolution. This is an important detail, when there is a large range in spatial detail in a field multiple shapelet decompositions may be necessary. The range between the minimum and maximum spatial resolution can be extended by increasing n_{max} which comes at the cost of computation time. An optimal centroid position is a necessary for a good decomposition using a minimal order of decompositions. Shapelets are variations on 2-D Gaussian functions, the best decompositions will be derived when the zero position is placed on the center of a Gaussian like source. Figure 6.1 shows a Hermite polynomial-based shapelet decomposition of a twin-lobed radio galaxy. The decomposition does reasonably well with $n = [15, 15]$ coefficients, the model error is $\lesssim 1\%$ to the original source. Compression is limited (15% of original image) because the source is made up to two Gaussian-like lobes.

6.1.1 Hermite Polynomials

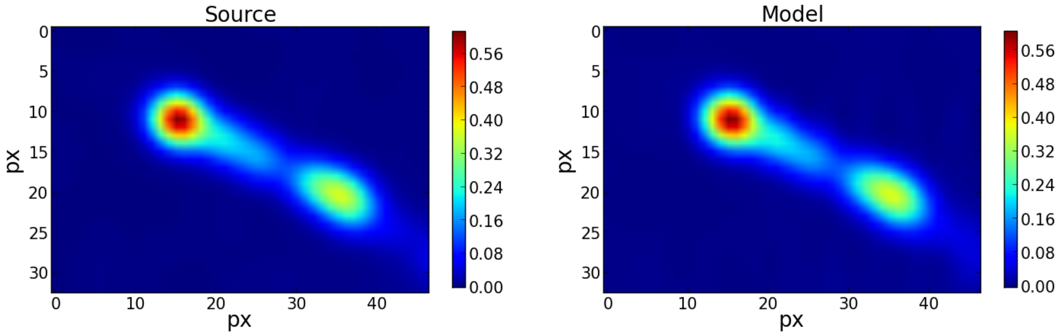
The Cartesian implementation of the shapelet basis set uses Hermite polynomials. The dimensional shapelet basis functions using Hermite polynomials is defined as

$$B_n(x;\beta) \equiv \beta^{-\frac{1}{2}} \left[2^n \pi^{\frac{1}{2}} n!\right]^{-\frac{1}{2}} H_n(\beta^{-1}x) e^{-\frac{x^2}{2\beta^2}} \tag{6.3}$$

Where β is the dimensional scale factor relating to angular scale on the sky. H_n is the n^{th} Hermite polynomial. The Hermite polynomial series is a set of orthogonal functions originally developed for probability, they are also used to represent the quantum harmonic oscillator states. This 1-D basis function is the product of a Gaussian and Hermite polynomial with a normalization factor. The 2-D basis function is

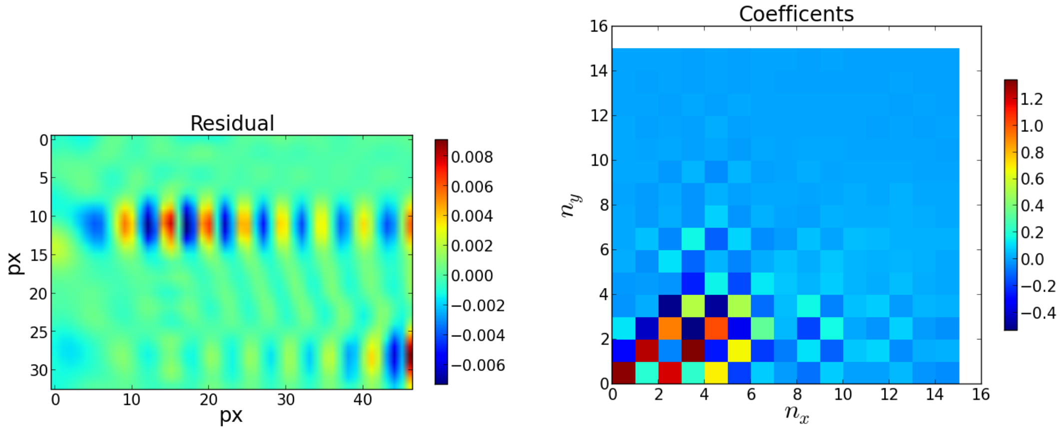
$$B_n(x, y; \beta_x, \beta_y) = B_n(x;\beta_x) \cdot B_n(y;\beta_y) \tag{6.4}$$

The first few 2-dimensional basis functions are shown in Figure 6.2. The grid like structure of the Cartesian formalism is apparent. Point sources embedded with an extended source are represented with the higher order functions.



(a) Source image

(b) Hermite shapelet model



(c) Residual (Source-Model)

(d) Model coefficient matrix

Figure 6.1: Shapelet model for a twin-lobed radio galaxy. A model (6.1b) is fit to the source image (6.1a) using a least-square minimization function ($\beta = [3.55, 4.43], x_c = [17.22, 26.27]$). The residual between the source and the model (6.1c) shows that the model fits to within $\sim 1\%$ of the source. The model is a $n = [15, 15]$ Hermite shapelet model (6.1d).

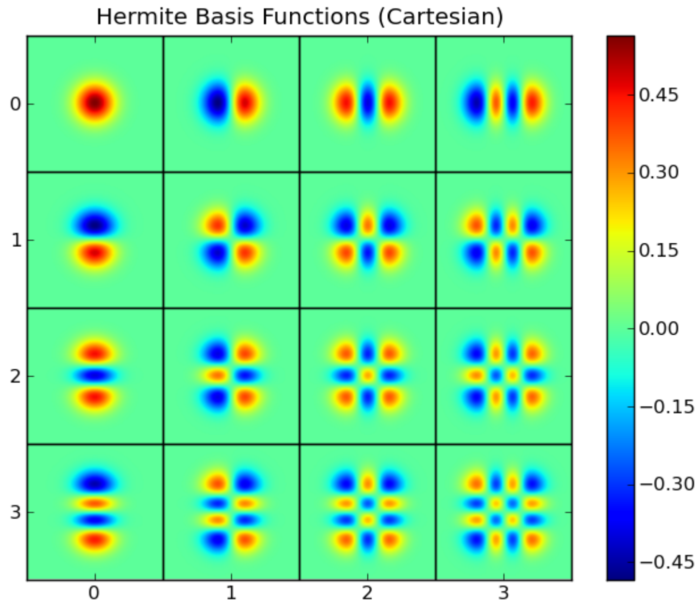


Figure 6.2: 2-dimensional Hermite (Cartesian) basis functions up to $n = 3$.

6.1.2 Laguerre Polynomials

A polar variant of shapelets uses Laguerre polynomials. The basis functions, $\chi_{n,m}(r, \theta; \beta)$, are intrinsically 2-dimensional. This formalization is more intuitive in terms of angular size and rotation. Since many astronomical sources have spherical symmetry the decomposition into polar forms is easier to understand. The Cartesian and polar forms are equivalent, a transformation can be applied to to convert from one to the other. The polar form tends to be easier to understand, the Cartesian form proves to derive better solutions and is usually the preferred basis set.

A 2-dimensional image in polar coordinates can be expressed using an infinite polar basis set, similar to Equation 6.2, as

$$f(r,\theta) = \sum_{n=0}^{\infty} \sum_{n}^{m=-n} f_{n,m}\chi_{n,m}(r,\theta;\beta) \tag{6.5}$$

Where $f_{n,m}$ is the coefficient for the (n, m) order shapelet. The basis functions are functions of (n, m) order Laguerre polynomials, Eq. 6.6. Unlike the Cartesian form the polar basis set is complex and only states where n and m are both even or odd are used.

$$\chi_{n,m}(r,\theta;\beta) \equiv \frac{-1^{(n-|m|)/2}}{\beta^{|m|+1}} \left[\frac{((n-|m|)/2)!}{\pi((n+|m|)/2)!}\right]^{1/2} L^{|m|}_{(n-|m|)/2} \frac{r^{2+|m|}}{\beta^2} e^{-r^2/2\beta^2} e^{-im\theta} \tag{6.6}$$

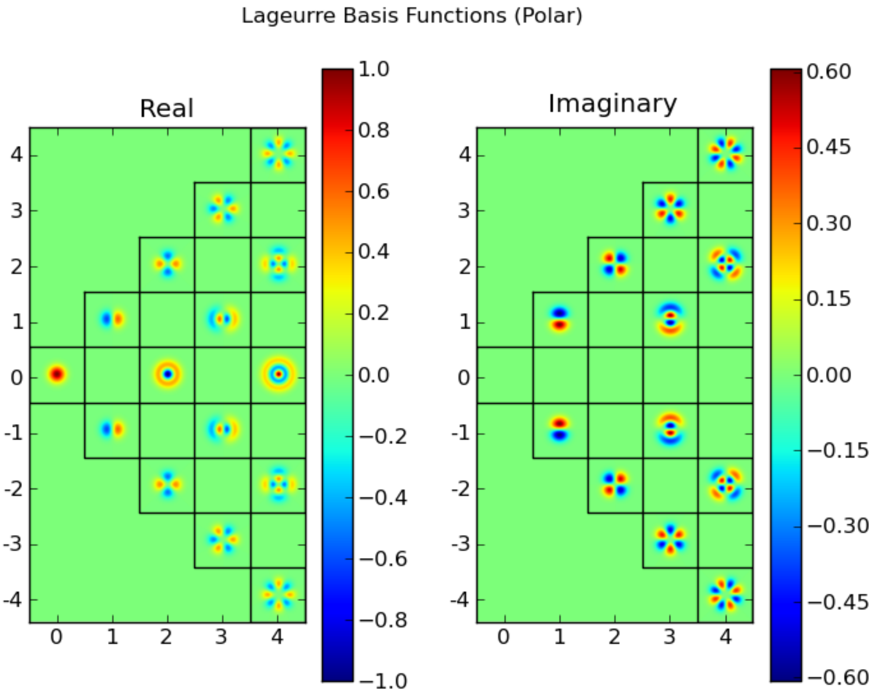


Figure 6.3: Polar shapelet basis functions up to $n = 4$ (x-axis), m (y-axis) ranges from $-n \leq m \leq n$. $m = 0$ values are real as shown in Eq. 6.6. Only functions where n and m are both either even or odd are valid.

As in the Cartesian form β is the scale factor, but only a single scale factor is used. The form of Equation 6.6 is the product of a Laguerre polynomial and a Gaussian with a normalization factor. The valid polar basis functions (up to $n = 4$) are shown in Figure 6.3.

6.1.3 Fourier Transform

Hermite polynomials are eigenfunctions of the Fourier transform and the Fourier transform of a Gaussian is another Gaussian. The Fourier transform of the basis function $B_n(x;\beta)$ from Equation 6.3 is

$$\mathcal{F}(B_n(k;\beta)) = i^n B_n(k;\beta^{-1}) \tag{6.7}$$

Up to a phase factor and a change in scale factor the Cartesian basis function is a Fourier transform of itself. Since a transformation can be made between the 2-dimensional Cartesian shapelet basis functions and polar shapelet basis functions then the Fourier transform of the polar form must also follow Equation 6.7.

A major computational cost of modelling extended sources for simulation or calibration is transforming to the uv plane from the image plane. By representing a source as a set of shapelet functions the transform only requires a evaluation of the shapelet functions at the uv positions. This is an $O(N_s M_{uv})$ operation where N_s is the number of shapelet basis functions used in modelling the source and M_{uv} is the number of pixels in uv space.

6.1.4 Lossy Data Compression

Modelling sources as shapelets can be used to compress the information of an image. This can be used to save storage space and improve access speed. This is particularly useful when simulating a field with many resolved or partially resolved sources. Partially resolved sources can be modelled with a few coefficients. This is a lossy process which introduces errors between source and model.

Choosing a good n_{max} has a significant effect on the model, as seen in Figure 6.4. The lowest order n values ($1 \leq n \leq 6$) are incomplete models and introduce ringing effects. By $n = 12$ the ringing has dampened and the choice of n becomes a choice of the model error. The data compression ratio scales as $O(n^2)$. With $n = 15$ the average error per pixel is $< 1\%$. There is a balance between data compression and expectable information loss.

6.2 Extended Sources

Resolved and extended sources are the bane of radio interferometry calibration. They are both the science goal and calibration impediment. When a source is unresolved all baselines observe the same level of flux. As the baselines become longer and the source becomes resolved the apparent flux drops off as those longer baselines are measuring a subset of the original unresolved source flux, Figure 6.5.

A source becomes resolved on a baseline when its angular scale $\theta_{src} \gtrsim \frac{\lambda_0}{D}$, λ_0 is the observing wavelength and D is the projected baseline length. A source falls into three regimes: unresolved, partially, and fully resolved.

Unresolved ($\theta_{src} < \frac{\lambda_0}{D}$): the source is point source like on all baselines. The flux is constant across baseline length.

Partially Resolved ($\theta_{src} \sim \frac{\lambda_0}{D}$): the source is only resolved on the longer baselines, and unresolved on the shorter baselines. The flux is constant across the shorter baselines but decreases on the longer ones.

Resolved ($\theta_{src} > \frac{\lambda_0}{D}$): the source is resolved on all baselines. There is a slope in flux across all baselines.

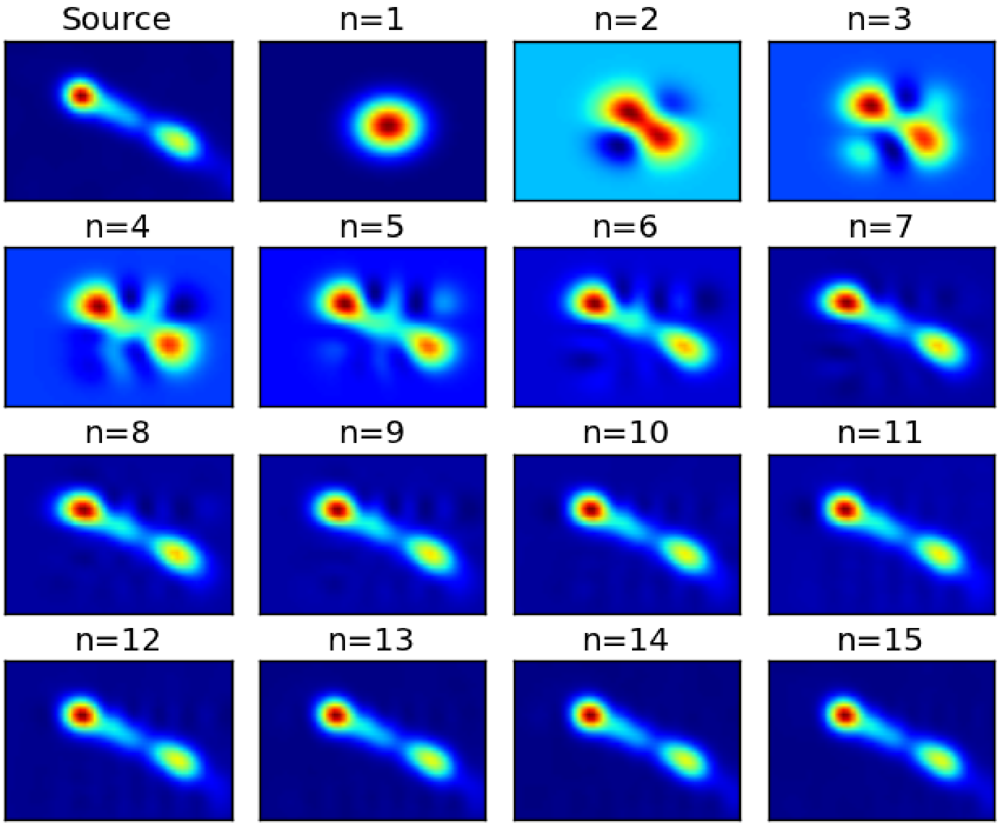


Figure 6.4: Hermite shapelet model using $1 \leq n \leq 15$ for the x and y dimensions, $n_x = n_y$ in these models. The original image is shown in the upper left corner. In the low order n plots the Hermite basis functions from Figure 6.2 can be seen.

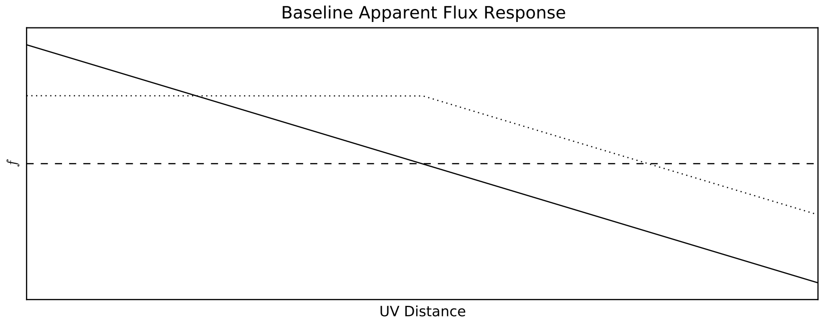


Figure 6.5: Apparent flux as a function of uv distance. An unresolved source (dashed) will have the same flux response on all baselines. A partially resolved source (dotted) will have a flat response on shorter baselines but longer baselines will resolve certain spatial components. Each baseline will measure different frequency components of a fully resolved source (solid).

The last regime, fully resolved, means a source in resolved in terms of the observing array and each baseline measures different spatial components. A source can always be further resolved with higher resolution arrays. What defines a source becomes a gray area.

Resolved and partially resolved sources are notoriously difficult to use as calibrators. The change in flux with baseline length means each baseline can not be treated the same when deriving complex gains. This is a major challenge in VLBI where most sources, besides pulsars, are at least partially resolved.

When imaging, the sky model is deconvolved from the image to get to the residual and lower level structure. Implementing a sky model which takes into account extended sources has proven a difficult challenge. Typically an extended source is modelled as a collection of point sources which, when convolved with the array PSF, produces a facsimile source in the image plane. Though in the image plane the model represents the source it has short comings in the uv plane which generally leads to an excess of residual flux.

Complex source ‘postage stamp’ images are used which can be properly transformed to the uv plane. These postage stamp images are images of a subset of the field which only contains the source of interest. Though postage stamp images have proven extremely effective in improving the dynamic range of CLEAN’d images, they too have limitations. Rotation and scaling require interpolation operations which introduce artifacts. These transformations are trivial using shapelets and does not introduce artifacts. Postage stamp image must be high enough resolution to model the source. This is fine for a small number for sources. But, large field of view, high resolution images which are being produced with new arrays there are many extended sources. Using analytical models, such as shapelets, greatly reduces the data footprint for individual sources.

6.2.1 Image Decomposition

An image I with n_{pix} pixels can be decomposed into k basis functions using a least square estimation [Lupton, 1993] (Chapter 11). The set of coefficients, $\hat{f}$, which minimize the least square estimation is determined with Equation 6.8.

$$\hat{f} = (M^T M)^{-1} M^T I \tag{6.8}$$

M is an $n_{pix} \times k$ matrix which is made up of k basis functions evaluated at n_{pix} pixels. The I matrix is the image flattened into a $n_{pix} \times 1$ matrix. The resulting solution, $\hat{f}$, is a $k \times 1$ matrix. The challenge of solving Eq. 6.8 is the $(M^T M)^{-1}$ matrix inversion. A number of scientific programming packages include optimized inversion functions.

Decomposition also relies on a good choice for the scale factor (β) and centroid position (x_c). These values are determined by either a priori knowledge about the source, a scalar minimizer across a range of parameters, and when needed, user input. The minimum scale factor will be at least the angular size of a pixel (extracted from FITS headers or with user input), but will likely vary by at least a few pixels in size. We also have foreknowledge that only a single source, or group of closely packed sources will be decomposed. So most of the postage stamp image will contain the source. An initial guess of $\beta_0 = \frac{1}{5}\theta_{img}$, where θ_{img} is the angular size of the postage stamp, has been found to produce reliable results. The initial guess to the centroid position is chosen to be the weighted flux mean, $x_{c,0} = \sum_{i=0}^{n} w_i x_i / \sum_{i=0}^{n} w_i$ where w_i is the flux value at the i^{th} pixel, of the image.

For the fits to β and x_c the value of n_{max} is held constant, Equation 6.9. This value is chosen based on the overall complexity of the source. If the source is only partially resolved and very Gaussian then n_{max} can be small ($\sim$ 3), this value will only grow with source complexity. Ideally we would just choose n_{max} to be very large, but the decomposition scales as $O(n_{max}^2)$. This initial choice of n_{max} will not necessarily be the final order of basis functions. This is only for fixing β and x_c to a good choice. Once they are picked a n_{max} fit will be performed.

$$\chi^2(\beta, x_c; n_{max}) = \sum_{n_{pix}} \frac{(I_M(x_p, \beta, x_c; n_{max}) - I(x_p))^2}{\sigma_p^2} \tag{6.9}$$

The centroid position and scale factor are fit in parallel, this is implemented using a switching downhill simplex method available in SciPy[2]. This function uses a random

---

[2] http://www.scipy.org/

number seed in the fit and is thus non-deterministic. Multiple runs of the minimizer will produce different solutions but will generally converge to a common result. Once the scale factor and centroid position has been determined, a range of n_{max} values are tested to determine the smallest order which can be used to represent the image within the desired error limits, Equation 6.10.

$$\chi^2(n_{max};\beta,x_c) = \sum_{n_{pix}} \frac{(I_M(x_p,n_{max};\beta,x_c) - I(x_p))^2}{\sigma_p^2} \tag{6.10}$$

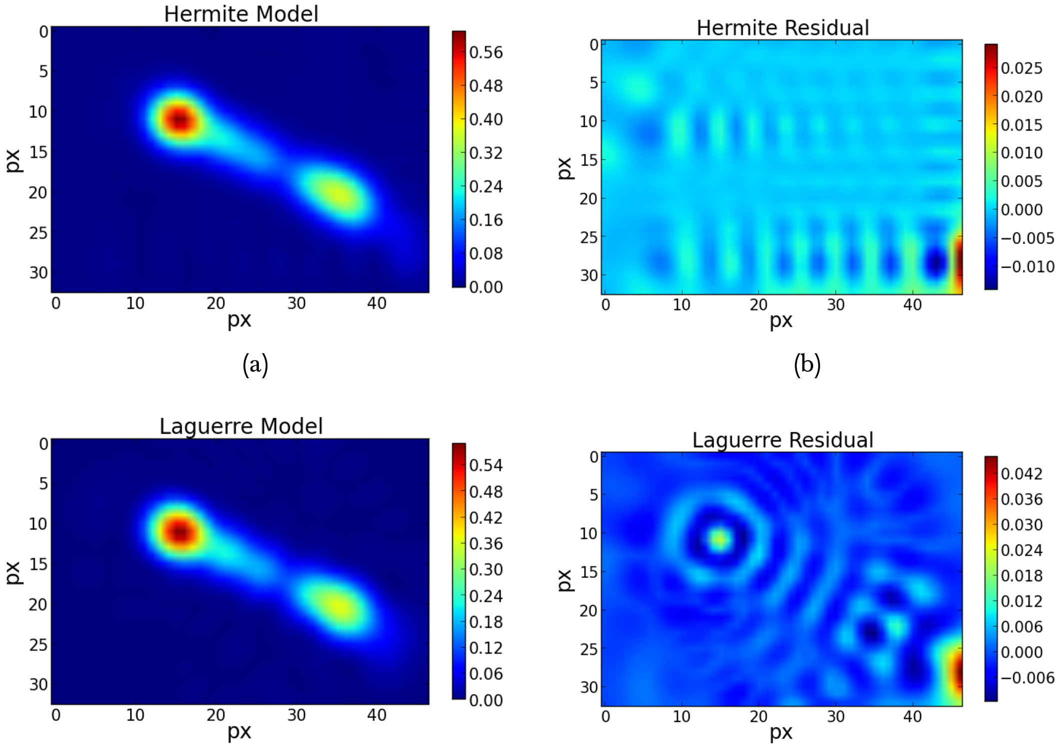
Shapelet decomposition using Hermite and Laguerre polynomials produce similar results given a fixed number of coefficients, Figure 6.6. But, they produce different artifacts in the residuals. The choice of polynomials depends on the source structure. Generally, a source which is spherical, such as an expanding supernova remnant or HII region can be better modelled with Laguerre polynomials. Asymmetric sources, such as lobed radio galaxies, tend to be better modelled with Hermite polynomials. A vast subset of sources can be equally well modelled with both basis sets, and it is often useful to decompose an image using both.

6.2.2 Transform to *uv* Space

Once an image has been decomposed into a set of shapelet functions the next step is to transform those functions to uv space. As noted in Equation 6.7, each basis function is unchanged under Fourier transform up to a phase and scale factor. These functions are evaluated at the observed uv positions which has the effect of convolving the array PSF with the decomposed image. At this point different weighting schemes can be used, though a natural weighting is used be default. Transformation back to the image domain will produce an image made up of the original decomposed image convolved with the array PSF over the observation.

It should be noted that the PSF of the original image has not been considered thus far. This is not considered a problem if the PSF has already been deconvolved or the image PSF is small compared to the PSF of the array being used for observation. If the image PSF is a factor, Chapter 3 of [Berge et al., 2012] covers correction solutions though this has not yet been implemented in the python code.

The centroid position can be placed anywhere on the sky by applying a phase offset based on the observing phase center. For the trivial case where the centroid is at the phase center no offset is required. Often the shapelet is added to the model data of a measurement set where the centroid position is different then the phase center. A phase offset is required based on the difference in angular distance and observing frequency.



(a) (b)

(c) (d)

Figure 6.6: Resulting models from Hermite (Cartesian) and Laguerre (Polar) polynomial decomposition (6.6a,6.6c). Both models use the same order of coefficients and produce similar models. But, the resulting residual (6.6b,6.6d) between source and model produce different artifacts.

6.2.3 Limitations of the Shapelet Model

Limitations arise in a number of extreme cases when using the shapelet formalism. High dynamic range structures, extremely widefield sources, high resolution and large images, and multi-source fields. The issues can be overcome with further extension of the formalism as they become necessary in future work.

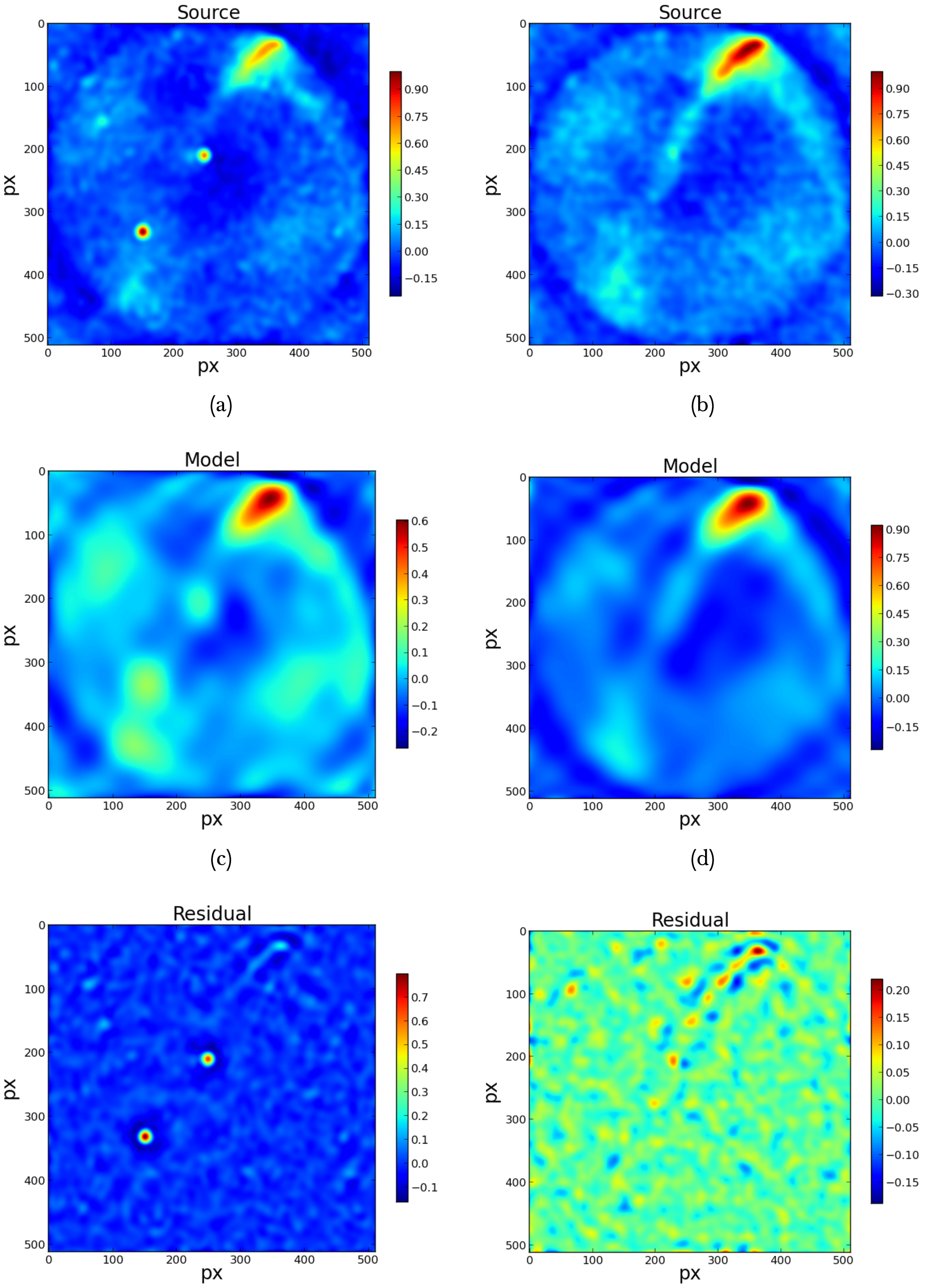
As noted earlier in the chapter, the shapelet basis set is made of functions which are perturbation of a Gaussian. This is an ideal basis set to model smoothly varying, Gaussian-like sources such as a distant radio galaxy or radio lobes. But if a field is made up of a large smooth region embedded with bright point sources then the shapelet decomposition can produce poor results, Figure 6.7. This is often the case in HII regions which have a smooth extended background with a number of point sources. The background is weaker than the point sources and the field can be thought of having two flux regions, one with bright point sources and the other being the weak background. Either the point sources can be deconvolved out before shapelet decomposition, as in Figure 6.7b, or multiple shapelet decompositions can be used to produce a good model. This can be done by spatially filtering the image into high and low spatial frequency components which will separate the point sources from the background. Or by decomposing small fields centered on the point source, subtracting those decompositions from the total image and then decomposing the residual image. In both case the resulting set of shapelet coefficients can be added back together in the image or uv domain.

Widefield images pose a problem in not just shapelet modeling but in imaging as projection effects smear and distort the image. Far from the phase center these effects are significant, see Chapter 5. If a source is imaged far from the phase center it will appear distorted. Any shapelet model derived from that image will also be distorted.

A very extended source, like the galaxy, can be decomposed into shapelets with a large enough n_{max}. But, the model will break down in uv space as the centroid is moved further from the observed phase center. Future work could be done to implement shapelet decomposition in a HEALPix format which does not have projection effects. Then the shapelet could be transformed to uv space without projection effects.

High resolution and large images pose a problem computationally as they will require a large n_{max} and larger matrix inversion. This is a processing problem which can be improved by code optimization, converting the code base to faster languages such as C, or using more computational hardware. A GPU-based implementation would be a fruitful project as that hardware is ideal for matrix operations.

Finally, fields with multiple sources are difficult to model. The choice of a good centroid is essential to the decomposition, if there are multiple sources it is not clear



(a) (b)

(c) (d)

(e) (f)

Figure 6.7: Shapelet model solution with and without point sources deconvolved from the source images. The extended source model is improved, Figure 6.7c versus Figure 6.7d, by first removing bright point sources which reduces the image dynamic range.

which is the ideal centroid position. A weighted average is chosen which generally produces poor results. As alluded to earlier in this section, a field could be broken up into smaller subfields which are combined after decomposition.

6.2.4 Python Software

As noted in the introduction to this chapter, a set of functions using the shapelet formalism has been written in python. The decomposition code is an adaptation of the IDL code presented by [Massey & Refregier, 2005]. The software is extended to include functionality specific to radio astronomy, such as the transformation to uv space, and reading/writing FITS and measurement set files. The core functionality is made of functions for decomposition, uv transformation, image processing and analysis, file I/O, and basis functions. A set of script then use these functions to generate and plot shapelet functions, read and write coefficient files, run decompositions using polar and Cartesian functions, transformation to uv space, and inserting uv data into measurement sets. The working code repository is hosted on Github[3].

6.3 Element Beam Patterns

Besides using shapelets for modelling extended source, they can be used to model primary beams. This provides a number of calibration and computational advantages. As the calibration improves and a higher dynamic range is desired, the beam pattern of each individual element will need to be included in the calibration model. The individual beam patterns will be perturbation of some theoretical or average pattern. Typically, when a beam pattern is included in calibration that pattern is based on an image or an analytical function. These both have limitations which, in some cases, can be improved upon with a shapelet-based model. Images have an intrinsic resolution, they are functions evaluated at finite points. Any other point in the beam must be determined by interpolation which both introduces error and is computationally expensive. Analytical models are created based on theoretical models of the beam. Perturbations on the theoretical model can be difficult to include into an analytical form. Often, a measured beam model will be fit to a set of basis functions such as Gaussians, cosines, or spherical harmonics. Shapelets fit into this domain, analytical functions fit to the measured data.

---

[3] https://github.com/griffinfoster/shapelets

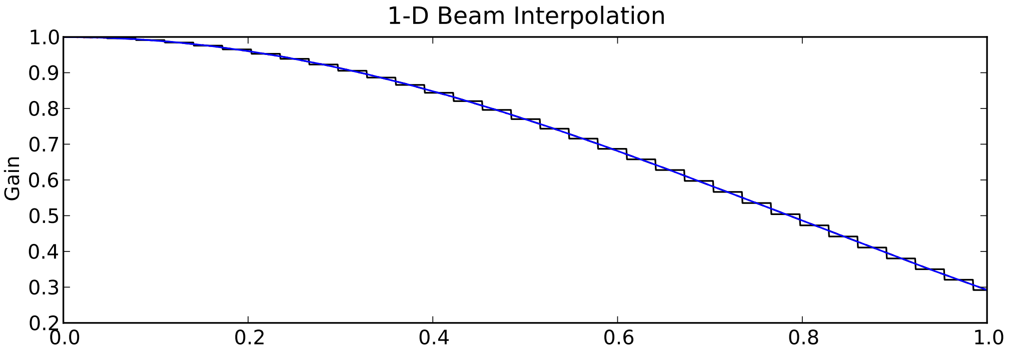
6.3.1 Image Based Beam Patterns

When an analytical beam model is not available an image-based model, whether voltage or power, is often used. This is the case when the beam has been measured using astronomical sources, holography, or artificial calibration sources. These patterns represent the actual beam pattern but they are also spatially quantized. To evaluate the beam response at a sub-pixel level an interpolation method must be applied to the image. Figure 6.8 shows a few interpolation methods along with their resulting error to the known pattern. As the interpolation method increases in complexity, none to linear to cubic, the residual error decreases. But each level of complexity increases the amount of processing required, which can drastically slow down calibration. Also, 2-D interpolation computational complexity scales much faster than 1-D schemes. As the required dynamic range increases so does the required accuracy in the beam pattern. Interpolation methods introduce a known level of error. Thus, the interpolation error must be below the beam pattern error set by the dynamic range goal. Shapelet-based models for beam patterns skip interpolation issues by representing the beam pattern with a set of smooth basis functions.

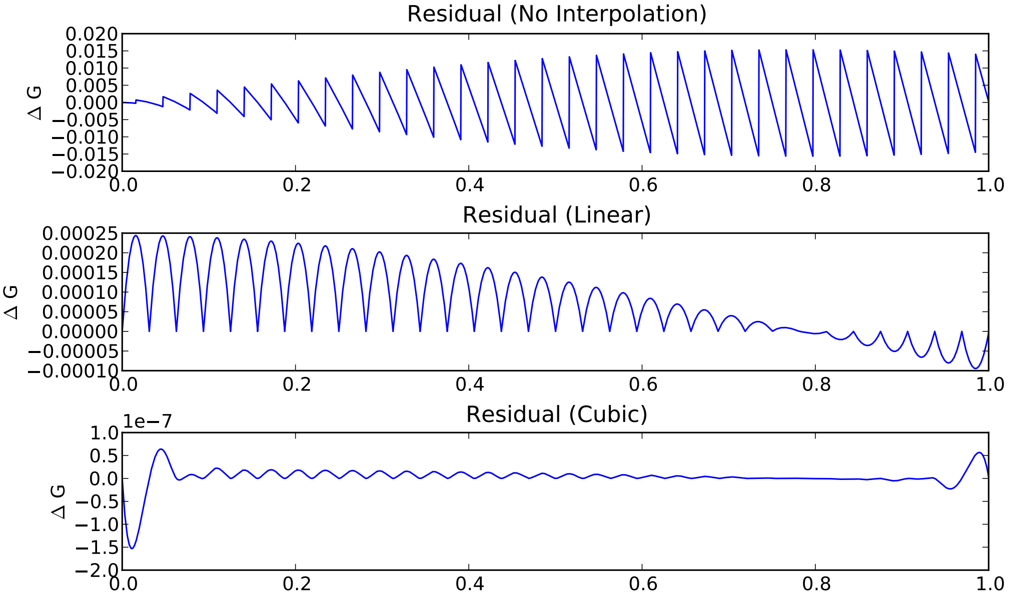
6.3.2 Shapelet Based Beam Patterns

Instead of interpolating the beam gain at a particular point every time a calibration is attempted the beam can be decomposed into shapelet functions once. This representation can then be quickly evaluated for a given position. The decomposition is the same as the extend source decomposition. Only the β scale factor needs to be fit for since the center of the beam pattern, x_c, is known. Since the beam pattern is static, only one decomposition (per frequency subband or range) is required. A large n_{max} decomposition, which is expensive in terms of computation, can be performed. This decomposition shifts the computation cost up front, but only ever needs to be done once. Then the shapelet functions can be quickly evaluated for each position, greatly speeding up calibration. And, avoiding image interpolation issues. Though, frequency interpolation will still be required unless a model is generated for each observing frequency subband.

Figure 6.9 shows a number of shapelet-based decompositions of the LOFAR LBA beam patterns at various observing frequencies. These beam patterns are low resolution images derived from the theoretical dipole model which covers the full, horizon to horizon field of view. The sharp edge effects at low altitudes make the beam pattern difficult to model at low altitudes. Within low altitude regions the beam can either be evaluated using an interpolation method or sources can be dropped from the calibration as they will be highly attenuated.

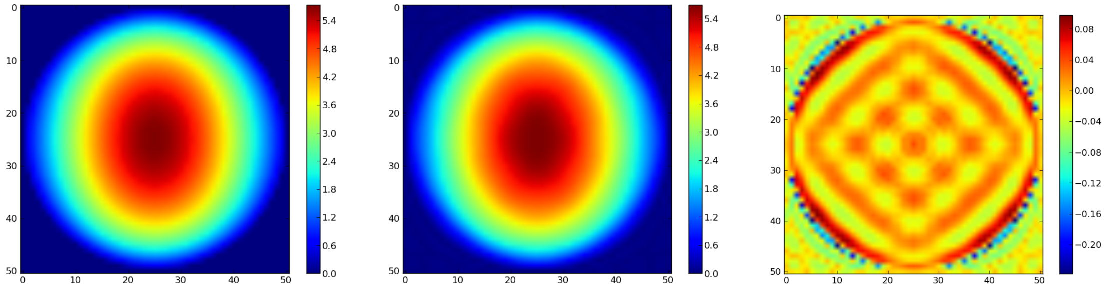


(a) 1-D slice of a $\cos^2$ beam pattern, with quantization effects from pixelization shown.



(b) Residual error introduced by using linear, cubic, and no interpolation. Linear interpolation reduces the gain error by two orders of magnitude, five orders for the cubic interpolation.

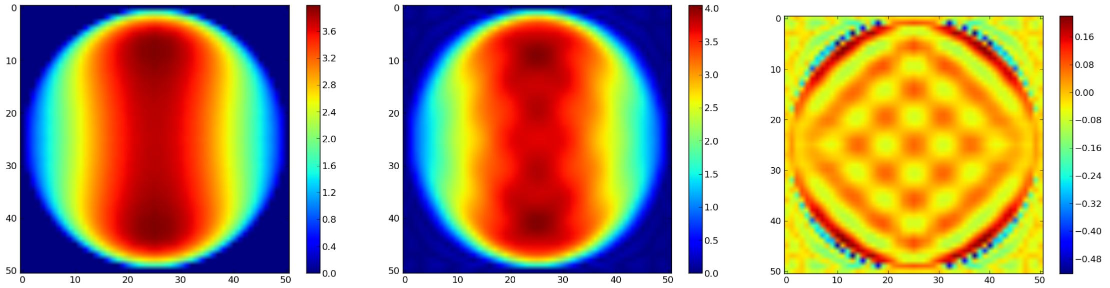
Figure 6.8: Response of interpolating a pixelized, 1-D beam shape using various methods: none, linear, cubic. Each method introduces error and characteristic patterns.



(a) LBA beam model, mode $(2, 0)$

(b) Polar shapelet model ($n_{max} = 22$).

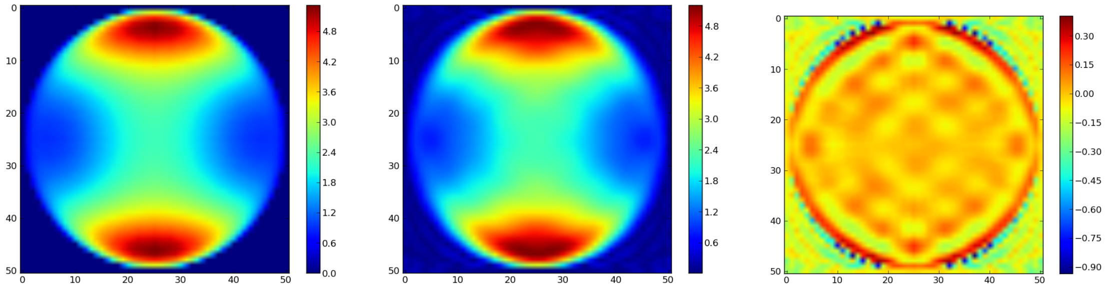
(c) Residual



(d) LBA beam model, mode $(6, 0)$

(e) Polar shapelet model ($n_{max} = 22$).

(f) Residual



(g) LBA beam model, mode $(8, 0)$

(h) Polar shapelet model ($n_{max} = 22$).

(i) Residual

Figure 6.9: Shapelet beam models of LBA dipoles at different observing frequencies. The model fits the image to $\leq 2\%$ across the majority of the beam. Sharp edge effects make the beam difficult to model at low altitudes.

6.4 Further Work

The shapelet package will continue to be an ever expanding toolkit, a wealth of features will be added to further improve its utility. The framework and code base has been developed to be as general as possible, with the scripts providing custom specialization. Decomposition of sources to shapelet functions is well automated. Further work is necessary to incorporate multiple shapelet objects into united data sets. Fields of interest will include many extend sources. Incorporating multiple shapelet-based objects into a single data set, with given phase offsets is being incorporated into the code base. This development goes hand in hand with the interface to the MeqTrees' environment. Currently, only a single shapelet object at the phase center in available in the simulation process. Expanding this to multiple objects is a key feature. Shapelet-based beam patterns will have a similar interface to MeqTrees, this addition requires setting up the beam Jones matrix of the measurement equation with a shapelet interface.

Shapelets offer an array agnostic method to model a field. Point source models rely of convolving groups of point sources with an array PSF to form the extended structure. Different arrays observing the same field, even the same array observing a field at a different time, will require a different sky model. A high resolution shapelet sky model can be convolved with the observing PSF to generate a unique sky model for a given observation but based on a known sky model.

Modelling extending source for calibration and for image deconvolution is a pervasive challenge which will only become a more serious issue as the sensitivity of radio arrays increases. Shapelets offer a unique method to quickly transfer from image to uv domains with a simple transform and have proven to be an invaluable tool for the calibration challenges which lie ahead. Moving beyond simple, point source based CLEAN models is a major challenge for the future of radio astronomy. Integrating shapelet models into the CLEAN cycle has great potential.

Chapter 7

Conclusion

This all would have been so much easier if two years ago I knew the things I know now.

-Charles Copley, 2012

In this era of large-N, low frequency radio astronomy we are observing the universe by re-evaluating a band of astronomy that has been greatly neglected since its golden age in the 1950s and 1960s. With our advancements in digital signal processing and computing we are expanding on old techniques which were once limited by the technology of the era. We are returning to the original concepts of radio astronomy, and redeveloping the field from first principles but with all the foreknowledge we have gained in the intervening years. This revolution in radio astronomy is leading not just to adaptations of the previous methods of observation but with entirely new methods. Often, it is difficult to determine when a field of science is undergoing a great era of advancement and development until long after that time has passed. But, it is clear, that where we stand in the field today is leaps and bounds to where we were only a few short years ago. And there is no doubt that some time in the future this will all happen again.

This thesis has attempted to cover a small portion of the recent developments in low frequency radio astronomy, with a focus on interferometric observations. As any thesis will, it began with a brief, much too brief, overview of the science goals at low frequencies. EoR is the great goal on the horizon, one of the considerable but still missing ages of the universe. Radio transients, these elusively variable sources in this stable universe, require new techniques and significant computation time but are part of a new time-variant universe we are just beginning to access. The fine structure and perturbations from large scale cosmic magnetism is an essential component of understanding galaxy evolution. All these science goals, along with others, require large-N, sensitive interferometric arrays which can access many octaves of bandwidth instantaneously to perform fast sky surveys. The currently in operation LOFAR and future SKA along with the many EoR focused arrays are pushing the bandwidth, time resolution, and sensitivity boundaries at the low frequencies. Key to these arrays is the digital FX correlator which computes the correlation of the complex signals between antenna baselines. The FX correlator architecture scales as $O(N^2)$. Historically the XF correlator came as an advancement of the original phase-switching power detectors. The FX correlator is mathematically equivalent to the XF design, computationally more efficient at modest to large N and has improved frequency characteristics. Modern correlators are typically implemented on FPGAs which offer a low power, efficient, and flexible platform for DSP. Though it should be noted that hybrid FPGA/GPU systems are becoming more attractive with the development of scientific GPU computing in recent years.

Signal correlation is a well understood process with a simple computational solution, but with large-N correlators issues arise out of scalability. Reducing the bandwidth and choosing the correct bit depth for each operation is an important aspect to correlator design. At low frequencies, human-generated RFI and the dynamic variability of the radio sky require large bit depth ADCs to accommodate the dynamic range. During the channelization stage, F stage, narrowband signal localization is performed with PFBs to contain RFI into small fractional bandwidth portions, opening up larger portions of the band. Bit growth during channelization can be effectively quantized down to ~ 4 bits without loss to the correlation signal integrity. This reduces the cost of the correlation operation and the bandwidth used in the cornerturn and interconnect. The antenna to frequency cornerturn is performed using an $O(n \log n)$ hierarchical structure based on high bandwidth network switches. In the cross correlation, the X stage, a pipelined architecture can be used with makes efficient use of multipliers at the cost of long memory delays. Or, a matrix architecture is used which minimizes memory resource usage while not fully using multiplier logic. The architecture is dependent on the limited resource on the specific FPGA. After correlation accumulation is performed to reduce the output data rate.

The FX correlator is just the first half of the digital backend. In the past the correlation matrix is written to disk for later processing. But with the increase in data rate due to the decrease in accumulation time and growth in N the correlation product must be processed in real time to further reduce the data to a rate which can be written to disk. Additionally, the scientific desire to improve the system calibration in the face of time-dependent variations requires on the fly calibration of signals.

A number of specialized topics in correlator design has been presented to further improve resource utilization. Ideally, each logical operation on a signal has a linear response which will mean quantization is a simple operation. But, with a priori knowledge of a signal's distribution a unique quantization scheme can be used to reduce the transport bandwidth with minimal loss to signal integrity. As multipliers are the key component to the cross correlation operation the choice of logical implementation is important. Multiplication can be implemented in specialized DSP logic, flexible 'soft' logic, or by using a logarithmic transform and adders. Depending on the bit width of the multiplication and available FPGA resources each of these methods have advantages. Finally, in low frequency systems which cover many octaves of bandwidth it is efficient to use a fractional channelization scheme to reduce the overall system bandwidth and computational cost at no loss to the scientific quality of the observation. These techniques offer an improvement in computational efficiency and data transport with minimal loss of information. The continued work towards low power systems will require careful instrument development based on the required scientific specifications.

Once the ground work for the scientific justification and technical challenges of correlation has been laid out, practical examples have been presented in the BEST-2 correlator system and the SEPcam station imager. The BEST-2 system is both a prototype for SEPcam and a fully functional instrument which is routinely used for observation. It is a 32 element, single polarization system which operates as an FX correlator, spatial FFT imager, and frequency-domain beamformer. The digital instrumentation is based on open source CASPER DSP libraries and hardware. The design uses streaming data transport from the F board to the X and S boards. Correlation is performed using a pipelined architecture and outputs to a receive computer over 10 Gb Ethernet. Initial observations were performed to test the analogue and digital system. Bright radio source were detected. Using point source calibrator sources, calibrated images were produced for a number of radio fields. This digital backend has proven a successful upgrade to the current system at Medicina Observatory. Additionally, the firmware developed and tested in the instrument has been reused in SEPcam.

The SEPcam station imager will be a FX correlator with a real time calibration and imaging backend for the LOFAR station at Chilbolton Observatory. The correlator is a 96 element, dual polarization system running on a single ROACH2 board. Correlations are output over 10 Gb Ethernet at 10 ms integrations. The board is interfaced to the current digital backend with a passive signal duplication board as to not interrupt normal station operations. SEPcam will run concurrently with the backend. The digital firmware has been verified in simulation and lab. The modularity of the design means the individual logic blocks can be tested independently, and combined into the larger system in stages. Correlation is performed using the same pipelined architecture as BEST-2 as the FPGA design is multiplier limited. The firmware is split into domain windows for synchronizing data streams, at each window edge is FIFO buffer. There is sufficient resource still available on the FPGA to allow for future expansion of functionality including further channelization and increasing the input bandwidth. Further work is required to integrate the correlator with the current backend. To achieve a real time system a data processing computer system will be required, this will attach to correlator outputs. SEPcam is a modular correlator system, this type of design will be similar to those used in phase 1 of the SKA. In fact, SEPcam computes nearly four times the number of baselines required for the low frequency band on SKA phase 1. SEPcam only processes a small portion of the full SKA band, though the modular design means that with duplication the full band can be covered. The next challenge will be in further reducing the system power consumption.

In a move toward a real time calibration and imaging system an automated pipeline has been developed for widefield, LOFAR single station correlation products. This is a parallel development path to the SEPcam correlator firmware. Using the existing station calibration correlator data products a measurement equation based model has been developed for both LBA and HBA observations. HBA observations require setting the tile element analogue beamformers to a specific 'all-sky' mode, which is not used during normal observations. The system calibration model takes into account a sky model, Class A radio sources are used at present but additional sources can be trivially added. The model also accounts for a beam model, the initial model is based on a half wavelength dipole analytical model but work has begun to use bright point sources to measure out an average beam. The model is then fit to the observation to produce complex gain solutions using the MeqTrees' environment. Snapshot images from a single integration are produced over a 24 hour period. Since the elements have an impressive field of view, w-projection effects would become a serious issue if combining observation in the UV domain. Instead, after individual images are produced they are interpolated onto a HEALPix sky map to remove projection effects. The resulting maps cover 24 hours of right ascension and $+90^\circ > \delta > -10^\circ$in declination. Using the large fractional bandwidth coverage of the array, initial measurements of source and galactic spectral index structure can be made. This calibration and imaging system is the first to generate sky maps for LOFAR stations. Continued development will lead to new understanding of the low frequency radio sky on extremely large scales.

Both the extend emission of synchrotron radiation in the galaxy and the element beam pattern pose challenges to calibration at low frequencies. Typically an extended source is modelled as multiple point sources in a sky model. And, the beam pattern as an image or analytical formula. As the size of N increases there will be more individual beam patterns and the resolution will increase leading to fields with more extended sources. Also, as the dynamic range specification increases the calibration must equally improve. Shapelet analysis is a method of modelling images as a set of functions to an arbitrary degree of accuracy. These have been proven to be useful in modelling galaxies in optical. Both a polar and Cartesian form has been used. The shapelet methods has been adapted for use in interferometric radio data, including transforming sources to to the UV domain. A software package has been developed to interact with images, FITS files, and measurement sets to generate shapelet-based models for simulation and calibration. An interface to MeqTrees has also been developed, and is undergoing testing. Beam pattens using shapelets functions have been used in simulation and calibration to reduce the errors introduced in image-based interpolation.

In our aim for vast dynamic range and sensitivity using basic antenna elements every portion of a system must be well understood. The errors from one point propagate through the system, only by knowing the system can we understand the effect we see in the end products. Hardware, both analogue and digital, and software are intertwined and feedback on each other. As the computing power increases we can afford to build larger N systems covering more bandwidth and processes wider fields of view. This requires new methods for calibration and imaging. To reduce the time to science and opens up new avenues of explorations. The days of recording an observation to disk and worrying about the result later are coming to an end. And, what lies ahead for radio astronomy will be like nothing we have seen before.

Appendix A

Sky Maps

Sky maps produced with the LOFAR station at Chilbolton Observatory. A full discussion of the map making process is discussed in Chapter 5. Each sky map is made up of 160 snapshot images spaced over 24 hours. Maps are masked at $\delta \leq -15°$. The flux scale for each map is based on calibrating Cygnus A and Cassiopeia A in each snapshot image to the flux presented in [Baars et al., 1977]. The flux of Cassiopeia A has been adjusted to 2012 values. A histogram color mapping is used. Maps are in a HEALPIX format with $n_{side} = 128$. The resolution of each map is dependent on frequency and the maximum baseline length in the LBA station, ~ 60 m.

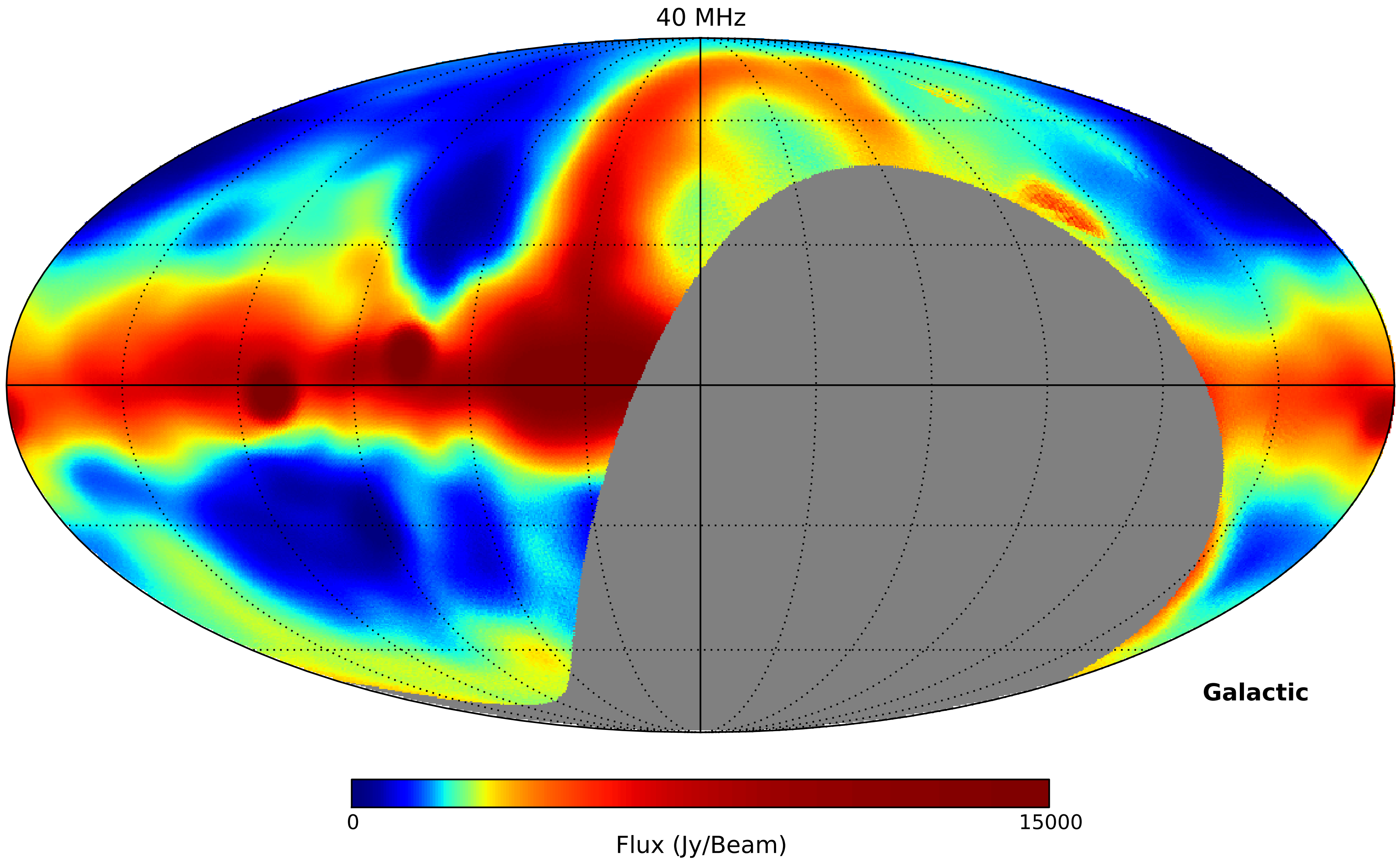


Figure A.1: Galactic coordinate projection of an equal area gridded sky map at 40 MHz using a 24 hour ‘snapshot’ observation from the LOFAR LBA at Chilbolton Observatory. The dynamic range of the map is ~ 8.

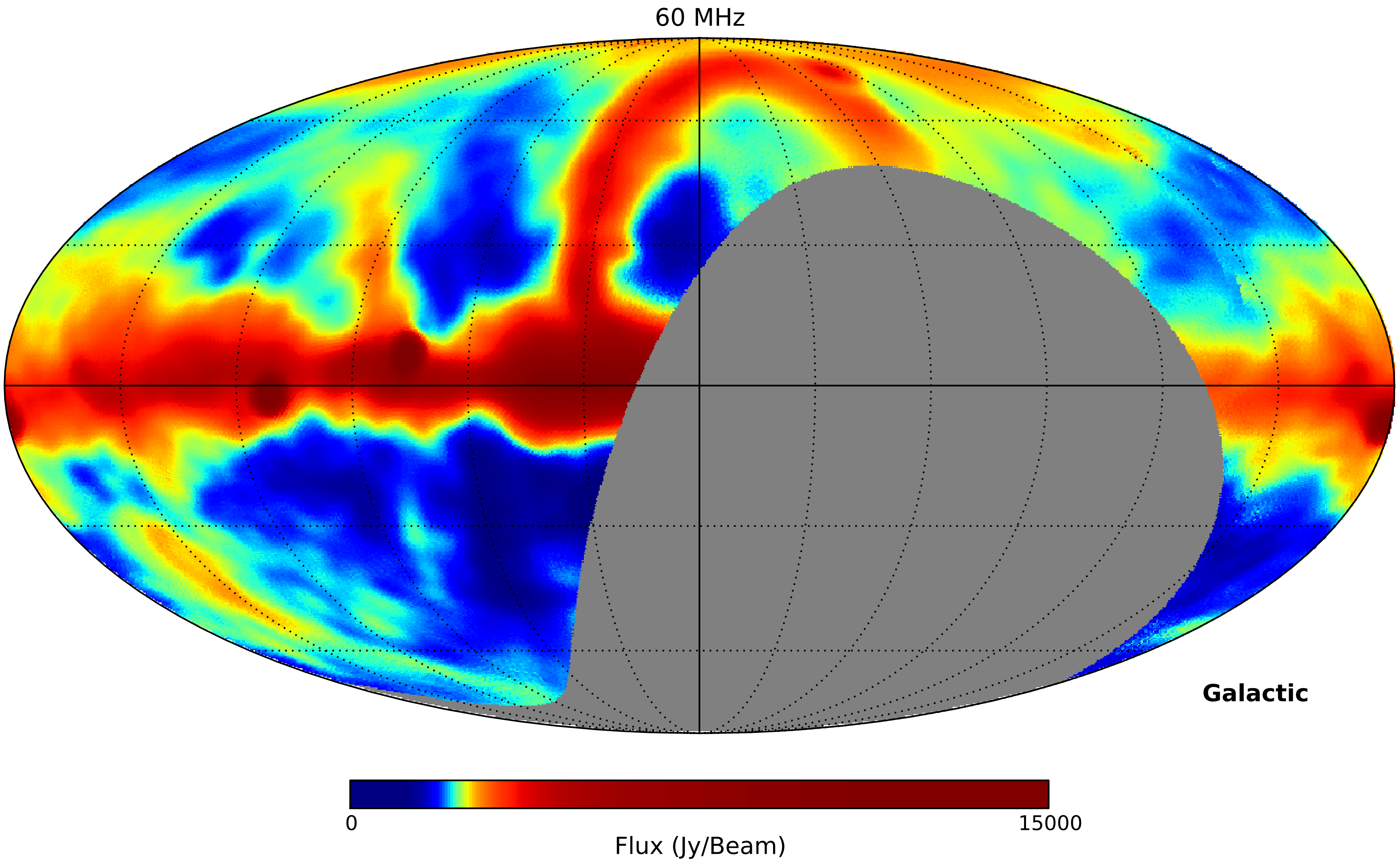


Figure A.2: Galactic coordinate projection of an equal area gridded sky map at 60 MHz using a 24 hour ‘snapshot’ observation from the LOFAR LBA at Chilbolton Observatory. The dynamic range of the map is ~ 14.

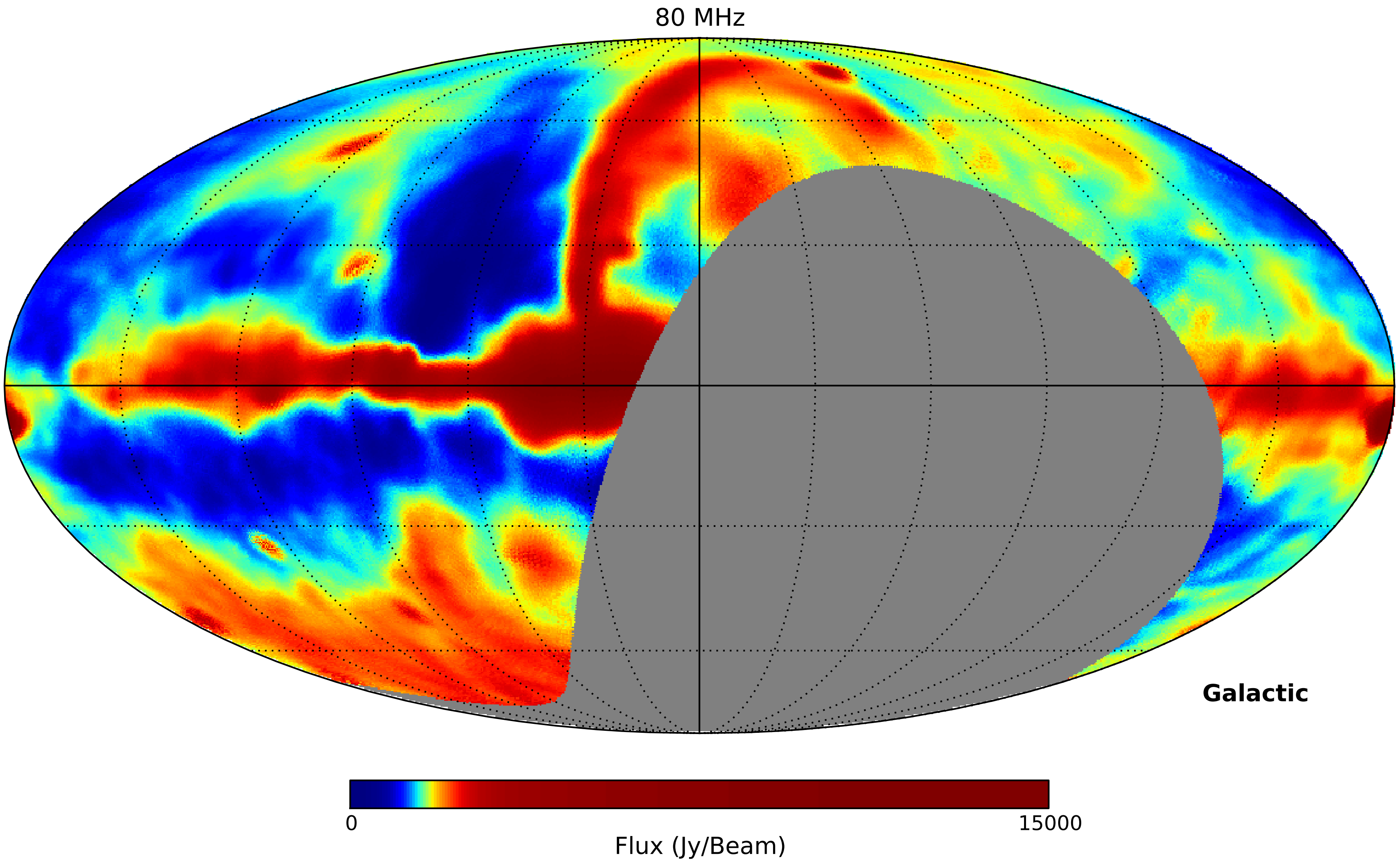


Figure A.3: Galactic coordinate projection of an equal area gridded sky map at 80 MHz using a 24 hour ‘snapshot’ observation from the LOFAR LBA at Chilbolton Observatory. The dynamic range of the map is ~ 15.

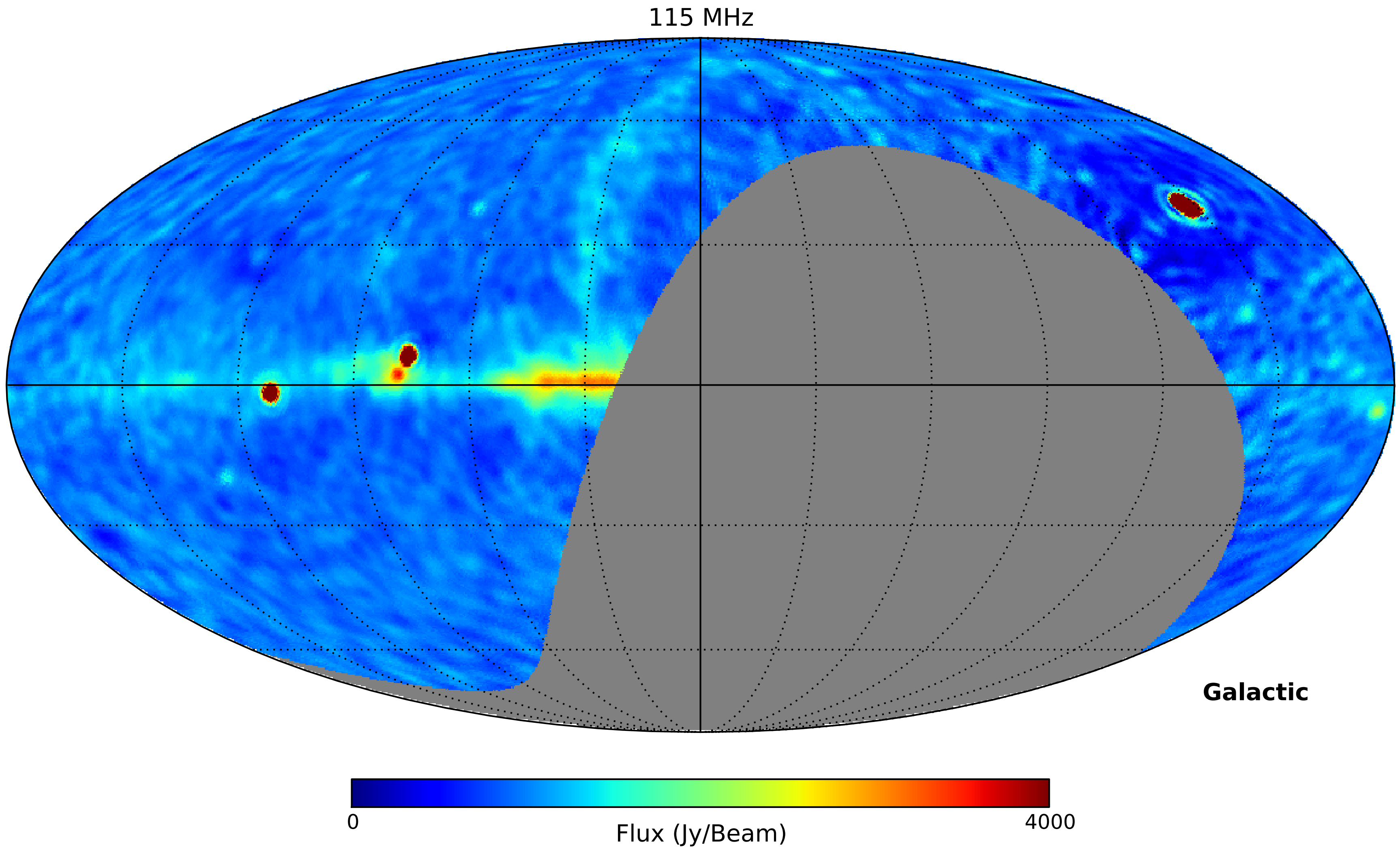


Figure A.4: Galactic coordinate projection of an equal area gridded sky map at 115 MHz using a 24 hour ‘snapshot’ observation from the LOFAR HBA at Chilbolton Observatory. A linear scale is used for the flux, with the dominate sources: Cygnus A, Cassiopeia A, and the sun have been clipped at 4000 Jy/Beam. The dynamic range of the map is ~ 20.

Appendix B

Thesis Statistics

Metadata for the thesis *Large-N Correlator Systems for Low Frequency Radio Astronomy*. Figure B.1 shows the aggregate total page, word, and character count used in this thesis. Over the first two and half months the thesis showed a fairly constant development. Throughout August development was flat due to conferences and travel. The final month and a half shows a tapering off in the development slope as I moved towards editing. Table B.1 shows a per chapter break down. Chapter 5 was at longest chapter in terms of pages, words, and characters. The average chapter length, excluding the conclusion chapter, was 27.5 pages ($\sigma = 5.25$). The five most frequently used unique words are: *data* (175), *correlator* (165), *source* (161), *frequency* (156) and *correlation* (152).

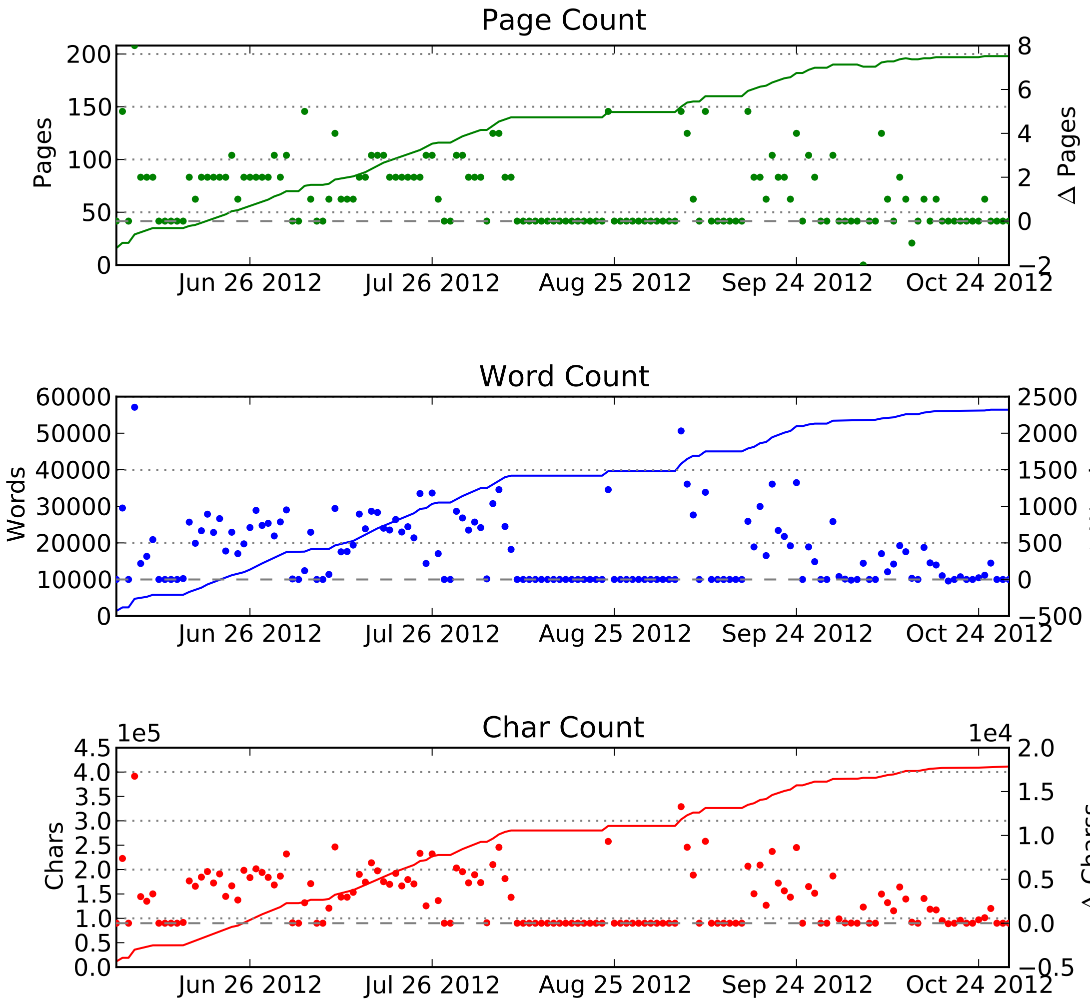


Figure B.1: Progression of page, word, and character count through writing period. The flat period throughout August was due to conference attendance and travel. Initial thesis work was started on the 4^{th} of June 2012 and submitted to the exam school on the 5^{th} of November 2012.

	Total	Ch. 1	Ch. 2	Ch. 3	Ch. 4	Ch. 5	Ch. 6	Ch. 7
Pages	198	32	29	21	29	34	20	5
Words	56209	11334	9905	5853	8905	10946	5175	1841
Characters	409713	77364	68087	44809	60448	86348	40390	11967
Figures	87	13	14	15	15	17	9	0

Table B.1: Aggregate thesis data.

References

[ALTERA, 2004] ALTERA (2004). *Implementing Multipliers in FPGA Devices.* ALTERA, application note 306, v3.0 edition.

[ALTERA, 2011] ALTERA (2011). *Logic Array Blocks and Adaptive Logic Modules in Stratix IV Devices.* ALTERA, siv51002-3.1 edition.

[Ananthakrishnan, 1995] Ananthakrishnan, S. (1995). The Giant Meterwave Radio Telescope / GMRT. *Journal of Astrophysics and Astronomy Supplement*, 16, 427.

[Baars et al., 1977] Baars, J. W. M., Genzel, R., Pauliny-Toth, I. I. K., & Witzel, A. (1977). The absolute spectrum of cas a: An accurate flux density and a set of secondary calibrators. *Astronomy & Astrophysics*, 61, 99–106.

[Baars et al., 1977] Baars, J. W. M., Genzel, R., Pauliny-Toth, I. I. K., & Witzel, A. (1977). The absolute spectrum of CAS A - an accurate flux density scale and a set of secondary calibrators. *A&A*, 61, 99–106.

[Backer, 2007] Backer, D. (2007). *Quantization with Four Bits.* Technical Report 15, CASPER.

[Balanis, 2005] Balanis, C. (2005). *Antenna Theory: Analysis and Design.* Wiley-Interscience.

[Bennett, 1962] Bennett, A. S. (1962). The preparation of the revised 3C catalogue of radio sources. *MNRAS*, 125, 75.

[Berge et al., 2012] Berge, J., Massey, R., & Refregier, A. (2012). *An Introduction to Shapelet Based Weak Lensing Image Processing.*

[Bracewell, 1984] Bracewell, R. N. (1984). Inversion of Non-Planar Visibilities. In J. A. Roberts (Ed.), *Indirect Imaging. Measurement and Processing for Indirect Imaging* (pp. 177).

[Brown, 1991] Brown, J. C. (1991). Calculation of a constant Q spectral transform. *Journal of the Acoustical Society of America*, 89(1), 425–434.

[Cane, 1978] Cane, H. V. (1978). A 30 MHz map of the whole sky. *Australian Journal of Physics*, 31, 561.

[Carilli & Rawlings, 2004a] Carilli, C. & Rawlings, S., Eds. (2004a). *Science with the Square Kilometer Array*, volume 48 of *New Astronomy Review*.

[Carilli & Rawlings, 2004b] Carilli, C. L. & Rawlings, S. (2004b). Motivation, key science projects, standards and assumptions. New A Rev., 48, 979–984.

[Carlson & Dewdney, 2000] Carlson, B. & Dewdney, P. (2000). Efficient wideband digital correlation. *Electronics Letters*, 36(11), 987 – 988.

[Carroll & Ostlie, 1996] Carroll, B. W. & Ostlie, D. A. (1996). *An Introduction to Modern Astrophysics*. Addison Wesley Longman.

[Chang & Refregier, 2002] Chang, T.-C. & Refregier, A. (2002). Shape Reconstruction and Weak Lensing Measurement with Interferometers: A Shapelet Approach. ApJ, 570, 447–456.

[Chikada et al., 1986] Chikada, Y., Ishiguro, M., Hirabayashi, H., et al. (1986). A very fast fft spectrum analyzer for radio astronomy. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.*, volume 11 (pp. 2907 – 2910).

[Clark et al., 2011] Clark, M. A., La Plante, P. C., & Greenhill, L. J. (2011). Accelerating Radio Astronomy Cross-Correlation with Graphics Processing Units. *ArXiv e-prints*.

[Cooley & Tukey, 1965] Cooley, J. & Tukey, J. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90), 297–301.

[Cordes, 2009] Cordes, J. (2009). *The Square Kilometer Array as a Radio Synoptic Survey Telescope: Widefield Surveys for Transients, Pulsars and ETI*. Technical Report 97, SKA Project Development Office.

[Cornwell et al., 2003] Cornwell, T., Golap, K., & Bhatnagar, S. (2003). *W Projection: A New Algorithm for Non-Coplanar Baselines*. Technical Report 67, NRAO eVLA.

[D'Addario, 2011] D'Addario, L. (2011). *Low-Power Correlator Architecture for the Mid-Frequency SKA*. Technical Report 133, SKA Project Development Office.

[de Oliveira-Costa et al., 2008] de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. (2008). A model of diffuse Galactic radio emission from 10 MHz to 100 GHz. *MNRAS*, 388, 247–260.

[de Vos et al., 2009] de Vos, M., Gunst, A. W., & Nijboer, R. (2009). The LOFAR Telescope: System Architecture and Signal Processing. *IEEE Proceedings*, 97, 1431–1437.

[Dewdney et al., 2010] Dewdney, P., bij de Vaate, J.-G., Cloete, K., et al. (2010). *SKA Phase 1: Preliminanry System Description.* Technical Report 130, SKA Project Development Office.

[Ellingson et al., 2009] Ellingson, S. W., Clarke, T. E., Cohen, A., et al. (2009). The Long Wavelength Array. *IEEE Proceedings*, 97, 1421–1430.

[Emerson, 1976] Emerson, R. (1976). *Biplex Pipelined FFT.* Technical Report 42-34, JPL Deep Space Network Progres Report.

[Ewen & Purcell, 1951] Ewen, H. I. & Purcell, E. M. (1951). Observation of a Line in the Galactic Radio Spectrum: Radiation from Galactic Hydrogen at 1,420 Mc./sec. Nature, 168, 356.

[Foster et al., 2012a] Foster, G., Hickish, J., Price, D., & Zarb Adami, K. (2012a). Instrumentation Design of a Spatial FFT and FX Correlator for the BEST-2 Array. Preprint.

[Foster et al., 2012b] Foster, G., Hickish, J., Price, D., & Zarb Adami, K. (2012b). Results from the BEST-2 Array Spatial FFT and FX Correlator. Preprint.

[Furlanetto et al., 2006] Furlanetto, S. R., Oh, S. P., & Briggs, F. H. (2006). Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. Phys. Rep., 433, 181–301.

[Górski et al., 2005] Górski, K. M., Hivon, E., Banday, A. J., et al. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ, 622, 759–771.

[Greisen, 1994] Greisen, E. (1994). *Non-linear Coordinate Systems in AIPS.* Technical Report 27, NRAO AIPS.

[Griffiths, 2005] Griffiths, D. (2005). *Introduction to Quantum Mechanics, 2/E Edition.* Addison-Wesley.

[Gunst & Schoonderbeek, 2007] Gunst, A. & Schoonderbeek, G. (2007). *LOFAR Station Architectural Design Document.* Technical Report LOFAR-ASTRON-ADD-013, ASTRON.

[Guzmán et al., 2011] Guzmán, A. E., May, J., Alvarez, H., & Maeda, K. (2011). All-sky Galactic radiation at 45 MHz and spectral index between 45 and 408 MHz. *A&A*, 525, A138.

[Hamaker et al., 1996] Hamaker, J. P., Bregman, J. D., & Sault, R. J. (1996). Understanding radio polarimetry. I. Mathematical foundations. A&AS, 117, 137–147.

[Haslam et al., 1981] Haslam, C. G. T., Klein, U., Salter, C. J., et al. (1981). A 408 MHz all-sky continuum survey. I - Observations at southern declinations and for the North Polar region. *A&A*, 100, 209–219.

[Haslam et al., 1982] Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. (1982). A 408 MHz all-sky continuum survey. II - The atlas of contour maps. A&AS, 47, 1.

[Hewish et al., 1968] Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. (1968). Observation of a Rapidly Pulsating Radio Source. Nature, 217, 709–713.

[Hickish, 2013] Hickish, J. (2013). *Realtime DSP in multi-element radio telescopes.* D.Phil in astrophysics, University of Oxford.

[Holler et al., 2007] Holler, C. M., Kaneko, T., Jones, M. E., Grainge, K., & Scott, P. (2007). A 6-12 GHz analogue lag-correlator for radio interferometry. *A&A*, 464, 795–806.

[Jansky, 1933] Jansky, K. (1933). Electrical disturbances apparently of extraterrestrial origin. *Proceedings of the IEEE*, 21, 1387.

[Jenet & Anderson, 1998] Jenet, F. A. & Anderson, S. B. (1998). The Effects of Digitization on Nonstationary Stochastic Signals with Applications to Pulsar Signal Baseband Recording. PASP, 110, 1467–1478.

[Jones, 1941] Jones, R. C. (1941). New calculus for the treatment of optical systems. I. Description and discussion of the calculus. *Journal of the Optical Society of America (1917-1983)*, 31, 488.

[Kant et al., 2007] Kant, G., Gunst, A., & Dorst, M. (2007). *High Band Antenna Architectural Design Document.* Technical Report LOFAR-ASTRON-ADD-019, ASTRON.

[Kassim et al., 2007] Kassim, N. E., Lazio, T. J. W., Erickson, W. C., et al. (2007). The 74 MHz System on the Very Large Array. ApJS, 172, 686–719.

[Landecker & Wielebinski, 1970] Landecker, T. L. & Wielebinski, R. (1970). The Galactic Metre Wave Radiation: A two-frequency survey between declinations +25^o and -25^o and the preparation of a map of the whole sky. *Australian Journal of Physics Astrophysical Supplement*, 16, 1.

[Lazio, 2008] Lazio, J. (2008). Radio wavelength transients: Current and emerging prospects. *Astronomische Nachrichten*, 329, 330.

[Lee, 2012] Lee, P. (2012). Personal communiqué.

[Lofgren, 1958] Lofgren, L. (1958). Analog multiplier based on the hall effect. *Journal of Applied Physics*, 29(2), 158–166.

[Lubberhuizen & Kooistra, 2007] Lubberhuizen, W. & Kooistra, E. (2007). *RSP Firmware Functional Specification*. Technical Report LOFAR-ASTRON-SDD-001, ASTRON.

[Lupton, 1993] Lupton, R. (1993). *Statistics in Theory and Practice*. Princeton, NJ, USA: Princeton University Press.

[Lyons, 2004] Lyons, R. G. (2004). *Understanding Digital Signal Processing (2nd Edition)*. Upper Saddle River, NJ, USA: Prentice Hall PTR.

[Manley et al., 2010] Manley, J., Welz, M., Parsons, A., Ratcliffe, R., & van Rooyen, R. (2010). *SPEAD:Streaming Protocol for Exchanging Astronomical Data*. Technical Report 2010/10/07, SKA South Africa.

[Massey & Refregier, 2005] Massey, R. & Refregier, A. (2005). Polar shapelets. *MNRAS*, 363, 197–210.

[Mitchell et al., 2010] Mitchell, D., Greenhill, L. J., Clark, M., et al. (2010). The Murchison Widefield Array. In *RFI Mitigation Workshop*.

[Mitchell, 1962] Mitchell, J. N. (1962). Computer multiplication and division using binary logarithms. *Electronic Computers, IRE Transactions on*, EC-11(4), 512 –517.

[Montebugnoli et al., 2009a] Montebugnoli, S., Bartolini, M., Bianchi, G., et al. (2009a). BEST back end. In *Proceedings of Wide Field Astronomy & Technology for the Square Kilometre Array (SKADS 2009). 4-6 November 2009. Chateau de Limelette, Belgium* (pp. 355–358).

[Montebugnoli et al., 2009b] Montebugnoli, S., Bianchi, G., Monari, J., et al. (2009b). BEST: Basic Element for SKA Training. In *Proceedings of Wide Field Astronomy & Technology for the Square Kilometre Array (SKADS 2009). 4-6 November 2009. Chateau de Limelette, Belgium* (pp. 331–336).

[Muller & Oort, 1951] Muller, C. A. & Oort, J. H. (1951). Observation of a Line in the Galactic Radio Spectrum: The Interstellar Hydrogen Line at 1,420 Mc./sec., and an Estimate of Galactic Rotation. Nature, 168, 357–358.

[Noordam & Smirnov, 2010] Noordam, J. E. & Smirnov, O. M. (2010). The MeqTrees software system and its use for third-generation calibration of radio interferometers. *A&A*, 524, A61.

[Oxford Electronic Research Centre, 2012] Oxford Electronic Research Centre (2012). Oskar-2. http://www.oerc.ox.ac.uk/ ska/oskar2/.

[Paciga et al., 2011] Paciga, G., Chang, T.-C., Gupta, Y., et al. (2011). The GMRT Epoch of Reionization experiment: a new upper limit on the neutral hydrogen power spectrum at z=8.6. *MNRAS*, 413, 1174–1183.

[Parsons et al., 2006] Parsons, A., Backer, D., Chang, C., et al. (2006). PetaOp/Second FPGA Signal Processing for SETI and Radio Astronomy. In *Asilomar Conference on Signals and Systems, Pacific Grove, CA* (pp. 2031–2035).

[Parsons et al., 2010] Parsons, A. R., Backer, D. C., Foster, G. S., et al. (2010). The Precision Array for Probing the Epoch of Re-ionization: Eight Station Results. AJ, 139, 1468–1480.

[Perini, 2009] Perini, F. (2009). Low noise design experience for the SKADS/BEST demonstrator. In *Proceedings of Wide Field Astronomy & Technology for the Square Kilometre Array (SKADS 2009). 4-6 November 2009. Chateau de Limelette, Belgium.* (pp. 341–345).

[Perini et al., 2009] Perini, F., Bianchi, G., Schiaffino, M., & Monari, J. (2009). BEST receiver experience: general architecture, design and integration. In *Proceedings of Wide Field Astronomy & Technology for the Square Kilometre Array (SKADS 2009). 4-6 November 2009. Chateau de Limelette, Belgium* (pp. 351–354).

[PMC-Sierra, 2004] PMC-Sierra (2004). *QuadPHY 10GX*. PMC-Sierra, pmc-2020263 edition.

[Poiesz et al., 2007] Poiesz, W., Kooistra, E., & Lubberhuizen, W. (2007). *RSP Firmware Design Description*. Technical Report LOFAR-ASTRON-SDD-018, ASTRON.

[Prasad & Wijnholds, 2012] Prasad, P. & Wijnholds, S. J. (2012). AARTFAAC: Towards a 24x7, All-sky Monitor for LOFAR. *ArXiv e-prints*.

[Price, 2012] Price, D. (2012). *Radio astronomy instrumentation for redshifted hydrogen line science*. D.Phil in astrophysics, University of Oxford.

[Reber, 1940] Reber, G. (1940). Cosmic static. *Proceedings of the IRE*, 28(2), 68 – 70.

[Refregier, 2003] Refregier, A. (2003). Shapelets - I. A method for image analysis. *MNRAS*, 338, 35–47.

[Romein et al., 2010] Romein, J. W., Broekema, P. C., Mol, J. D., & van Nieuwpoort, R. V. (2010). The lofar correlator: implementation and performance analysis. *SIGPLAN Not.*, 45(5), 169–178.

[Ryle, 1952] Ryle, M. (1952). A New Radio Interferometer and Its Application to the Observation of Weak Radio Stars. *Royal Society of London Proceedings Series A*, 211, 351–375.

[Ryle & Vonberg, 1947] Ryle, M. & Vonberg, D. D. (1947). Relation Between the Intensity of Solar Radiation on 175 Mc./s. and 80 Mc./s. *Nature*, 160, 157–159.

[Santos et al., 2005] Santos, M. G., Cooray, A., & Knox, L. (2005). Multifrequency Analysis of 21 Centimeter Fluctuations from the Era of Reionization. ApJ, 625, 575–587.

[Sault et al., 1996] Sault, R. J., Hamaker, J. P., & Bregman, J. D. (1996). Understanding radio polarimetry. II. Instrumental calibration of an interferometer array. A&AS, 117, 149–159.

[Semiconductors, 2009] Semiconductors, A. (2009). AT84AD001C dual 8-bit 1 gsps adc data sheet.

[Smirnov, 2011a] Smirnov, O. M. (2011a). Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism. *A&A*, 527, A106.

[Smirnov, 2011b] Smirnov, O. M. (2011b). Revisiting the radio interferometer measurement equation. II. Calibration and direction-dependent effects. *A&A*, 527, A107.

[Smirnov, 2011c] Smirnov, O. M. (2011c). Revisiting the radio interferometer measurement equation. III. Addressing direction-dependent effects in 21 cm WSRT observations of 3C 147. *A&A*, 527, A108.

[So & Brodersen, 2008] So, H. K.-H. & Brodersen, R. (2008). A unified hardware/software runtime environment for FPGA-based reconfigurable computers using BORPH. *Transactions on Embedded Computing Systems*, 7(2), 1–28.

[Taylor et al., 1999] Taylor, G. B., Carilli, C. L., & Perley, R. A., Eds. (1999). *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*.

[Thomas et al., 2009] Thomas, R. M., Zaroubi, S., Ciardi, B., et al. (2009). Fast large-scale reionization simulations. *MNRAS*, 393, 32–48.

[Thompson et al., 2007] Thompson, A. R., Emerson, D. T., & Schwab, F. R. (2007). Convenient formulas for quantization efficiency. *Radio Science*, 42, 3022.

[Thompson et al., 2001] Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W., Eds. (2001). *Interferometry and Synthesis in Radio Astronomy, 2nd Edition*. Wiley.

[UniBoard, 2012] UniBoard (2012). Uniboard wiki. http://www.radionet-eu.org/fp7wiki/doku.php?id=jra:uniboard.

[Urry et al., 2007] Urry, W., Wright, M., Dexter, M., & MacMahon, D. (2007). *ATA Memo: The ATA Correlator*. Technical Report 73, University of California, Berkeley.

[van Cappellen et al., 2007] van Cappellen, W., Ruiter, M., & Kant, G. (2007). *Low Band Antenna Architectural Design Document*. Technical Report LOFAR-ASTRON-ADD-009, ASTRON.

[van de Hulst, 1982] van de Hulst, H. (1982). The origin of radio waves from space. In W. Sullivan (Ed.), *Classics in Radio Astronomy*, Studies in the History of Modern Science (pp. 302). Springer.

[Van Veen & Buckley, 1988] Van Veen, B. & Buckley, K. (1988). Beamforming: a versatile approach to spatial filtering. *ASSP Magazine, IEEE*, 5(2), 4 –24.

[Weinreb, 1963] Weinreb, S. (1963). *A Digital Spectral Analysis Technique and Its Application to Radio Astronomy*. PhD in electrical engineering, Massachusetts Institute of Technology.

[Wiener, 1930] Wiener, N. (1930). Generalized harmonic analysis. *Acta Mathematica*, 55, 117–258.

[XILINX, 2003] XILINX (2003). *Using Embedded Multipliers in Spartan-3 FPGAs.* XILINX, xapp467 (v1.1) edition.

[XILINX, 2010] XILINX (2010). *Aurora 8B/10B Protocol Specifications.* XILINX, sp002 (v2.2) edition.

[XILINX, 2011a] XILINX (2011a). *LogiCORE IP Multiplier.* XILINX, ds255 (v11.2) edition.

[XILINX, 2011b] XILINX (2011b). *Virtex-6 FPGA GTX Transceivers.* XILINX, ug366 (v2.6) edition.

[XILINX, 2012a] XILINX (2012a). *Virtex-5 FPGA User Guide.* XILINX, ug190 (v5.4) edition.

[XILINX, 2012b] XILINX (2012b). *Virtex-5 FPGA XtremeDSP Design Considerations.* XILINX, 193 (v3.5) edition.

[XILINX, 2012c] XILINX (2012c). *Virtex-6 Family Overview.* XILINX, ds150 (v2.4) edition.

[Zarka, 1998] Zarka, P. (1998). Auroral radio emissions at the outer planets: Observations and theories. J. Geophys. Res., 103, 20159–20194.

[Zaroubi, 2012] Zaroubi, S. (2012). The Epoch of Reionization. *ArXiv e-prints.*