Low-Frequency Interferometry: Design, Calibration, and Analysis Towards Detecting the Epoch of Reionization

by

Aaron Robert Parsons

B.A. (Harvard University) 2002
M.A. (University of California, Berkeley) 2006

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Astrophysics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Donald Backer, Chair
Professor Geoff Bower
Professor William Holzapfel
Professor Chung-Pei Ma

Fall 2009

The dissertation of Aaron Robert Parsons is approved:

Chair Date

Date

Date

Date

University of California, Berkeley

Fall 2009

Low-Frequency Interferometry: Design, Calibration, and Analysis Towards Detecting the Epoch of Reionization

Aaron Robert Parsons

Abstract

Low-Frequency Interferometry: Design, Calibration, and Analysis Towards Detecting the Epoch of Reionization

by

Aaron Robert Parsons
Doctor of Philosophy in Astrophysics

University of California, Berkeley

Professor Donald Backer, Chair

Low-frequency interferometry provides us with the possibility of directly observing, via red-shifted 21cm emission, the ionization of the primordial intergalactic medium by radiation from the first stars and black holes. Building such interferometers presents daunting technical challenges related to the cross-correlation, calibration, and analysis of data from large antenna arrays with wide fields-of-view in an observing band below 200 MHz.

Addressing cross-correlation data processing, I present a general-purpose correlator architecture that uses standard 10-Gbit Ethernet switches to pass data between flexible hardware modules containing Field Programmable Gate Array chips. These chips are programmed using open-source signal processing libraries developed to be flexible, scalable, and chip-independent. This work reduces the time and cost of implementing a wide range of signal processing systems, and facilitates upgrading to new generations of processing technology. This correlator architecture is supporting the incremental build-out of the Precision Array for Probing the Epoch of Reionization.

Targeting calibration concerns, I present a filtering technique that can be applied to individual baselines of wide-bandwidth, wide-field interferometric data to geometrically select regions on the celestial sphere that contain primary calibration sources. The technique relies on the Fourier transformation of wide-band frequency spectra

from a given baseline to obtain one-dimensional "delay images", and then the transformation of a time-series of delay images to obtain two-dimensional "delay/delay-rate images." These filters are augmented by a one-dimensional, complex CLEAN algorithm has been developed to compensate for data-excision effects related to the removal of radio frequency interference. This approach allows CLEANed, source-isolated data to be used to isolate bandpass and primary beam gain functions. These techniques are applied to PAPER data as a demonstration of their value in calibrating a new generation of low-frequency radio interferometers with wide relative bandwidths and large fields-of-view.

Finally, I describe PAPER's overall architecture and summarize two PAPER deployments: a 4-antenna array in of Western Australia and an 8-antenna array in Green Bank, WV. After reporting on system characterization and data analysis techniques, I present an all-sky map synthesized between 139 MHz and 174 MHz using data from both arrays that reaches down to 80 mJy (4.9 K, for a beam size of 2.15e-5 steradians at 154 MHz), with a 10 mJy (620 mK) thermal noise level that indicates what would be achievable with better foreground subtraction. I calculate angular power spectra (C_ℓ) in a cold patch and determine them to be dominated by point sources. Although the sample variance of foregrounds dominates errors in these power spectra, I measure a thermal noise level of 310 mK at $\ell = 100$ for a 1.46-MHz band centered at 164.5 MHz. This sensitivity level is approximately three orders of magnitude in temperature above the expected level of 21cm fluctuations associated with reionization.

To my partner, son, parents, brother, grandmother, aunts, uncles, and cousins. . . to all of my loving family.

Contents

List of Figures

List of Tables

Acknowledgments

Firstly, I would like to thank my parents Ken and Shirley for their love, support, and for all of the unique experiences they have given me to draw strength and direction from. Tinkering in the shop with my dad, solving math problems in the car with my mom, and growing up on a farm have been better preparation for being a radio astronomer than I could have believed. Thank you, too, to my brother Reid for his friendship through the years and for being the one person who can appreciate how far we have come from where we started and how amazing the journey has been. Thank you to my brother-in-law David for sharing a passion for astronomy and education. Thank you to my sister-in-law Emmie, for sharing her warmth and compassion, and to my parents-in-law Stanley and Diana for sharing their home and humor. And thank you to my partner Sarah, for sharing everything.

I am especially indebted to Dan Werthimer, who took a chance on me seven years ago, turned me into a radio astronomer, and set me on the path that has led to this thesis. Your generosity has provided me with more opportunities than I have known what to do with and your warmth keeps you surrounded with a dynamo of collaborators and friends that keeps generating new opportunities.

Don Backer is other great financier of my journey to becoming a bonafide astronomer. Thank you for taking me in at such an opportune time in the development of PAPER, for encouraging me to branch out in new directions, and for having confidence in me as I have sometimes struggled to find my footing.

Thank you to the other members of my thesis committee: Bill Holzapfel, Geoff Bower, and Chung-Pei Ma. Your help in making this document a reality has been

very much appreciated.

I would like to acknowledge the fantastic work and super-human efforts of the PAPER team: Don Backer, Griffin Foster, Jason Manley, Rich Bradley, Erin Benoit, Chaitali Parshare, Nicole Gugliucci, James Aguirre, Danny Jacobs, and Chris Carilli.

Thanks also to my CASPER comrades for their friendship and labors: Daniel Chapman, Henry Chen, Pierre Droz, Terry Filiba, Jason Manley, Arash Parsa, and Andrew Siemion. I would also like to recognize Jeff Mock, who inspired me with his skills as an FPGA engineer, and who we all lost far too soon.

I appreciate the great ideas, instruments, and support to come out of the Berkeley Radio Astronomy Lab through the efforts of its talented members: Matt Dexter, David MacMahon, Oren Milgrome, Lynn Urry, and Mel Wright.

Thank you to the students, professors, and staff of the Berkeley Astronomy Department for everything that makes this department one of the best and friendliest in the world. I would like to specially thank Dexter Stewart for her wonderful smile as she helped me with all things administrative. I also want to thank the scientists and staff of Arecibo Observatory, and especially Chris Salter and Tapasi Gosh, for their support, advice, and friendship during my two years en la Isla del Encanto.

I would like to acknowledge the students, faculty, and sponsors of the Berkeley Wireless Research Center (supported by NSF Infrastructure grant 0403427), and the significant efforts by staff at NRAO's Green Bank and Charlottesville sites in supporting PAPER. Thank you to David Herne, Merv Lynch, R. Beresford & T. Sweetnam of CSIRO for their aid in the PWA-4 deployment, and J. Richards and the Western Australia government for their support.

And to my friends in California, Puerto Rico, and around the world: a big THANK YOU for always being there.

Funding through graduate school was provided by the University of California, Berkeley and by the National Science Foundation through AST grants 0804508, 0505354, 0607838, and 0619596.

Chapter 1

Introduction

How and when did the universe make the transition from being predominantly neutral to highly ionized? The transition in the primordial intergalactic medium (IGM) that resulted from the formation of the first luminous objects in the universe is a key period of cosmic structure formation, and represents a major frontier of cosmic evolution yet to be explored. Recent observations of the Gunn-Peterson (GP) trough in the absorption of Lyman-α emission from the most distant quasars and of large-scale polarization in the Cosmic Microwave Background (CMB) indicate that reionization was a complex process, with significant variance in space and time, starting perhaps as high as $z < 14$, with the last vestiges of the the neutral IGM being etched-away by $z < 6$ (Fan et al. 2006; Page et al. 2007). Unfortunately, both of these ground-breaking results are limited in diagnostic capabilities: the GP effect saturates at low neutral fractions, and the CMB polarization is an integral measure of Thompson scattering back to recombination at $z \sim 1000$.

Recently, the prospect of using redshifted emission from the 21cm hyperfine transition of neutral hydrogen to trace the state of the IGM through reionization has attracted much attention (Furlanetto et al. 2006). First-generation experiments targeting a statistical detection of this signal could constrain the time, duration, and the magnitude of fluctuations caused by cosmic reionization. Such results would significantly impact our understanding of the dominant processes at work during this

period in the development of our universe. Constraining the redshift range of cosmic reionization would provide important bounds on the role of dark matter halos in speeding gravitational collapse and structure formation, while the magnitude and power spectrum of redshifted 21cm EoR fluctuations contains information about the relative contributions of massive population I stars and early quasars to heating and ionizing the IGM. Furthermore, the duration of reionization will advance our understanding of the influence that this heating and ionization exerted over the formation of new ionizing sources (Mao et al. 2008). More importantly, a first-generation detection of the EoR signal will pave the way for deeper, next-generation probes both of the IGM through the "Dark Ages" ($100 > z > 20$) and of EoR as a function of redshift and direction. Such a 3-dimensional probe would rival the CMB in its rich dependence on cosmological parameters (McQuinn et al. 2006), and would provide a unique window into the formation of structure in the early universe.

However, daunting challenges must be addressed in the design of an experiment to detect redshifted 21cm emission in the 100 to 200MHz band expected to encompass cosmic reionization. Among the challenges to be addressed is the technical challenge of correlating the large ($N > 100$) numbers of antennas needed to probe reionization on a variety of size scales, the stringent calibration requirements for removing astrophysical foregrounds such as point sources and galactic synchrotron radiation that outshine the expected EoR fluctuations by 5 to 7 orders of magnitude in temperature, and the challenge removing such foregrounds as seen through a distorting ionosphere that fluctuates on ~ 10 second timescales.

This thesis comprises work on the technical development and calibration of a low-frequency, wide-field array for detecting the Epoch of Reionization using redshifted 21cm emission, and on analyzing data from this array to explore the dominant astrophysical foregrounds that interfere with a detection of cosmic reionization.

1.1 A Scalable Correlator

A new generation of radio telescopes is achieving unprecedented levels of sensitivity and resolution, as well as increased agility and field-of-view, by employing high-performance digital signal processing hardware to phase and correlate large numbers of antennas. The computational demands of these imaging systems scale in proportion to BMN^2, where B is the signal bandwidth, M is the number of independent beams, and N is the number of antennas. The specifications of many new arrays lead to demands in excess of tens of PetaOps per second.

In Chapter 2[1], I present a general-purpose correlator architecture that was developed to meet this challenge. This architecture relies on standard 10-Gbit Ethernet switches to pass data between flexible hardware modules containing Field Programmable Gate Array (FPGA) chips. These chips are programmed using open-source signal processing libraries we have developed to be flexible, scalable, and chip-independent. This work reduces the time and cost of implementing a wide range of signal processing systems, with correlators foremost among them, and facilitates upgrading to new generations of processing technology. We present several correlator deployments, including a 16-antenna, 200-MHz bandwidth, 4-bit, full Stokes parameter application deployed on the Precision Array for Probing the Epoch of Reionization.

1.2 Calibration of Low-Frequency, Wide-Field Radio Interferometers Using Delay/Delay-Rate Filtering

In Chapter 3[2], I present a filtering technique that can be applied to individual baselines of wide-bandwidth, wide-field interferometric data to geometrically select regions on the celestial sphere that contain primary calibration sources. The tech-

[1]Chapter 2 is nearly identical to the text of Parsons et al. (2008)
[2]Chapter 3 is nearly identical to the text of Parsons & Backer (2009)

nique relies on the Fourier transformation of wide-band frequency spectra from a given baseline to obtain one-dimensional "delay images", and then the transformation of a time-series of delay images to obtain two-dimensional "delay/delay-rate images." Source selection is possible in these images given appropriate combinations of baseline, bandwidth, integration time and source location. Strong and persistent radio frequency interference (RFI) limits the effectiveness of this source selection owing to the removal of data by RFI excision algorithms. A one-dimensional, complex CLEAN algorithm has been developed to compensate for RFI-excision effects. This approach allows CLEANed, source-isolated data to be used to isolate bandpass and primary beam gain functions. These techniques are applied to data from the Precision Array for Probing the Epoch of Reionization (PAPER) as a demonstration of their value in calibrating a new generation of low-frequency radio interferometers with wide relative bandwidths and large fields-of-view.

1.3 The Precision Array for Probing the Epoch of Reionization: 8 Station Results

The Precision Array for Probing the Epoch of Reionization (PAPER) is being developed to detect 21cm emission from the early Universe, when the first stars and galaxies were forming. In Chapter 4[3], I describe the overall experiment strategy and architecture, and summarize two PAPER deployments: a 4-antenna array in the low-RFI environment of Western Australia, and an 8-antenna array at our prototyping site in Green Bank, WV. From these activities I report on system performance, including primary beam model verification, dependence of system gain on ambient temperature, measurements of receiver and overall system temperatures, and characterization of the RFI environment at each deployment site.

Chapter 4 illustrates an all-sky map synthesized between 139 MHz and 174 MHz using data from both arrays that reaches down to 10 mJy (620 mK, for a beam size

[3]Chapter 4 is nearly identical to Parsons et al. (2009)

of 2.15e-5 steradians at 154 MHz) RMS noise in the coldest parts of the northern sky. Angular power spectra (C_ℓ) calculated in this cold patch are shown to be dominated by point sources, but with signs of galactic synchrotron emission contributions at lower frequencies and angular wavemodes. Although the sample variance of foregrounds dominates errors in these power spectra, a thermal noise level of 310 mK is measured at $\ell = 100$ for a 1.46-MHz band centered at 164.5 MHz. This sensitivity level is approximately three orders of magnitude in temperature above the expected level of 21cm fluctuations associated with reionization.

1.4 Conclusion

In Chapter 5, I discuss the next steps down the path towards detecting cosmic reionization. For first-generation experiments such as PAPER, limited collecting area will force us to carefully choose an antenna configuration that optimizes sensitivity to the EoR signal relative to astrophysical foregrounds. These foreground are in the process of being characterized now, and with more information about them and our ability to model and remove them with accurate array calibration, we will soon have a better idea of how to best approach this optimization.

Finally, looking forward to second-generation reionization experiments, I enumerate some of the technical challenges that will have to be addressed with regard to signal processing instrumentation for correlation and real-time calibration and imaging. Realizing that such challenges are not unique to reionization experiments, I advocate for extending the architectural concepts of modularity and packetization presented in Chapter 2 to broader classes of digital instruments. Such solutions will hopefully take advantage of the burgeoning community that has formed around the Center for Astronomy Signal Processing and Electronics Research (CASPER) to share engineering resources among the many large radio astronomy facilities that are under development.

Chapter 2

A Scalable Correlator Architecture Based on Modular FPGA Hardware, Reuseable Gateware, and Data Packetization

2.1 Introduction

Radio interferometers, which operate by correlating the signals from two or more antennas, have many advantages over traditional single-dish telescopes, including greater scalability, independent control of aperture size and collecting area, and self-calibration. Since the first digital correlator built by Weinreb (Weinreb 1961), the processing power of these systems has been tracking the Moore's Law growth of digital electronics. The decreasing cost per performance of these systems has influenced the design of many new radio antenna array telescopes. Some next-generation array telescopes at meter, centimeter and millimeter wavelengths are: the LOw Frequency ARray (LOFAR), the Precision Array for Probing the Epoch of Reionization (PAPER), the Murchison Widefield Array (MWA), the Long Wavelength Array (LWA), the Expanded Very Large Array (EVLA), the Allen Telescope Array (ATA), the Karoo

Array Telescope (MeerKAT), the Australian Square Kilometer Array Demonstrator (ASKAP), the Atacama Large Millimeter Array (ALMA). and the Combined Array for Research Millimeter-wave Astronomy (CARMA). This paper presents a novel approach to the intense digital signal processing requirements of these instruments that has many other applications to astronomy signal processing.

While each generation of electronics has brought new commodity data processing solutions, the need for high-bandwidth communication between processing nodes has historically lead to specialized system designs. This communication problem is particularly germane for correlators, where the number of connections between nodes scales with the square of the number of antennas. Solutions to date have typically consisted of specialized processing boards communicating over custom backplanes using non-standard protocols. However, such solutions have the disadvantage that each new generation of digital electronics requires expensive and time-consuming investments of engineering time to re-solve the same connectivity problem. Redesign is driven by the same Moore's Law that makes digital interferometry attractive, and is not unique to the interconnect problem; processors such as Application-Specific Integrated Circuits (ASICs) and Field Programmable Gate Arrays (FPGAs) also require redesign, as do the boards bearing them, and the signal processing algorithms targeting their architectures.

Our research is aimed at reducing the time and cost of correlator design and implementation. We do this, firstly, by developing a packetized communication architecture relying on industry-standard Ethernet switches and protocols to avoid redesigning backplanes, connectors, and communication protocols. Secondly, we develop flexible processing modules that allow identical boards to be used for a multitude of different processing tasks. These boards are applicable to general signal processing problems that go beyond correlators and even radio science to include, e.g., ASIC design and simulation, genomics, and research into parallel processor architectures. General purpose hardware reduces the number of boards that have to be redesigned and tested with each new generation of electronics. Thirdly, we create parametrized signal processing libraries that can easily be recompiled and scaled for

each generation of processor. This allows signal processing systems to quickly take advantage of the capabilities of new hardware. Finally, we employ an extension of a Linux kernel to interface between CPUs and FPGAs for the purposes of testing and control, presenting a standard file interface for interacting with FPGA hardware.

This paper begins with a presentation of the new correlator design architecture in §2.2. The hardware to implement this architecture follows in §2.3, and the FPGA gateware used in the hardware is summarized in §2.4. Issues concerning system integration are given in §2.5, and performance characterization of subsystems are given in §2.6. Results from our first deployments of the packetized correlator are displayed in §2.7. Our final section summarizes our progress and points to a number of directions we are pursuing for the next generation of scalable correlators based on modular hardware, reuseable gateware and data packetization. An appendix gives a glossary of technical acronyms since this paper makes heavy use of abbreviated terms.

2.2 A Scalable, Asynchronous, Packetized FX Correlator Architecture

Correlators integrate the pairwise correlation between complex voltage samples from polarization channels of array antenna receivers at a set of frequencies. Once instrumental effects have been calibrated and removed, the resultant correlations (called visibilities) represent the self-convolved electric field across an aperture sampled at locations corresponding to separations between antennas. These visibilities can be used to reconstruct an image of the sky by inverting the interferometric measurement equation:

$$V_\nu(u,v) = \iint G_{i,\nu} G^*_{j,\nu} I_\nu(\ell,m) e^{-2\pi i(u\ell+vm+w(\sqrt{1-\ell^2-m^2}-1))} d\ell dm \qquad (2.1)$$

I_ν represents the sky brightness in angular coordinates (ℓ, m), and (u, v, w) correspond to the separation in wavelengths of an antenna pair relative to a pointing direction. For antennas with separate polarization feeds, cross-correlation of polarizations yields components of the four Stokes parameters that characterize polarized radiation, here

defined in terms of linear polarizations $(\parallel, \perp)$ for all pairs of antennas A and B (Rybicki & Lightman 1979):

$$I = A_\parallel B_\parallel^* + A_\perp B_\perp^* \qquad Q = A_\parallel B_\parallel^* - A_\perp B_\perp^*$$
$$U = A_\parallel B_\perp^* + A_\perp B_\parallel^* \qquad V = A_\parallel B_\perp^* - A_\perp B_\parallel^*$$

I measures total intensity, V measures the degree of circular polarization, and Q and U measure the amplitude and orientation of linear polarization.

The problem of computing pairwise correlation as a function of frequency can be decomposed two mathematically equivalent but architecturally distinct ways. The first architecture is known as "XF" correlation because it first cross-correlates antennas (the "X" operation) using a time-domain "lag" convolution, and then computes the spectrum (the "F" operation) for each resulting baseline using a Discrete Fourier Transform (DFT). An alternate architecture takes advantage of the fact that convolution is equivalent to multiplication in Fourier domain. This second architecture, called "FX" correlation, first computes the spectrum for each individual antenna (the F operation), and then multiplies pairwise all antennas for each spectral channel (the X operation). An FX correlator has an advantage over XF correlators in that the operation that scales as $O(N^2)$ with the number of antennas, N, is a complex multiplication as opposed to a full convolution in an XF correlator (D'Addario 2001; Yen 1974).

Though there are mitigating factors (such as bit-growth for representing the higher dynamic range of frequency-domain data) that favor XF correlators for small numbers of antennas (Thompson et al. 2001), FX correlators are more efficient for larger arrays. Since scalability to large numbers of antennas is one of the primary motivations of our correlator architecture, we have chosen to develop FX architectures exclusively.

2.2.1 Scalability With Number of Antennas and Bandwidth

The challenge of creating a scalable FX correlator is in designing a scalable architecture for factoring the total computation into manageable pieces and efficiently

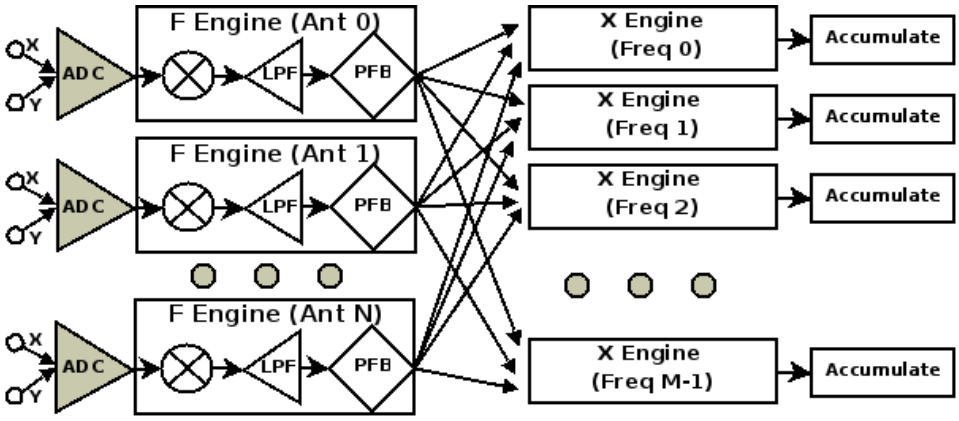

Figure 2.1: In a simplistic FX correlator, the signals from N antennas are first decomposed into M frequency channels (F operation) and then cross-multiplied (X operation). Different channels are never cross-multiplied, making them natural units for X engine processing. Thus, each X engine handles all baselines for one frequency channel.

bringing together data in each piece for computation. Traditionally, the spectral decomposition (in F engines) has been scaled to arbitrary bandwidths by using analog mixers and filters to divide the operating band of each antenna into the widest subbands that can be processed digitally using existing technology. Within correlation of a given subband, the complexities of computation and of data distribution both scale linearly with bandwidth and quadratically with the number of antennas. It is imperative that the arrangement of cross-multiplication engines (hereafter referred to as X engines) minimize data replication/retransmission, even as X engines expand to encompass many boards. Fortunately, each frequency channel of an FX correlator is computationally independent, providing a natural boundary for dividing computation among processing nodes.

Figure 2.1 illustrates a simplistic architecture for an FX correlator that takes advantage of the computational independence of channels to avoid unnecessary data transmission; the total X computation has been factored into X engines that cross-multiply all antenna pairs for a single frequency channel. This architecture is overly simplistic, since an X engine's performance can be equated to an aggregate input

bandwidth that it can handle. For the sake of efficiency, an X engine processor should receive as many channels as it has capacity to process. In this case, the number of X engines is given by:

$$\#\text{ X Engines} = \frac{(\text{Antenna Bandwidth}) \times (\#\text{ Antennas})}{\text{X Engine Processing Bandwidth}} \tag{2.2}$$

Multiplexing channels into X engines makes cross-multiplication complexity independent of the number of channels. There are three potential bottlenecks for scaling this architecture: the complexity of interconnecting F engines and X engines, the bandwidth into individual X engines, and the amount of computation in an X engine relative to the size of a processing chip/board/system. Each of these bottlenecks warrants further discussion.

The potential bottleneck of connecting N antenna-based F engines to M channel-based X engines is highlighted by the criss-crossed lines in Figure 2.1. Historically, this bottleneck has been addressed with custom backplanes and transmission protocols. However, our group has taken the novel approach of using high-performance, commercially available, 10-Gbit/s Ethernet (10GbE) switches to solve this problem. As will be discussed, these switches currently have the bandwidth and switching capacity to handle large correlators, and represent a negligible fraction of the total cost of correlator hardware. Furthermore, switching technology is driven by commercial applications and by Moore's Law, making it likely that future switches will continue increasing in number of ports and bandwidth per port.

A second potential bottleneck concerns how data rates and numbers of X engines scale with antenna bandwidth. It is important that we consider various bandwidth cases, owing to the variety of science applications driving large, next-generation systems. For example, correlators for large arrays of low-bandwidth antennas will need to multiplex data into higher bandwidth processors, while arrays with larger bandwidths will face the opposite problem. In our architecture, we make the reasonable assumption that the number of frequency channels always exceeds the number of antennas. This assumption ensures that the per-port bandwidth into an X engine never exceeds what is transmitted per antenna. Multiple channels may then be mapped into an

X engine up to its computational capacity (allowing efficient resource utilization for low-bandwidth arrays), and additional X engines may be added for high-bandwidth applications. Antenna bandwidths requiring transmission above 10 Gbits/s can be accommodated by connecting F engines to multiple 10GbE ports. Frequency channels are then assigned to each port, which connect separate switches and sub-networks of X engines. In this way, bandwidths may be scaled up to the transmission capability of an F processor by increasing the number of subnets, and not switch complexity.

The third and final potential bottleneck concerns how the sizes of individual X engines scales with the number of antennas. Both large and small numbers of antennas pose scaling problems. The size of an X engine responsible for computing all baseline cross-multiples with a fixed input data rate scales as $O(N)$, while the number of X engines required to accommodate the expanding data bandwidth with increasing numbers of antennas also scales as $O(N)$, accounting for the $O(N^2)$ scaling of computing in a correlator. For sufficiently large N, the size of an X engine can exceed the size of any processing chip or board. Our solution has been to develop an X engine whose pipelined architecture allows it to be split across multiple processors with simple point-to-point connectivity. This allows many processors to be chained together from a switch port to meet the computational demands of an X engine. Scaling to small N is equally challenging, because the aggregate correlator bandwidth decreases as $O(N)$, while computational complexity scales down as $O(N^2)$. As a result, we can find that the number of X engines that fit onto a chip/board exceeds the rate at which data can be received. The threshold where this problem is encountered can be changed by designing processors with greater connectivity, but once hardware is fixed, there is no other recourse but to accept a certain inefficiency for low numbers of antennas. While this is a fundamental limitation of our architecture, the cost of small correlators is typically dominated by development (not hardware), so a certain architectural inefficiency can be accommodated for the savings it affords in development time.

2.2.2 Globally Asynchronous Locally Synchronous Systems

Packetized data transmission simplifies the cross-connect problem inherent to correlators, but this comes at the price of global synchronicity. Packetized communication is fundamentally asynchronous: data can arrive scrambled, delayed, or not at all. Locally-synchronous X engine processing must therefore transition from being timing-driven (with throughput tied to an FPGA clock, for example) to being asynchronously data-driven. Though data buffers and control signals complicate development, Globally Asynchronous Locally Synchronous (GALS) design facilitates system integration and leads to robust design (Chapiro 1984; Plana et al. 2007). Processors run at clock rates above the data rate, using local oscillators that can drift with temperature. By allowing for non-transmission of data, individual components can fail without causing global failure–an important feature for large systems where components may fail regularly during operation. GALS design also insulates processing architectures from decisions regarding sample rates and antenna bandwidths, allowing for greater operational flexibility. Finally, individual processing elements may be redesigned and upgraded in a GALS system without affecting the overall architecture, facilitating early adoption of new technology.

Data-driven processing on locally synchronous processors like FPGAs requires controlling propagation through the processing pipeline. However, routing control signals to every multiplier, accumulator, and logic element in a pipeline can lead to excessive routing and gating demands. To avoid this, we have implemented a window-based processing architecture for algorithms where the results derived from one set of data samples are computationally independent from the next. In this architecture, processing elements are allowed to run freely at their native rate without being enabled/disabled, but are only provided data when an entire window of data has been buffered. These windows of data are provided synchronously with the inherent window boundaries of the processing element, and an entire output window is flagged as valid. Internally, a processor processes both valid and invalid data–it is only the external buffering system that keeps track of data validity. This technique is applicable to many common operators such as cross-multipliers, DFTs, and accumu-

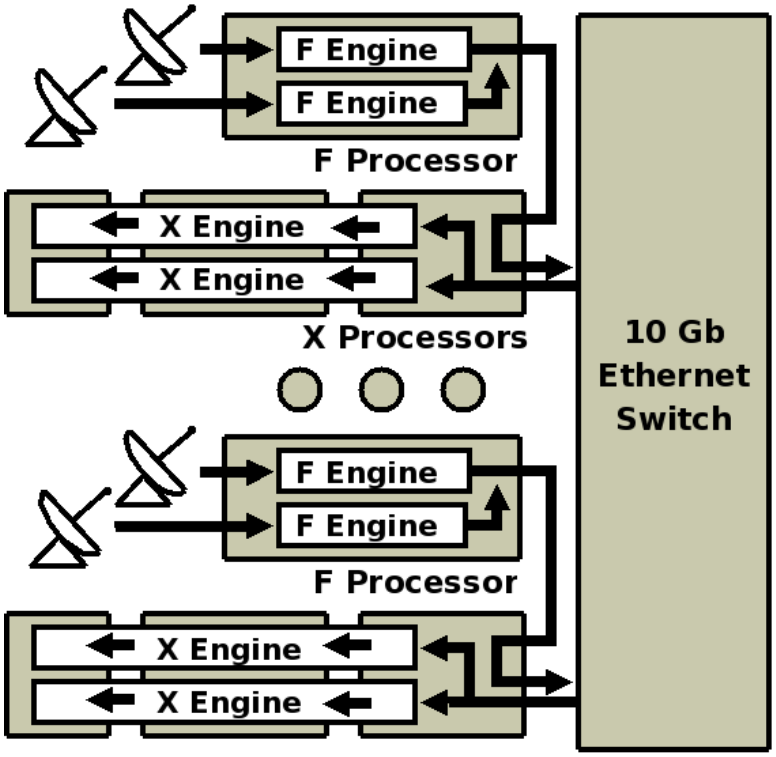

Figure 2.2: Data bandwidth per antenna is equal to the processing bandwidth of an X processor in this example application. Transmitted data is routed through an X processor to take advantage of bidirectionality of 10GbE ports, thereby halving the number of ports on the switch.

lators. Finite Impulse Response (FIR) filtering is an operation notable for not being window-based.

2.2.3 Example Applications

Perhaps the best method for demonstrating the flexibility and scalability of our correlator architecture is through example applications. To illustrate techniques for using hardware and ports efficiently, we will map processing into fictitious hardware that corresponds roughly in capability to the CASPER (Center for Astronomy Signal Processing and Engineering Research)[1] hardware discussed in Section 2.3.

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

Our first example (Fig. 2.2) illustrates an antenna signal bandwidth sufficiently low so that data from 2 polarization channels of 2 antennas can be transmitted over one 10GbE connection. Assuming that the number of antennas evenly divides the number of frequency channels, and that the processing bandwidth of an X engine matches the data bandwidth of one antenna, there will be the same number of X engines as F engines, and each X engine will receive $1/N^{th}$ of the total bandwidth, where N is the number of antennas. F engine transmission and X engine reception are combined on a single port to make use of the bi-directionality of 10GbE. This optimization halves the size of the switch needed. Multiple X processors can be chained together from a single 10GbE port using point-to-point connections. For cases where the number of antennas does not evenly divide the number of frequency channels, one can adjust packet transmission to drop remainder channels so that the band may be equally divided among X engines.

A second example (Fig. 2.3) illustrates a case where the bandwidth from a single F engine exceeds the transmission capacity of a 10GbE link. Here, data can be split by frequency channel across two ports. Since different channels are never cross-multiplied, each of these links goes to a separate subnet of switched X engines. Thus, two smaller (and often less expensive per port) switches may be substituted for one large one. Each X engine still receives the same bandwidth as in the previous example, although this now represents a smaller fraction of the total bandwidth. Note that the same X processor used in the first example functions here without modification. Only the number of X engines and the transmission pattern has changed.

A final example (Fig. 2.4) explores the case where the capacity of an X processor and a 10GbE link both exceed the data bandwidth. In this case, multiple F engines can (but do not have to) be chained together to minimize the number of switched ports. As should be the case, only half as many X engines (as compared to Fig. 2.2) are necessary for a given number of antennas. X processors operate in the same configuration as before, oblivious to changes in F engines.

These examples highlight the flexibility of the hardware and gateware for targeting a number of applications. One shortcoming they also illustrate is how the

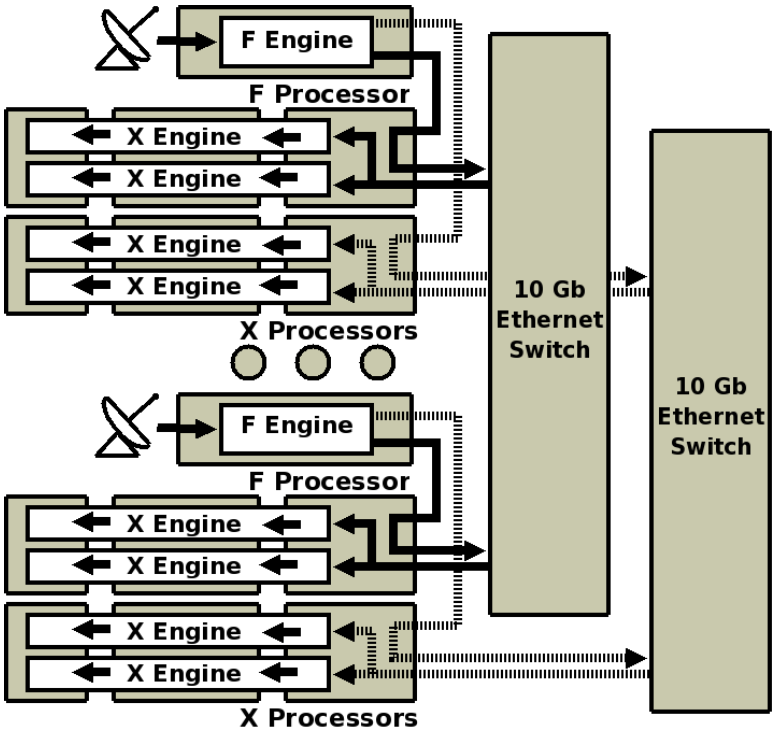

Figure 2.3: Data bandwidth per antenna can exceed what can be carried over 10GbE. Here, the frequency band has been spread across ports by channel, so that each half of transmission occurs on an isolated subnet. This is possible because different channels are never cross-multiplied in an FX correlator.

cabling between components differs for different bandwidths. Therefore the different bandwidth operations are not as easily reconfigured as might be desired for varying science goals on a given telescope. Research is ongoing to improve the rapid reconfigurability that is an essential specification for the most general radio interferometer array applications.

2.3 Modular, FPGA-based Processing Hardware

A flexible and scalable correlator architecture is of limited use without equally dynamic processing hardware that can support a variety of configurations. FPGAs provide a unique combination of flexibility and performance that make them well-

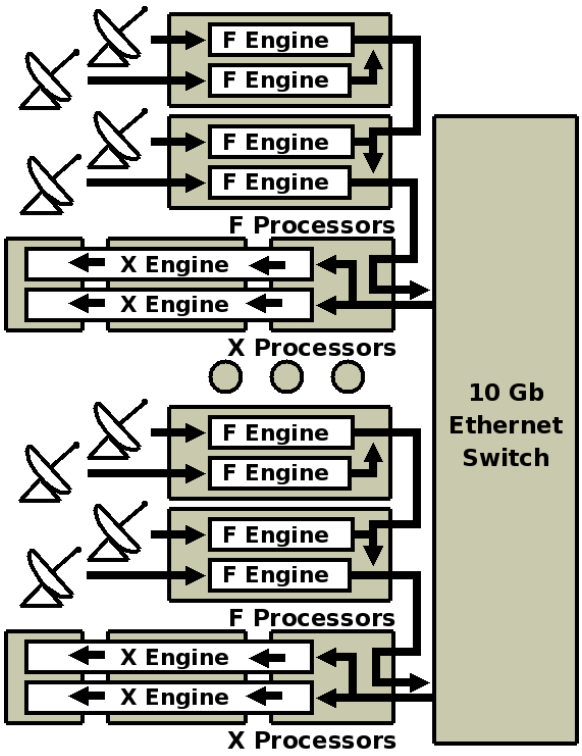

Figure 2.4: When the processing bandwidth of an X engine exceeds the antenna bandwidth by at least a factor of 2, half as many X processors are needed for a given number of antennas. X processors operate independently of data bandwidth; the same design handles this and the previous two cases (Figs. 2.2 and 2.3). Only the number of X processors and the data transmission pattern have changed.

suited for moderate-scale signal processing applications such as correlators and spectrometers (Parsons et al. 2006). A primary goal of the CASPER group has been development of multipurpose processing modules that can be of general use to the astronomy signal processing community, and beyond. We seek to minimize the effort of redesigning and upgrading hardware by modularizing processing hardware, by minimizing the number of different modules in a system, and by employing industry-standard interconnection protocols.

Hardware modularity is the idea that boards should have consistent interfaces in order to be connectible with an arbitrary number of heterogeneous components to meet the computing needs of an application ("computing by the yard"), and that

Board	Board Cost	Cost with FPGAs	Gops per Sec	Power (W)
IBOB	$400	$2700	70	30
BEE2	$5000	$23500	500	150
ROACH*	$1000	$3200	400	50
ADC (1Gs/s×2)	$200	$200	N/A	2
ADC (3Gs/s)*	$1000	$1000	N/A	5

Table 2.1: The cost, processing capability, and power consumption of CASPER hardware are summarized above. Parameters for boards flagged with a * have been estimated from prototype versions.

upgrading/revising a component does not change the way in which components are combined in the system. Minimization of hardware reproduction costs is often used to motivate the design of specialized hardware for large-scale correlators. However, the longer development times inherent to such solutions, and the necessity of targeting specific components from the outset, suggest that a modular solution, initiated nearer to the deployment date, will employ newer technology that costs less and uses less power per operation. The predicted economy of mass-producing specially-designed hardware must be tempered by its expected devaluation by Moore's Law over the course of correlator development. This devaluation makes the argument that hardware modularity can reduce the overall system cost, even for large-scale systems, by reducing development time.

In current correlator systems, we rely on two CASPER FPGA-based processing boards; Internet Break-Out Boards (IBOBs) are generally used for implementing per-antenna F engine processing, and second-generation Berkeley Emulation Engines (BEE2s) implement X engine processing. Work is progressing on a new board, the Reconfigurable Open Architecture for Computing Hardware (ROACH), that will provide a single-board solution to both F and X processing. The cost, processing capacity, and power consumption of these boards are summarized in Table 2.1.

IBOBs (Fig. 2.5) can interface to two Analog-to-Digital Converter (ADC) boards, each capable of digitizing two streams at 1 Gsamples/sec or a single stream at 2 Gsam-

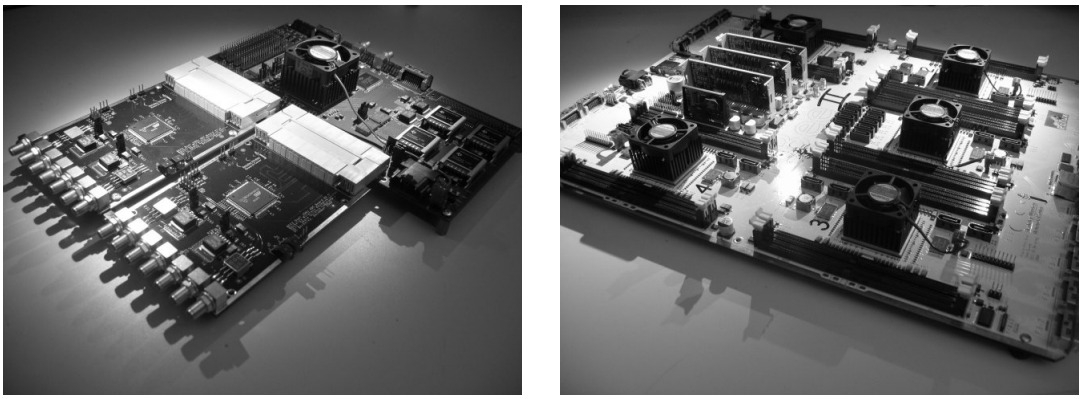

Figure 2.5: Our correlator architecture relies on modular FPGA-based processing hardware developed by our group to combine flexibility, upgradeability, and performance. Illustrated above are: (left) IBOB and ADC FPGA/digitizer modules (right) The Berkeley Emulation Engine (BEE2) FPGA board

ples/sec using an Atmel AT84AD001B dual 8-bit ADC chip. This data is processed by a Xilinx XC2VP50 FPGA containing 232 18×18-bit multipliers, two PowerPC CPU cores, and over 53,000 logic cells. Two ZBT SRAM chips provide 36 Mbits of extra buffering, and two 10GbE-compatible CX4 connectors provide a standard interface for connecting to other boards, switches, and computers. A detailed discussion of ADC signal fidelity is presented in Section 2.6. We are developing a second ADC board that allows four signal sampling at 200 Msample/sec.

The BEE2 board Chang et al. (2005) (Fig. 2.5) was originally designed for high-end reconfigurable computing applications such as ASIC design, but has been conscripted for astronomy applications in a collaboration between the BWRC[2], the UC Berkeley Radio Astronomy Laboratory, and the UC Berkeley SETI group. The 500 Gops/sec of computational power in the BEE2 is provided by 5 Xilinx XC2VP70 Virtex-II Pro FPGAs, each containing 328 multipliers, two PowerPC CPU cores capable of running Linux, and over 74,000 configurable logic cells. Each FPGA connects to 4 GB of DDR2-SDRAM, and four 10GbE-compatible CX4 connectors, and all FPGAs share a 100-Mbps Ethernet port. The size and connectivity of the BEE2 board

[2]Berkeley Wireless Research Center http://bwrc.eecs.berkeley.edu

make it suitable for implementing X engine processing in our correlator architecture.

The ROACH board is being developed in collaboration with MeerKAT and NRAO,[3] and is scheduled for release in the third quarter of 2008. It is intended as a replacement for both IBOB and BEE2 boards. A single Xilinx Virtex-5 XC5VSX95T FPGA containing 94,000 logic cells and 640 multiplier/accumulators provides 400 Gops/sec of processing power and is connected to a separate PowerPC 440EPx processor with a 1 GbE network connection. The board contains 4 GB of DDR2 DRAM and two 36Mbit QDR SRAMs, four 10GbE-compatible CX4 connectors, and two interfaces that allow the use of the current ADC boards, or a new 3 Gsamples/sec (6 Gsamples/sec dual-board interleaved) ADC. The scale, economy, and peripheral interfaces of this board will make it appropriate for both F and X engine processing, and will enable a single-board correlator architecture.

2.4 Gateware

Efficient, customizable signal processing libraries are another important component of a flexible and scalable correlator architecture. Towards this goal, our group has designed a set of open-source libraries[4] for the Simulink/Xilinx System Generator FPGA programming language. These libraries abstract chip-specific components to provide high-level interfaces targeting a wide variety of devices. Signal processing blocks in these libraries are parametrized to scale up and down to arbitrary sizes, and to have selectable bit widths, latencies, and scaling. Though the design principles of parametrization and scalability have added complexity to the initial design of these libraries, it dramatically enhances their applicability and potential for longevity as hardware evolves. It also decreases testing time by allowing developers to debug scale models of systems that derive from the same parametrization code and are behaviorally similar to larger systems. In this section, we present several components of our libraries vital to the design of flexible correlators.

[3]The National Radio Astronomy Observatory (NRAO) is owned and operated by Associated Universities, Inc. with funding from the National Science Foundation

[4]Available at http://casper.berkeley.edu

2.4.1 A Digital Down-Converter

The rising speed of ADCs has enabled digitization to occur increasingly early in the antenna receiver chain. We are thus replacing analog electronics commonly known as intermediate frequency processor (gain, band definition) and baseband mixer (conversion to zero frequency and filtering). There are numerous advantages to doing this. Digital mixing allows dynamically selecting an operating frequency within the digitized band while ensuring perfect sine-cosine phasing in the local oscillator (LO) mixing frequency. Digitizing a wider bandwidth than will be ultimately processed makes analog filtering less critical; inexpensive filters with slow roll-offs can be used, and passband rippling can be corrected. Finally, digital filtering allows flexibility and control in selecting passband shapes and adjusting fine delays. One can even split out several bands from the same signal. The issue of quantization levels and other digital artifacts needs to be carefully addressed.

Our library provides a digital down-conversion core with a runtime-selectable mixing frequency. Using a discretely sampled sine wave in an addressable lookup table, we can approximate nearly any mixing frequency by rounding a wide accumulation register (incremented every clock) to the nearest address in the lookup table. Digital sine waves have an accuracy dictated by the number of bits used to represent a value; a lookup table need only have enough samples to achieve comparable accuracy. The fact that the derivative of $\sin(x)$ reaches a maximum magnitude of 1 allows the sampling interval of a sine wave to be simply equated to the accuracy of a coefficient over that time interval. As a result, a lookup table only need be addressed with the same bit-width as the sample width to implement an arbitrary mixing frequency.

Our library also contains a decimating FIR filter. Digital filters have advantages over analog filters by being reprogrammable and by providing exact, calculable passbands. This filter is often used for suppressing harmonics of the mixing frequency and for steepening the rolloff of cheaper analog filters, but it has also been relied upon for implementing IF sub-band selection digitally. In practice, one must weigh the need for performance and flexibility against the cost of FPGA resources compared to analog filters. As an example, the response of the FIR filter used in various

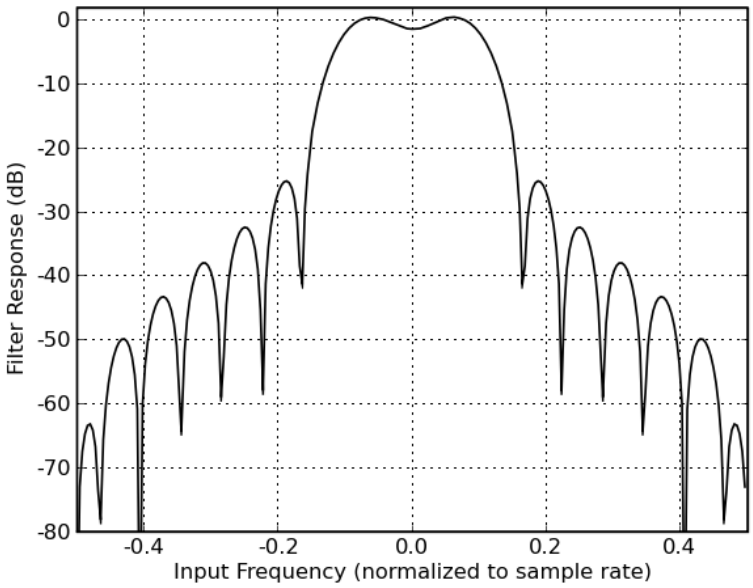

Figure 2.6: This example response an the FIR filter in a digital down-converter, illustrates the 16 tap low-pass design used in the correlator deployments presented later.

correlator designs is shown in Figure 2.6. Since the exact shape of this filter can be calculated, it is possible to remove passband ripple post-channelization because of the large dynamic range available in output of our FFT core.

2.4.2 A Polyphase Filter Bank Front-End

The Polyphase Filter Bank (PFB) (Crochiere & Rabiner 1983; Vaidyanathan 1990) is an efficient implementation of a bank of evenly spaced, decimating FIR filters. The PFB algorithm decomposes these filters into a single polyphase convolution followed by a DFT. Since DFTs have been highly optimized algorithmically, this results in an extremely efficient implementation. Equivalently, the PFB may be regarded as an improvement on the Fast Fourier Transform (FFT) that uses a front-end polyphase FIR filter to improve the frequency response of each spectral channel (Fig. 2.7). This improvement comes at the cost of buffering an additional window

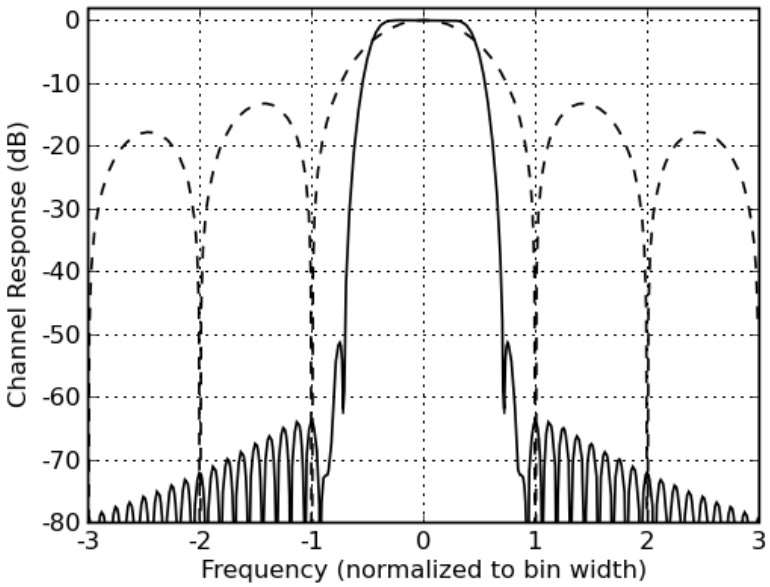

Figure 2.7: The response of a frequency channel in an 8-tap Polyphase Filter Bank (solid) using a Hamming window is compared to an equivalently sized Discrete Fourier Transform (dashed). This particular PFB, implemented for 2048 channels, is used in the correlator deployments presented in Section 2.7.

of samples and adding a complex cross-multiplication for each additional tap in the polyphase FIR. This PFB implementation has seen widespread use in the astronomy community in 21 cm hydrogen surveys (Heiles et al. 2004), pulsar surveys (Demorest et al. 2004), antenna arrays (Bradley et al. 2005), Very Long Baseline Interferometry, and other applications.

Our core is parametrized to use selectable windowing functions, allowing adjustment of the out-of-band rejection and passband ripple/rolloff. Blackman and Tukey (Blackman & Tukey 1958) provides a summary of the characteristics and trade-offs of various windows. Each polyphase FIR tap, at the cost of increased buffering and additional multipliers, increases filter steepness by adding samples (in increments of the number of channels) to the time window used in the PFB. For fixed-point implementations, a practical upper limit to the number of PFB taps is set by the number of bits used to represent filter coefficients; the sinc function's 1/x tapering ceases to

be representable when $\pi T > \pi + 2^{B+1}$ where T is the number of taps, and B is the coefficient bit width. Finally, the width of a PFB channel is tunable by adjusting the period of the sinc function, forcing adjacent bandpass filters to overlap at a point other than the -3 dB point. Note that this causes power to no longer be conserved in the Fourier transform operation.

2.4.3 A Bandwidth-Agile Fast Fourier Transform

The computational core of our FFT library is an implementation of a radix-2 biplex pipelined FFT (Rabiner & Gold 1975) capable of analyzing two independent, complex data streams using a fraction of the FPGA resources of commercial designs (Dick 2000). This architecture takes advantage of the streaming nature of ADC samples by multiplexing the butterfly computations of each FFT stage into a single physical butterfly core. When used to analyze two independent streams, every butterfly in this biplex core outputs valid data every clock for 100% utilization efficiency.

The need to analyze bandwidths higher than the native clock rate of an FPGA led us to create a second core that combines multiple biplex cores with additional butterfly cores to create an FFT that is parametrized to handle 2^P samples in parallel (Parsons 2009). This FFT architecture uses only 25% more buffering than the theoretical minimum, and still achieves 100% butterfly utilization efficiency. This feat is achieved by decomposing a 2^N channel FFT into 2^P parallel biplex FFTs of length 2^{N-P}, followed by a 2^P channel parallel FFT core using time-multiplexed twiddle-factor coefficients.

Finally, we have written modules for performing two real FFTs with each half of a biplex FFT using Hermitian conjugation. Mirroring and conjugating the output spectra to reconstitute the negative frequencies, this module effects a 4-in-1 real biplex FFT that can then be substituted for the equivalent number of biplex cores in a high-bandwidth FFT. Thus, our real FFT module has the same bandwidth flexibility as our standard complex FFT.

Dynamic range inside fixed-point FFTs requires careful consideration. Tones are folded into half as many samples through each FFT stage, causing magnitudes to grow

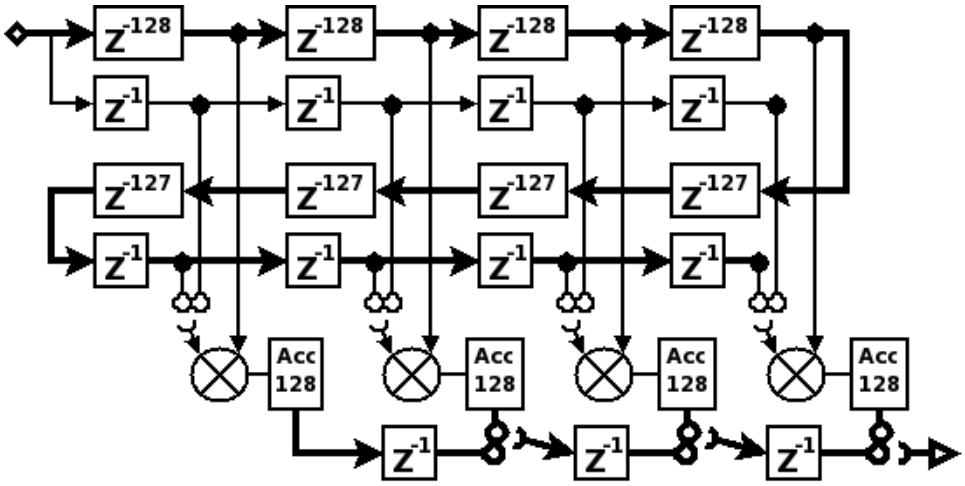

Figure 2.8: This X engine schematic illustrates the pipelined flow of data that allows it to be split across multiple FPGAs and boards. With continuous data input, all multipliers (with the possible exception of the final stage for even values of N_{ant}) are used with 100% efficiency.

by a factor of 2 for narrow-band signals, and $\sqrt{2}$ for random noise. To avoid overflow and spectrum corruption, our cores contain optional downshifts at each stage. In an interference-heavy environment, one must balance loss of SNR from downshifting signal levels against loss of integration time due to overflows. A good practice is to place time-domain input into the most-significant bits of the FFT and downshift as often as possible to avoid overflow and minimize rounding error in each butterfly stage. However, it is also best to avoid using the top 2 bits on input since the first 2 butterfly stages can be implemented using negation instead of complex multiplies, but the asymmetric range of 2's complement arithmetic can allow this negation to overflow.

2.4.4 A Cross-Multiplication/Accumulation (X) Engine

Our FX correlator architecture employs X engines to compute all antenna cross-multiples within a frequency channel, and multiple frequencies are multiplexed into the core as dictated by processor bandwidth; the complex visibility V_{ij} (Eq. 2.1) is

the average of the product of complex voltage samples from antenna i and antenna j with the convention that the voltage $j > i$ is conjugated prior to forming product. In collaboration with Lynn Urry of UC Berkeley's Radio Astronomy Lab we have implemented a parametrized module (Fig. 2.8) for computing and accumulating all visibilities for a specified number of antennas. An X engine operates by receiving N_{ant} data blocks in series, each containing T_{acc} data samples from one frequency channel of one antenna. The first samples of all blocks are cross-multiplied, and the $N_{ant}(N_{ant}+1)/2$ results are added to the results from the second samples, and so on, until all T_{acc} samples have been exhausted. Accumulation prevents the data rate out of a cross-multiplier from exceeding the input data rate. An X engine is divided into stages, each responsible for pairing two different data blocks together: the zeroth stage pairs adjacent blocks, the first stage pairs blocks separated by one, and so on. As the final accumulated results become available, they are loaded onto a shift register and output from the X engine.

However, as a new window of $N_{ant} \times T_{acc}$ samples arrives, some stages, behaving as described above, would compute invalid results using data from two different windows. To avoid this, each stage switches between cross-multiplying separations of S to separations of $N_{ant} - S$, which happen to be valid precisely when separations of S would be invalid. As a result, there need be only $floor(N_{ant}/2+1)$ stages in an X engine. Every T_{acc} samples, each stage outputs a valid result, yielding $N_{ant} \times floor(N_{ant}/2+1)$ total accumulations; for even values of N_{ant}, $N_{ant}/2$ of the results from the last stage are redundant. All other multiplier/accumulators are 100% utilized. Each stage also computes all polarization cross-multiples (Eq. 2.2) using parallel multipliers.

When one X engine no longer fits on a single FPGA, it may be divided across chips at any stage boundary at the cost of a moderate amount of bidirectional interconnect. The output shift register need not be carried between chips; each FPGA can accumulate and store the results computed locally. In order for the output shift register's $floor(N_{ant}/2+1)$ stages to clear before the next accumulation is ready, an X engine requires a minimum integration length of: $T_{acc} > floor(N_{ant}/2+1)$. In current hardware, a practical upper limit on T_{acc} is set by the 2×4 Mbit of SRAM

storage available on the IBOB. For 2048 channels with 4-bit samples, and double buffering for 2 antennas, 2 polarizations, this limit is $T_{acc} \leq 128$. Longer integration requires an accumulator capable of buffering an entire vector of visibility data, and typically occurs in off-chip DRAM. The maximum theoretical accumulation length in correlator is determined by the fringe rate of sources moving across the sky, and is a function of observing frequency, maximum antenna separation, and (for correlators with internal fringe rotation) field-of-view across the primary beam.

Cross-multiplication comes to dominate the total correlator processing budget for large numbers of antennas. As a result, care must be taken both to reduce the footprint of a complex multiplier/accumulator and to make full and efficient use of the resources on an FPGA processor. The number of bits used to carry a signal should be minimized while retaining sufficient dynamic range to distinguish signal from noise. We have chosen to focus on 4-bit multipliers in current applications, and the subjects of dynamic equalization and Van Vleck correction generalized to 4 bits are explored in Section 2.6 for optimizing signal-to-noise ratios (SNR) in our correlators. To make full use of FPGA resources, we construct 4-bit complex multipliers using distributed logic, dedicated multiplier cores, and look-up tables implemented in Block RAMs.

It is possible to perform the bulk of an N-bit complex multiply in an M-bit multiplier core by sign-extending numbers to $2N$ bits and combining them into two M-bit, unsigned numbers. Multiplying $(a + bi)(c + di)$, these representations are $(2^{M-2N}a_s + b_s)$ and $(2^{M-2N}c_s + d_s)$, where $n_s = 2^{2N} + n$. The bits corresponding to $ac, ad + bc, bd$ may be selected from the product, provided that the sign-extension to $2N$ bits shifts $a + d$ beyond the bits occupied by ad. This yields the constraint:

$$6N - 1 < M \tag{2.3}$$

The 18-bit multipliers in current Xilinx FPGAs can efficiently perform 3-bit complex multiplies, but fall short of 4 bits.

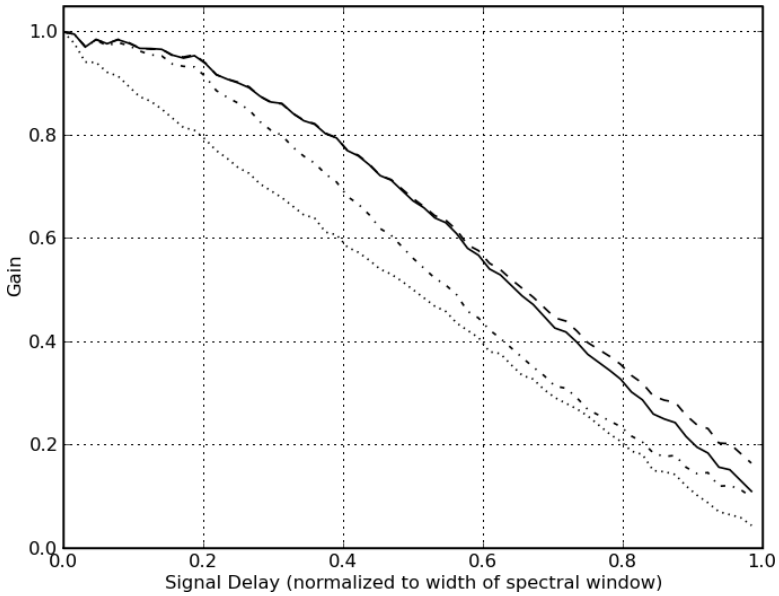

Figure 2.9: Cross-correlation of noise decreases as a function of signal delay between antenna inputs. PFBs operate on a wider window of data compared to DFTs, and use non-flat sample weightings, yielding a different correlation response versus signal delay compared to the standard result presented in Thompson et al. (2001) Thompson et al. (2001). Graphed are the responses of PFBs with 8 taps (solid), 4 taps (dashed), 2 taps (dot dashed), and the response of a DFT (dotted).

2.5 System Integration

2.5.1 F Engine Synchronization

Though we have touted GALS design principles for X engine processing, digitization and spectral processing within F engines must be synchronized to a time interval much smaller than a spectral window to avoid severe degradation of correlation response (Fig. 2.9). This attenuation effect, resulting from the changing degree of overlap of correlated signals within a spectral window, can be caused by systematic signal delay between antennas, as well as by source-dependent geometric delay; FX correlators with insufficient channel resolution experience a narrowing of the field of view related to channel bandwidth. This effect has been well explored for FX correlators employing DFTs (see Chapter 8 of Thompson et al. (2001)), but Polyphase

Filter Banks show a different response owing to a weighting function that extends well beyond the number of samples used in a DFT. Given a standard form for PFB sample weighting of $\text{sinc}\left(\frac{\pi t}{N\tau_s}\right) W\left(\frac{t}{2TN\tau_s}\right)$, where N is the number of output channels, T is the number of PFB taps, τ_s is the delay between time-domain samples, and W is an arbitrary windowing function that tapers to 0 at ± 1, the gain versus delay $G(\tau)$ of a PFB-based FX correlator is given by:

$$G(\tau) = \int_{-\infty}^{\infty} \left[\text{sinc}\left(\frac{\pi t}{N\tau_s}\right) W\left(\frac{t}{2TN\tau_s}\right)\right] \times \left[\text{sinc}\left(\frac{\pi(t-\tau)}{N\tau_s}\right) W\left(\frac{t-\tau}{2TN\tau_s}\right)\right] \, dt$$

For the purpose of F Engine synchronization, we rely on a one-pulse-per-second (1PPS) signal with a fast edge-rate provided synchronously to a bank of F processors running off identical system clocks. This signal is sampled by the system clock on each processor, and provided alongside ADC data. A slower, asynchronous "arm" signal is sent from a central node to each F engine at the half second phase to indicate that the next 1PPS signal should be used to generate the reset event that synchronizes spectral windows and packet counters. This ensures that samples from different antennas entering X engines together were acquired within one or two system clocks of one another. The degree of synchronization is determined by the difference in path lengths of 1PPS and the system clock from their generators to each F engine. This path length can be determined from celestial source observations using self-calibration, and barring temperature effects, will be constant for a correlator configuration following power-up.

2.5.2 Asynchronous, Packetized "Corner Turner"

The choice of the accumulation length T_{acc} in X engines determines the natural size of UDP packets in our packet-switched correlator architecture. For current CASPER hardware where channel-ordering occurs in IBOB SRAM, T_{acc} is constrained by the available memory to an upper limit of 128 samples for 2048-channel dual-polarization, 4-bit, complex data, yielding a packet payload of 256 bytes. A header containing 2 bytes of antenna index and 6 bytes of frequency/time index is added to each packet to enable packet unscrambling on the receive side. The frequency/time

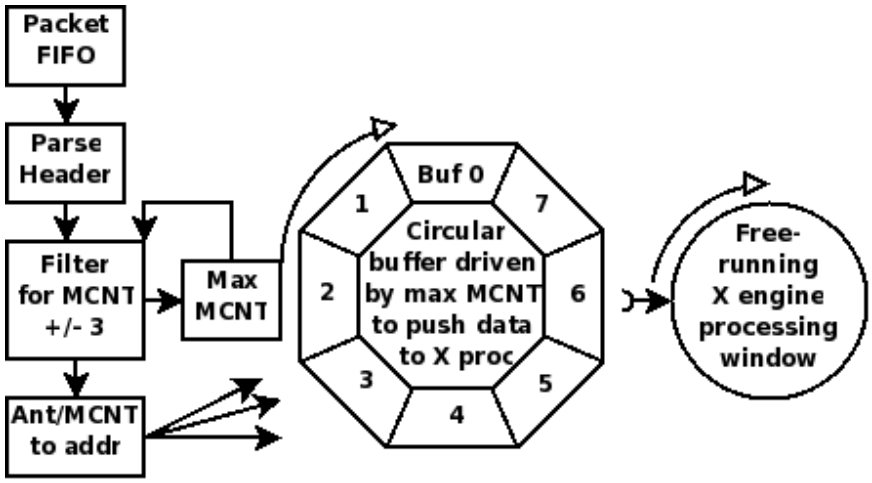

Figure 2.10: Before transmission, each F engine packet is tagged with an antenna number and master counter (MCNT) encoding time and frequency. Received packets are filtered to the narrow range of MCNTs, and maximum MCNT slides smoothly up as packets are received. A free-running X engine processes available windows when it is ready. This architecture allows data to be processed at a lower data rate than the FPGA clock rate without requiring every element in the pipeline to have a enable signal.

index (hereafter referred to as the master counter, or MCNT) is a counter that is incremented every packet transmission. The lower bits count frequencies within a spectrum, and the rest count time. Combined with the antenna index, MCNT completely determines the time, frequency, source, and destination of each packet; MCNT maps uniquely to a destination IP address.

Packet reception (Fig. 2.10) is complicated by the realities of packet scrambling, loss, and interference. A circular buffer holding N_{win} windows worth of X engine data stores packet data as they arrive. The lower bits of MCNT act as an address for placing payloads into the the correct window, and the antenna index addresses the position within that window. When data arrives $N_{win}/2$ windows ahead of a buffered window, that window is flagged for readout, and is processed contiguously on the next window boundary of the free-running X engine. Using packet arrival to determine when a window is processed allows a data-rate dependent time interval for all packets to arrive, but pushes data through the buffer in the event of packet loss. On readout,

the buffer is zeroed to ensure that packet loss results in loss of signal, rather than the introduction of noise. F engines can be intentionally disconnected from transmission without compromising the correlation of those remaining.

Packet interference occurs when a well-formed packet contains an invalid MCNT as a result of switch latency, unsynchronized F engines, or system misconfiguration. Such packets must be prevented from entering the receive buffer, since they can lead to data corruption; one would prefer that a misconfigured F engine antenna result in data loss for that antenna, rather than data loss for the entire system. To ensure this behavior, incoming packets face a sliding filter based on currently active MCNTs. Packets are only accepted if their MCNT falls within the range of what can currently be held in the circular buffer. As higher MCNTs are received and accepted, old windows are flagged for read out, freeing up buffer space for still higher MCNTs. This system forces MCNTs to advance by small increments and prevents the large discontinuities indicative of packet interference. In the eventuality that a receive buffer accidentally locks onto an invalid MCNT from the outset, a time-out clause causes the currently active MCNT to be abandoned for a new one if no new data is accepted into the receive buffer.

A final complication comes when implementing a bidirectional 10GbE transmission architecture such as the one outlined in Figure 2.2. Commercial switches do not support self-addressed packet transmission; they assume that the transmitter (usually a CPU) intercepts these packets and transfers them to the receive buffer. On FPGAs, this requires an extra buffer for holding "loopback", and a multiplexer for inserting these packets into the processing stream. A simple method for this insertion would be to always insert loopback packets, if available, and otherwise to insert packets from the 10GbE interface. However, there is a maximum interval over which packets with identical MCNTs can be scrambled before the receive system rejects packets for being outside of its buffer. This simple method has the undesirable effect of including switch latency in the time interval over which packets are scrambled, causing unnecessary packet loss. Our solution is to pull loopback packets only after packets with the same MCNT arrive through the switch.

2.5.3 Monitor, Control, and Data Acquisition

The toolflow we have developed for CASPER hardware provides convenient abstractions for interfacing to hardware components such as ADCs, DRAM, and 10 GbE transceivers, and allows specified registers and BRAMs to be automatically connected to CPU-accessible buses. On top of this framework, we run BORPH–an extension of the Linux operating system that provides kernel support for FPGA resources (So & Brodersen 2006; So 2007). This system allows FPGA configurations to be run in the same fashion as software processes, and creates a virtual file system representing the memories and registers defined on the FPGA. Every design compiled with this toolflow comes equipped with this real-time interface for low- to moderate-bandwidth data I/O. By emulating standard file-I/O interfaces, BORPH allows programmers to use standard languages for writing control software. The majority of the monitor, control, and data acquisition routines in our correlators are written in C and Python. For 8-16 antenna correlators, the bandwidth through BORPH on a BEE2 board is sufficient to support the output of visibility data with 5-10s integrations.

For correlators with more antennas or shorter integration times, the bandwidth of the CPU/FPGA interface is incapable of maintaining the full correlator output. This limitation is being overcome by transmitting the final correlator output using a small amount of the extra bandwidth on the 10GbE ports already attached to each X engine. After accumulation in DRAM, correlator output is multiplexed onto the 10GbE interface and transmitted to one or more Data Acquisition (DA) systems attached to the central 10GbE switch. These systems collect and store the final correlator output. With a capable DA system, the added bandwidth through this output pathway can be used to attain millisecond integration times, opening up opportunities for exploring transient events and increasing time resolution for removing interference-dominated data.

The capabilities of correlators made possible by our research are placing new challenges on DA systems (Wright 2005). There is a severe (factor of 100) mismatch between the data rates in the on-line correlator hardware and those supported by the off-line processing. Members of our team are currently pursuing research on how

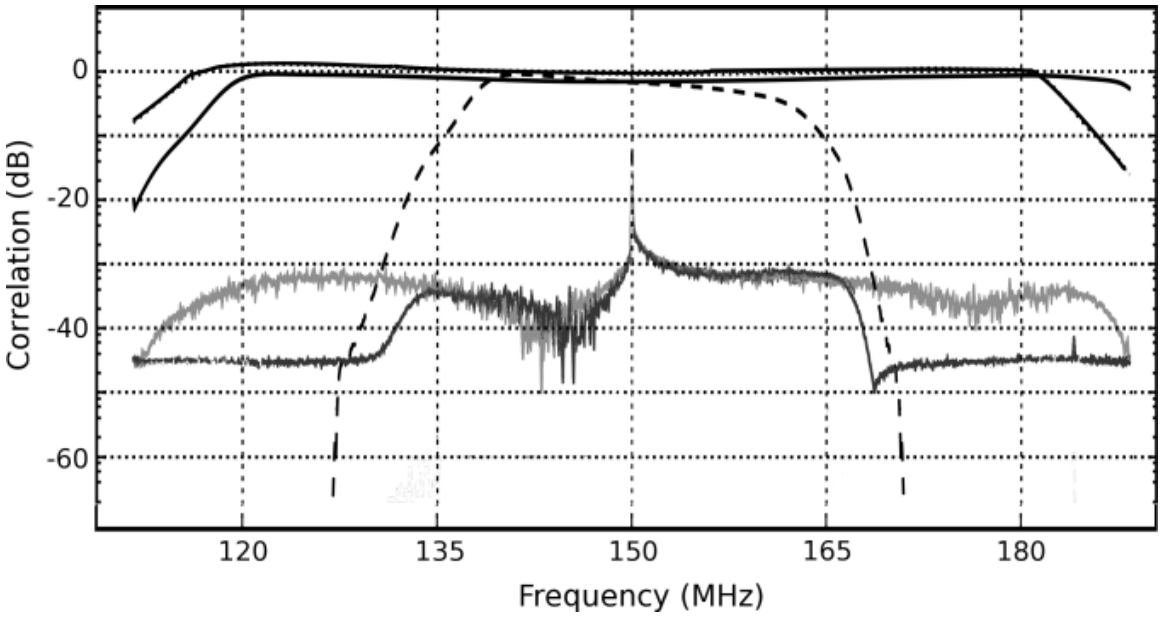

Figure 2.11: Uncorrelated noise sources with similar bandpass shapes were input to two channels of one ADC board (solid black) and a third noise source with a narrower passband was input to to a second ADC board (dashed black) in the "Pocket Correlator" system. Crosstalk levels between signal inputs on the same ADC board (light gray) and between ADC boards sharing an IBOB (dark gray) peak at −28 dB.

this can be resolved both for correlators and for generic signal processing systems using commercially available compute clusters. For correlators, our group is currently exploring how to implement calibration and imaging in real-time to reduce the burden of expert data reduction on the end user, and to make best use of both telescope and human resources.

2.6 Characterization

2.6.1 ADC Crosstalk

Crosstalk is an undesirable but prevalent characteristic of analog systems wherein a signal is coupled at a low level into other pathways. This can pose a major threat

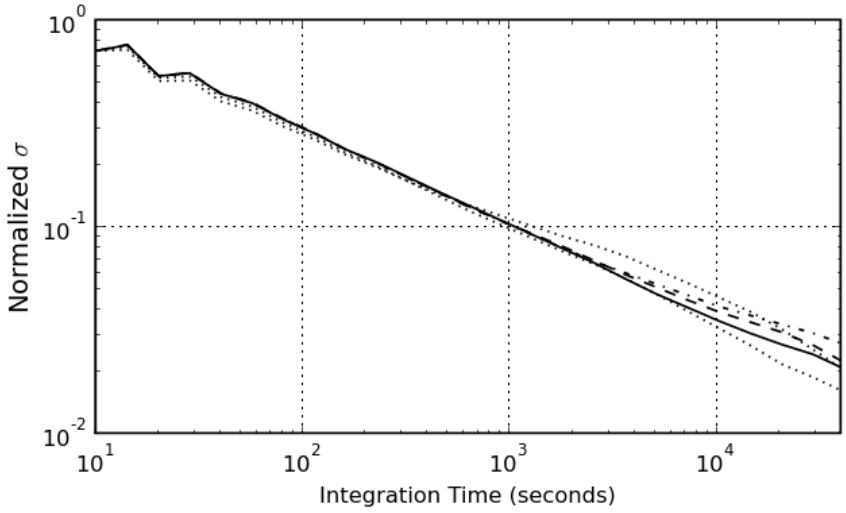

Figure 2.12: Measurements of the standard deviation versus integration time of the correlation between independent noise sources into the same ADC board show that crosstalk exhibits stability over a period of 1 day for all frequency channels Although phase switching may still be desireable, this stability allows crosstalk to be calibrated and removed after correlation.

to sensitivity in systems that integrate noise-dominated data to reveal low-level correlation. For CASPER hardware, we have examined crosstalk levels between signal inputs sharing an ADC chip, and between different ADC boards on the same IBOB. Figure 2.11 illustrates a one-hour integration of uncorrelated noise of various bandwidths input to the "Pocket Correlator" system (see Section 2.7). Between inputs of the same ADC board, a coupling coefficient of ~ 0.0016 indicates crosstalk at approximately -28 dB. This coupling is a factor of 5 higher than the -35 dB isolation advertised by the Atmel ADC chip, and is most likely the result of board geometry and shared power supplies. Crosstalk between inputs on different ADCs also peaks at the -28 dB level, but shows more frequency-dependent structure.

Crosstalk may be characterized and removed, provided that its timescale for variation is much longer than the calibration interval. Figure 2.12 demonstrates that for integration intervals ranging from 7.15 seconds to approximately 1 day (the limit

of our testing), crosstalk amplitudes and phases vary around stable values in a lab test that, when subtracted, yield noise that integrates down with time. Even though crosstalk is encountered at the −28 dB level, its stability allows suppression to at least −62 dB. This stability has allowed crosstalk to be removed post-correlation, and we have until recently deferred adding phase switching. Developments along this line are proceeding by introducing an invertible mixer (controlled via a Walsh counter on an IBOB) early in the analog signal path, and removing this inversion after digitization. Phase switching must be coupled with data blanking near boundaries when the inversion state is uncertain. Blanking will be most easily implemented by intentionally dropping packets of data from F engine transmission, and by providing a count of results accumulated in each integration for normalization purposes.

2.6.2 XAUI Fidelity and Switch Throughput

CASPER boards are currently configured to transmit XAUI protocol over CX4 ports as a point-to-point communication protocol and as the physical layer of 10GbE transmission. Because the Virtex-II FPGAs used in current CASPER hardware do not fully support XAUI transmission standards xil (2004, 2005), current devices can have sub-optimal performance for certain cable lengths. We expect the new ROACH board, which employs Virtex-5 FPGAs, to have better performance in this regard. For cable lengths supported in current hardware, we tested XAUI transmission fidelity using matched Linear Feedback Shift Registers (LFSRs) on transmit and receive. Error detection was verified using programmable bit-flips following transmitting LFSRs. Over a period of 16 hours, 573 Tb of data were transmitted and received on each of 8 XAUI links. During this time, no errors were detected, resulting in an estimated bit-error rate of $2.2 \cdot 10^{-16}$ Hz. We also tested the capability of two Fujitsu switches (the XG700 and the XG2000) for performing the full cross-connect packet switching required in our FX correlator architecture. By tuning the sample rate inside F engines of an 8-antenna (4-IBOB) packetized correlator, we controlled the transmission rate per switch port over a range of 5.96 to 8.94 Gb/s. In 10-minute tests, packet loss was zero for both switches in all but the 8.94 Gb/s case. Packet loss in this final case

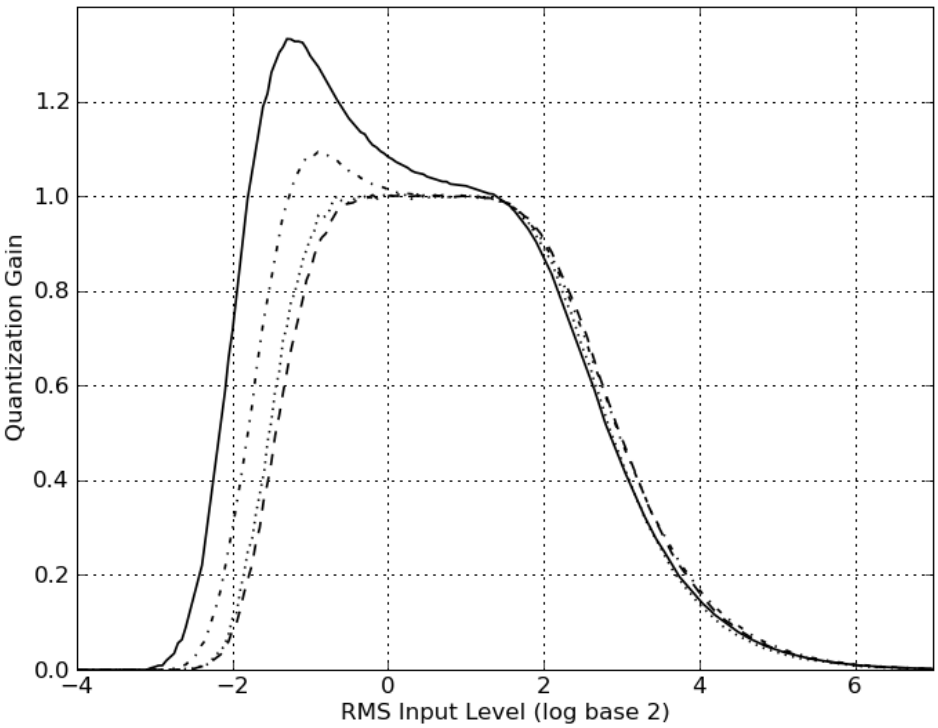

Figure 2.13: Illustrated above is the relative gain through a 4-bit, 15-level quantizer as a function of input signal level (log base 2). Plotted are gain curves for the cross-correlation of two gaussian noise sources with correlation levels of 100% (solid), 80% (dot-dashed), 40% (dotted), and 20% (dashed).

was traced to intermittent XAUI failure as a result of imperfect compliance with the XAUI standard, as described previously. Overheating of FPGA chips in the field has also been reported as a source of intermittent operation.

2.6.3 Equalization and 4-Bit Requantization

Correlator processing resources can be reduced by limiting the bit width of frequency-domain antenna data before cross-multiplication. However, digital quantization requires careful setting of signal levels for optimum SNR and subsequent calibration to a linear power scale (Thompson et al. 2001; Jenet & Anderson 1998). Correlators using 4 bits represent an improvement over their 1 and 2 bit predeces-

sors, but there are still quantization issues to consider. The total power of a 4-bit quantizer has a non-linear response with respect to input level as shown in Figure 2.13. In currently deployed correlators, we perform equalization (per channel scaling) to control the RMS channel values before requantizing from 18 bits to 4 bits. This operation saturates RFI and flattens the passband to reduce dynamic range and to hold the passband in the linear regime of the 4-bit quantization power curve. Equalization is implemented as a scalar multiplication on the output of each PFB using 18-bit coefficients from a dynamically updateable memory. These coefficients allow for automatic gain control to maintain quantization fidelity through changing system temperatures.

2.7 Deployments and Results

2.7.1 A Pocket Correlator

The "Pocket Correlator" (Fig. 2.14) is a single IBOB system that includes F and X engines on a single board for correlating and accumulating 4 input signals. Each input is sampled at 4 times the FPGA clock rate (which runs up to 250 MHz), and a down-converter extracts half of the digitized band. This subband is decomposed into 2048 channels by an 8-tap PFB, equalized, and requantized to 4 bits. With all input signals on one chip, X processing can be implemented directly as multipliers and vector accumulators, rather than as X engines. Limited buffer space on the IBOB permits only 1024 channels (selectable from within the 2048) to be accumulated. Output occurs either via serial connection (with a minimum integration time of 5 seconds) or via 100-Mbit UDP transmission (with a minimum integration time in the millisecond range). This system can act as a 2-antenna, full Stokes correlator, or as a 4-antenna single polarization correlator.

The Pocket Correlator is valuable as a simple, stand-alone instrument, and for board verification in larger packetized systems. It is being applied as a stand-alone instrument in PAPER, the ATA, and the UNC PARI observatory. A 4-antenna,

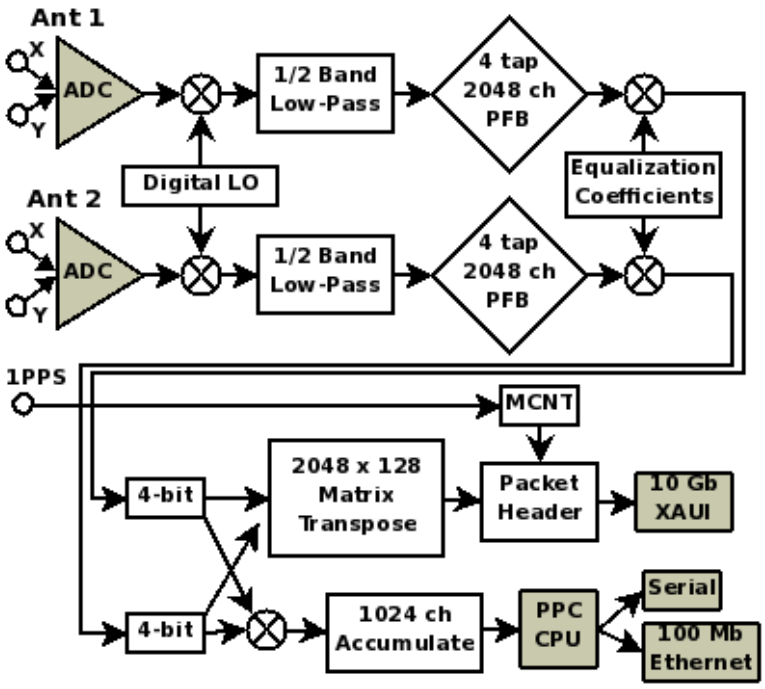

Figure 2.14: This IBOB design serves a dual purpose as a stand-alone "Pocket Correlator" and an F processor in a 16 antenna packetized correlator deployment. Note the parallel output pathways for each function.

single polarization deployment of the PAPER experiment in Western Australia in 2007 used the Pocket Correlator to collect the data used to produce a 150 MHz all-sky map illustrated in Figure 2.15. In addition to demonstrating the feasibility of post-correlation crosstalk removal, this map (specifically, the imperfectly removed sidelobes of sources) illustrates a problem that will require real-time imaging to solve for large numbers of antennas.

2.7.2 An 8-Antenna, 2-Stokes, Synchronous Correlator

This first generation multi-board correlator demonstrated the functionality of signal processing algorithms and CASPER hardware, but preempted the current packetized architecture–it operated synchronously. This version of the correlator was most heavily limited by X engine resources, all of which were implemented on a single

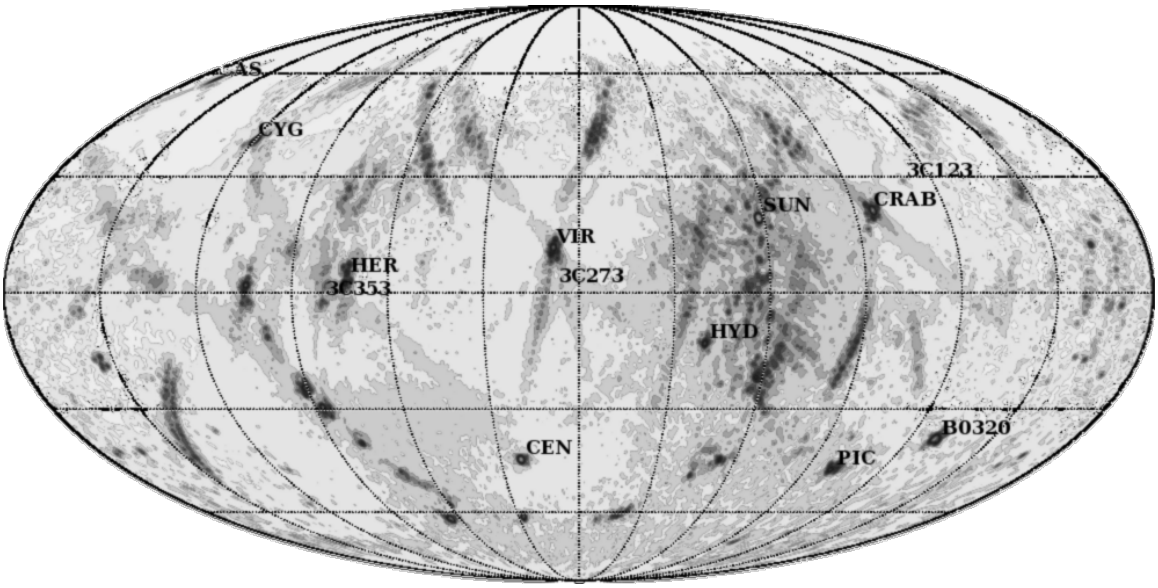

Figure 2.15: This all-sky image, made using a 75-MHz band centered at 150 MHz with the "Pocket Correlator" as part of the PAPER experiment in Western Australia, achieves an impressive 10,000:1 signal-to-noise ratio using 1 day of data.

FPGA to simplify interconnection. The total number of complex multipliers in the X engines of an N_{ant} antenna array is: $N_{cmac} = floor(N_{ant}/2 + 1) \times N_{ant} \times N_{pol}$; the limited number of multipliers on a BEE2 FPGA only allowed for supporting half the polarization cross-multiples. This system was an important demonstration of the basic capabilities of our hardware and software, and provided a starting point for evolving a more sophisticated system. Deployments of this system at the NRAO site in Green Bank as part of the PAPER experiment, and briefly at the Hat Creek Radio Observatory for the Allen Telescope Array, are being supersede by the packetized correlator presented in the next section.

2.7.3 A 16-Antenna, Full-Stokes, Packetized Correlator

This packetized FX correlator is a realization of the architecture outlined in Figure 2.2, with F processing for 2 antennas implemented on each IBOB, and matching X processors implemented on each corner FPGA of two BEE2s. Each F processor

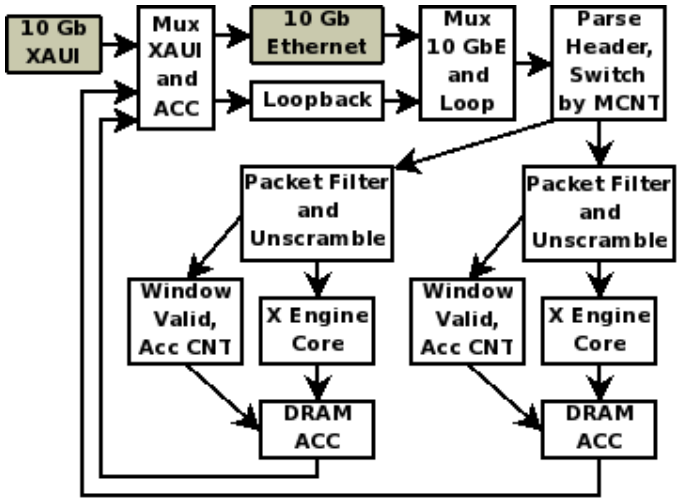

Figure 2.16: A BEE2-based X processor in a packetized correlator transmits data from an F engine over 10GbE and stores self-addressed packets in a "loopback" buffer. These streams are merged on the receive side, and packets are distributed to two X engines. Accumulation occurs in DRAM buffers, and the results are packetized and output over the same 10GbE link. A data aquisition system connects to the same switch as the X engines.

is identical to a Pocket Correlator (Fig. 2.14), but branches data from the equalization module to a matrix transposer in IBOB SRAM to form frequency-based packets. Packet data for each antenna are multiplexed through a point-to-point XAUI connection to a BEE2-based X processor, and then relayed in 10GbE format to the switch. The number of channels in this system is limited to 2048 by memory in IBOB SRAM for transposing the 128 spectra needed to meet bandwidth restrictions between X engines and DRAM-based vector accumulators.

The X processor in this packetized correlator implements the transmit and receive architecture illustrated in Figure 2.16 for two X engines sharing the same 10GbE link. Each X engine's data processing rate is determined by packets arriving in its own receive buffer, and results are accumulated in separate DRAM DIMMs. The accumulated output of each X processor is read out of DRAM at a low bandwidth and transmitted via 10GbE packets to a CPU-based server where all visibility data is collected and written to disk in MIRIAD format (Sault et al. 1995) using interfaces

from the Astronomical Interferometry in PYthon (AIPY) package[5].

The clocks for the BEE2 FPGAs are asynchronous 200-MHz oscillators, and IBOBs run synchronously at any rate lower than this. Packet transmission is statically addressed so that all each X engine processes every 16th channel. We use 8 ports of a Fujitsu XG700 switch to route data. This system is is scalable to 32 antennas before two X engines no longer fit on a single FPGA. For larger systems, the number of BEE2s will scale as the square of the number of antennas, and the number of IBOBs will scale linearly. A 32-antenna, 200-MHz correlator on 16 IBOBs and 4 BEE2s is now working in the lab, and a 16-antenna version using 8 IBOBs and 2 BEE2s has been deployed to the NRAO site in Green Bank with the PAPER experiment.

2.8 Conclusion

By decreasing the time and engineering costs of building and upgrading correlators, we aim to reduce the total cost of correlators for a wide range of scales. Small- and medium-scale correlators with total cost dominated by development clearly stand to benefit from our research. It is less clear if the cost of large-scale correlators can be reduced by the general-purpose hardware used in our architecture. Though minimization of replication cost favors the development of specialized parts, there are two factors that can make a generic, modular solution cost less.

The first factor to consider is time to deployment. Even if the monetary cost of development is negligible in the budget of a large correlator, the cost of development time can be significant. If a custom solution takes several years to go from design to implementation, the hardware that is deployed will be out of date. Moore's Law suggests that when a custom solution taking 3 years to develop is deployed, there will exist processors 4 times more powerful, or 4 times less expensive for the equivalent system. The cost of a generic, modular system has to be tempered by the expected savings of committing to hardware closer to the ultimate deployment date.

The second factor is the cost of upgrade. Many facilities (including the ATA) are

[5]http://pypi.python.org/pypi/aipy

beginning to appreciate the advantages of designing arrays with wider bandwidths and larger numbers of antennas than can be handled by current technology. Correlators may then be implemented inexpensively on scales suited to current processors, and upgraded as more powerful processors become available. Modular solutions facilitate this methodology.

2.9 Glossary of Technical Terms

- ADC - Analog to Digital Converter
- ASIC - Application-Specific Integrated Circuit processor
- BEE2 - Berkeley Emulation Engine, rev. 2
- BORPH - Berkeley Operating system for Re-Programmable Hardware
- BRAM - Block RAM: Random Access Memory inside an FPGA
- CX4 - 10GbE-compatible industry standard connector
- CPU - Central Processing Unit
- DDR2 - Double-Data-Rate 2 type of off-FPGA Synchronous DRAM
- DIMM - Dual Inline Memory Module
- DFT - Discrete Fourier Transform
- DRAM - Dynamic Random Access Memory
- FFT - Fast Fourier Transform algorithm
- FIR - Finite Impulse Response digital filter
- FPGA - Field Programmable Gate Array processor
- FX - Correlator architecture implemented as frequency channelization, then cross-multiplication

- GALS - Globally Asynchronous, Locally Synchronous system architecture
- GB - GigaByte
- IBOB - Internet Break-Out Board
- LFSR - Linear Feedback Shift Register
- LO - Local Oscillator
- MCNT - Master Counter
- PFB - Polyphase Filter Bank
- PowerPC - a specific CPU architecture
- QDR - Quad-Data-Rate type of off-FPGA SRAM
- ROACH - Reconfigurable, Open Architecture for Computing Hardware
- SNR - Signal-to-Noise Ratio
- SRAM - Static Random Access Memory
- UDP - User Datagram Protocol Ethernet packetization
- XAUI - X (ten) Attachment Unit Interface point-to-point transmission protocol
- XF - Correlator architecture implemented as cross-multiplication, then frequency channelization
- 1PPS - 1 Pulse Per Second clock signal
- 10GbE - 10 Gigabit per second Ethernet communication standard

Chapter 3

Calibration of Low-Frequency, Wide-Field Radio Interferometers Using Delay/Delay-Rate Filtering

3.1 Introduction

The rapid growth of the capabilities of digital signal processing is enabling a new generation of interferometric arrays based on large numbers of antennas and/or wide instantaneous frequency coverage. Current examples of such arrays include the Expanded Very Large Array[1] (EVLA), the Allen Telescope Array[2] (ATA), the LOw Frequency ARray[3] (LOFAR), the Precision Array for Probing the Epoch of Reionization[4] (PAPER), the Murchison Widefield Array[5] (MWA), the Long Wavelength Array[6] (LWA), the Karoo Array Telescope[7] (KAT), and the Australia Square Kilometer Array Prototype[8] (ASKAP). The increase in the number of elements in these

[1] http://www.aoc.nrao.edu/evla
[2] http://ral.berkeley.edu/ata
[3] http://www.lofar.org
[4] http://astro.berkeley.edu/~dbacker/eor
[5] http://haystack.mit.edu/ast/arrays/mwa
[6] http://lwa.unm.edu
[7] http://www.kat.ac.za
[8] http://www.atnf.csiro.au/projects/askap

arrays is a result of requirements for larger collecting areas and a trend towards smaller individual antenna elements. This trend reflects an evolving curve of array cost versus antenna element size (Weinreb & D'Addario 2001) whose minimum is shifting towards smaller antennas (with diameter d) as the cost of array correlation (scaling as $\sim$d^{-4}) drops relative to the materials cost of producing the array antennas (scaling approximately as $\sim$d $^{0.7}$).

The smaller individual elements of new arrays have larger fields-of-view (FoVs) that result in faster surveying speeds. However, when many parameters are poorly known, a larger FoV complicates early array calibration by decreasing the extent to which a single source dominates the correlated flux between antennas. Without isolation of sources, self-calibration cannot be performed as a direct computation using raw data, but rather must rely on a priori models of the sky and primary beam response pattern to divide out the baseline-dependent interference pattern (Cornwell & Fomalont 1989). This interference pattern is illustrated by the sum over sources in the basic measurement equation for interferometric response to a set of point sources:

$$V_{ij}(\nu,t) = G_{ij}(\nu,t)\sum_n A_{ij}(\nu,\vec{s}_n(t))S_n(\nu)e^{2\pi\nu i(\tau_{g,ijn}(\nu,t)+\tau_{e,ij})} \tag{3.1}$$

where i,j are antenna indices, ν is radio frequency, t is time, G is the complex frequency-dependent electronics gain, A is the antenna beam gain in the source direction $\vec{s}$ with unit normalization toward the zenith, S is the source flux, τ_g is the geometric delay for baseline i,j in the direction $\vec{s}_n$, and τ_e is the non-geometric, relative electrical path delay. Both G,A and τ_g,τ_e are typically factorized into antenna-based gains and delays, respectively. Accurate solutions for the internal degrees of freedom in an array, especially for such parameters as the spatial variation of antenna beam gains, requires access to a variety of calibration sources. Wide-field arrays have the problem that their FoV nearly always includes sources so bright that their sidelobes conceal lesser sources useful for calibration. The need to remove sources whose phase and amplitude solutions are not of interest in order to access sources that are useful for calibration can be a time-consuming and distracting process.

However, advances in feed design and processing bandwidth are also increasing

the amount of frequency data available in the latest interferometers. Wide-bandwidth data with sufficient channel resolution make the delay transform–the Fourier transform of a frequency spectrum–a powerful tool defining delay patterns (see §2.2 of Thompson et al. (2001)) that separate sources on the sky. In the following sections, we discuss techniques for using the delay (D) transform and its analog along the time axis, the delay-rate (DR) transform, to separate the fluxes of strong celestial sources. Through the construction of delay/delay-rate (DDR) filters, we demonstrate how sources may be isolated to facilitate self-calibration in wide-FoV, wide fractional bandwidth interferometers. Following the development of these calibration techniques, we demonstrate their application to data from PAPER (Bradley et al. 2005), a low-frequency, non-phase-tracking, dipole array whose steradian FoV and 100-MHz bandwidth motivated this work.

3.2 The Delay (D) Transform

The frequency spectrum of visibilities on a measured baseline as a function of time reflects an interference pattern between the complex vectors corresponding to each coherently added point source in the primary beam (Eq. 3.1). As discussed in the previous section, the calibration process may be significantly simplified provided that the several interfering components present in the data of each channel may be separated from one another. This so-called "source separation" should ideally be tunable in its precision, so that coarse separation may be achieved using poorly characterized parameters, and increasingly accurate calibration improves the achievable separation. For a single baseline, there are only two parameters available for separating sources: frequency and time. In the next two sections, we will discuss techniques for using both of these parameters to separate the flux of point sources.

Within the spectrum of a baseline at a given time, each source exhibits a linearly varying phase versus frequency, reflecting the geometric group delay associated with the projection of the baseline in the direction of the source.

$$\tau_g(\nu, t) \equiv \frac{b_x}{c} \cos\delta \cos H(t) - \frac{b_y}{c} \cos\delta \sin H(t) + \frac{b_z}{c} \sin\delta \tag{3.2}$$

where (b_x, b_y, b_z) are baseline components with units of length in the radial, eastern, and northern polar directions, respectively, δ is the source declination, and $H \equiv h - \alpha$ is the source hour angle as a function of sidereal time h. The geometric delay is frequency-independent, and can be extracted using a Fourier transform between frequency (F)-domain and delay (D)-domain coordinates:

$$\begin{aligned}\hat{V}_{ij}(\tau, t) &= \int_{-\infty}^{\infty} G_{ij}(\nu, t) \left[\sum_n A_{ij}(\nu, \vec{s}_n(t)) S_n(\nu) e^{2\pi\nu(\tau_{g,ijn} + \tau_{e,ij})} \right] e^{-2\pi i \nu \tau} \, d\nu \\ &= \hat{G}_{ij}(\tau, t) * \sum_n \left[\hat{A}_{ij}(\tau, \vec{s}_n(t)) * \hat{S}_n(\tau) * \delta_D \left(\tau_{g,ijn} + \tau_{e,ij} - \tau \right) \right]\end{aligned} \tag{3.3}$$

As illustrated above, this D transform maps the flux from each interfering source to the corresponding delay, which will typically include a systematic, non-geometric delay $\tau_{e,ij}$ owing to the relative electrical signal path delays between antennas i and j. While this procedure is effective as a first step in source separation, the D transform does not result in a one-to-one mapping of the celestial sphere to delay coordinates; sources that lie in a plane perpendicular to the baseline vector share the same geometric delay (see Fig. 3.1). Furthermore, frequency-dependent interferometer gains create a convolution kernel that spreads the gain of a source in D domain, resulting in an effective delay resolution. For a flat passband, this resolution is approximately related to the bandwidth sampled by the interferometer $\Delta\tau \sim \frac{1}{\Delta\nu}$, and translates to a ring of finite width at the intersection of the celestial sphere with a plane of constant delay. Given the finite D-domain resolution, we will hereafter assume that $\hat{V}_{ij}(\tau, t)$ is sampled in "delay bins" of width $\Delta\tau$.

The ring on the sky defined by a delay bin centered on $\tau_0 = \tau_{g,ij} + \tau_{e,ij}$ can be translated into coordinates of right ascension and declination (α, δ) using Eq. 3.2. In celestial coordinates, the width and orientation of a delay ring evolve with time, as expressed using partial derivatives of Eq. 3.2 with respect to α and δ:

$$\begin{aligned}\Delta\tau =& \frac{\partial \tau_g}{\partial \alpha} \Delta\alpha + \frac{\partial \tau_g}{\partial \delta} \Delta\delta \\ =& - \left[\frac{b_x}{c} \cos\delta \sin H + \frac{b_y}{c} \cos\delta \cos H \right] \Delta\alpha \\ &+ \left[-\frac{b_x}{c} \sin\delta \cos H + \frac{b_y}{c} \sin\delta \sin H + \frac{b_z}{c} \cos\delta \right] \Delta\delta\end{aligned} \tag{3.4}$$

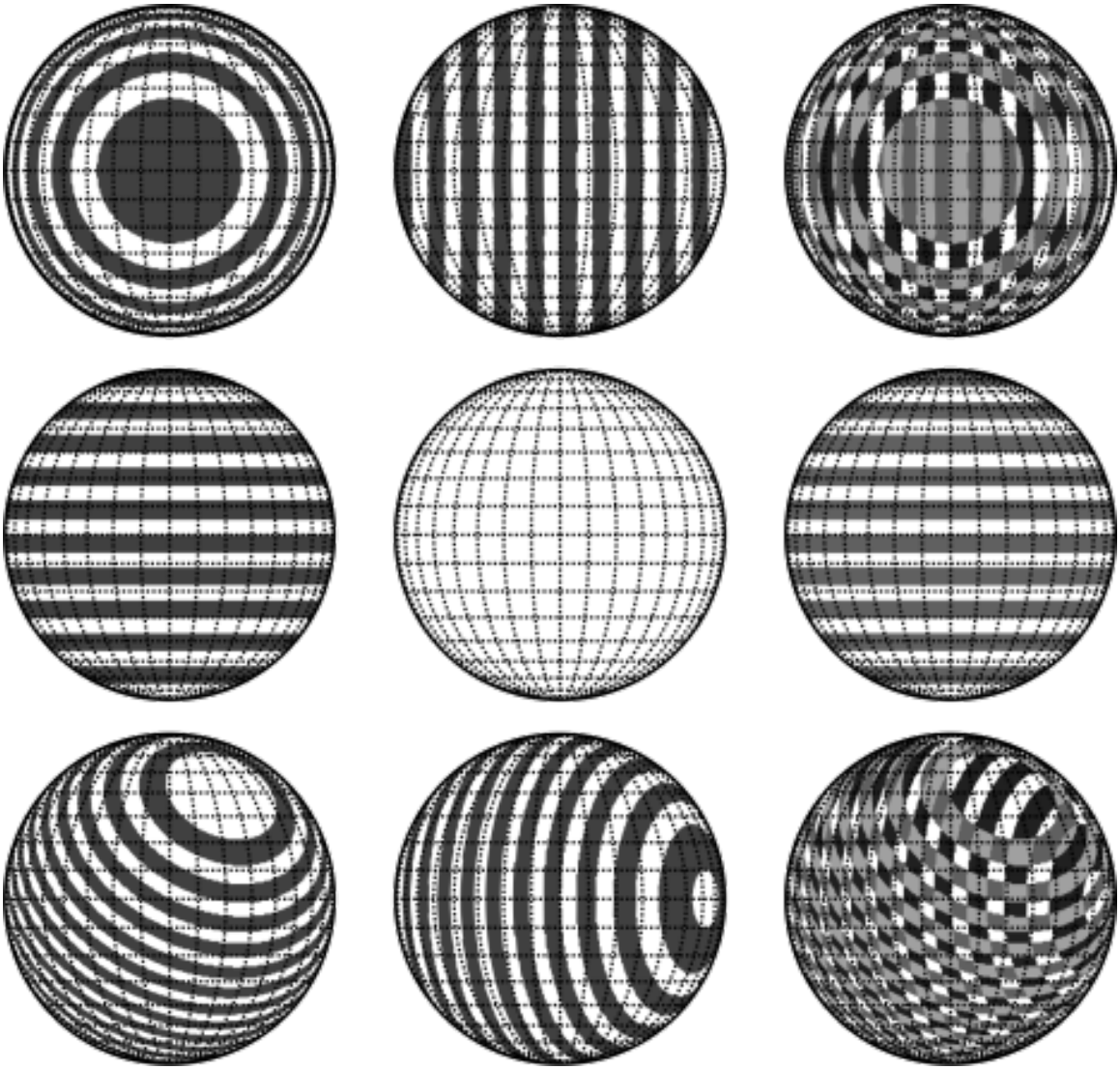

Figure 3.1: Plotted above are (left to right) delay, delay-rate, and combined contours projected onto a celestial hemisphere centered on zero hour angle (horizontal axis) and zero declination (vertical axis). These contours illustrate regions of constant width in delay/delay-rate (DDR) coordinates for baselines that are oriented (respectively, top to bottom) in the equatorial plane pointing out of the page, parallel to the polar axis, and 25° west of north and tangent to 40° S latitude. DDR filters select for intersections between these contours (third column). The width of a selected region depends on the orientation of a baseline with respect to a source, and for most orientations, there exists another "lobe" of sensitivity where isocontours of delay and delay-rate intersect for a second time.

Given the time-variable orientation of the delay ring containing the fixed point on the celestial sphere (α_0, δ_0), and looking ahead to the next section where we will show how the flux in a delay ring can be further localized using delay-rates, we will ignore the parametric nature of a delay ring and simply compute an average full-width in $\Delta\alpha, \Delta\delta$ corresponding to $\Delta\tau$ by averaging Eq. 3.4 over $H_0 = h - \alpha_0 \in [-\frac{\pi}{2}, \frac{\pi}{2})$:

$$\begin{aligned} \langle \Delta\alpha \rangle =& \Delta\tau / \left| \frac{2b_y}{\pi c} \cos\delta \right| \\ \langle \Delta\delta \rangle =& \Delta\tau / \left| -\frac{2b_x}{\pi c} \sin\delta + \frac{b_z}{c} \cos\delta \right| \end{aligned} \tag{3.5}$$

The angular width θ_D of a delay filter may be approximated to order-of-magnitude as

$$\theta_D(\text{deg}) \sim \frac{30}{\Delta\nu(\text{MHz}) \left|\vec{b}\right| (\text{km})}, \tag{3.6}$$

where $\Delta\nu$ represents the bandwidth used in the delay transform and $\left|\vec{b}\right|$ is the length of the baseline involved. For a 100-MHz bandwidth and a 1-km baseline, a delay filter has a resolution of approximately 20 arcminutes.

3.3 Delay-Rate (DR) Filtering

The delay of a source on the celestial sphere changes with time owing to the rotation of the b_x, b_y components of a baseline with the Earth. From Eq. 3.2, the rate-of-change of delay of a source at (α_0, δ_0) is given by:

$$\frac{\partial\tau_g}{\partial t} = -\omega_\oplus \left(\frac{b_x}{c} \sin H_0 + \frac{b_y}{c} \cos H_0 \right) \cos\delta_0 \tag{3.7}$$

where $\omega_\oplus$ is the rotation rate of the Earth. By phasing visibilities to a source with time variable phase $2\pi\nu\tau_g(t)$ using the current best calibration parameters, it is possible to stop the fringe of a source so that a Fourier transform of the time axis over the

interval $[t_0, t_1]$ will add visibilities coherently into an area near zero delay-rate:

$$\begin{aligned}\hat{V}_{ij}(\nu, f) &= \int_{t_0}^{t_1} G_{ij}(\nu, t) \sum_n \left[A_{ij}(\nu, \vec{s}_n(t)) S_n(\nu) e^{2\pi\nu(\tau_{g,ijn} + \tau_{e,ij} - \tau_0 - \tau)}\right] e^{-2\pi i f t}\, dt \\ &= \hat{G}_{ij}(\nu, f) * \sum_n \left[\hat{A}_{ij}(\nu, f_n) * \hat{S}_n(\nu) * \int_{t_0}^{t_1} \delta_D(\tau_{gr,ijn} + \tau_{e,ij} - \tau_0 - \tau) e^{-2\pi i f t}\, dt\right]\end{aligned} \tag{3.8}$$

A filter near zero delay-rate along the f axis (the Fourier complement to t) has a resolution determined by $\Delta\tau\Delta f \sim \frac{\Delta\tau}{\Delta t}$ (the resolution of a delay filter divided by time window used in the DR transform). Note that any time-variable gain (for example, the non-tracking primary beam of a PAPER dipole) enters as a convolution kernel along the delay-rate axis. Such a filter restricts flux to a ring where the celestial sphere intersects the plane parallel to $(\frac{b_x}{c}\sin H_0 + \frac{b_y}{c}\cos H_0)$ and the Earth's rotational axis (see Fig. 3.1). As in the case of the D transform, this ring has a time-variable width and orientation with respect to a fixed point on the celestial sphere that can be described using partial derivatives of Eq. 3.7:

$$\begin{aligned}\Delta\tau\Delta f =& \frac{\partial^2\tau_g}{\partial\alpha\partial t}\Delta\alpha + \frac{\partial^2\tau_g}{\partial\delta\partial t}\Delta\delta \\ =& \omega_\oplus\left(\frac{b_x}{c}\cos H - \frac{b_y}{c}\sin H\right)\cos\delta\Delta\alpha \\ &+ \omega_\oplus\left(\frac{b_x}{c}\sin H + \frac{b_y}{c}\cos H\right)\sin\delta\Delta\delta\end{aligned} \tag{3.9}$$

Similarly, the average full width of this filter in celestial coordinates around the phase center (α_0, δ_0) can be described by averaging Eq. 3.9 for $H_0 = h - \alpha_0 \in [-\frac{\pi}{2}, \frac{\pi}{2})$:

$$\begin{aligned}\langle\Delta\alpha\rangle =& (\Delta\tau\Delta f)/\left|\frac{2\omega_\oplus b_x}{\pi c}\cos\delta\right| \\ \langle\Delta\delta\rangle =& (\Delta\tau\Delta f)/\left|\frac{2\omega_\oplus b_y}{\pi c}\sin\delta\right|\end{aligned} \tag{3.10}$$

An order-of-magnitude estimate of the angular width θ_{DR} of a delay-rate filter is given by

$$\theta_{DR}(\text{deg}) \sim \frac{100}{\Delta\nu(\text{MHz})\left|\vec{b}\right|(\text{km})\Delta t(\text{hr})}, \tag{3.11}$$

where $\Delta\nu$ represents the bandwidth used in the delay transform, $\left|\vec{b}\right|$ is the length of the baseline involved, and Δt is the time interval used in the delay-rate transform. Using a 1-km baseline, a bandwidth of 100MHz, and 1 hour of data, a delay-rate filter has a resolution of approximately 1 degree.

3.4 A Combined Delay/Delay-Rate (DDR) Filter

By phasing visibility data for a baseline (possibly using imperfect calibration) to a point (α_0, δ_0) and performing Fourier transformations along both the frequency and time axes, it is possible to apply a DDR filter near the origin in delay/delay-rate space that selects for a restricted area near the phase center. The fundamental resolution of this filter is determined by the minimum bounds placed by Equations 3.5 and 3.10, but wider filters may be constructed by selecting multiple bins along both the τ and f axes. After a filter has been applied in DDR domain (either to null or extract a region near the phase center), one can then apply the inverse Fourier transformations to return to the frequency-time (FT) domain. The data may then be unphased from the specified source, if desired, to return to the original phase center. This filtering process is described using matrix operator notation as:

$$V'_{ij}(\nu,t) = \phi^{-1}(\nu,t) \cdot F^{-1}_{\nu\tau} \cdot F^{-1}_{tf} \cdot G(\tau,f) \cdot F_{ft} \cdot F_{\tau\nu} \cdot \phi(\nu,\tau) \cdot V_{ij}(\nu,t) \tag{3.12}$$

where ϕ is the phasor to a point on the celestial sphere, F represents a Fourier transform, and G is a gain function representing the spatial filter desired.

The DDR filtering process requires that the width of the convolution kernels associated with the frequency-dependent electronics gain, antenna beam gain, and source flux spectra in Eqs. 3.3 and 3.8 be small compared to the filter width needed to isolate strong sources. While this may often be the case for smoothly varying functions characterizing the response of analog systems and the wide-band emission of celestial sources, the excision of faulty data, particularly data containing radio-frequency interference (RFI), challenges this assumption by introducing sharp features into an otherwise smoothly varying function. The effects of nulling data in a spectrum before

constructing a "delay image" bear many similarities to the effects of incompletely sampling an aperture in traditional synthesis imaging. In fact, the effects of constructing a delay image with an incompletely sampled passband can be compensated for by using the same deconvolution techniques used in synthesis imaging.

A variant of the CLEAN (Högbom 1974) algorithm, adapted to a complex function in one dimension for a celestial sky dominated by a handful of point sources, is a particularly fast and robust algorithm for deconvolving the effect of passband gaps (Roberts et al. 1987). In one-dimensional, complex CLEAN, a fraction of the largest magnitude feature (by bin) of the "dirty image"–the Fourier transform of the spectrum containing nulled data–is iteratively propagated to a model deconvolved image after being divided by the gain of the "dirty beam". This model is then used to derive residuals between the predicted dirty image and the actual one, and these residuals are used as the dirty image in the next iteration. In the DDR imaging case, the dirty beam consists of the Fourier transform of the sampling function reflected in the data. This deconvolution process can be illustrated as a modification of Eq. 3.12:

$$V'_{ij}(\nu,t) = \phi^{-1}(\nu,t)\cdot F^{-1}_{\nu\tau}\cdot F^{-1}_{tf}\cdot G(\tau,f)\cdot \tilde{S}^{-1}_{\tau}(f)\cdot F_{ft}\cdot \tilde{S}^{-1}_{t}(\tau)\cdot F_{\tau\nu}\cdot \phi(\nu,\tau)\cdot S(\nu,t)\cdot V_{ij}(\nu,t) \tag{3.13}$$

where S represents a sampling function whose singularity (having multiplied some data by zero) makes it non-invertible, but whose effects can nonetheless be undone in approximation, represented by $\tilde{S}^{-1}$, via an iterative one-dimensional deconvolution along the delay and delay-rate axes. In order for complex CLEAN to converge, it is vital that the estimated gain of the dirty beam reproduce the phase of the main lobe of the complex kernel. For a mostly sampled aperture, it is a reasonable assumption that the phase of the peak response of the kernel can be taken as the phase of the overall gain for estimating updates to the clean image and for incorporating the final residuals of the CLEAN process. These residuals represent what is left after the CLEAN process has converged to a specified tolerance.

The computational complexity of the CLEAN algorithm scales between $O(N)$ and $O(N^2)$ with the number of data samples, depending on the number of image-domain pixels whose magnitude exceeds the specified termination tolerance. Since

the computational complexity of the Fourier transform operation involved in DDR imaging scales as $O(N\ log(N))$, the relative computational expense of CLEANing is sensitive to this tolerance. Provided that one specifies a termination tolerance that matches the degree to which a few strong point sources dominate data, the additional computational expense of the CLEAN operation is negligible. The open-source software toolkit Astronomical Interferometry in PYthon[9] (AIPY) contains an implementation of DDR imaging, with the option of using one-dimensional complex CLEAN along both delay and delay-rate axes to remove the effects of nulled data.

A priori knowledge of the shape of the passband can be incorporated into the model of the dirty beam to decrease the footprint of sources in the deconvolved image. However, direction-dependent gains and frequency-dependent source fluxes pose the same problems as in standard imaging (Bhatnagar et al. 2008; Conway et al. 1990)– namely that the kernels of these effects change per source, so that deconvolution cannot be performed using a single dirty beam. Provided that sources are sufficiently separated in DDR domain so that the convolution kernels representing the passband, source spectrum, and primary beam do not adversely affect source isolation, the kernels of these functions can be used to reconstruct the corresponding FT-domain functions. After nulling interfering sources, deconvolving by a sampling function, and extracting a swath in DDR domain around the source at phase center, the remaining point-spread function reveals the effects of the convolution kernels associated with the frequency-dependent electronics gain, antenna beam gain, and source flux spectra in Eqs. 3.3 and 3.8. Having preserved these functions (smoothed by a factor determined by the size of the swath extracted), one can transform DDR data back into FT domain to reveal their combined effect. Given a model source spectrum, one has direct access to each baseline's response to that source versus time and frequency. This process is modeled in Figure 3.2 and applied to real data in Figure 3.3.

After using DDR filtering to isolate a source while retaining information about the bandpass and primary beam, classic single-source self-calibration (Jennison 1958; Pearson & Readhead 1984) can be used to produce the closure phase and amplitude

[9]http://pypi.python.org/pypi/aipy

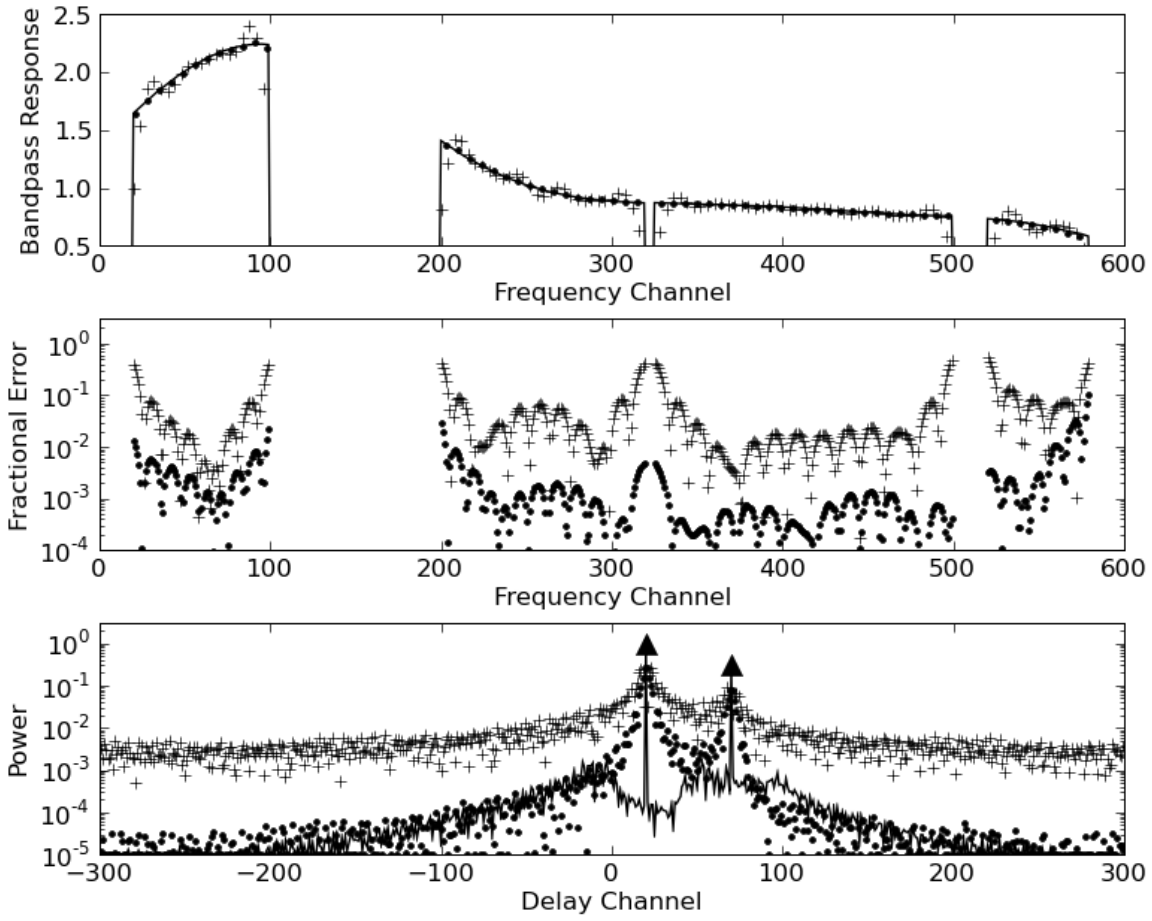

Figure 3.2: The above plots demonstrate a model application of 1-dimensional CLEAN in combination with delay-domain filtering to isolate a source and estimate passband response. A smooth passband (solid, top) has been nulled at various locations to reflect the removal of RFI-tainted data. The delay transform of this spectrum yields a "dirty delay image" (pluses, bottom) that differs substantially from the model (triangles, bottom). From this dirty image, the maximal region around the strongest source is extracted and Fourier transformed back into frequency (F) domain to obtain an estimate of the bandpass shape (pluses, top and middle). However, this estimate can be improved dramatically by CLEANing the dirty delay image by the delay transform of the sampling function (dots, bottom) before filtering to a single source and transforming into F domain (dots, top and middle). Finally, the dirty delay image may be deconvolved by the product of the sampling function and the estimated bandpass to yield a more accurate delay image (solid, bottom).

quantities from which antenna-based calibration parameters can be deduced. Because DDR filtering is sensitive to the orientation of a baseline relative to strong celestial sources, the degree to which sources may be isolated varies between baselines. For this reason, it may be necessary to manually exclude certain baselines from the self-calibration process at times when source separation is particularly problematic. For arrays consisting of many more than 4 antennas, this necessity does not significantly impact the accuracy of self-calibration.

3.5 Shortcomings of DDR Filters

The pair of rings defined by the delay and delay-rate bins specified in a filter intersect at two points on the celestial sphere (see Fig. 3.1). As a function of time, one of these points of intersection remains centered on the specified phase center while the other swings around the sky in a pattern that depends on the orientation of the baseline and the location of the phase center. Thus, a defined filter achieves the desired result averaged over time, but at any given time it sees two separate areas on the sky with equal weight. For some baseline orientations, one of these areas can be attenuated by the primary beam, but this is not generally the case.

The effect of having a "double-lobed" response has minimal impact on source-isolation procedures. An effective method for isolating a source consists of applying a series of narrowly tailored nulling filters aimed at other strong sources in the field-of-view, followed by a relatively coarse extracting filter aimed at the desired source. This technique has the advantage of maintaining sufficient range around the desired phase center for accommodating imperfect calibration and for reconstructing beam and passband shapes, while minimizing the probability that a secondary lobe of this wide filter sweeps across another strong source.

The complex effects of the secondary lobe of DDR filters suggest exercising caution when constructing maps using data where such filters have been used to excise certain sources. Though such filters can indeed be used effectively in these situations, one should remain cognizant of the fact that an attenuating filter of changing

size has been swept across wide regions of the sky at a variable rate. However, with increasing numbers of baselines at unique orientations, the relative contribution of any single secondary lobe of a DDR filter decreases. As a result, this technique may have particularly useful applications to spatial imaging with large arrays of widefield antennas.

3.6 Application to PAPER Calibration

The Precision Array for Probing the Epoch of Reionization (PAPER) is an experiment aimed at detecting fluctuations in 21cm emission from neutral hydrogen at redshifts $z = 7$ to 11 as it is ionized in the first epoch of star formation (Bradley et al. 2005). To this end, interferometric arrays of dipole elements have been deployed at the NRAO site in Green Bank[10], which we call PGB, and at the proposed Murchison Radio-astronomy Observatory (MRO) site in Western Australia[11], which we call PWA. These arrays are being steadily expanded in a series of deployments with the ultimate goal of correlating more than one hundred antennas in the low-interference environment of Western Australia. Currently, PAPER has deployed a 4-element array at MRO (PWA-4) and a 16-element array in Green Bank (PGB-16), with a typical antenna spacing of 200m.

The PAPER dipole element has a broad (125-185 MHz) frequency response owing to a modified sleeved dipole design, with a smooth, single-lobe primary beam to facilitate the exacting calibration that is necessary for this experiment. Each antenna maintains a fixed pointing toward the zenith as the sky rotates through–all PAPER data is taken as a "drift scan". The full-width at half-maximum (FWHM) of the primary beam is nearly 60°, and extends from horizon to horizon without a null. The wide FoV and large relative bandwidth of the PAPER experiment have complicated progress toward an accurate first-order calibration for the reasons outlined in §3.1.

[10]The National Radio Astronomy Observatory (NRAO) is owned and operated by Associated Universities, Inc. with funding from the National Science Foundation.

[11]We acknowledge the Wajarri-Yamatji people of Australia as the Native Title Claimants of the purposed MRO lands and thank them for allowing scientific activity on the site.

In particular, the RFI environment at the Green Bank site is such that wide swaths in frequency and time must be excised before the data are usable for astronomical purposes. The preceding DDR filtering and CLEANing techniques were designed precisely to combat this problem so that source fluxes can be separated to facilitate self-calibration.

In Figure 3.3, we demonstrate the application of a delay filter to an integration from one baseline of PGB-16. The substantial sidelobes in D domain that result from RFI excision smear Cyg A and Cas A (the two dominant sources) together, corrupting the attempt to isolate Cyg A via a delay filter and to use that source to calibrate system gain. By applying one-dimensional CLEAN using the kernel that results from the D transform of the sampling function, the sidelobes in D domain can be reduced dramatically. With cleanly separated spikes attributable to Cyg A and Cas A, the flux of Cyg A may be extracted with greater fidelity. By extracting a swath around the desired source, we preserve a convolving kernel in D domain that retains information about the source spectrum and smooth bandpass function. A smooth estimate of the passband may be obtained by Fourier transforming this kernel back into F domain and dividing by a known source spectrum. Notice in Figure 3.3 how a smooth passband resembling the auto-correlation spectrum of one of the antennas has been constructed from the response of Cyg A. The difference between the auto-correlation and cross-correlation spectrum of Cyg A is attributable to galactic synchrotron emission that is resolved out by the baseline being used.

There are cases for which pure delay filtering is inadequate for separating sources. The left side of Figure 3.4 illustrates a waterfall plot of the delay spectrum of a baseline over the course of four hours, during which the delay tracks of Cas A (center) and Cyg A (right to left) cross as the sources drift through the primary beam. In this case, a naive D-domain filter for suppressing Cyg A and extracting Cas A cuts a swath out of the derived spectrum for Cas A (see Fig. 3.5) and corrupts the attempted derivation of the primary beam shape as Cas A drifts through it. But as the right side of Figure 3.4 illustrates, the DR transform provides another axis that can be used to separate the fluxes of the sources involved. Filtering in DDR-domain, we are able to track

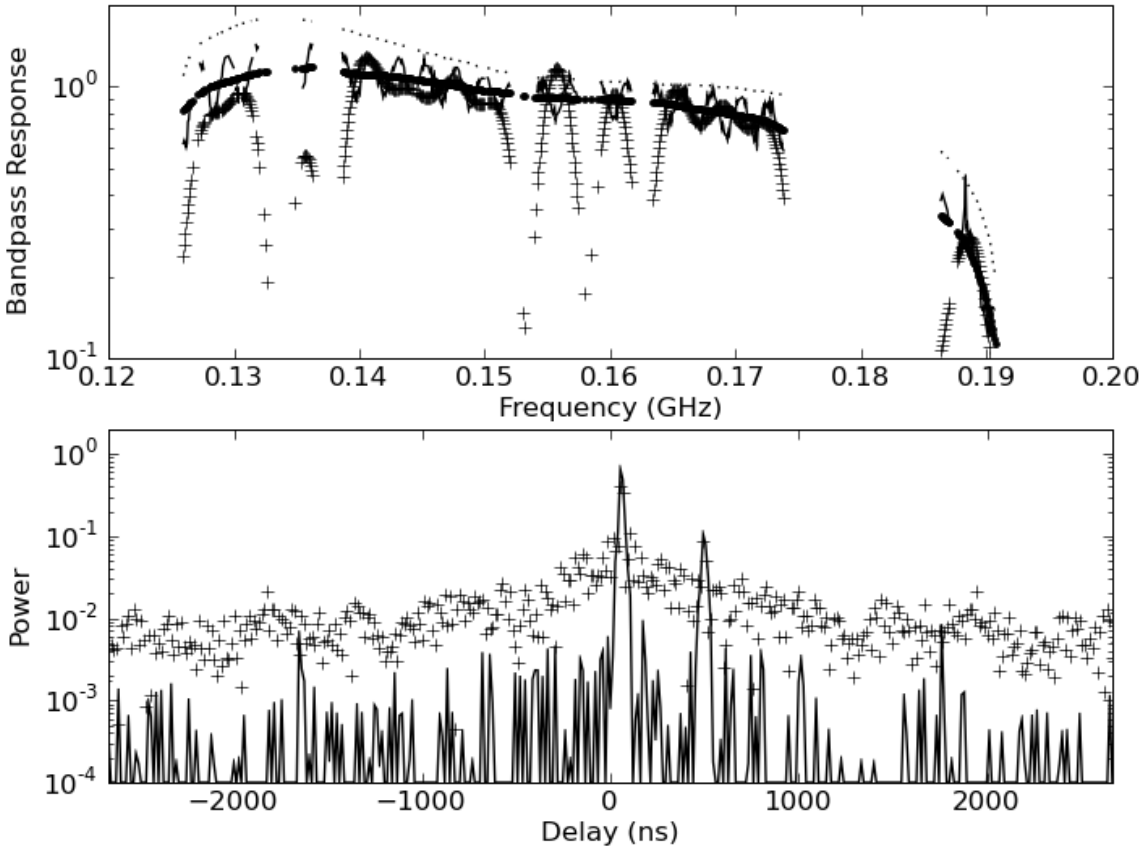

Figure 3.3: Illustrated above are the results of the same passband extraction process described in Figure 3.2, as applied to a baseline spectrum recorded by the PAPER array in Green Bank, WV. The magnitude of the baseline spectrum (thin solid, top) exhibits an interference pattern between Cyg A and Cas A (spikes at 0, 500ns, respectively, in the lower plot). Extracting a region around Cyg A in the dirty delay image (pluses, bottom) yields a poorly estimated passband (pluses, top). However, CLEANing the delay spectrum by the sampling function (solid, bottom) produces a passband estimate (thick, top) that compares favorably with a downscaled auto-correlation spectrum of one of the antennas (dotted, top). The difference between them is attributable to galactic synchrotron emission in the auto-correlation spectrum that is resolved out by the baseline in question.

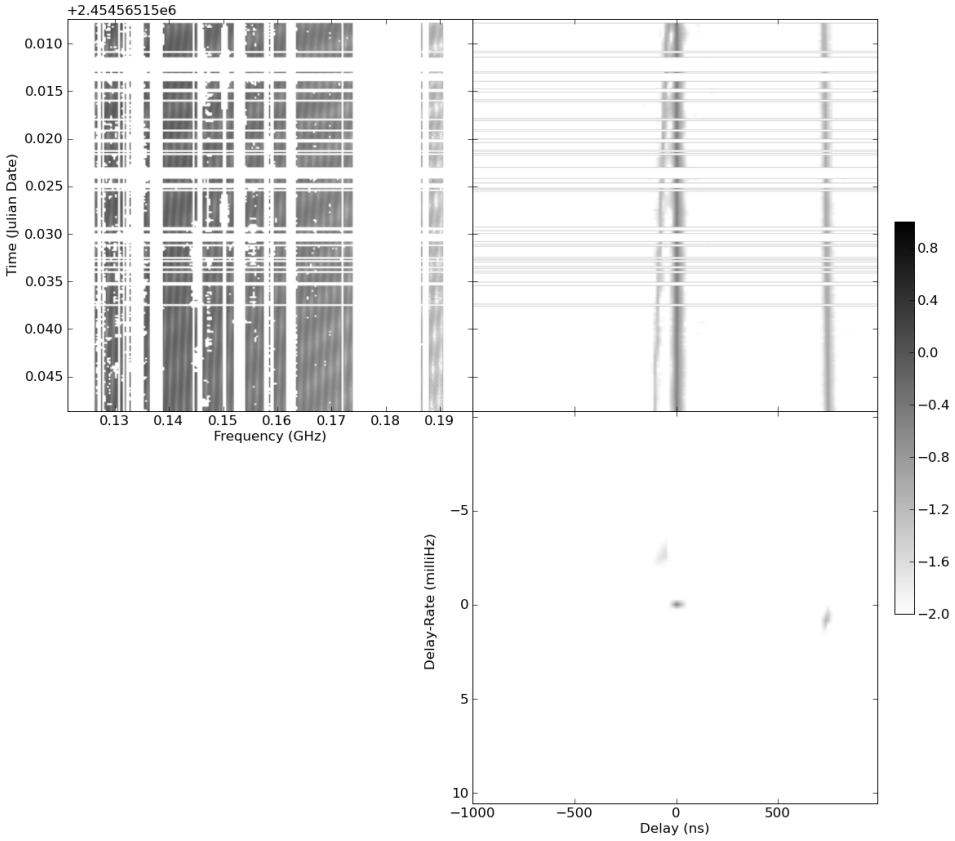

Figure 3.4: Plotted on the top-left is a 1-hour time-series of spectral data, recorded by a single baseline of the Green Bank PAPER array, that has been phased to Cas A. Beating fringes between Cas A, Cyg A, and Tau A create amplitude ripples across the passband. Data excised due to RFI appear as white. The plot on the top-right illustrates the same data, delay (D) transformed. Flux from Cas A now forms a central stripe; Cyg A appears nearby, drifting toward negative delay; Tau A is at high positive delay. The effects of flagged channels have been removed by CLEAN, but entire flagged integrations still appear as white. Finally, the bottom-right plot shows the result of applying both D and delay-rate (DR) transforms. Although nearby in D-domain, Cas A and Cyg A now are clearly separated along the DR axis. Cyg A and Tau A appear blurred because of their changing DRs relative to Cas A over the period of an hour. This blurring illustrates why data should be phased to a region of interest before applying DDR filters. Note that because PAPER is a drift-scan array, source amplitudes change with time. The effects of this are described in §3.3. The above figures are plotted in units of $\log_{10}$(flux-density) with a fixed offset.

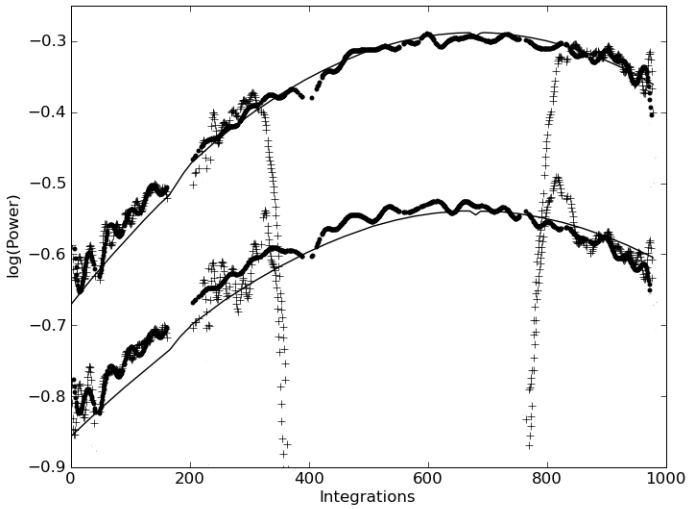

Figure 3.5: Using four hours of data from baseline described in Figure 3.4, the above plot illustrates the response of two channels (143 MHz on top and 165 MHz on bottom) using data in which Cyg A as been filtered out with a narrow filter and Cas A has been isolated with a broader filter. Using a strictly delay (D)-domain filter (pluses) to remove Cyg A results in a drop-out in the response of Cas A when the two sources cross in the D domain. However, a combination delay/delay-rate filter (dots) retains information about the smooth gain variation as Cas A drifts through the primary beam. Predicted models of the beam response at 143 MHz (solid top) and 165 MHz (solid bottom) are provided for reference.

the spectrum of Cas A through its intersection in the D domain with Cyg A, and compare it to a beam model predicted through computer simulation. As was the case for delay filtering, we can preserve information about changing source flux (in this case, caused by drifting through the primary beam) by selecting a swath around the source in DR domain containing the kernel that represents the Fourier transform of the time-dependent gain.

Finally, having used source isolation to self-calibrate PGB-8 data, we demonstrate an application of the same DDR filtering techniques to 2-dimensional aperture synthesis imaging. Figure 3.6 demonstrates how a wide-field image of Cyg A and Cas A (top) can be filtered to remove Cyg A by applying the appropriate DDR filter to

data from each baseline. Though this technique can be very effective in removing broad sidelobes associated with a strong celestial source, one must beware of the effects of the secondary lobes of these filters (§3.5), which in this case have increased the sidelobes associated with Cas A (Fig. 3.6, bottom). The secondary lobe of one of the baselines has swept across the region of sky containing Cas A, leaving a grating interference pattern that represents the contribution that visibilities measured by that baseline should have made to the image, as modeled by the dirty beam used for deconvolving the dirty 2-dimensional image. The absence of this contribution appears as the missing fringe pattern with an inverted sign. Though such filtering can prove valuable in removing strong sources to reveal weaker ones lost in sidelobes, it is clear that the resultant images are corrupted as a result, and that accurate imaging will require the visibilities predicted from these corrupted images to be compared with the measured ones, possibly using iterative sky/visibility modeling to converge on an accurate image.

3.7 Conclusion

We have described a technique for isolating source fluxes by Fourier transforming FT-ordered data from an individual baseline into DDR domain, applying filters, and then performing the inverse transforms. The ability to isolate sources vastly simplifies the calibration of interferometric arrays with wide fields-of-view where self-calibration would otherwise require an accurate a priori sky model to account for multiple strong sources within the primary beam. Imprecise calibration and wide filters can be used initially to select wide areas around sources that encompass calibration errors and allow improved calibrations to be obtained. As calibration improves, these filters can be more narrowly tailored to allow the extraction of bandpass and beam functions from their corresponding DDR-domain kernels, and ultimately such filters can even be used in combination with traditional aperture synthesis imaging to eliminate the sidelobes of strong sources and reveal weaker celestial sources.

While such filters have many desirable qualities such as minimal reliance on

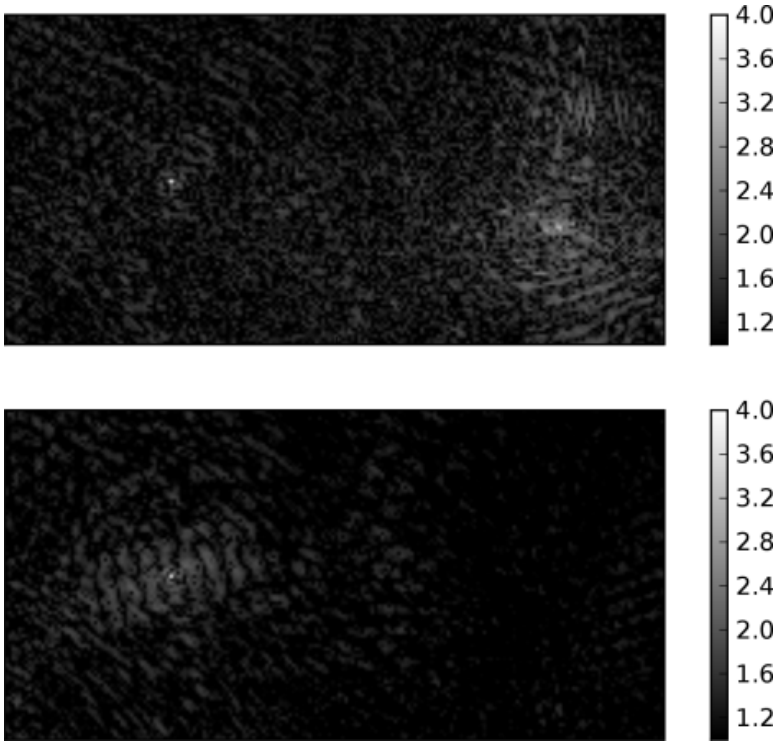

Figure 3.6: Delay/delay-rate (DDR) filters may also be used in combination with traditional two-dimensional aperture synthesis imaging to null bright sources whose sidelobes interfere with imaging other regions. The above $45^\circ \times 90^\circ$ images of Cas A (left, in $\log_{10}$(Jy/beam)) illustrate how Cyg A (right) can be filtered out of the data of all baselines used for imaging, resulting in the complete suppression of the source, and an attenuation of a surrounding region of sky whose width is baseline-dependent. However, the secondary lobe of a DDR filter (§3.5) causes distortions in other regions of the image that must be weighed against the advantages of suppressing bright sources. In the lower panel, a grating sidelobe of Cas A has been introduced as the secondary lobe of one of the 28 baselines used in imaging swept across the region. With increasing numbers of antennas, the relative contributions of the secondary lobes of DDR filters decreases in imaging, making this an attractive imaging technique for large arrays.

prior calibration and geometric widths that are independent of frequency, the fact that such filters have a time-variable orientation with respect to the celestial sphere and a secondary response lobe that sweeps across a large area on the sky limit their usefulness for precision imaging. However, combined with a modeling process that

feeds estimates of the sky from such images into model visibilities that are compared with measured data, DDR filters may constitute an important imaging tool even for high dynamic-range applications.

Chapter 4

The Precision Array for Probing the Epoch of Reionization: 8 Station Results

4.1 Introduction

The Epoch of Reionization (EoR) marks the transition of the primordial intergalactic medium (IGM) from a neutral to a highly ionized state as a result of radiation from the first stars and massive black holes (Loeb & Barkana 2001). This phase transition represents a key benchmark in the history of cosmic structure formation and a major frontier of cosmic evolution yet to be explored. Observations of Gunn-Peterson (GP) absorption by the IGM towards distant quasars (Becker et al. 2001; Fan et al. 2006) and large-scale Cosmic Microwave Background (CMB) polarization from Thompson scattering (Page et al. 2007) have constrained cosmic reionization between redshifts $6 < z < 14$. However, deeper exploration of reionization via these probes faces fundamental limitations: the GP-effect saturates at low neutral fractions and CMB polarization is an integral measure of the Thompson optical depth to recombination.

The most incisive probe of EoR is direct observation of the neutral IGM using

the hydrogen 21cm line (Furlanetto et al. 2006; Barkana & Loeb 2005a). The rich astrophysics traced by HI and the intrinsic three-dimensionality of the signal make this approach especially appealing (Barkana & Loeb 2005b; Loeb & Zaldarriaga 2004). However, the challenges of exploring reionization with red-shifted 21cm emission in an observing band below 200 MHz are daunting. EoR detection experiments require unprecedented levels of instrumental calibration and foreground characterization. The brightness temperatures of polarized galactic synchrotron emission, continuum point-sources, and galactic/extra-galactic free-free emission can exceed the expected $\sim$10 mK fluctuations of the 21cm EoR signal by more than 5 orders of magnitude (Zahn et al. 2007; Santos et al. 2005). Wide fields-of-view (FoVs), large fractional bandwidths, large numbers of antennas, significant RFI environments, and ionospheric variation all present challenges for next-generation low-frequency arrays.

The Precision Array to Probe the Epoch of Reionization (PAPER) is a first-generation experiment focused on statistical EoR detection, but with sufficient sky coverage and sensitivity to detect the very rare, largest-scale structures formed at the end of reionization. PAPER represents a focused effort to overcome the substantial technical challenges posed by using large, meter-wave interferometric arrays to detect the 21cm EoR signal.

In §4.2, we present the principle deployments of the PAPER instrument. We then outline the architectures of the analog (§4.3) and digital (§4.4) signal paths and describe the calibration pipeline (§4.5) that is applied to the data. Finally, we present observational results in §4.6 that include an all-sky map and angular power spectra measured toward a colder patch of the synchrotron sky.

4.2 PAPER Deployments

PAPER is being developed as a series of deployments of increasing scope to address the instrumentation, calibration, and foreground characterization challenges that must be surmounted in order to detect a 21cm signal from EoR. By characterizing and optimizing each component in the array with careful engineering, we hope to

		PGB-4	PWA-4	PGB-8	PWA-64
Deployment Date		2004	2007	2006-8	(2009)
N_{ant}		4	4	8	64
N_{pol}		1	1	1	4
Ω_{B}	(str)	.96	.96	0.43	0.43
$\Delta\nu_{\text{corr}}$	(MHz)	100	150	150	100
N_{chan}		256	2048	2048	≥ 2048
τ_{int}	(s)	32	7.16	14.32	≤ 7.16
τ_{obs}	(days)	1	3	14	90
d_{max}	(m)	100	150	300	600

Table 4.1: Described above are the characteristics of the three principal PAPER deployments to date, along with a planned 64 antenna deployment, at at the NRAO facility in Green Bank (PGB) and near the proposed Murchison Radio Observatory site in Western Australia (PWA).

reduce the complexity of data calibration and analysis. Our staged approach allows for a systematic investigation of observational challenges, with a capacity for adaptation as the characteristics of our instrument and of interfering foregrounds are better understood.

There have been three principal PAPER deployments, whose characteristics, along with a planned 64 antenna deployment, are outlined in Table 4.1; N_{ant} is the number of antennas in each deployment; N_{pol} is the number of polarization cross-multiples computed in the correlator; Ω_{B} denotes the solid angle of the primary beam at 150 MHz; $\Delta\nu_{\text{corr}}$ is the correlated bandwidth; N_{chan} is the number of frequency channels computed over the correlated bandwidth; τ_{int} is the integration time per visibility; τ_{obs} is the longest continual operation; d_{max} is the approximate maximum baseline length. In its initial 2004 deployment at the NRAO facilities in Green Bank[1] the PAPER instrument consisted of four sleeved dipoles above planar ground-screens arranged on an east-west line with integrated differential amplifiers. This minimum-redundancy array, dubbed PGB-4, had a longest baseline of 100 meters and provided initial field experience and a means by which to evaluate our analog and digital elec-

[1]The National Radio Astronomy Observatory (NRAO) is owned and operated by Associated Universities, Inc. with funding from the National Science Foundation.

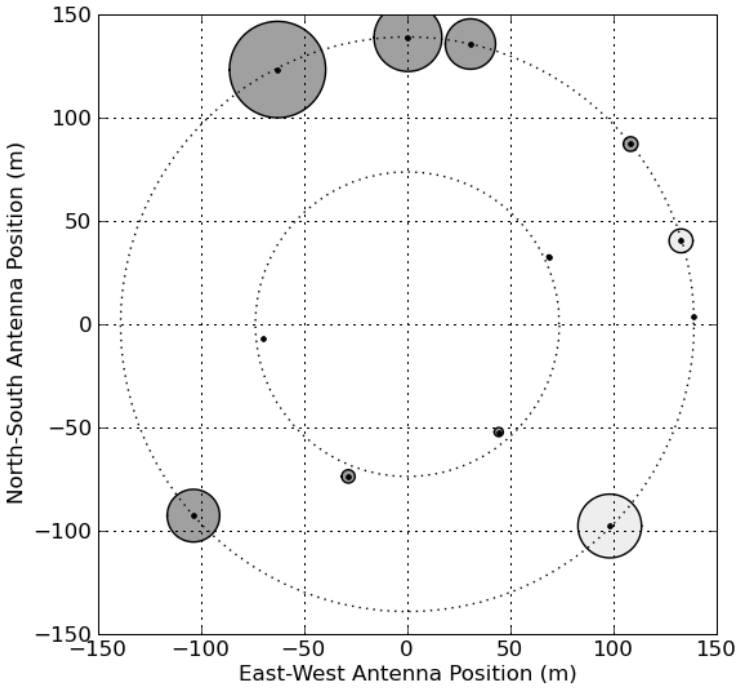

Figure 4.1: Illustrated above are the antenna configurations, in topocentric coordinates, for the PGB-8 deployment (outer, dotted) and the PWA-4 deployment (inner, dotted). Centered on each antenna are circles whose radii represent antenna elevations scaled by a factor of 10, with light/dark gray indicating positive/negative elevation with respect to a fiducial antenna.

tronics. PGB-4 established PAPER's basic architecture as that of a transit array of zenith-pointing dipoles connected to a central correlator via fixed-length cables running above ground.

The second PAPER deployment, PGB-8, began in 2006 and underwent continual development until it was upgraded to PGB-16 in October 2008. PGB-8 consisted of eight antennas deployed on a 300m diameter circle (Fig. 4.1) at the NRAO Green Bank Galford Meadow site, with a small hut for rudimentary climate control built at the center of the array to house the receivers and correlator. As discussed in §4.3, antenna signals propagate to the hut over 75 ohm coaxial cable. The correlator itself went through several cycles of improvement paralleling the development of the scalable correlator architecture described in §4.4. In its final state, the PGB-8

correlator employed a packetized correlator (PaCo-8) that processed 2048 spectral channels across 150 MHz of bandwidth, computing all 4 Stokes parameters, although the analog system only supported one polarization. In early 2008, we improved the forward gains of antennas by adding 45° side reflectors to the ground-screens.

Our third deployment, PWA-4, was near the Murchison Radio Observatory (MRO) site[2] in Western Australia in July 2007. This path-finding array gave us firsthand experience with the logistics of deploying and operating an array at a very remote site. The success of this deployment gave us confidence in the low RFI levels and data quality that can be expected from a much larger array at this site. PWA-4 consisted of four antennas with planar ground-screens (no side reflectors) arranged in a trapezoid pattern with a maximum baseline of 150m. The 4-input "Pocket Correlator" (PoCo) described in §4.4 correlated signals from these four elements.

A fourth deployment, PGB-16, is underway. A total of sixteen sleeved dipole antennas using ground-screens with side reflectors have been deployed at Galford Meadow. Initial operations are proceeding in single-polarization mode using PaCo-8, but will evolve to dual-polarization with a PaCo-16 correlator.

4.3 The Analog System

The levels of instrumental calibration and foreground characterization that will be required to model and remove signals interfering with an EoR detection are unprecedented in the 100-MHz to 200-MHz band expected to encompass reionization. With this in mind, we have taken care that each stage of our analog system (see Fig. 4.2) exhibits smooth responses as a function of frequency and direction, thereby minimizing the number of parameters needed to describe these responses and limiting the magnitude of errors introduced by imperfect calibration.

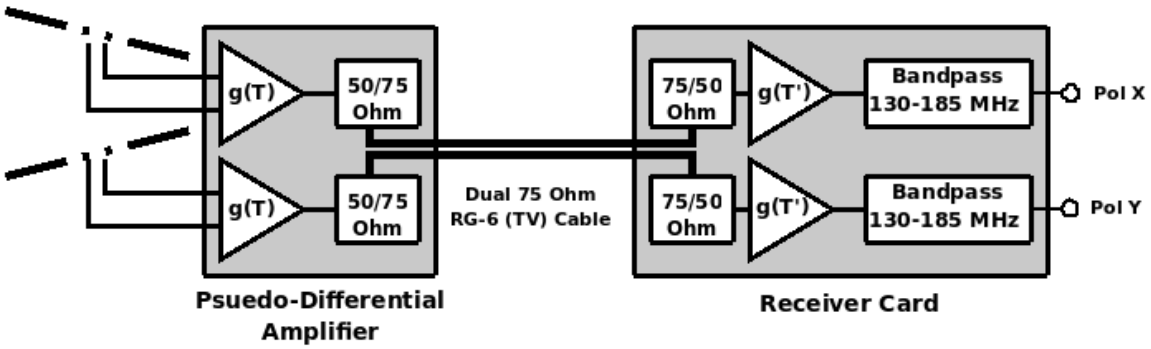

Figure 4.2: The PAPER analog signal path flows from crossed dipole elements attached to a Pseudo-Differential Amplifier (PDA), through RG-6 coaxial cable that runs over the surface of the ground, to a Receiver Card (RC) that filters signals to a 130-MHz to 185-MHz band before transmitting them to the correlator. Gain elements in both the PDA and RC are sensitive to ambient temperature. An enclosure is under development to maintain RCs at a constant temperature, so that the only temperature-dependent gain that must be modeled is that of the PDA for each antenna.

4.3.1 Antenna Design

PAPER antennas are designed as rugged dual-polarization versions of the sleeved dipole (Johnson 1993). Crossed dipoles made from copper tubing are encased between two thin aluminum disks, creating a dual-resonance structure that broadens the antenna's frequency response. The dimensions of the tubing and disks have been tuned for efficient operation over a 120-MHz to 170-MHz band. This design produces a spatially smooth primary beam pattern that evolves slowly with frequency.

The antenna design includes a grounding structure that alleviates the gain variations that result from the effects of climatic conditions on the dielectric properties of earth ground. PGB-4 and PWA-4 deployments employed simple wire-mesh ground planes supported by a wooden framework (Fig. 4.3, top). This design was improved to include planar wire-mesh reflectors that attach to the original ground-screen (now supported by a steel framework) and rise outward at a 45° angle (Fig. 4.3, bottom), essentially becoming a dual-polarization trough reflector (Hall 1988). This design

[2]We acknowledge the Wajarri-Yamatji people of Australia as the Native Title Claimants of the purposed MRO lands and we thank them for allowing scientific activity on the site.

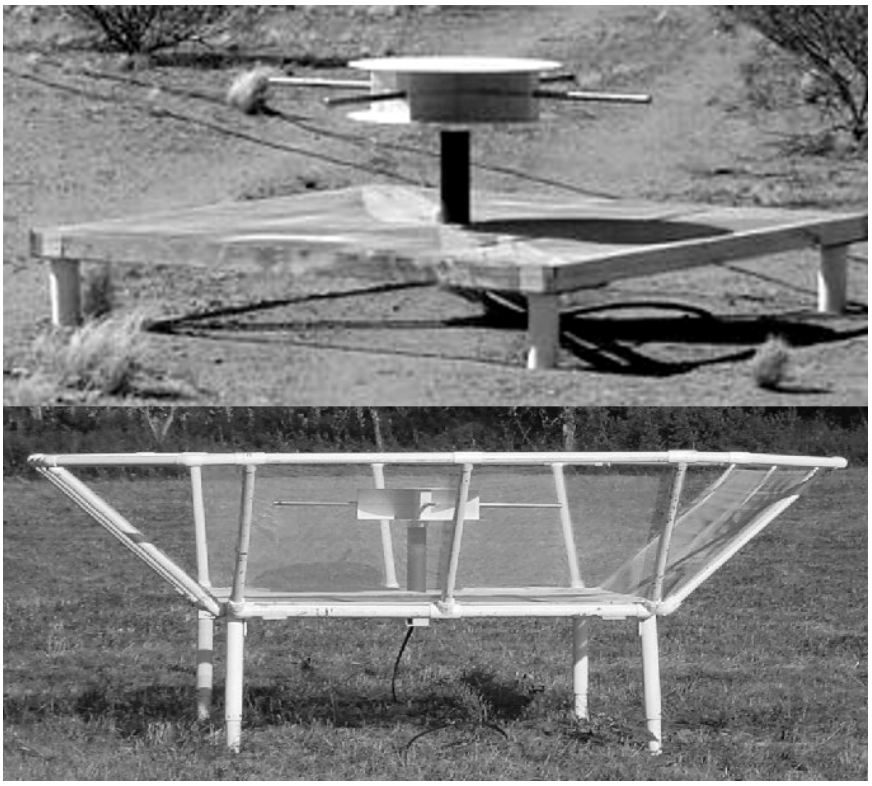

Figure 4.3: PAPER antenna elements are dual-polarization sleeved dipoles mounted above grounding structures. These elements have been designed for smooth spatial and spectral responses to facilitate calibration. For the PWA-4 deployment (top) this grounding structure was a flat, 2m×2m ground-screen mounted on a wooden frame. In the later PGB-8 deployment (bottom), this structure was upgraded to include side reflectors that narrow the size of the primary beam to more closely match the size of colder patches in the synchrotron sky.

produces a primary beam whose angular size more closely matches the angular size of colder patches in galactic synchrotron emission, effectively reducing PAPER's sky-noise-dominated system temperature. Reduction of horizon gain also mitigates susceptibility to some RFI sources. We have ensured that these side-reflectors have not compromised the spatial and spectral smoothness of the primary beam.

The smoothness of the primary beam as a function of frequency and direction enable it to be effectively parametrized using low-order spherical harmonics and fre-

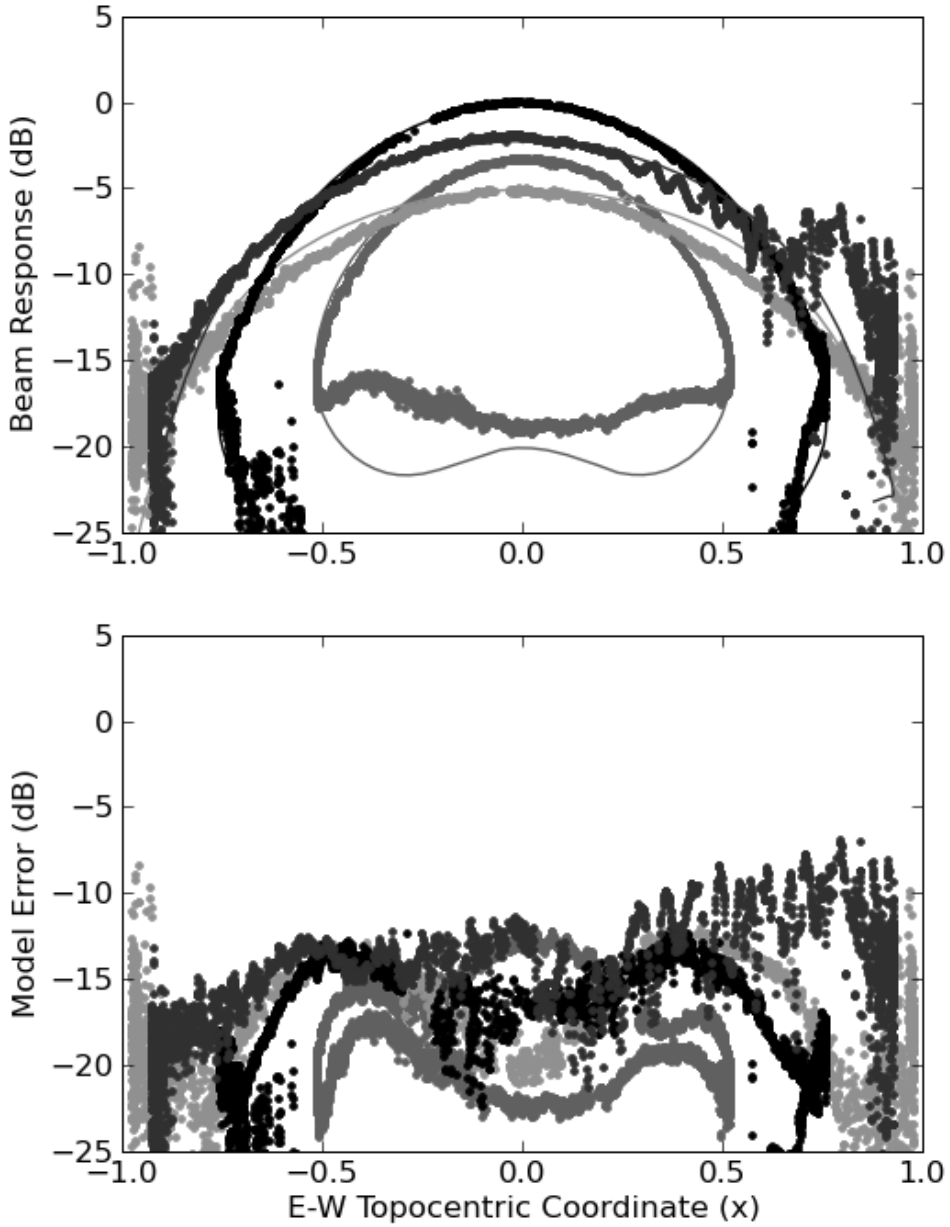

Figure 4.4: Shown in the top plot are the predicted (lines) and measured (dots) responses of the PGB-8 primary beam integrated between 138 MHz and 174 MHz towards (top to bottom at center of plot) Cygnus A, Taurus A, Cassiopeia A, and Virgo A. Responses are relative to the primary beam's zenith response. The lower plot shows the residual response once a model of the primary beam has been subtracted. Note that the measured beam response of Taurus A is complicated by sidelobes of the Sun at $x > 0.2$.

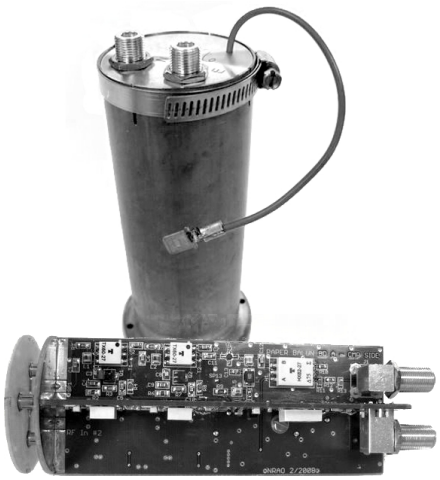

Figure 4.5: The dual-polarization pseudo-differential amplifier, pictured in its casing (upright) and with the casing removed (bottom), is housed in the riser suspending the dipole elements above the ground screen. The gain of this amplifier is sensitive to temperature, with a coefficient of -0.02 dB/K.

quency polynomials according to the equation:

$$a_\nu(\hat{s}) = \sum_{k=0}^{7} \nu^k \left[\sum_{\ell=0}^{8} \sum_{m=0}^{\ell} a_{\ell m}(k) Y_{\ell m}(\hat{s}) \right], \tag{4.1}$$

where the frequency-dependence of the antenna response is modeled by a seventh-order polynomial and the spatial variation of each polynomial coefficient is described by $a_{\ell m}$ coefficients of low-order spherical harmonic functions $Y_{\ell m}(\hat{s})$. This first-order beam model smoothly interpolates to any chosen pointing and frequency, and has been assumed for all of the analysis presented in this paper. As shown in Figure 4.4, the model predicts the perceived flux densities of sources with approximately 93% accuracy for zenith-angles less than 45°. Errors in the beam model are primarily the result of the spectral response of the primary beam being steeper as a function of zenith angle than what is predicted by the model. This systematic effect is currently impeding the accurate measurement of spectral indices of calibrator sources.

4.3.2 Pseudo-Differential Amplifier Design

A pseudo-differential amplifier (PDA, Fig. 4.5) is housed within the riser suspending each dipole element above its ground-screen. Amplifiers employ three cascaded gain stages where the initial stage consists of a differential amplifier based on NEC NE46100 bipolar junction transistors. This circuit features an inherent direct-current pathway from input to ground, making the amplifier tolerant of electrostatic charge build-up that can pose a serious problem in dry climates (Jackson & Farrell 2006). The total ∼32 dB gain through the PDA unit is sensitive to ambient temperature with a coefficient $H = -0.02$ dB/K over the range 5–50°C. This coefficient enters into the voltage gain g_ν of an antenna as

$$g_\nu(t) = g_\nu(t_0) H \left(T(t) - T(t_0)\right), \tag{4.2}$$

where $T(t)$ is the ambient temperature as a function of time t and t_0 represents a fixed time at which static gain calibration is performed. Although not corrected for in the PWA-4 and PGB-8 systems, gain variation with temperature can be partially offset by recording the average ambient temperature of the array and applying a correction to all recorded visibilities. We are currently developing a system for using one antenna input as a "gain-o-meter" by swapping dipoles for a load attached at the front of the PDA unit and using the gain variation of this input to normalize the gains of the other antenna inputs. A more advanced system could also record the ambient temperature at each PDA to allow for per-antenna gain correction.

We have estimated the receiver temperature of the antenna/PDA system by modeling the frequency-dependent auto-correlation power P_ν as a function of time t:

$$P_\nu(t) = |g_\nu(t)|^2 \, k_\mathrm{B} \left[T_\mathrm{sky}(t) + T_\mathrm{rx}\right] \sqrt{\Delta\nu} \tag{4.3}$$

where $g_\nu(t)$ is defined in Equation 4.2; $T_\mathrm{sky}(t)$ is the temperature of a model galactic synchrotron sky weighted by the primary beam of an antenna; T_rx is the receiver temperature; $\Delta\nu$ is the bandwidth of a channel. T_sky was modeled by scaling a 408-MHz sky map (Haslam et al. 1982) to 162 MHz using a spectral index of -2.52 (Rogers & Bowman 2008). As the galactic synchrotron sky rotates through the

primary beam, the perceived $T_{\rm sky}$ varies, providing a modulation that enables the separation of sky and receiver temperatures. After applying a first-order correction for the temperature-dependence of amplifier gain using measured temperatures from a nearby weather station, we fit an average receiver temperature of 110 K for the PGB-8 system.

4.3.3 Signal Transmission

Coaxial cables transport antenna signals from each PDA unit to a central processing location. Because these cables are not buried, they must be rugged to withstand harsh environmental conditions and the occasional chew from local fauna. We chose RG-6, 75 ohm cable with a polyethylene jacket for its stable propagation characteristics as a function of temperature and humidity and for its low cost. Signal attenuation over a 150m cable run is approximately 12 dB at 150 MHz, with a slope of +0.034 dB/MHz. This cable also contains a wire suitable for delivering DC power to PDA units.

On the receiving end, dual-channel receiver boards consisting of amplifier stages and band-limiting filters prepare signals for digitization in the correlator. The high gain of these receiver cards poses a significant regenerative feedback concern not found in heterodyne systems (Slurzberg & Osterheld 1961). This issue was mitigated by mounting the receivers inside a special shielded enclosure (Bradley 2006). The gain of the amplifiers in the receiver cards is also sensitive to temperature, with a coefficient of -0.045 dB/K. To avoid a second temperature dependence in signal gain, a thermal enclosure with thermoelectric heat pumps will be employed to stabilize the temperature of all receiver boards. The "gain-o-meter" approach described in §4.3.2 will also provide a first-order correction.

4.4 The Digital System

A series of real-time digital FX correlators employing Field-Programmable Gate Array (FPGA) processors addresses the growing digital signal processing (DSP) needs

of visibility computation in PAPER deployments. These correlators are based on the architecture described in Parsons et al. (2008), wherein DSP engines transmit packetized data through 10-Gbit Ethernet (10-GbE) links to commercial switches that are responsible for routing data between boards. This architecture, along with a set of analog-to-digital converters and modular FPGA-based DSP hardware and a software environment for programming, debugging, and running them, were developed in collaboration with the Center for Astronomy Signal Processing and Electronics Research (CASPER, Parsons et al. 2006). The flexibility of this correlator design shortens development time, allowing a series of correlators of increasing scale to be developed parallel to PAPER's incremental build-out.

4.4.1 CASPER Hardware and Gateware

The generic FX correlator architecture (Fig. 4.6) on which PAPER correlators are based consists of a set of modules responsible for digitizing, down-converting, and channelizing antenna inputs ("F" Engines), followed by a set of signal processing modules that cross-multiply all antenna and polarization samples for each frequency and accumulate the results ("X" Engines). The problem of transmitting data from every F engine to every X engine is solved by packetizing data according to the 10-GbE protocol and then using a commercial switch to sort data streams. This approach, unique to the CASPER architecture for correlators of this size and bandwidth, avoids custom backplanes and communication protocols that are tailored to a single application, which are the traditional solution to the cross-connect problem in correlators.

Signals from PAPER antennas are digitized by Atmel AT84AD001B dual 8-bit ADC chips capable of digitizing two streams at 1 Gsample/sec. While overrated for this application, this chip has been used in many CASPER applications and is extensively tested. The wide bandwidth of this ADC allows the entire PAPER bandwidth to be Nyquist-sampled as a single, real voltage and then digitally mixed to baseband with perfect in-phase and quadrature-phase components. Two ADC boards connect to each of the CASPER IBOB processing boards used as F engines in PAPER's cor-

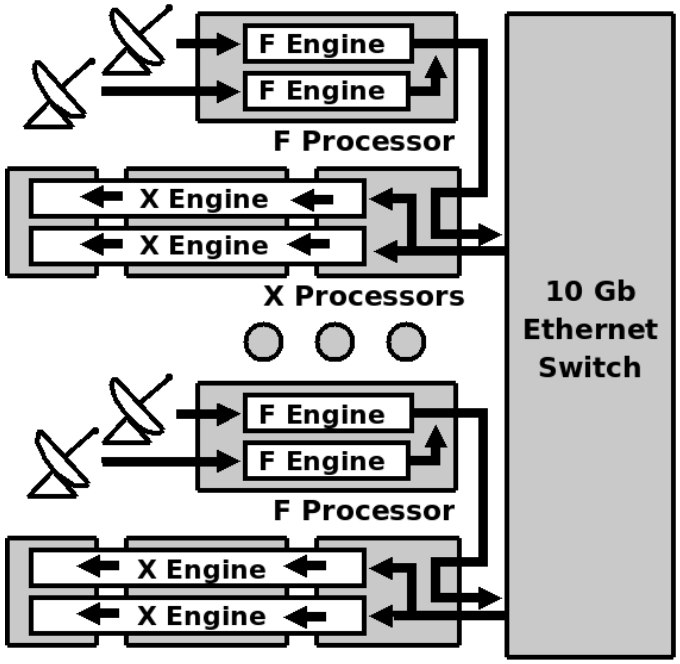

Figure 4.6: PAPER's packetized correlator (PaCo) follows the frequency–cross-multiply (FX) architecture developed by the Center for Astronomy Signal Processing and Electronics Research (CASPER). PaCo currently employs IBOB boards for spectral processing (F engines, above) and cross-multiplication in BEE2 boards (X engines, above); see Parsons et al. (2008) for details. A 10-GbE switch is used to route data between boards, so that data for a subset of channels from all antennas are collected at each X engine, where all cross-multiples are computed.

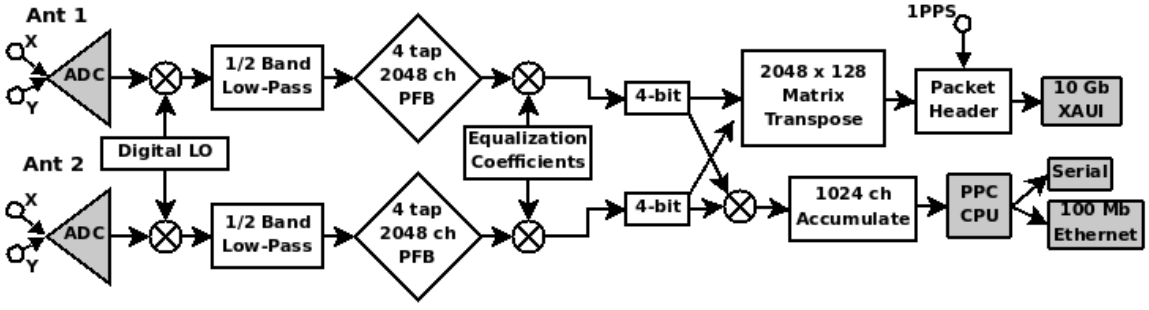

Figure 4.7: The digital processing in an F engine of PAPER's packetized correlator (Fig. 4.6) also doubles as a stand-alone "Pocket Correlator" (PoCo) for correlating four signal inputs. In both designs, data is re-quantized to 4 bits after spectral decomposition. To avoid signal distortion associated with bit truncation, signals are equalized before truncation and the inverse equalization is applied in post-processing.

relator. On the IBOB, one FPGA is responsible for the signal flow outlined in Figure 4.7–namely, digitally down-converting four digitized signals, spectrally decomposing the resulting baseband signals using a Polyphase Filter Bank (PFB) (Crochiere & Rabiner 1983; Vaidyanathan 1990), and re-quantizing spectral data to 4 bits. PAPER correlators equalize antenna spectra before re-quantization to ensure optimal linearity of the output power with respect to input power, as described in §4.5.1. Output data are passed briefly to the BEE2 board (described below) where they are formatted into 10-GbE packets and transmitted to a commercial switch.

A 12-port Fujitsu CX600 10-GbE switch is employed for routing packets between F and X engines. Each port on this switch carries data for two antennas. The performance of this switch was evaluated over a period of 16 hours, with pseudo-random data transmitted and received on each data link. During this time, no errors were detected, placing a limit on the bit-error rate of transmission at 2.2e-16 bits/s. Furthermore, the asynchronous correlator architecture employed by PAPER is tolerant of dropped packets, so errors that do occur do not cause system failures.

We rely on CASPER's BEE2 boards for implementing X engine processing. Each of four FPGAs on a BEE2 operates independently to unscramble packets received from the 10-GbE switch. Two X engine cores inside each BEE2 FPGA then compute cross-multiplications for all N antennas for $1/N^{\mathrm{th}}$ of the frequency channels and the results are accumulated for a selectable time interval. Output of accumulated data over a 100-Mbit Ethernet connection currently imposes a 14-second minimum integration time. We are currently migrating this output to 10-GbE, which will allow shorter integrations that more adequately resolve ionospheric fluctuations.

4.4.2 PoCo and PaCo-8

The "Pocket Correlator" (PoCo, Fig. 4.8, top) used with the PWA-4 deployment is a single IBOB system for correlating and accumulating four input signals (Fig. 4.7). ADCs sample each analog input at 600 MHz and a 75-MHz to 225-MHz band is extracted digitally. A 4-tap PFB decomposes this sub-band into 2048 channels, which are then equalized to remove per-channel scaling differences before being re-quantized

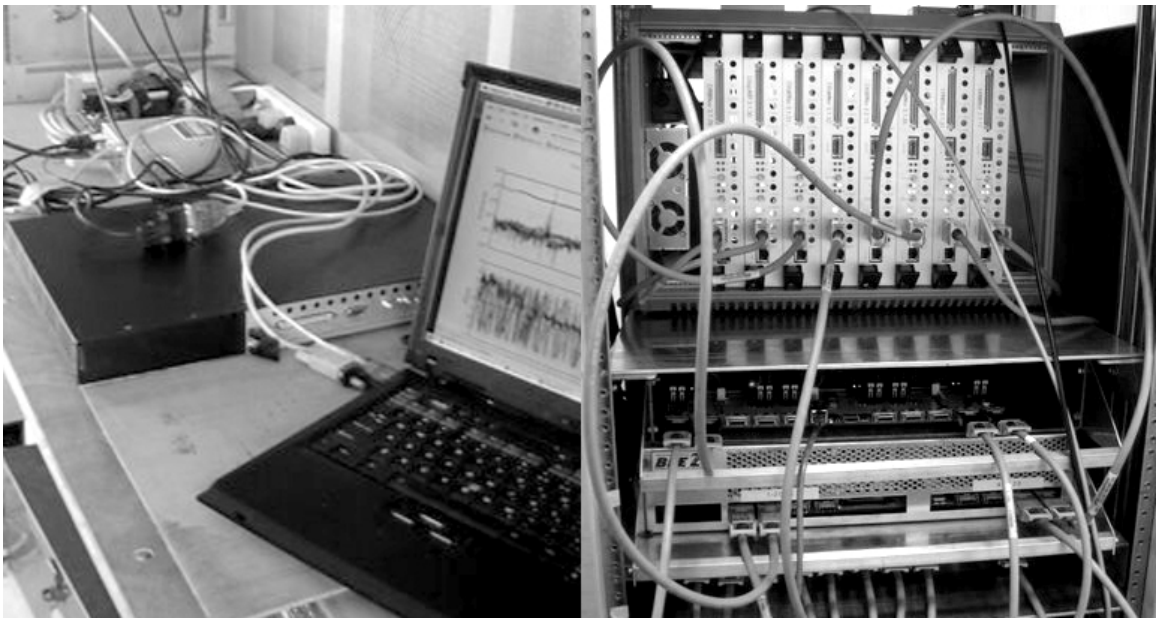

Figure 4.8: For the PWA-4 deployment, a "Pocket Correlator" (left) was employed for correlating four single-polarization antenna inputs. Data were recorded in MIRIAD format on a USB disk attached to a laptop. For PGB-8 and larger deployments, a packetized correlator (right) connects multiple F engine processors through a 10-GbE switch to X engine processors and output data are collected by a server for recording.

to 4 bits. Limited buffer space on the IBOB permits only 1024 channels (selectable from within the 2048) to be accumulated. Accumulated visibilities are output via serial connection to a host computer after being integrated for 7.16 seconds.

The 8-antenna packetized correlator (PaCo-8, Fig. 4.8, bottom) used with the PGB-8 deployment employs four IBOBs and one BEE2 board that communicate through a Fujitsu XG700 10-GbE switch, following the architecture outlined in Figure 4.6. Each IBOB in this system operates identically to PoCo, but branches data from the equalization module to a matrix transposer for forming frequency-based packets. Packet data for each antenna are multiplexed through a point-to-point connection to a BEE2 FPGA and then relayed in 10-GbE format to the switch. A central CPU on each BEE2 board collects all visibility data after it has been accumulated for 14.32 seconds and transmits it over 100 Mb Ethernet to a server where it is written to disk in MIRIAD format (Sault et al. 1995). In field deployments, correlators are housed

inside a special shielded enclosure (Bradley 2006) to mitigate self-interference.

4.5 The Calibration Pipeline

The greatest challenge of using low-frequency interferometry to detect the cosmic reionization lies in characterizing the celestial sky and modeling the system response such that foregrounds to the EoR signal can be effectively suppressed. Various methods for estimating the effects of instrumental calibration on a statistical EoR detection have been analyzed (Morales 2005; Morales et al. 2006; Bowman et al. 2006), but improving the calibration and stability in early instruments remains an open problem (Yatawatta et al. 2008; Bowman et al. 2007). Wide fields-of-view, large fractional bandwidths, strong RFI environments, and ionospheric variation all complicate the calibration of next-generation low-frequency arrays such as PAPER, the Murchison Widefield Array[3] (MWA), the Long Wavelength Array[4] (LWA), and the LOw Frequency ARray[5] (LOFAR). Obtaining an accurate sky model requires accurate models of primary beams, receiver passbands, gain variation, and array geometry. However, characterization of these often depends on the availability of an accurate sky model. The PAPER approach to this problem has been to emphasize precision, with the goal of characterizing system components to within 1% in order to facilitate further model refinement via sky modeling and self-calibration.

Although many standard tools exist for calibrating interferometric data, we are developing an open-source software project called Astronomical Interferometry in PYthon[6] (AIPY) that modularizes interferometric data reduction to facilitate the exploration of new algorithms and calibration techniques. The niche targeted by this package are developing arrays whose evolving instrumental characteristics require flexible, adaptable tools. AIPY uses the dynamically interpreted Python programming language to wrap together interfaces to packages such as MIRIAD (Sault et al.

[3]http://haystack.mit.edu/ast/arrays/mwa
[4]http://lwa.unm.edu
[5]http://www.lofar.org
[6]http://pypi.python.org/pypi/aipy

1995) and HEALPix (Górski et al. 2005) using a common numerical array interface. AIPY also includes object-oriented implementations of a wide range of algorithms useful for radio interferometry, including W-Projection (Cornwell et al. 2003), CLEAN (Högbom 1974), Maximum-Entropy deconvolution (Cornwell & Evans 1985; Sault 1990), aperture synthesis imaging, faceted map-making, visibility simulation, and coordinate transformations. AIPY relies heavily on 3rd party, public-domain modules for scientific computing, allowing its development to focus on interferometry-specific functionality.

AIPY analysis centers around parametrizing interferometric measurement equations for the instrument in question. However, before data from a PAPER correlator is ready to be compared to the measurement equation shown in Equation 4.4, it requires gain linearization, RFI excision, and crosstalk removal.

4.5.1 Gain Linearization

Mitigating the effects of data quantization in the PAPER correlator requires that power levels be carefully set for optimal SNR and that output data be converted to a linear power scale (see Ch. 8 of Thompson et al. (2001)). Correction factors for gain non-linearities in digital systems employing 1-bit and 2-bit quantization of voltage samples are well-known. The non-linear power response of the PAPER correlator, with its 4-bit correlation, is still significant. Our approach to gain linearization is detailed in Parsons et al. (2008). Inside the correlator, optimal SNR during requantization is ensured by applying an equalization function that shifts data samples to the region of maximal linearity in the 4-bit quantization response curve. This equalization function is updated hourly to adjust for changing mean power levels as the galactic synchrotron sky traverses the dipole beam. The first steps of PAPER calibration are to deduce an average 4-bit value from the accumulated output of the correlator, to invert the known quantization response for this value (Fig. 13 in Parsons et al. (2008)), and then to remove the equalization function used in the correlator.

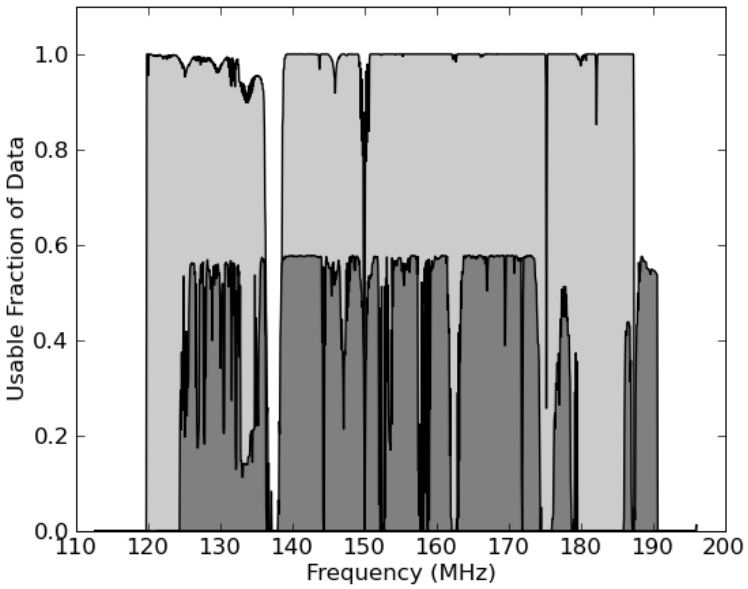

Figure 4.9: Above are illustrated the fraction of usable integrations after flaggin RFI for each frequency channel over 3 days of observation with PWA-4 (light) and PGB-8 (dark). Approximately 40% of PGB-8 integrations are unusable owing to saturation of ADC inputs by strong RFI. The RFI environment at PWA is exquisite, with fewer channels rendered unusable by continuous narrow-band transmissions and almost no saturation events.

4.5.2 RFI Excision

As a general solution to identifying data points that depart from an otherwise smooth function, we employ an algorithm that iteratively fits a smooth polynomial to visibility amplitudes, subtracts that polynomial to identify and exclude outlying points, and repeats. This algorithm identifies outlying points by finding the mean and standard deviation of the logarithm of all data points and defining a threshold in units of the standard deviation. The effectiveness of this technique is improved by first subtracting simulated visibilities using accurate models of the array response and the celestial sky.

This RFI excision algorithm is applied to spectra of cross-correlation data to identify narrow-band interference and to time-series in auto-correlation data to identify integrations where strong RFI such as aircraft communication saturates the ADC

input. Such saturation events result in wide-band "drop-outs" where signal is lost across all frequencies. In both cases, detection of an interference event in a given channel/integration on any baseline results in that data being flagged for all baselines. Figure 4.9 illustrates the usable fraction of data averaged over a full day for the PGB-8 and PWA-4 deployments. The RFI environment in Western Australia represents a marked improvement over that of Green Bank, with fewer channels occupied by continuous RFI and almost no ADC saturation events. Such events cause approximately 40% of PGB-8 integrations to be unusable.

4.5.3 Crosstalk Removal

The inadvertent introduction of a correlation between two independent inputs via a cross-coupling pathway is unavoidable in radio interferometers. Many systems use phase switching to help remove crosstalk introduced along the signal path from an antenna to the ADC. PAPER does not currently implement phase switching, although we are exploring options for doing so. However, in the absence of (or in addition to) phase switching, there are other techniques for removing crosstalk from visibility data, provided that such crosstalk is adequately stable with time. PAPER's crosstalk is modeled as an additive component of measured visibilities that varies slowly on the timescale of several hours. Because this crosstalk is stable on a timescale much longer than the inverse of the typical fringe-rate of a source over the shortest baseline and because its power is concentrated at low delays, PAPER crosstalk is easily suppressed using frequency-based time-averages or delay/delay-rate filters (Parsons & Backer 2009). As with RFI, crosstalk removal is facilitated by the subtraction of simulated visibilities.

4.5.4 PAPER's Measurement Equation

At the heart of the PAPER analysis model is a measurement equation for the single polarization visibility response ($V_\nu(t)$) of a baseline at channel center frequency

ν and time t:

$$V_\nu(t) = G_\nu(t) \sum_n A_\nu(\hat{s}_n) S_\nu(\hat{s}_n) e^{2\pi i \nu(\vec{b}\cdot\hat{s}_n + \tau) + \phi} \tag{4.4}$$

where $\hat{s}_n$ is a time-dependent unit vector pointing in the direction of a source n whose flux density is S; $G_\nu(t) \equiv g_{i\nu}(t) g^*_{j\nu}(t)$ is the frequency-dependent baseline response, which is expressed in terms of the voltage gain of each antenna ($g_{ij,\nu}$; Eq. 4.2) for constituent antennas indexed by i, j; $A_\nu(\hat{s}_n) \equiv a_{i\nu}(\hat{s}_n) a^*_{j\nu}(\hat{s}_n)$ is the direction-dependent response of a baseline's primary beam, expressed in terms of the electric field response of each antenna ($a_{ij,\nu}$; Eq. 4.1); $\vec{b}$ represents the baseline vector separating antenna positions in units of light propagation time; τ is the relative electronic delay of antenna signals into the correlator; and ϕ is a relative instrumental phase between antenna signals.

The design of our antenna and analog electronics leads to antenna-based gains (a_{ij}, g_{ij}) that can be modeled by smooth functions in time and frequency with a modest number of parameters. The degrees of freedom in these parameters are overconstrained by wide-band visibility data, even when using a sky model that includes only a small number of point sources. However, the larger FoV of PAPER elements has complicated early calibration by decreasing the extent to which a single source dominates the correlated signal between antennas. Without isolation of sources, self-calibration cannot proceed as a direct computation using raw data, but rather must proceed by fitting models of the array and the observed sky to remove baseline-dependent interference patterns (Cornwell & Fomalont 1989). To address this problem, we employ a two-tiered calibration process wherein initial coarse calibration that employs source-isolation techniques is followed by multi-source, least-squares fitting to visibilities for the parametrization described in (4.4).

4.5.5 First-Order, Time-Independent Self-Calibration

Phase calibration involves solving for the position and electronic signal delay associated with each antenna. The parameter space associated with phase calibration contains many local minima separated by wavelength increments projected toward

calibrator sources. To speed the phase calibration process, we rely on theodolite-surveyed antenna positions. When initial antenna positions include errors greater than a wavelength, we have developed a procedure for fitting positions that is reasonably robust against converging to non-global minima. First, a single source is isolated in data either by selecting a time when one source is dominant, or by using coarse delay/delay-rate filtering (Parsons & Backer 2009). Baseline components b_x, b_y and a throw-away phase term are then fit for a pair of nearby antennas using a narrow set of frequency channels and a range of time over which fringes wrap several times. Next, data including a few other sources are used to separate b_z out of the remaining phase term. As a final step in single-baseline phase calibration, wide-bandwidth data are used to account for the final phase as a combination of electronic signal delay (τ) and a small, frequency-independent phase term (ϕ).

Following accurate phase calibration for a subset of antennas, another antenna is added and the procedure outlined above is repeated using data for all baselines that connect this antenna to the calibrated subset. With phase calibration for four antennas, preliminary antenna-based gain calibration is possible. PAPER visibilities have been flux-calibrated using Cygnus A with a value of 1.09e4 Jy at 150 MHz and a spectral index of -0.69, which are derived from the data in Baars et al. (1977). The beam model described in §4.3.1 provided a first-order correction for the effects of the primary beam on the Cygnus A spectrum. In addition to an initial gain model that incorporated the measured responses of the analog and digital filters in the PAPER pipeline, fourth-order polynomials in frequency were fit for the bandpass functions ($g_{i\nu}$) of antennas in the array.

At each stage, the residuals remaining after flux-calibration can help to highlight baselines with poorer fit scores–a hallmark of convergence to a non-global minimum. Once all problematic baselines have been eliminated, simultaneous fitting of phase and gain parameters can proceed until time-variable effects such as ionospheric distortion and changing primary beam response towards strong sources dominate residual visibilities.

4.5.6 Excision of Strong Point-Sources

Strong point-sources, while useful for initial calibration, complicate further calibration by degrading sensitivity toward other positions on the sky. This is especially true for small arrays, since the sidelobe response toward a source scales approximately as $\sqrt{N_{vis}}$, where N_{vis} is the number of independent locations in the uv-plane that are being phased and summed. For PWA-4 and PGB-8, sources stronger than a few hundred Jy can dominate the response of a synthesized beam, even when using earth rotation synthesis.

Deep removal of the strongest sources is essential for improving primary beam models using sources at a variety of declinations. A time-invariant model of source position, flux density, and spectral index is moderately effective at removing these sources, depending on the accuracy of current beam models and the activity of the ionosphere. For PGB-8, we were able to achieve ∼95% suppression of sources using static modeling. However, the residual flux densities of the strongest five sources (Sun, Cygnus A, Cassiopeia A, Virgo A, Taurus A) were still strong enough to introduce noticeable imaging sidelobes. Improving source removal required fitting for offsets in source position and flux spectrum as a function of time. We attribute these variations to ionospheric refraction (Fig. 4.10) and defects in the primary beam model, respectively. All of these components were necessary to achieve source suppression at a level of 1e-3 to 2e-4, depending on a source's position within the primary beam.

4.6 Observational Results

Results were obtained using 3-day observations with a single east-west linear polarization for both PWA-4 and PGB-8 deployments. For both sets of data, the five strongest sources visible from the northern hemisphere (Sun, Cygnus A, Cassiopeia A, Virgo A, and Taurus A) were removed following the procedure described in §4.5.6. Residuals associated with these sources reflect deviations of their spectra (as viewed through an imperfectly calibrated primary beam) from a strict power-law. The Sun was further suppressed via the application of a delay/fringe-rate filter (§4.5.5). Al-

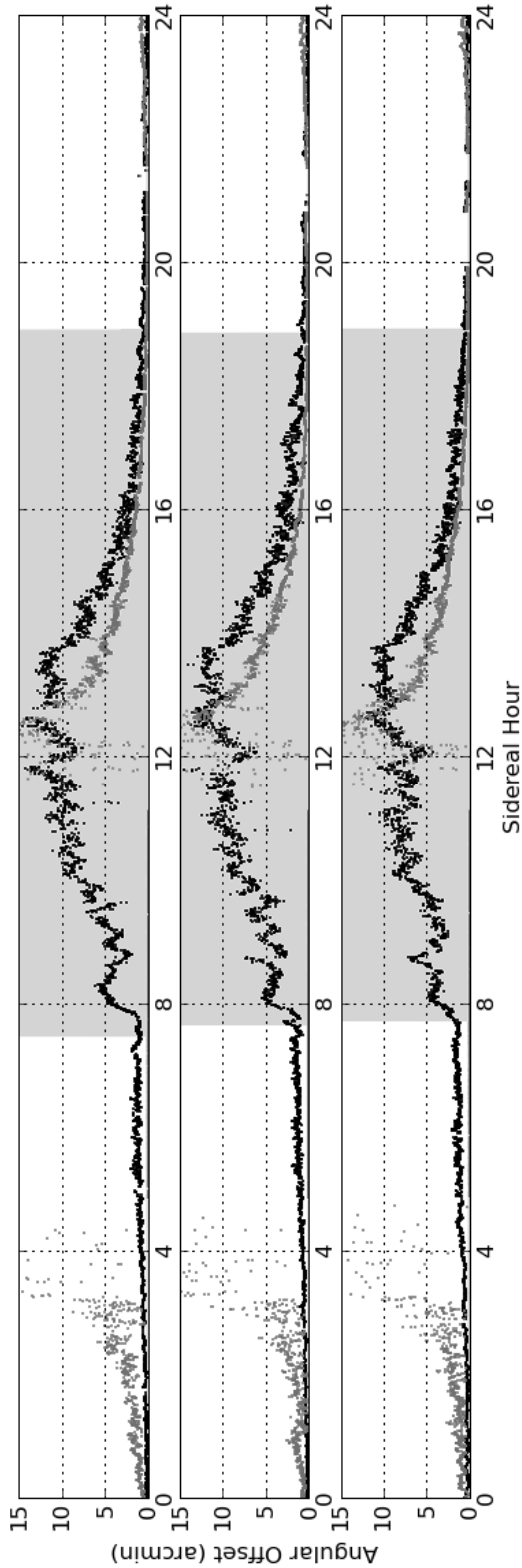

Figure 4.10: Illustrated above are the magnitudes of the angular offsets, primarily in the zenith direction, of Cygnus A (gray dots) and Cassiopeia A (black dots) fit as a function of time relative to their cataloged positions for 3 days of observations (top to bottom) with PGB-8. Both sources exhibit greater variability in position when the Sun was below the horizon (gray shaded area), presumably as a result of greater ionospheric variability at those times. The offsets of both sources also exhibit a smoothly varying component that repeats day-to-day, which may be attributed to refraction from spherical and wedge components of the ionosphere (Komesaroff 1960).

though these filters potentially corrupt wider areas of the map, the strength and variability of the Sun is such that their application greatly improves imaging of the daytime sky.

4.6.1 PWA-4/PGB-8 All-Sky Map

The all-sky map in Figure 4.13 illustrates average flux density per beam area over a band between 138.8 MHz and 174.0 MHz. The northern hemisphere was imaged using PGB-8 data with peak beam response at declination +38.5°. The southern hemisphere was imaged with data from PWA-4, which employed both a more compact array configuration and a broader primary beam. Peak beam response in the southern hemisphere lies at declination −26.7°. Imaging over the full 35.2-MHz bandwidth was performed by generating full maps in 1.46-MHz intervals and then summing maps. Each map was generated using 200 facets, with phase centers equally spaced around the sphere. While each imaged facet employed W-projection (Cornwell et al. 2003) to correct for the curvature of the sky, facet imaging was nonetheless required to incorporate differences in data weighting between phase centers. This weighting optimizes SNR by accounting for the array's direction-dependent gain toward that location as it drifts through the primary beam.

Thermal noise and point-source sidelobes both contribute to the noise level seen in this all-sky map. As mentioned in §4.6, only the strongest five sources have been removed from this map prior to imaging. Image-domain CLEAN deconvolution was used to achieve modest suppression of sidelobes from other sources. The complexities of wide-field imaging with drift-scan data make further post-imaging deconvolution impractical, since CLEAN components of gridded visibilities may not be directly applied to ungridded visibility data. Our plan is to use imaging as a pathway for identifying sources that may then be fit and removed from raw visibilities. With this in mind, we have devoted little additional effort toward suppressing sidelobes of imaged sources. From the map in Figure 4.13, we measure an RMS flux density of noise between point sources of 80 mJy (4.9 K using a synthesized beam area of 2.15e-5 steradians at 156.4 MHz). As we discuss in §4.6.2 and in §4.6.3, this noise

level exceeds thermal noise by a factor of eight, and is dominated by point-source sidelobes, with only a small contribution from galactic synchrotron emission at lower frequencies.

The position-dependent response of primary beams in each array is evidenced by changing noise levels in the map as a function of declination. Noise levels in the map also change with right-ascension because system temperature depends on the average temperature of galactic synchrotron emission across the primary beam. To illustrate the levels of thermal noise in the all-sky map, we repeated the map-making process detailed above, but with alternating integrations added with opposite signs, so that celestial sources with slow time variation were heavily suppressed relative to thermal noise. The map generated by this technique, illustrated in Fig. 4.14, shows that for the PGB-8 portion of the map, noise levels range from 10 mJy to 50 mJy (620 mK to 3.1 K using a synthesized beam area of 2.15e-5 steradians at 156.4 MHz) in the declination range of peak sensitivity of the array.

4.6.2 Thermal Noise Contributions to PGB-8 Map

The map statistics in the direction of a colder patch of the galactic synchrotron were evaluated using 1.46-MHz facet images with phase centers at $(11^h20, 30°00)$, $(11^h20, 40°00)$, $(12^h00, 30°00)$, and $(12^h00, 40°00)$. Each frequency interval and phase center was imaged twice with independent sets of alternating (odd and even time index) 14-second integrations from the 3-day PGB-8 observation described in 4.6.1. For each facet, these two images were summed and differenced to produce images of the celestial sky and of thermal noise, respectively. For the following analysis, image statistics were evaluated independently over 10° diameter circular patches around each phase center and averaged over facets afterward.

Thermal noise levels were inferred from measured RMS pixel values according to the equation:

$$\sigma_{\mathrm{px}} = \left\langle \frac{2k_B T_{\mathrm{rms}} \Omega_s}{\lambda^2} \right\rangle = \frac{2k_B T_{\mathrm{sys}} \Omega_B}{\lambda^2 \sqrt{\Delta\nu \tau N(N-1)}}, \tag{4.5}$$

where Ω_s is the solid angle of a synthesized beam; Ω_B is the solid angle of the primary

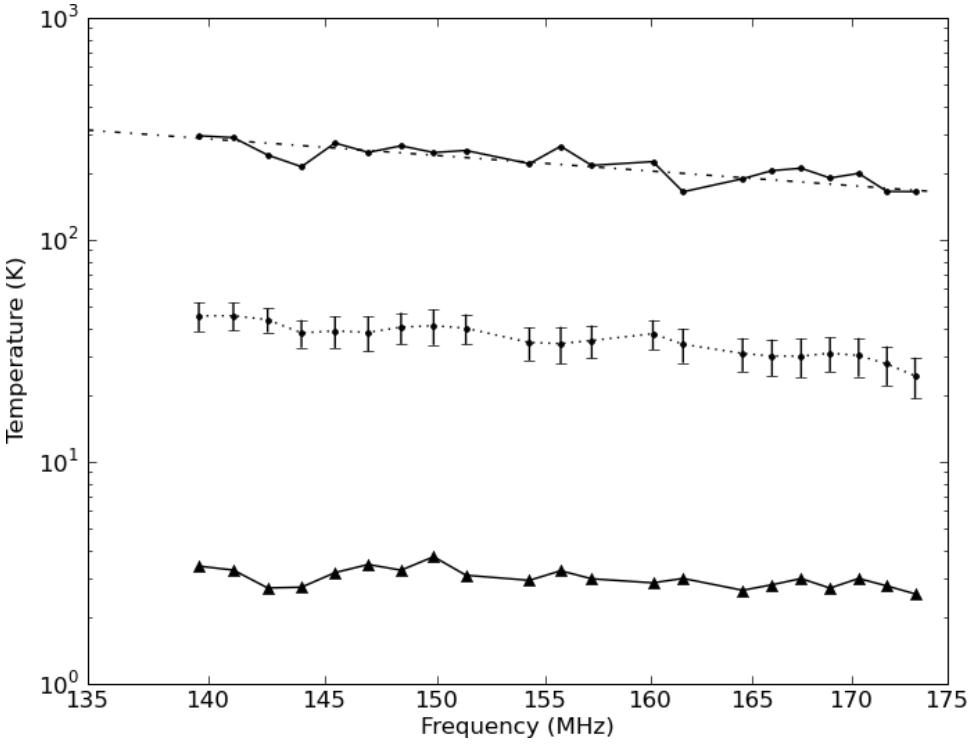

Figure 4.11: We imaged a colder patch of synchrotron sky using 3 days of PGB-8 data grouped into 1.46-MHz bands, adding integrations with alternating signs so that thermal noise could be estimated from RMS pixel values (solid, triangles). From these values, a system temperature was inferred (solid, dots) using Equation 4.5. Comparison with a predicted synchrotron spectrum for this patch of sky (dot-dashed) indicates that sky-noise dominates the system temperature of PGB-8. The average sky temperature (dotted, 2σ error bars) measured over these facets is substantially lower than the predicted synchrotron spectrum because large-scale structure is resolved out by the partially filled aperture of PGB-8.

beam; $\Delta\nu$ is the bandwidth of data used in the image; τ is the integration time; N is the number of antennas in the array; and $< ... >$ denotes the RMS value. The thermal noise images produced by differencing integrations yield the $T_{\rm rms}$ estimates plotted in Fig. 4.11 for each 1.46-MHz band.

From each measured $T_{\rm rms}$ we also inferred the system temperature $T_{\rm sys}$ present before integration in the correlator. A comparison of this temperature with the $T_{\rm sky}$ predicted for a colder patch of the galactic synchrotron (see §4.3.2) shows that the PGB-8 system temperature is dominated by sky-noise. Finally, the standard deviation of pixel values in the summed facet images exceeds thermal noise, implying contributions from celestial point sources, their sidelobes, and galactic synchrotron emission. The temperature associated with this pixel distribution reflects a sky temperature as perceived by the partially filled aperture of PGB-8. This perceived temperature is approximately a factor of 6 lower than the temperature of the galactic synchrotron because sparse samplings of short uv-spacings cause much of the emission from this source to be resolved out.

4.6.3 Angular Power Spectra

The angular power spectra illustrated in Figure 4.12 are measured using the PGB-8 data described in §4.6. Our analysis follows the techniques developed in White et al. (1999) for calculating angular power spectra with interferometric data in the flat-field approximation. To make this approximation, gridded visibility data projected towards each of the four phase centers defined in §4.6.2 were limited to a 10° diameter FoV by applying a circular boxcar windowing function W, whose effect in the uv-plane is to convolve by the complementary Fourier kernel $\tilde{W}(\vec{u})$. Within each facet, angular power spectra were calculated using the equation:

$$C_\ell = \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2$$

$$C(u) \simeq \left(\frac{\lambda^2}{2k_{\rm B}\Omega_{\rm W}}\right)^2 \frac{\oint \left|V(\vec{u}) * \tilde{W}(\vec{u})\right|^2 d^2u}{2\pi u}, \qquad (4.6)$$

where $\Omega_{\rm W}$ represents the solid angle spanned by W; and $V(\vec{u})$ represents the measured visibility as a function of position $\vec{u} \equiv (u, v)$ in the uv-plane; $\oint$ denotes integration around a ring of $|\vec{u}| = u$. The relationship between uv-coordinates and angular wavenumber is defined by $\ell \simeq 2\pi\,|\vec{u}|$. In practice, $V(\vec{u})$ is accompanied by sample weights $B(\vec{u})$, so that Equation 4.6 becomes:

$$C(u) \simeq \left(\frac{\lambda^2}{2k_{\rm B}\Omega_{\rm B}}\right)^2 \frac{\oint \left|(B(\vec{u})V(\vec{u})) * \tilde{W}(\vec{u})\right|^2 d^2u}{\oint \left|B(\vec{u}) * \tilde{W}(\vec{u})\right|^2 d^2u}, \tag{4.7}$$

where $\Omega_{\rm B}$ represents the solid angle spanned by the primary beam.

For each facet, separate $C(u)$ spectra were formed in 1.46-MHz intervals by integrating around rings in the uv-plane that were logarithmically spaced in radius and width. Within each ring, data were weighted by the square of the primary beam response toward phase center for optimal SNR. We estimated the contribution of thermal noise to the error in each $C(u)$ measurement by generating power spectra for thermal noise images. This was done using the technique of adding integrations with alternating signs described in the previous section. The error from sample variance was much larger than the error from thermal noise, owing to the limited aperture sampling of the PGB-8 array and the limited area on the sky employed for this analysis. Error contributions from sample variance were estimated by measuring the standard deviation of all $C(u)$ measurements in rings of constant u in the uv-plane for each facet, and dividing this by the square root of the number of independent $C(u)$ measurements made. We estimated independent measurements by calculating the size of a coherence patch in the uv-plane that arises from $\tilde{W}(\vec{u})$ and then degrading the pixel resolution in the uv-plane to a corresponding size. At the lower resolution, each pixel sample was considered an independent measurement, with the caveat that for a real-valued sky, only half of the samples in a $|\vec{u}| = u$ ring are independent.

PGB-8 angular power spectra are dominated by unclustered point sources, as indicated by the approximate ℓ^2 power-law in $\ell(\ell+1)C_\ell/2\pi$ of the angular power spectra plotted in Figure 4.12. However, at lower radio frequencies for $\ell < 100$, we see a flattening of these spectra to a shallow, negative power-law that shows qualitative

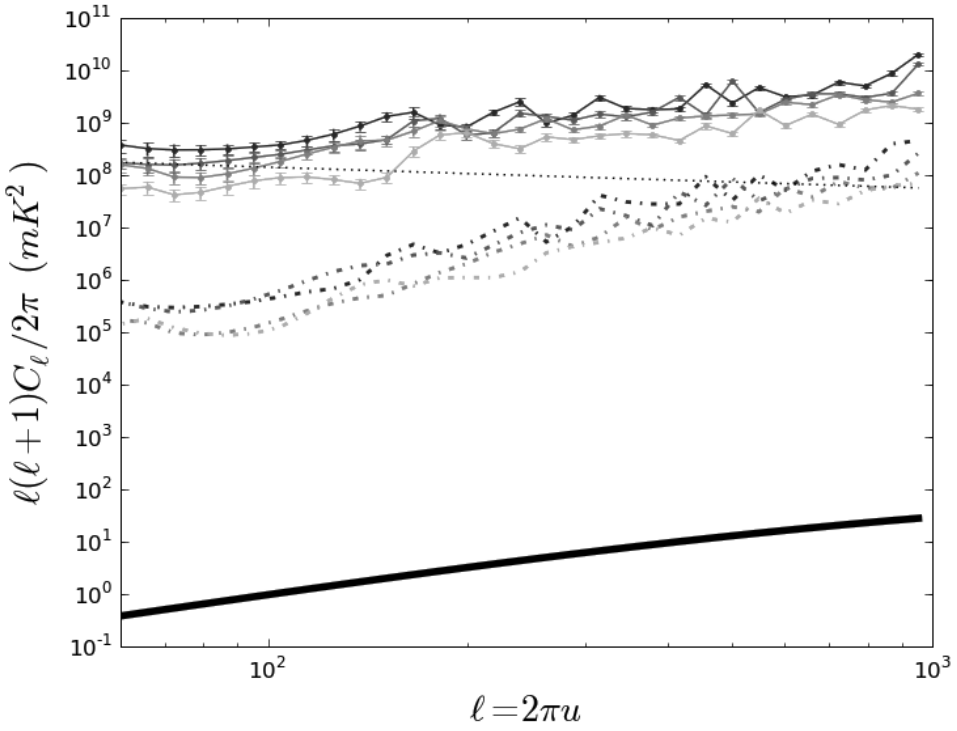

Figure 4.12: Shown above are angular power spectra measured at 146.9, 155.7, 164.5, and 173.3 MHz (solid, top to bottom at $\ell = 100$, respectively, with 2σ error bars), averaged over 1.46-MHz intervals for four 10° facets near $11^h40, 35°00$. These spectra contain contributions from both point sources and galactic synchrotron emission. Dominant errors are from cosmic variance, which has been estimated from the standard deviation of measurements at the same ℓ in the uv-plane. Dot-dashed lines indicate angular power spectra of "noise images", generated by adding integrations with alternating signs, and are indicative of the thermal noise level in these measurements. Also plotted: a fiducial reionization power spectrum at $z = 9.2$ (ν =140 MHz, thick solid) and a predicted galactic synchrotron spectrum at 147 MHz (dotted) (Tegmark et al. 2000; Santos et al. 2005).

agreement with the $C_{\ell,\mathrm{sync}} \propto \ell^{-2.4}$ model adopted by Tegmark et al. (2000) and Santos et al. (2005). Progress in modeling point sources and in subtracting spectrally smooth emission associated with synchrotron sources will facilitate the suppression of these foregrounds, with a fundamental limit placed by the thermal limit of these observations (Fig. 4.12, dot-dashed), which we measure to be 310 mK for $\ell = 100$ at 160 MHz. Overall, our measurements indicate that over the range $10^2 < \ell < 10^3$, the fundamental sensitivity level of these observations is between two and three orders of

magnitude in temperature above the fiducial model of reionization at $z = 9.2$ (ν =140 MHz) used by Santos et al. (2005).

4.7 Conclusion

PAPER is following an incremental build-out strategy that emphasizes the characterization and optimization of instrumental performance at each stage of development. We have given special attention to designing antenna elements whose smooth spatial and spectral responses are conducive to calibration that will be required to model and remove strong foregrounds to the 21cm EoR signal. Facilitating the incremental build-out of PAPER is a series of digital correlators that take advantage of the scalability and reusability of CASPER's packet-switching correlator architecture. We are developing post-correlation calibration, imaging, and analysis pipelines as part of the open-source AIPY software package. This package emphasizes the needs of low-frequency interferometry and facilitates experimentation with new analysis architectures and calibration techniques.

With results from the PWA-4 and PGB-8 deployments in Western Australia and Green Bank, respectively, we demonstrate a first level of calibration that is sufficient for modeling strong sources to a level where ionospheric refraction and temperature-dependent gains must be taken into account. This calibration relies heavily on a static model of the primary beam. Next steps in calibration will focus on the first-order removal of temperature-dependence in antenna gains and on measuring deviations of the primary beam of each antenna from the computed model.

Data from the PWA-4 and PGB-8 deployments have been used to generate an all-sky map that attains a thermal noise level of 10 mJy/Beam (corresponding to 620 mK, for a 2.15e-5 steradian synthesized beam size at 154.4 MHz) integrated across a 138.8-MHz to 174.0-MHz band. This achievement represents a first iteration in a cycle of improvement wherein sky models are used to improve array calibration, which in turn allows increasingly accurate sky models to be generated. In analyzing the noise characteristics of this map, we find the system temperature of the PGB-8

array to be consistent with a model of galactic synchrotron emission. Angular power spectra generated from these data indicate that point sources are currently the dominant foreground to the EoR signature, with evidence for contributions from galactic synchrotron emission at lower radio frequencies and angular wavemodes. Errors in C_ℓ measurements are dominated by sample variance, but by differencing integrations, we measure the thermal noise level at $\ell = 100$ and 160 MHz to be 310 mK–a sensitivity threshold that lies between two and three orders of magnitude in temperature above fiducial reionization models. These results demonstrate the need for a next level of calibration, modeling, and foreground suppression that will be pursued in the next PAPER deployments.

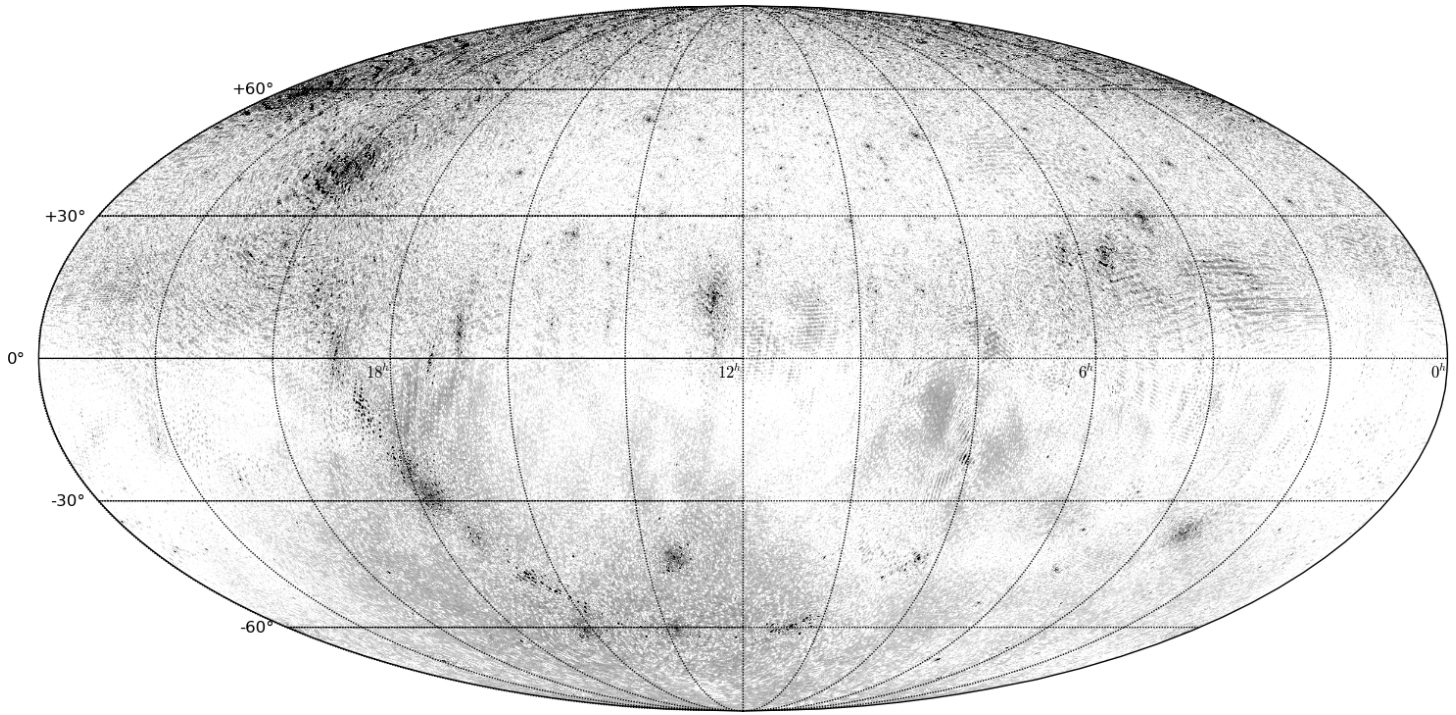

Figure 4.13: Illustrated above is an all-sky equatorial (J2000) map (see §4.6.1) averaged between 138.8 MHz and 174.0 MHz, in units of $\log_{10}(Jy/Beam)$ ranging from -1 (white) to 1 (black). Northern hemisphere data were obtained with data from PGB-8 and southern hemisphere with data from PWA-4, both with 3 days of observation using a single east-west linear polarization. Models of Cygnus A ($19^h59, 40°44$), Cassiopeia A ($23^h23, 58°49$), Taurus A ($05^h35, 22°01$), Virgo A ($12^h31, 12°23$), and the Sun ($1^h10, 7°27$) have been subtracted/filtered from visibilities prior to imaging. In the northern hemisphere, SNR peaks near ($12^h00, 40°00$); sensitivity degrades with declination because of declining primary beam response and with right ascension because of increasing levels of sky-noise from galactic synchrotron emission. In the region of peak SNR, measured temperatures reach a minimum of 4.9 K, as measured by the incompletely sampled aperture of PGB-8. Thermal noise in this same area is measured to be 620 mK (see Fig. 4.14), indicating that point-source sidelobes and diffuse galactic synchrotron emission dominate the noise-floor in this map

Figure 4.14: Plotted above (top, gray-scale) is a noise map obtained by imaging the PGB-8 data used to produce Figure 4.13, but with consecutive integrations added with alternating signs. The map is in units of $\log_{10}(Jy/Beam)$ ranging from -2 (white) to 0 (black). Flux densities in this map are indicative of thermal noise in the northern hemisphere of Figure 4.13. The lower line-graph illustrates a the cut through the map at declination $+38.5°$, where PGB-8 is optimally sensitive. Flux densities in this cut range from 10 mJy to 50 mJy, corresponding to temperatures of 620 mK and 3.1 K at 156.4 MHz, using a synthesized beam size of 2.15e-5 steradians.

Chapter 5

Conclusion

5.1 Calibration and Foreground Removal for First-Generation EoR Arrays

As discussed in Chapter 4, numerous astrophysical foregrounds interfere with the direct detection of a global EoR signature, including polarized galactic synchrotron emission, continuum point-sources, and galactic/extra-galactic free-free emission. The design of interferometers targeting EoR detection are strongly influenced by the estimated angular power-spectra and frequency dependence of these foregrounds (Wang et al. 2006; Bowman et al. 2008). Because there are few measurements of such foreground properties in this frequency band, foreground characterization will be a vital step en route to any EoR detection.

The calibration work described in Chapter 4 results in refinements to a celestial sky model dominated by point sources and galactic synchrotron emission. The preliminary sky model generated using data from 4- and 8-antenna deployments of the PAPER experiment illustrates the degree to which sidelobes from point sources create a noise floor that interferes with the detection of cosmic reionization. Subtracting models from measured data will substantially suppress these foregrounds, but will leave residuals that result from calibration errors and incorrect compensation for ionospheric distortion. Furthermore, point-source confusion will prohibit additional

model refinement by direct imaging once the flux limit is reached where multiple sources reside within each synthesized beam. We will need to investigate the degree to which this confusion noise may be further suppressed using models derived from existing source catalogs and how the power spectrum of residual confusion noise may be differentiated from the predicted power spectrum of the EoR signal as a function of frequency.

Galactic synchrotron emission has a smoothly varying frequency spectrum, but we do not know the degree to which Faraday screens introduce polarization structure onto the galactic synchrotron or the degree to which calibration errors will couple this polarization structure into a frequency dependence in the power spectrum. Confusion noise will also have a frequency-dependent power spectrum as a result of the changing shape of a synthesized beam with frequency. EoR experiments must work to characterize these effects, as they have direct impact on how an interferometric array should be configured for maximum sensitivity to cosmic reionization. Since the 3-dimensional power spectrum of EoR fluctuations may be probed perpendicular to the line-of-sight using interferometric samples of the angular power spectrum and parallel to the line-of-sight using the Fourier transform of frequency data (Morales 2005), an EoR experiment has considerable flexibility in choosing an angular scale and inverse-frequency scale. Results from EoR simulations and measured foreground characteristics must be used to optimize antenna placement within an array, focusing limited collecting area towards detecting specific components of the EoR power spectrum that differentiate it from foreground contamination.

5.2 Next-Generation Arrays

Next-generation EoR arrays will face many technical challenges in their construction. Among the requirements are large collecting areas and large fields-of-view for survey speed, a large frequency range for redshift coverage, low cost, and precise calibration. Survey speed considerations and the decreasing cost of digital computing favor arrays with large numbers of small-aperture antenna elements. However, the

correlators in use for all existing large radio interferometers have computing costs that scale quadratically with the number of elements N, in the large-N limit. Novel techniques are being explored for griding electric field data before it is presented to the correlator so that FFT techniques can be employed to change the scaling of computational complexity in correlators from N^2 to $N \log N$. The FFT Telescope (Tegmark & Zaldarriaga 2008) and the MOFF correlator (Morales 2008) are two examples of such architectures that become computationally efficient for dense arrays. There is little doubt that developing technologies for building large-N correlators at reasonable cost is critical, since the capabilities of next-generation EoR arrays will be determined by the availability of digital computing power and high-bandwidth interconnect.

Calibrating and imaging the correlator output also poses a substantial computational burden for these arrays. The need to resolve time-variable ionospheric conditions drives correlators to shorter integration times that increase data rates. For large arrays, these data rates can reach levels where the traditional data reduction path of data storage and off-line post-processing is no longer viable. Heightened time-dependent calibration requirements, wide-field imaging with non-coplanar arrays, and real-time RFI mitigation techniques increase the computational complexity of post-processing. As a result, next-generation arrays will find that calibration and imaging require digital processing comparable in complexity to correlators (Wright 2005), while the algorithmic complexity of real-time imaging and calibration suggests that CPU- or GPU-based cluster-computing solutions may be appropriate (Wayth et al. 2007).

The pace of advances in digital signal processing (DSP) technology dictates that digital processors in these systems will need to be upgraded multiple times over the operational lifetimes of these arrays. The development of technologies that facilitate consistent, scalable instrument architectures, interoperability between families of DSP hardware, and interoperability between hardware generations may help ease such transitions, allowing new digital technology to be quickly co-opted to expand the science capabilities and power efficiency of these arrays. As shown in Chapter 2, solutions employing general-purpose digital hardware, packetized communication

protocols, and high-level DSP libraries are viable solutions for high-performance astronomy signal processing and have the property of being modular and upgradeable. A growing community of radio astronomers and electrical engineers are investing in shared solutions centered around these technologies, finding that the platform-independent, open-source approach reduces development time, risk, and cost to a given project, and enhances opportunities for innovative approaches owing to the rapid dissemination of information and techniques.

5.3 Concluding Remarks

Low-frequency interferometry is undergoing a period of exciting development as a new generation of facilities are coming online to address extraordinarily well-defined science goals. It is a rare thing to be engaged in the early development of a field–the tomography of cosmic reionization–that, if the cosmic microwave background (CMB) is any model, will attract an immense amount of interest as it matures. As with CMB analysis, it is to be expected that sophisticated analysis techniques will be explored to tease out the subtle and complex information that is embedded in the geometry of reionization. However, using low-frequency interferometers to study cosmic origins may be unique in the degree of algorithmic and technical sophistication needed very early in the signal processing chain. While we can begin to look ahead to techniques for constraining reionization models using three-dimensional power spectra, it is clear that a great part of the work lies much closer to the instrument–in designing scalable signal processing architectures, implementing real-time data analysis and calibration systems, and in re-evaluating the basic techniques of interferometry.

With basic development of low-frequency interferometers being driven by reionization experiments and Square Kilometer Array (SKA) prototypes, it is easy to get excited about the potential for finding new and unheralded phenomena along the way. Reionization experiments will be particularly likely to uncover the unexpected because of the incredible level of foreground removal that they will need to achieve. With point sources and synchrotron emission suppressed by five to seven orders of

magnitude, there is very little room for stray signals to hide. Along with this exciting possibility also comes the sobering realization that any number of expected or unexpected foreground signals may pose a substantial threat to the science objectives in question.

That said, first-generation reionization experiments are making quick progress, as are several SKA prototype arrays. While it may seem premature to contemplate second-generation arrays while first-generation versions are still in their developmental stages, doing so is necessary to maintain the substantial momentum that has accumulated behind the development of low-frequency interferometers. Continuing to develop interest, funding, and community organization will be vital for keeping this exciting field moving at such a rapid pace.

Bibliography

2004, RocketIO Tranceiver User Guide (UG024 v2.5), Xilinx, http://www.xilinx.com

2005, Virtex-II Pro and Virtex-II Pro X Platform FPGAs: Functional Description (DS083-2 v4.5), Xilinx, http://www.xilinx.com

Baars, J. W. M., Genzel, R., Pauliny-Toth, I. I. K., & Witzel, A. 1977, A&A, 61, 99

Barkana, R. & Loeb, A. 2005a, ApJ, 624, L65

—. 2005b, ApJ, 626, 1

Becker, R. H., Fan, X., White, R. L., Strauss, M. A., Narayanan, V. K., Lupton, R. H., Gunn, J. E., Annis, J., Bahcall, N. A., Brinkmann, J., Connolly, A. J., Csabai, I., Czarapata, P. C., Doi, M., Heckman, T. M., Hennessy, G. S., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., McKay, T. A., Munn, J. A., Nash, T., Nichol, R., Pier, J. R., Richards, G. T., Schneider, D. P., Stoughton, C., Szalay, A. S., Thakar, A. R., & York, D. G. 2001, AJ, 122, 2850

Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, A&A, 487, 419

Blackman, R. B. & Tukey, J. W. 1958, The measurement of power spectra (Dover Publications Inc.)

Bowman, J. D., Barnes, D. G., Briggs, F. H., Corey, B. E., Lynch, M. J., Bhat, N. D. R., Cappallo, R. J., Doeleman, S. S., Fanous, B. J., Herne, D., Hewitt, J. N., Johnston, C., Kasper, J. C., Kocz, J., Kratzenberg, E., Lonsdale, C. J., Morales,

M. F., Oberoi, D., Salah, J. E., Stansby, B., Stevens, J., Torr, G., Wayth, R., Webster, R. L., & Wyithe, J. S. B. 2007, AJ, 133, 1505

Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, ApJ, 638, 20

—. 2008, ArXiv:0807.3956

Bradley, R. 2006, A Low Cost Screened Enclosure for Effective Control of Undesired Radio Frequency Emissions, NRAO EDIR 317, http://www.gb.nrao.edu/electronics/edir

Bradley, R., Backer, D., Parsons, A., Parashare, C., & Gugliucci, N. E. 2005, in Bulletin of the American Astronomical Society, Vol. 37, 1216–+

Chang, C., Wawrzynek, J., & Brodersen, R. W. 2005, IEEE Design and Test of Computers, 22, 114

Chapiro, D. M. 1984, PhD thesis, Stanford Univ., CA.

Conway, J. E., Cornwell, T. J., & Wilkinson, P. N. 1990, MNRAS, 246, 490

Cornwell, T. & Fomalont, E. B. 1989, in Astronomical Society of the Pacific Conference Series, Vol. 6, Synthesis Imaging in Radio Astronomy, ed. R. A. Perley, F. R. Schwab, & A. H. Bridle, 185–+

Cornwell, T. J. & Evans, K. F. 1985, A&A, 143, 77

Cornwell, T. J., Golap, K., & Bhatnagar, S. 2003, W-Projection: A New Algorithm for Non-Coplanar Baselines, EVLA Memo 67

Crochiere, R. & Rabiner, L. R. 1983, Multirate Digital Signal Processing (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1983. 336 p.)

D'Addario, L. 2001, Correlators: General Design Considerations, ATA Memo 24, Allen Telescope Array, http://ral.berkeley.edu/ata/memos/

Demorest, P., Ramachandran, R., Backer, D., Ferdman, R., Stairs, I., & Nice, D. 2004, in Bulletin of the American Astronomical Society, 1598–+

Dick, C. 2000, High-Performance 1024-Point Complex FFT/IFFT V2.0, http://www.xilinx.com

Fan, X., Carilli, C. L., & Keating, B. 2006, ARA&A, 44, 415

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, Phys. Rep., 433, 181

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, ApJ, 622, 759

Hall, G., ed. 1988, The ARRL Antenna Book, 15th Edition (American Radio Relay League, Newington, Connecticut, 1988.), 16

Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1

Heiles, C., Goldston, J., Mock, J., Parsons, A., Stanimirovic, S., & Werthimer, D. 2004, in Bulletin of the American Astronomical Society, 1476–+

Högbom, J. A. 1974, A&AS, 15, 417

Jackson, T. L. & Farrell, W. M. 2006, Parallel Processing Symposium, International, 44, 2942

Jenet, F. A. & Anderson, S. B. 1998, PASP, 110, 1467

Jennison, R. C. 1958, MNRAS, 118, 276

Johnson, R. C. 1993, Antenna Engineering Handbook, 3rd Edition (New York, McGraw-Hill, 1993.), 18–23

Komesaroff, M. M. 1960, Australian Journal of Physics, 13, 153

Loeb, A. & Barkana, R. 2001, ARA&A, 39, 19

Loeb, A. & Zaldarriaga, M. 2004, Physical Review Letters, 92, 211301

Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., & Zahn, O. 2008, Phys. Rev. D, 78, 023529

McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, ApJ, 653, 815

Morales, M. F. 2005, ApJ, 619, 678

—. 2008, ArXiv:0812.3669

Morales, M. F., Bowman, J. D., Cappallo, R., Hewitt, J. N., & Lonsdale, C. J. 2006, New Astronomy Review, 50, 173

Page, L., Hinshaw, G., Komatsu, E., Nolta, M. R., Spergel, D. N., Bennett, C. L., Barnes, C., Bean, R., Doré, O., Dunkley, J., Halpern, M., Hill, R. S., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Peiris, H. V., Tucker, G. S., Verde, L., Weiland, J. L., Wollack, E., & Wright, E. L. 2007, ApJS, 170, 335

Parsons, A. 2009, IEEE Signal Processing Letters, accepted

Parsons, A., Backer, D., Chang, C., Chapman, D., Chen, H., Crescini, P., de Jesus, C., Dick, C., Droz, P., MacMahon, D., Meder, K., Mock, J., Nagpal, V., Nikolic, B., Parsa, A., Richards, B., Siemion, A., Wawrzynek, J., Werthimer, D., & Wright, M. 2006, in Asilomar Conference on Signals and Systems, Pacific Grove, CA, 2031–2035

Parsons, A., Backer, D., Siemion, A., Chen, H., Werthimer, D., Droz, P., Filiba, T., Manley, J., McMahon, P., Parsa, A., MacMahon, D., & Wright, M. 2008, PASP, 120, 1207

Parsons, A. R. & Backer, D. C. 2009, Submitted to AJ. ArXiv:0901.2575

Parsons, A. R., Backer, D. C., Bradley, R. F., Aguirre, J. E., Benoit, E. E., Carilli, C. L., Foster, G. S., Gugliucci, N. E., Herne, D., Jacobs, D. C., Lynch, M. J., Manley, J. R., Parashare, C. R., Werthimer, D. J., & Wright, M. C. H. 2009, Submitted to AJ. ArXiv:0904.2334

Pearson, T. J. & Readhead, A. C. S. 1984, ARA&A, 22, 97

Plana, L. A., Furber, S. B., Temple, S., Khan, M., Shi, Y., Wu, J., & Yang, S. 2007, IEEE Des. Test, 24, 454

Rabiner, L. R. & Gold, B. 1975, Theory and application of digital signal processing (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1975. 777 p.)

Roberts, D. H., Lehar, J., & Dreher, J. W. 1987, AJ, 93, 968

Rogers, A. E. E. & Bowman, J. D. 2008, AJ, 136, 641

Rybicki, G. B. & Lightman, A. P. 1979, Radiative processes in astrophysics (New York, Wiley-Interscience, 1979. 393 p.)

Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575

Sault, R. J. 1990, ApJ, 354, L61

Sault, R. J., Teuben, P. J., & Wright, M. C. H. 1995, in Astronomical Society of the Pacific Conference Series, Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes, 433–+

Slurzberg, M. & Osterheld, W. 1961, Essentials of Radio-Electronics, 2nd Edition (New York, McGraw-Hill, 1961.), 595

So, K. H. 2007, PhD thesis, Berkeley Wireless Research Center, UC Berkeley, CA.

So, K. H. & Brodersen, R. W. 2006, in 16th International Conference on Field Programmable Logic and Applications (Madrid, Spain), 349–354

Tegmark, M., Eisenstein, D. J., Hu, W., & de Oliveira-Costa, A. 2000, ApJ, 530, 133

Tegmark, M. & Zaldarriaga, M. 2008, ArXiv:0805.4414

Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2001, Interferometry and Synthesis in Radio Astronomy, 2nd Edition (New York, Wiley-Interscience, 2001. 692 p.)

Vaidyanathan, P. P. 1990, Proc. IEEE, 78, 56

Wang, X., Tegmark, M., Santos, M. G., & Knox, L. 2006, ApJ, 650, 529

Wayth, R., Dale, K., Greenhill, L. J., Mitchell, D. A., Ord, S., & Pfister, H. 2007, in Bulletin of the American Astronomical Society, Vol. 38, 744–+

Weinreb, S. 1961, Proc. IEEE, 49, 1099

Weinreb, S. & D'Addario, L. 2001, Cost Equation for the SKA, SKA Memo 1, Square Kilometre Array, http://www.skatelescope.org/PDF/memos

White, M., Carlstrom, J. E., Dragovan, M., & Holzapfel, W. L. 1999, ApJ, 514, 12

Wright, M. 2005, Real Time Imaging, SKA Memo 60, Square Kilometre Array, http://www.skatelescope.org/PDF/memos

Yatawatta, S., Zaroubi, S., de Bruyn, G., Koopmans, L., & Noordam, J. 2008, ArXiv:0810.5751

Yen, J. L. 1974, A&AS, 15, 483

Zahn, O., Lidz, A., McQuinn, M., Dutta, S., Hernquist, L., Zaldarriaga, M., & Furlanetto, S. R. 2007, ApJ, 654, 12