

# Square Kilometre Array Low Atomic commercial off-the-shelf correlator and beamformer

Grant A. Hampson,\* John D. Bunton<sup>ORCID</sup>, David Humphrey,  
Keith J. Bengston, Guillaume Jourjon, Andrew B. Bolin<sup>ORCID</sup>,  
Yuqing Chen, Euan R. Troup<sup>ORCID</sup>, Giles C. Babich  
and Jason C. van Aardt<sup>ORCID</sup>

CSIRO, Space and Astronomy, Marsfield, New South Wales, Australia

**Abstract.** The Square Kilometre Array Low is a next generation radio telescope, consisting of 512 antenna stations spread over 65 km, to be built in Western Australia. The correlator and beamformer (CBF) design is central to the telescope signal processing. CBF receives 6 Tera-bits-per-second (Tbps) of station data continuously and processes it in real time with a compute load of 2 Peta-operations-per-second (Pops). The correlator calculates up to 22 million cross products between all pairs of stations, whereas the beamformers (BFs) coherently sum station data to form more than 500 beams. The output of the correlator is up to 7 Tbps, and the BF 2 Tbps. The design philosophy, called “Atomic COTS,” is based on commercial off-the-shelf (COTS) hardware. Data routing is implemented in network switches programmed using the Programming Protocol-Independent Packet Processors (P4) language and the signal processing occurs in COTS field-programmable gate array (FPGA) cards. The P4 language allows routing to be determined from the metadata in the Ethernet packets from the stations. That is, metadata describing the contents of the packet determines the routing. Each FPGA card inputs a fraction of the overall bandwidth for all stations and then implements the processing needed to generate complete science data products. Generation of complete science products in a single FPGA is named here as Atomic processing. A Tango distributed control system configures the multitude of processing modes as well as maintaining the overall health of the CBF system hardware. The resulting 6 Tbps in and 9 Tbps out, 2 Pops Atomic COTS network attached accelerator occupies five racks and consumes 60 kW. © *The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: [10.1117/1.JATIS.8.1.011018](https://doi.org/10.1117/1.JATIS.8.1.011018)]

**Keywords:** antenna arrays; correlators; beam steering; accelerator architectures; communication switching; field-programmable gate arrays.

Paper 21103SS received Aug. 31, 2021; accepted for publication Dec. 22, 2021; published online Feb. 2, 2022.

## 1 Introduction

Building correlators and beamformers (CBFs) for a large interferometer has always been a challenge because of the high data rates and the need for real-time processing. The Square Kilometre Array<sup>1</sup> (SKA) is the next generation radio telescope with a “Mid” frequency array in South Africa and a “Low” frequency array in Australia. Here, the design of the SKA Low CBF is described. The challenge is a real-time system with continuous input data rates of 6 terabits per second (Tbps), equivalent compute of 2 Peta-operations per second (Pops) and output data rates that can exceed the input data rate. The solution used to meet this computing and communications challenge is described in this paper.

In the past century, application-specific integrated circuits (ASICs) were required to meet the requirements for correlators.<sup>2-4</sup> At the turn of the century, ASIC non-recurring engineering (NRE) costs continued to increase, and field-programmable gate arrays (FPGAs) started to replace them. Bespoke FPGA boards,<sup>5,6</sup> and FPGA/ASIC boards,<sup>7</sup> were integrated into a fixed

---

\*Address all correspondence to Grant A. Hampson, [grant.hampson@csiro.au](mailto:grant.hampson@csiro.au)

interconnection routing structure with each FPGA/ASIC performing a specific function. More recently, systems have been built using commercial components, including graphics processing units (GPUs) and network switches. For example, CASPER<sup>8</sup> technology has been used with network switches to connect bespoke FPGA boards together in MeerKAT,<sup>9</sup> while CHIME<sup>10</sup> and LEDA<sup>11</sup> combine GPUs and bespoke FPGA boards. COBALT<sup>12</sup> is an example of a fully GPU system.

In this paper, we describe a further advance that has become possible with the advent of commercial off-the-shelf (COTS) FPGA boards (Xilinx Alveo<sup>13</sup>) aimed for use in data centers. It builds on the CASPER approach to correlators: switches for data routing and FPGAs boards for computation. Also leveraged is the considerable support provided with COTS hardware; for example, free Xilinx software is provided for board monitoring, hardware configuration, and register interfaces. Use of the free software also allows an easy upgrade path to newer Alveo boards, and the free tools include everything needed to build FPGA images, without the need for expensive licenses. Combined with the fact that no time is used in hardware development, this can shorten the time to completion, in some cases by years. Compared with GPU implementations, the on board 100 GbE links avoid the use of network interface cards and leave the peripheral component interconnect express (PCIe) bus free for monitoring and control rather than requiring it for high-speed data transfers.

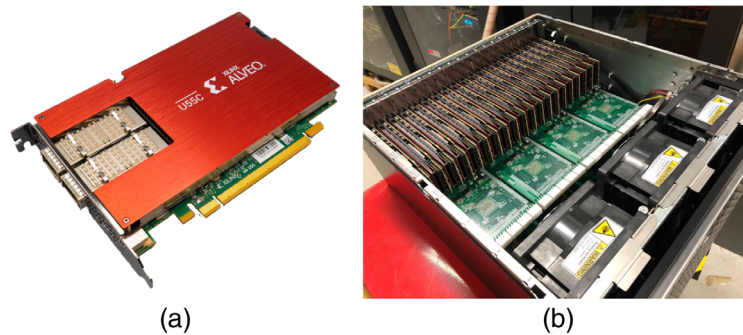
CASPER used standard network switches, but there can be a steep learning curve for new systems. An investigation of the Programming Protocol-Independent Packet Processors<sup>14</sup> (P4) language and hardware showed a metadata driven approach was possible. This made the P4 switch just an extension of what was done in a previous FPGA system with dedicated interconnection that the authors were familiar with. Data routing is now controlled through easy to design Yet Another Markup Language (YAML) files.

A final piece of the puzzle is the adoption of Atomic processing to avoid underutilization of compute resources in the FPGA. Even with two 100 GbE ports, the Alveo board has a lower input/output (I/O) per unit of compute than the CASPER SKARAB board. By Atomic processing, we mean all required processing on input data occurs in a single Alveo. An end-to-end BF or correlator incorporating filterbanks (FBs) and wavefront correction is implemented in a single FPGA for a subset of the bandwidth. This is enabled by high-bandwidth memory (HBM) directly attached to the FPGA that is on the Alveo board. Multiple operations are undertaken on a single FPGA with intermediate data buffered in the HBM between operations. The buffering allows operations to be asynchronous, which significantly eases the design constraints compared with previous FPGA designs where data synchronization between and within FPGAs was critical. We call the overall approach Atomic COTS as we achieve Atomic processing on COTS hardware.

We suggest that this approach is applicable to all correlators and BFs. Indeed, the authors are already using it for the BF of the new cryogenically cooled phased array feed being built for the Parkes 64 m dish. To assist other users, CSIRO is currently developing an open source Alveo reference design and is a beta tester for Precision Time Protocol (PTP) to enable timing synchronization between Alveos. The hardware is low in cost relative to bespoke hardware, readily available with short lead times, and easy to use within the supplied software and firmware environments. Although the I/O capacity of the Alveo card is lower than previous bespoke designs, the use of Atomic processing and P4 switching technology maximizes the compute per unit of I/O, allowing the full capabilities of the hardware to be utilized.

## 2 SKA Low BF and Correlator

COTS hardware results in a significant cost saving compared with bespoke hardware, both in capital cost as well as firmware and software development cost. The Xilinx COTS board chosen is shown in Fig. 1, together with similar hardware bought for the Parkes CryoPAF that is being used to verify common aspects of the two systems. The board is a standard PCIe interface card with no integrated fan, as the cooling fans are part of the server (also shown in Fig. 1) This is an advantage in the CBF as the expected life of the system is over 10 years, and the server fans can be replaced without halting processing in the Alveo. The Alveo PCIe cards come with driver software to monitor card health status, as well as for writing and reading registers to the FPGA



**Fig. 1** (a) Xilinx Alveo U55C COTS board containing a Virtex® UltraScale+™ HBM FPGA and two 100 GbE ports and (b) server containing three cooling fans and 20 PCIe slots each with a half-height previous-generation Alveo U50 board used for initial testing monitoring and control function, 100 GbE links and PTP.

shell. Users write code into a firmware kernel that talks to the PCIe shell, HBM and 100 GbE interfaces.

With Atomic processing (all correlator or BF processing in a single Alveo), it is necessary to route a small part of the total bandwidth (called a frequency channel) for all telescope elements to an Alveo U55C. For this function, COTS network switches programmable with the P4<sup>14</sup> language are used. This avoids the black box nature of standard network switches as well as providing greater control of packet routing and telemetry. With the P4 language, the header information of each data packet is read, and this is used to directly route data to the correct Alveo. Thus, a given frequency channel for all telescope elements is reliably routed to a single Alveo for processing. The only legacy Ethernet function needed is Address Resolution Protocol (ARP), which is required to keep connections to external standard Ethernet switches active.

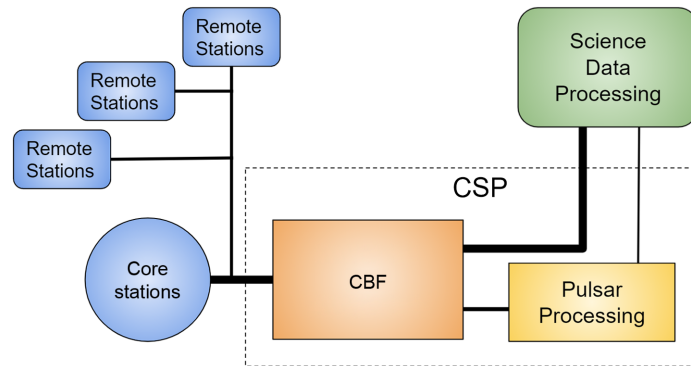
For SKA Low, the resulting Atomic COTS CBF meets the challenge of Low CBF processing. The hardware consists of 21 4U servers, each with 20 Alveo U55C PCIe boards, as shown in Fig. 1 (the 21st server is for redundancy and testing). Initial testing was with the Alveo U50LV. It was found sufficient computationally, but its dissipation limit of 75 W was too marginal, so a seamless upgrade to the U55C was made. Each Alveo connects via a single 100 GbE link to one of 11 second-layer 64-port P4 switches. These second-layer P4 switches connect to nine first-layer P4 switches that connect to stations and next level science processes. The complete system occupies five racks and is estimated to consume 60 kW of power, comprising P4 switches 15 kW, Alveo 32 kW, servers 11 kW, and rack cooling 2 kW. Power estimates are derived from Xilinx estimates cross checked against a pulsar timing (PST) BF implementation running in Alveo U50LVs. The use of COTS hardware and the supplied Xilinx software results in a considerable saving in time, cost, and effort compared with the previous bespoke solution.

It is interesting to compare with a similar earlier design. The LEDA correlator has a very similar FB (8 k versus 4 k channels), half the number of antennas, and about one-fifth the bandwidth (57 MHz). This makes the total correlation compute requirement about 1/20 of the SKA Low correlator. But as FBs scale linearly and correlation scales quadratically, the true scaling is closer to 1/15. The SKA Low correlator uses about half of the system hardware: ~30 kW and ~2.5 racks. Scaling LEDA (9 kW and 1 rack) to SKA Low correlator capabilities would result in a system that uses 135 kW and occupies ~15 racks. Much of the improvement compared with LEDA is due to technology improvement. But even so, it appears that the Atomic COTS approach presented here results in a compact and energy-efficient design.

### 3 SKA Low Signal Processing

Signal processing in the SKA Low telescope has three main components (Fig. 2):

1. The core and remote stations, including station beamforming across the 256 antennas in a station and the generation of calibration data for the station BF.



**Fig. 2** SKA Low signal processing components. The CBF reside in CSP and receive station data from the 512 core and remote stations. The correlator capability generates visibilities that are sent to SDP and beamform generates tied array beams for PST and PSS.

2. Central signal processing (CSP) whose functions are correlation and beamforming between stations, and pulsar search (PSS) and PST on the beams from the BF within CSP.
3. Science data processing (SDP) is a supercomputer whose main task is to take the data from the correlator and generate images. For pulsar processing, it analyzes pulsar and transient candidates.

### 3.1 SKA Low Stations

The stations for the SKA Low are located in the Murchison Shire of Western Australia more than 600 km north of Perth. The receptors are 131,072 log periodic antennas grouped into 512 stations, each with 256 dual-polarization antennas. The antennas of the station are beamformed to generate up to 48 independent beams with a total bandwidth of 300 MHz (this can be thought of as 48 6.25 MHz station beamlets, 11.6 Gbps with 8-bit precision data). Each beamlet can point independently, with an independently selected sky frequency. In forming the beamlet, any number of weighted station antennas can be used. If the full station (i.e., all 256 antennas) is not used to form a beamlet, it is designated as a substation beamlet. An astronomy beam is a set of beamlets all with the same pointing and differing substation will have different apertures. The flexibility of the system contributes to the complexity of both understanding and implementing the processing. Examples of the myriad of station configurations are:

- wide band: all beamlets pointing in the same direction and covering the full 300 MHz, effectively one 300 MHz beam
- fly's eye: 48 beams (beamlets) all pointed in different directions at the same frequency
- wide field-of-view: eight separate substation (apertures) beams all covering the same frequency band and all looking in the same direction.

A further complication is subarrays that are composed of an arbitrary set of stations/substations with a common set of beams. Up to 16 simultaneous subarrays are allowed. Subarrays are processed independently, and any change to a subarray must not affect the operation of any other subarray. In the correlator each beamlet of a subarray is correlated against every corresponding beamlet for all stations (or substations) in the subarray. A subarray capability can use any part of currently unallocated resources, from any station.

### 3.2 SKA Low CBF

Station beams from stations and substations are all processed by the Low CBF subsystem. With the Atomic COTS, the input is flexible; it can accommodate the original 40 GbE links<sup>15</sup> with data for two stations, as well as new standard 100 GbE link with either four or six stations per link. The BF in CBF coherently sums data from up to 512 stations or substations to form 16 PST beams with up to 300 MHz of bandwidth, as well as 500 PSS beams with 118 MHz of bandwidth

(a 250 beam 236-MHz mode also exist). PST beams are used to accurately time the pulse arrival times of known pulsars. PSS beams are used to search for new pulsars and transients, such as fast radio bursts.

The correlator forms integrated complex cross products between all pairs of corresponding signals (same frequency and beam direction) across all stations or substations in a subarray. Each cross product is a visibility and measures a Fourier component of the common field of view of the two stations (or substations) being correlated. As correlations are formed between all pairs of signals from the stations the compute requirements are proportional to  $BW \times N^2$ , where  $N$  is the number of signals correlated and  $BW$  is the bandwidth. With 512 stations, there are 1024 corresponding signals (each station provides dual polarization data) which produce 524,800 pairs (this includes the autocorrelation: the signal with itself). For the “standard” correlator, this is implemented on 55,296 separate 5.4 kHz frequency channels every 0.85 s; for a 300 MHz bandwidth, the output data rate is 34 G visibilities/s (or 2.7 Tbps using 80-bits per complex visibility, that includes two floats plus quality metrics) which is sent to SDP. With substations, the number of possible signals increases as each station can be split into multiple substations. For example, splitting each station into four substations results in up to 2048 substations, and the number of possible signal pairs increases to 8.4 million. In addition, the integration time decreases to 0.28 s. To keep the data rate manageable, the bandwidth decreases as the number of substations increases to keep the data rate below 5.76 Tbps (to avoid saturation of the visibility link to SDP).

A final function of the correlator is zoom modes, where smaller sections of the 300 MHz sky bandwidth are processed at higher frequency resolutions. For this function, part or all of the beamforming hardware is repurposed to allow up to 64 zoom windows each with 1728 frequency channels. Each zoom band has a separately selectable center frequency and frequency resolution. The available choices of frequency resolution are 14, 28, 56, 113, 226, 452, 904, and 1808 Hz.

The overall capabilities required of the CBFs are summarized in Table 1. The standard correlator is always implemented and is needed for calibration of all other capabilities. The other capabilities use shared resources that are dynamically allocated. The number of Alveo boards needed for each capability is determined by a combination of I/O, memory, and compute limitations. For example, PST is I/O limited, and the correlator is HBM limited.

A full array has all (sub)stations and bandwidth allocated to it. These can be divided amongst multiple subarrays, where each subarray has a subset of the (sub)stations and bandwidth. Additionally, subarray bandwidth can be allocated to multiple independent beams. For example, there are three subarrays each of 100 MHz bandwidth, one subarray could be a sensitive high-resolution subarray with all stations, one subarray could be PSS configured with only the core antennas allocated to provide maximal field of view, and the final subarray could be PST

**Table 1** Summary of the SKA Low CBF capabilities (output data products) and resource usage. The full correlator (5.4 kHz) is always operational to enable calibration. Other capabilities share the remaining 192 Alveo. The maximum bandwidth and number of stations for each capability are listed.

Capability (frequency resolution)	Number of Alveo boards needed	Bandwidth	Maximum number of (sub)stations
Correlator standard (5.4 kHz) always operational	192	300 MHz	512
		$\frac{300 \text{ MHz} \times 512^2}{\text{max \#substations}^2}$	720, 1024, 1440, 2048, or 3376
Zoom (14 to 1808 Hz)	192	64 zoom windows	512
	128	42 zoom windows	512
	64	21 zoom windows	512
PSS Beamformer (14.5 kHz)	128	118 MHz@500 beams or 236 MHz@250 beams	512
PST Beamformer (3.6 kHz)	64	300 MHz@16 beams	512

configured as eight beam pairs all with different pointings. These subarrays must operate independently and not affect other subarrays even during the process of allocation and deallocation.

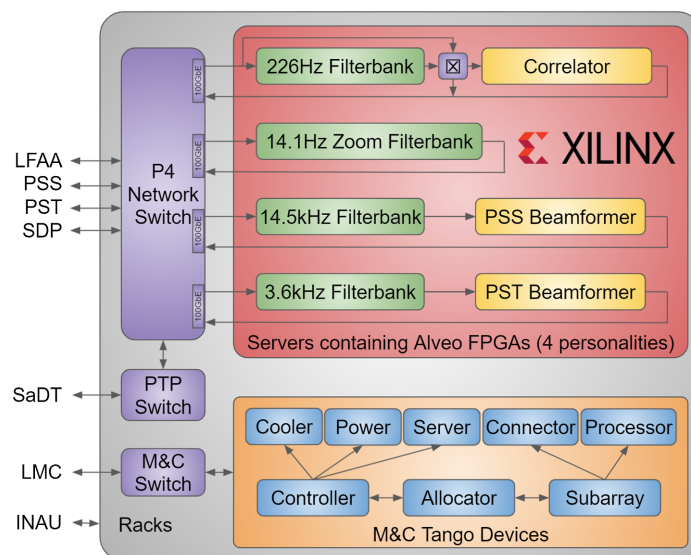
#### 4 CBF System Processing Overview

A broad overview of the CBF is shown in Fig. 3. The heart of the system is 20 servers containing 400 Xilinx Alveo U55 PCIe cards. An additional server with 20 Alveo provides redundancy, the ability to test new functionality, and system upgrades without affecting science operations. Each Alveo card can have one of four personalities: correlator, PSS BF, PST BF, or zoom FB. The number of Alveo allocated to each personality is listed in Table 1. The correlator personality uses approximately 1.2 Pops, and the two pulsar BFs require 0.8 Pops for a total of 2 Pops [which excludes wavefront correction and radio frequency interference (RFI) processing].

A P4 switch network is used to route the self-describing Ethernet packets from the stations to the appropriate Alveo. Each station produces 384 frequency channels with a channel spacing of 781 kHz, to provide a maximum bandwidth of 300 MHz. For the 192 correlator Alveo, the P4 switch network routes two 781 kHz station channels of data for all stations to each Alveo. For other functions, the Alveo personalities are allocated dynamically, and the mix of personalities and desired observations informs the duplication of the station data to different Alveo personalities.

Noting that all Ethernet links are bidirectional allows science data products generated by the Alveo to be routed back through the P4 network switch. This satisfies the data transport requirements of visibilities to SDP, and beams to PSS and PST. As an example, each PST Alveo generates a fractional part of the total bandwidth for any given beam. This bandwidth is aggregated to a single output by the P4 network switch so that each PST output link contains full bandwidth data for a single beam.

PTP packets are an additional system input, coming from the Signal and Data Transport interface. This is distributed via the P4 switches, which are PTP enabled, to the Alveo. This provides a common time across all Alveo which allows sequencing of the output packets to avoid congestion in the P4 switches.



**Fig. 3** High level overview of SKA Low CBF (networking, processing, and control) showing all SKA interfaces. Each Alveo contains one of the four available personalities and receives station data via the P4 network switch. The switch also transports science data products to the upstream science capabilities (imaging and transient/pulsar). M&C Tango devices are pieces of software that control the various components in the system (connector = P4 switch and processor = Alveo boards).

Control of the system comes from the CSP Local Monitor and Control interface to the CBF system. All hardware within the CBF system is connected to a local Ethernet switch. The CBF Monitor and Control (M&C) uses the Tango distributed control system<sup>16</sup> with a predefined SKA control hierarchy. Each Tango device has a state, attributes and executes commands. The functions of the key predefined Tango devices are:

- The controller device is in charge of system power and overall health.
- The allocator assigns resources to subarrays and defines data paths in the P4 switch.
- The subarray devices (up to 16) set the configuration and execution of signal processing.

Specific Tango devices that control a specific part of the CBF system include:

- The cooler device controls rack level liquid-to-air cooling.
- The power device controls the distribution of single-phase power within the racks.
- The server device controls the servers that house the Alveo.
- The processor controls the signal processing functions within the Alveo.
- The connector controls the P4 switch data routing.

## 5 P4 Switch Network

The station data ingest of CBF consists of  $128 \times 100$  GbE inputs, which is connected through a network to 420 Alveo to produce science data. With this many connections, it requires a two-layer switch network routing 196,608 different packet entities, with each packet being individually routed to between two and four different Alveos. To accomplish this, P4 switches have been selected. These switches are fully programmable in the P4 computer language. P4 is a domain-specific (P4 switches or targets) language optimized for network data forwarding. There are many types of P4 switches, but for SKA Low a 64-port 100 GbE Tofino switch has been selected with nine switches in the first layer and 11 in the second layer (a total of 20 switches).

Unlike standard network switches, P4 does not directly support multiple protocols. Instead, the programmer defines the header protocol and names in the P4 language. This is in contrast to OpenFlow switches as used in the MeerKAT correlator.<sup>17</sup> OpenFlow is layered on top of the Transmission Control Protocol and is a protocol for controlling the flow of standard Ethernet packets. In the context of P4 switches, the user can easily add domain-specific protocols, such as Streaming Protocol for Exchanging Astronomical Data (SPEAD),<sup>18</sup> by creating a header definition in P4 language. This definition represents the various fields of a given protocol header/footer and the user can associate a unique name to each of these fields for inline manipulation when the switch receives a packet.

As a result, P4 switches, on receiving a packet, decode the headers and routes the packet according to the information read. In the case of CBF, the packets are User Datagram Protocol with a metadata header that includes *subarray\_id*, *station\_id*, *substation\_id*, *beam\_id*, and *frequency\_id*, from the SPEAD protocol. The P4 switch parses this metadata and with a match-action table determines the port or ports to which the packet is to be routed. Note, while Internet Protocol or Media Access Control addresses are not needed to route the data, only the values in a few of the metadata fields, although the P4 switch is still capable of using this information for interoperability with the rest of the telescope infrastructure. The use of match-action tables supports the existence of multiple independent subarrays, where a subarray is a subset of the stations and beams.

To bring a subarray into existence, match-action tables are generated and loaded to the appropriate physical switch. In the P4 switch, input packets are matched against the entries in the table and if there is a match the action specified by the match is taken. Only packets whose metadata match are routed. Routing for a given subarray is stopped by removing the corresponding match-action tables. Note that in a similar manner telemetry data can be generated for each of the rules which provide deep insight into the status of the system.

As resources cannot be shared between subarrays, it is seen that a new subarray can only be composed of resources not already used in a subarray (beams, frequency, and station). These free resources can be allocated to the subarray and routed by building and installing the appropriate

match-action tables. All subarrays are independent and can be brought into existence at independent times as well as being deallocated independently.

The overall development of the basic components for routing packets based on SPEAD headers are straightforward for networking engineers with knowledge of software development. It took us <3 months to implement and validate at line rate (100 Gbps) the routing using the *frequency\_id* from the SPEAD header. We were able to associate and retrieve metrics associated with routing tables installed on the switch with a little more work. The crux of the difficulty with using P4 switches is similar to what we would have had to do with traditional or OpenFlow switches, and mainly consists in developing a modular and efficient management and control framework. However, on this front, P4 switches have an edge over other networking technologies as they are not tied to a given framework or tool. Indeed, we are leveraging generic Remote Procedure Calls to interact and control the switch, which allows for greater flexibility.

## 6 Alveo Personalities

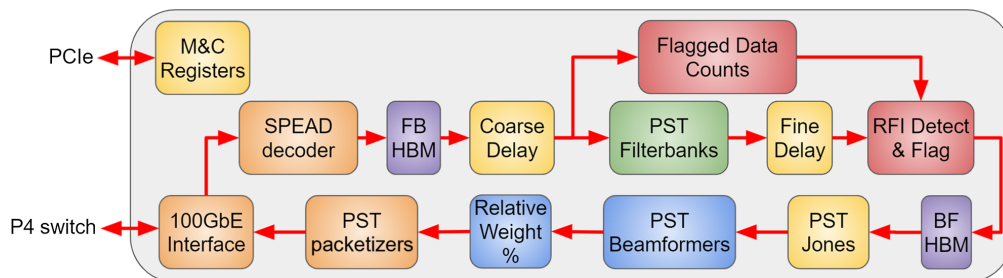
The CBF must generate a number of different science products: standard and zoom visibilities for imaging, and beams with different requirements for PSS and PST. To generate these products, Alveo cards are personalized to generate a single product, and there are four basic personalities. This has the advantage that development and debugging of the FPGA code is independent of the others, which simplifies development.

### 6.1 PST and PSS BFs

The PST BF coherently sums the signal from the stations within a 10 km radius to form beams. As the extent of this array is 20 km, the beams are very narrow, but the targets are point-like, and it is sufficient to have the beam pointing at the target. The functions needed to implement this beamforming are shown in Fig. 4. It is seen that the HBM buffers the data in two locations: input to FBs and input to the BF. Compared with previous FPGA designs, we have implemented at CSIRO this buffering breaks the FPGA processing into three sections, where each section is asynchronous to the others. In previous designs, there was insufficient buffering, and changes in timing in one section could cause following sections to fail. This buffering has meant it has become much easier to write and debug the FPGA firmware.

The three sections are station input, FB/wavefront align, and BF, with HBM providing the buffer between each. The input buffer isolates the FPGA processing from the vagaries of data transport in a switched network, checks data validity and lost packets, and records packet metadata including timestamps.

Between the next two buffers sits, an oversampled polyphase FB to channelize the data to a bandwidth that allows a simple complex weight and add BF. It also brings the station wavefronts into alignment by applying a delay. The delay is in two steps: coarse and fine. Coarse delay selects a time sample as first input to the FB and provides delay to a precision of one sample (1.08  $\mu$ s). At the output of the FB, a phase slope with frequency is applied to provide delays that



**Fig. 4** Functional diagram of the PST BF signal chain for a single Alveo. HBM provides a buffer and corner turn function and coarse and fine delay bring signals to a common wavefront. Input data may be flagged, and summaries of these inform flagging at the FB output. Relative weights provide information on relative SNR for a given beam output.



are a very small fraction of  $1.08 \mu\text{s}$ . All Alveo personalities include the input and FB/wavefront correction stages.

The final function is RFI detection and flagging. Input data may be corrupted, lost, or flagged at the station level. Flagged or missing data can degrade the FB channel response and as such, the input data to the FB are quality checked and an estimate of degradation found. If the degradation is too high, the output of the FB is flagged to ensure the data are not used in future processing. Following the FB, a running power average is calculated (at  $\sim\text{kHz}$  resolution), and any data greater than a selectable multiple of the mean will be interpreted as RFI and flagged. The multiple used is determined experimentally and will be different for different science cases. Any data flagged as interference do not contribute to astronomy output products. A data quality metric for each output product is also produced to inform the astronomer. This includes how much data were flagged.

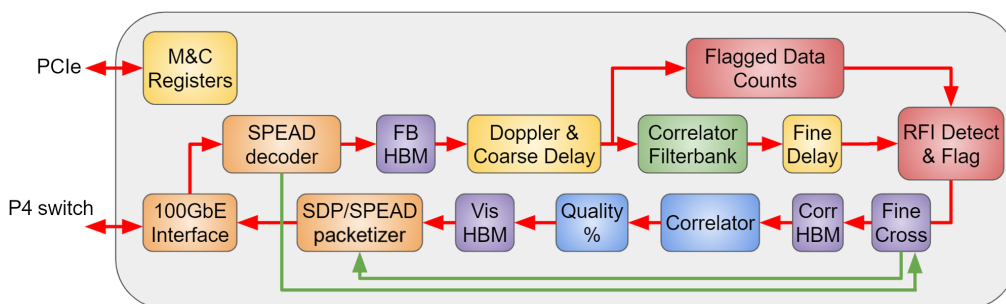
The final processing section is the BF and includes a Jones matrix correction (PST Jones) for every input to each beam to ensure very high polarization purity. Only unflagged data are included in the sum for each beam. To allow correction for changes in signal-to-noise ratio (SNR) and signal levels that this produces, the percentage of unflagged data is calculated and the percentage included in the output data packets.

The PSS BF functionality is very similar to the PST BF except for the Jones matrix correction. For PSS, an on-station beam correction is applied after the FB and off-axis beam correction after the BF. The PSS BF Alveo generates 500 dual polarization beams on a bandwidth of 0.92 MHz, whereas the PST BF Alveo generates 16 dual polarization beams on a bandwidth of 4.7 MHz. Also, beamforming is at different frequency resolutions for the two BFs.

## 6.2 Correlator

The correlator is an FX (frequency domain transformation ‘‘F’’ followed by cross correlation ‘‘X’’) correlator where the use of the polyphase FB<sup>19</sup> to channelize the data provides high compute efficiency.<sup>20,21</sup> The polyphase FB approach also allows efficient BFs<sup>22</sup> (previous section) and search for extraterrestrial intelligence (SETI) searches.<sup>23</sup> The functional diagram of the correlator Alveo personality is shown in Fig. 5. It is very similar to the PST BF (Fig. 4) with the Jones matrix correction removed and the BF replaced by a correlator. However, the FB is quite different. The BFs have 64 (PSS) and 256 (PST) channel FBs. In the correlator, this is increased to 4096 channel FBs with 3456 channels processed by the correlator (as the station channel is oversampled by 32/27). The correlator FB base frequency resolution is 226 Hz and following the correlator channels are aggregated to provide multiple frequency resolutions. Standard observing averages over 24 of these channels giving a frequency resolution of 5.4 kHz. Zoom mode resolutions from 226 Hz to 1.8 kHz are obtained by averaging one to eight channels.

The correlator is very flexible. It must handle the division of the station data among up to 16 subarrays, which must all be processed independently. As well, it must be able to switch between standard 5.4 kHz correlation and zoom band correlations. The internal function of the correlator is based on a multiplier and memory-efficient correlation cell.<sup>24</sup> The correlation cell implements



**Fig. 5** Functional diagram of correlator signal chain for a single Alveo. In addition to common BF functions, the correlator applies a local Doppler correction to each station. For specialty functions (fine zooms and substations), part of the processing can be bypassed using the fine cross module.

a small part of the correlation. By appropriately sequencing the correlation cell, subarrays of any size can be processed. After the correlation, cell is a frequency accumulator that provides multi-frequency resolution capabilities. The required sequence of operations is programmed into the correlator Alveo by the control system making the system very flexible.

### 6.2.1 Atomic exemptions

For the BFs and standard correlator, the operation of the Alveo cards is Atomic: the input data are station data, and the outputs are science-ready products for SDP, PSS, and PST. Basic correlator operation encompasses processing up to 512 stations at full bandwidth and implementation of zoom windows with frequency resolutions of 226 Hz to 1.8 kHz. But the correlator is also required to process up to 3376 substations and zoom windows from 14 to 113 Hz. These modes are not Atomic and require data transfer between Alveo cards.

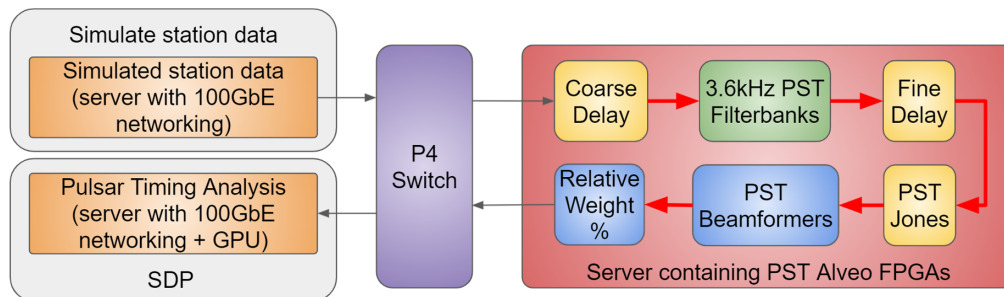
In the case of high substation numbers, the HBM is insufficient to store the input data, so the FB and wavefront correction must be done over multiple Alveo cards for a single frequency channel (781 kHz) from the stations. To facilitate this, the “Fine Cross” function shown in Fig. 5 is included. This allows FB data to be output to other Alveo cards and for that data to be received.

For the fine zoom modes to achieve a resolution of 14 Hz, a 65,536 channel FB is needed, which is difficult to fit in with the correlator function. Instead, the required correlator FB and wavefront correction for these resolutions are implemented on a separate Alveo card and input to the correlator personality via the P4 switches and the Fine Cross to the correlator HBM buffer.

### 6.3 Implementation Status

We have prototyped most of the Alveo firmware blocks required for the CBFs, including the basic correlation cell, FBs, 100 GbE interfaces, and monitoring and control software. In addition to testing individual functional blocks in simulation, we have tested a complete PST BF on a single Alveo card, using the test setup shown in Fig. 6. As real station data are not yet available, simulated station data are generated offline in a system model and stored, then streamed at real-time rates into the P4 switch. To keep the amount of simulated data manageable, this test was conducted with 64 stations and 1.5 MHz of bandwidth. This test was able to demonstrate all the functionality in the BF, including FBs, coarse and fine delays, and BF, resulting in reconstruction of a simulated pulsar based on low SNR station data. The current PST BF firmware is for 48 beams and two channels with 41% of FPGA compute resources used. It is estimated a final 16 beam system on six channels and 512 stations will use 47% of compute resources and approximately 36% of the FPGA memory and logic.

To further test the P4 switch performance, a P4 switch was used to take 96 Gbps input stream and reduplicate it 14 times with modified metadata. This relied on the P4 functionality of being able to read header metadata and modify the data before generating an output packet.



**Fig. 6** Test setup for the PST BF. A single server generates 64 station test data for the PST Alveo, which is then routed through the P4 switch. The Alveo implements wavefront correction on station data on the PST FB. The BF generates 48 beams each with separate Jones matrix polarization corrections. The percentage of valid data calculated and beam data are output to the P4 switch, which transfers this to a separate server for verification.

The 14 links at 96 Gbps simulated data from 112 stations. This was passed to a second P4 switch programmed to route data for 1/14 of the bandwidth for all stations to its outputs. Packet counters in the P4 verified basic functionality and the capture of metadata in a connected server for a single 100 GbE link verified packets were correctly routed. For connection to legacy switches, ARP functionality has also been demonstrated.

## 7 Conclusion

Presented here is possibly one of the most complex radio astronomy correlators ever conceived. With the introduction of substations, any station can act as multiple stations, and the correlator is not one but many, catering for arrays from 512 to more than 3000. The concept of subarraying allows any grouping of stations or substations so multiple correlator sizes are simultaneously called for. Within a subarray, multiple beams with different pointings are allowed, and simultaneous zoom bands are permitted on any of the subarrays all of which must operate independently.

The SKA Low correlator is a challenge due to high data rates, significant compute, and the need for real-time processing. Compared with previous systems where highly interconnected FPGAs were used, this design is based on COTS hardware: P4 switches and Alveo U55C cards. The P4 switch implements all data routing and unlike standard network switches route data based on packet metadata. The P4 switches provide data for all stations for a small part of the bandwidth to each Alveo card. The Alveo card then implements all signal processing and produces output packets that contain the final CBF science data. The processing is Atomic, and the system is characterized as “Atomic COTS.”

We have shown that Atomic COTS meets this challenge, where P4 switches allow flexible routing and all required processing on input data is contained within a single COTS signal processor. It is a practical and cost-effective approach to implement correlation and beamforming systems that are processing continuous input data rates of many Tbps. It is inherently scalable and can rapidly adapt to technology improvements.

## References

1. A. M. McPherson et al., “Square kilometer array project status report,” *Proc. SPIE* **10700**, 107000Y (2018).
2. A. R. Thompson et al., “The very large array,” *Astrophys. J. Suppl.* **44**, 151–167 (1980).
3. W. E. Wilson et al., “The Australia telescope compact array broad-band backend: description and first results,” *Mon. Notices Royal Astronom. Soc.* **416**, 832–856 (2011).
4. A. Baudry and J. Webber, “The ALMA 64-antenna correlator: main technical features and science modes,” in *XXXth URSI General Assembly and Sci. Symp.*, pp. 1–4 (2011).
5. W. E. Wilson et al., “The Australia telescope compact array broad-band backend: description and first results,” *Mon. Not. R. Astron. Soc.* **416**(2), 832–856 (2011).
6. G. W. Schoonderbeek et al., “UniBoard2, A generic scalable high-performance computing platform for radio astronomy,” *J. Astron. Instrum.* **08**(2), 1950003 (2019).
7. R. Perley et al., “The expanded very large array,” *Proc. IEEE* **97**(8), 1448–1462 (2009).
8. J. Hickish et al., “A decade of developing radio-astronomy instrumentation using CASPER open-source technology,” *J. Astron. Instrum.* **5**(4), 1641001 (2016).
9. J. Jonas and the MeerKAT Team, “The MeerKAT radio telescope,” in *Proc. MeerKAT Sci.: On the Pathway to the SKA*, Stellenbosch (2016).
10. K. Bandura, “ICE-based custom full-mesh network for the CHIME high bandwidth radio astronomy correlator,” *J. Astron. Instrum.* **5**(4), 1641004 (2016).
11. J. Kocz et al., “Digital signal processing using stream high performance computing: a 512-input broadband correlator for radio astronomy,” *J. Astron. Instrum.* **4**(1 and 2) 1550003 (2015).
12. P. C. Broekema et al., “COBALT: a GPU-based correlator and beamformer for LOFAR,” *Astron. Comput.* **23**, 180–192 (2018).
13. Xilinx, Inc., “Breathe new life into your data center with Alveo adaptable accelerator cards,” Xilinx WP499, 2018, [https://www.xilinx.com/support/documentation/white\\_papers/wp499-alveo-intro.pdf](https://www.xilinx.com/support/documentation/white_papers/wp499-alveo-intro.pdf).

14. P. Bosshart et al., “P4: Programming Protocol-Independent Packet Processors,” *ACM SIGCOMM Comput. Commun. Rev.* **44**(3), 88–95 (2014).
15. G. Comoretto et al., “The signal processing chain of the low frequency aperture array,” *Proc. SPIE* **11445**, 1144571 (2020).
16. P. Verdier, J. L. Pons, and F. Poncet, “Tango control system management tool,” in *Proc. ICALEPCS2011*, Grenoble (2011).
17. M. J. Slabber et al., “MeerKAT data distribution network,” *Proc. SPIE* **10707**, 107070H (2018).
18. J. Manley et al., “SPEAD: Streaming Protocol for Exchanging Astronomical Data,” 2010, <https://casper.astro.berkeley.edu/astrobaki/images/9/93/SPEADsignedRelease.pdf>.
19. R. Schafer and L. Rabiner, “Design and simulation of a speech analysis-synthesis system based on short-time Fourier analysis,” *IEEE Trans. Audio Electroacoust.* **21**(3), 165–174 (1973).
20. J. D. Bunton, “An improved FX correlator,” ALMA memo 342 (2000).
21. J. D. Bunton, “SKA correlator advances,” *Exp. Astron.* **17**, 251–259 (2004).
22. J. D. Bunton and R. Navarro, “DSN deep-space array-based network beamformer,” *Exp. Astron.* **17**, 299–305 (2004).
23. A. M. Peterson, K. S. Chen, and I. R. Linscott, *The Multichannel Spectrum Analyzer*, M. D. Papagiannis Ed., IAU Publications, pp. 373–383 (1985).
24. W. Kamp, N. Abel, and G. Comoretto, “Complex multiply accumulate cells for the square kilometre array correlators,” in *Int. Conf. ReConFigurable Comput. and FPGAs (ReConFig)*, pp. 1–6 (2018).

**Grant A. Hampson** is a team leader in the Signal Processing Technologies at CSIRO Space and Astronomy. He received his PhD in computing “Implementing Multi-Dimensional Digital Hardware Beamformers” from Monash University in 1997. His passion is to design and build world class radio astronomy instrumentation and enjoys working at the cutting edge of processing, memory, and communications technology. He was the team leader for the groundbreaking Australian SKA Pathfinder (ASKAP), which utilizes a 36-beam phased array feed.

**John D. Bunton** is a senior principal research scientist at CSIRO Space and Astronomy. He received his BSc, BE (Hons.), and PhD degrees from the University of Sydney in 1973, 1975, and 1983, respectively. He is an author of more than 80 journal papers. He was the project engineer at the ASKAP radio telescope. His current research interests include radio astronomy correlators and signal processing. He is a senior member of IEEE.

**Guillaume Jourjon** is a senior scientist at CSIRO. He received his PhD from the University of New South Wales and Toulouse University of Science in 2008. Prior to his PhD, he received an Engineer degree from ISAE. He is an author of more than 70 research papers including in prestigious conferences and journals. His research areas of interest are related to distributed computing, software defined network, and in-network computing in the context of radio-astronomy.

Biographies of the other authors are not available.