Draft version 2021-09-03
Typeset using LaTeX **preprint** style in AASTeX63

ngVLA Electronics Memo #10: A SCREAM-Compatible ngVLA Cross-Correlation Engine: Key Requirements Review and Option Trade-Off Study

Nolan Denman[1]

[1] *Central Development Laboratory, National Radio Astronomy Observatory*

Contents

Note This document includes a description of selected key XE requirements and a trade-off study of the architecture and hardware choices, and is intended to satisfy the Conceptual Design Review (CoDR) requirements concerning down-selection of design options (Küsel 2021a,b). For a description of the conceptual design for the XE, please see Denman et al. (2021a).

Note This document is solely concerned with ngVLA operations in synthesis imaging mode; pulsar observation modes are addressed in separate documents (Denman et al. 2021b,c).

1. TASKS REQUIRED OF THE X-ENGINE

This document provides an overview of an ngVLA Cross-Correlation Engine (X-Engine or XE) design which is compatible with the SCREAM correlator design; the system presented may be tweaked for compatibility with any correlator which can supply channelized data in an appropriate format. Figure 1 provides an overview of data and control flows within the X-Engine, and the tasks which are or may be performed.

1.1. *Input from the B&C Nodes*

The X-Engine receives channelized data from the previous stage of the correlator, accompanied by a selection of metadata (potentially including a flag stream). Details of the interconnection such as routing and packet handling are not considered in this document; it is assumed to be an Ethernet switched network with sufficient bandwidth.

1.1.1. *Data Itself*

The input data to the correlator X-Engine is the packetized output from the B&C nodes, as collated and sent by the B&C Master FPGA. Delay corrections have already been applied and packets arrive in approximately chronological order.

The X-Engine receives a data packet[1] with a total complex bit depth of N_{bits}, N_{pol} polarization components, and N_{chan} frequency channels with total size

$$\left(\frac{N_{bits}}{16}\right)\left(\frac{N_{pol}}{2}\right)\left(\frac{N_{chan}}{16k}\right) * 64\,\mathrm{kiB} \tag{1}$$

from each of the N_{ant} receivers and over a total bandwidth of $\Delta\nu$, once every

$$\left(\frac{20\,\mathrm{GHz}}{\Delta\nu}\right)\left(\frac{N_{chan}}{16k}\right) * 0.82\,\mathrm{microseconds} \tag{2}$$

[1] The term 'packet' refers here to a single time-interval's observations for all frequency channels; it will consist of multiple network protocol packets.

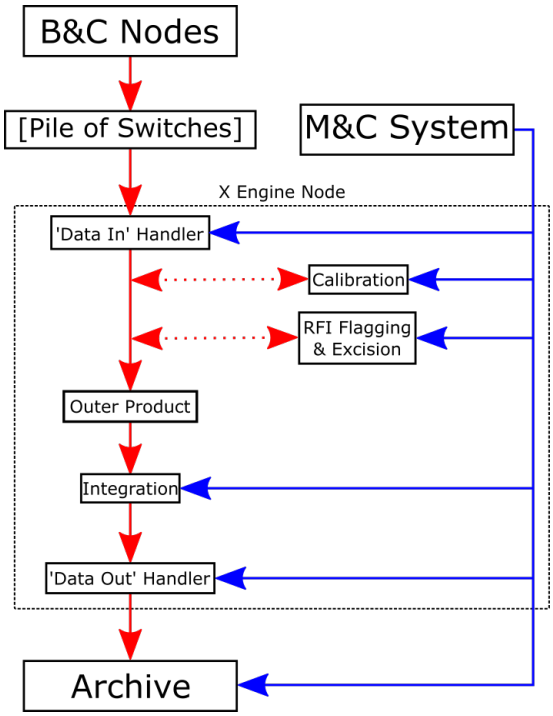

Figure 1. An illustration of data and control flows within the X-Engine. Operations marked with a dotted line may not be required in the final version.

for a total input bandwidth of

$$\left(\frac{\Delta\nu}{20\,\mathrm{GHz}}\right)\left(\frac{N_{bits}}{16}\right)\left(\frac{N_{pol}}{2}\right)\left(\frac{N_{ant}}{263}\right) * 21\,\mathrm{TB/s} \tag{3}$$

1.1.2. *Metadata*

In addition to the beam data itself, the B&C nodes will produce (or relay) metadata required by the X-Engine or later processing stages. Most immediately relevant to the X-Engine's operation are metadata with receiver ID, timing, and frequency information.

1.1.3. *Flags*

The topic of high-time-resolution Radio-Frequency Interference (RFI) mitigation within the ngVLA correlator is currently under examination (Rau et al. 2019; Selina et al. 2020a). In the event that high-time-resolution flagging is implemented prior to the X-Engine, the data may be accompanied by a set of flags of comparable dimensions. These are considered separately from the general metadata due to their potentially large size.

1-bit flags for each of the input time-samples will increase the input rate by a few percent; higher-bit-depth flags have proportionally greater data transport requirements. These flags must be carefully propagated; the interaction of flagging and correlation/integration has not yet been defined.

This is independent of, and possibly in addition to, any RFI excision which is implemented in the X-Engine itself (§2.3).

1.2. *Input from the M&C System*

In addition to the data and metadata arriving from the B&C nodes, observing and processing parameters must be supplied by the ngVLA monitor and control (M&C) system (Koski et al. 2019). The details of interaction between the XE and M&C systems are not yet defined.

1.3. *Correlation*

The core of the X-Engine is the cross-correlation stage, in which we compute the product of the channelized electric field measurements from each pair of receivers to produce the sky-conjugate visibilities.

The computational requirements are most easily expressed in terms of complex multiply-and-accumulate operations or "CMACs", performed by devices referred to as "CMAC units". A single CMAC may also be viewed as (or replaced by) four multiplications and four additions of the real component values.

For n receivers, correlation produces $n(n+1)/2$ visibilities, each of which must be computed once per time-step for each frequency channel. We therefore require

$$\left(\frac{N_{pol}}{2}\right)\left(\frac{N_{ant}}{263}\right)\left[\left(\frac{N_{pol}}{2}\right)\left(\frac{N_{ant}}{263}\right)+1\right]\left(\frac{N_{chan}}{16k}\right) * 2.3 \cdot 10^9 \tag{4}$$

CMAC operations once every

$$\left(\frac{20\,\mathrm{GHz}}{\Delta\nu}\right)\left(\frac{N_{chan}}{16k}\right) * 0.82\,\mathrm{microseconds} \tag{5}$$

For comparison, the CHIME X-engine (built with GPU hardware purchased in mid-2016) performs $2.1 \cdot 10^9$ CMACs of 4+4-bit correlation once every 2.56 microseconds.

1.4. *Integration*

The resulting visibilities are time-averaged prior to readout. For 8+8-bit components, the product may potentially exceed a 16+16-bit value due to the summation (although not if components are independently accumulated). The complex visibilities are initially integrated by a factor equal to the dot product length (if one is used, see §2.6); assuming this is less than 2^{15}, 32+32-bit accumulators should be sufficient for this intermediate integration. Large downstream integration factors may require higher-bit-depth accumulation, depending on the observation parameters.

1.5. *Output*

Note Version B.04 of the ngVLA Preliminary System Requirements (Selina et al. 2020c) introduced the following constraint: "CON104: Maximum Data Rate: The maximum data rate from the correlator shall not exceed 132 GB/s.".

1.5.1. *Metadata*

In addition to metadata present in the input (§1.1.2), which may be relayed with the data, the XE itself may generate metadata concerning observing parameters and results. These have not yet been specified; this will be informed by downstream requirements.

1.5.2. *Visibilities*

For data with a readout cadence Δt on the scale of seconds, the total output rate is relatively modest:

$$\left(\frac{N_{pol}}{2}\right)\left(\frac{N_{ant}}{263}\right)\left[\left(\frac{N_{pol}}{2}\right)\left(\frac{N_{ant}}{263}\right)+1\right]\left(\frac{N_{chan}}{16k}\right)\left(\frac{1\,\mathrm{second}}{\Delta t}\right)\left(\frac{N_{bits}}{64}\right) * 18\,\mathrm{GB/s} \tag{6}$$

Readout cadences near the ~ 1 ms level result in unmanageably high output data bandwidths (see **Note** above). In order to support these cadences, it will be necessary to place limits on the total bandwidth, number of channels, bit depth, and/or number of antennas. If the number of samples integrated is smaller than the number of receivers in the array, it is preferable (from a data rate perspective) to simply record the channelized voltage streams and perform the correlation offline as part of later processing stages.

For a test/example implementation in which the correlation parameters are more-or-less fixed, the output data structure may be very simple; assuming that all products are present in each output packet, the timing information and thread/channel IDs would be sufficient to make the data useable. A more detailed output data structure will be required for later implementations.

2. FEATURES AND EXTENSIONS OF THE X-ENGINE

2.1. *Sub-Array Operations*

The ngVLA is expected to operate most frequently in a 'sub-array' mode, in which subsets of antennas are configured and operate independently to improve the overall occupancy of the instrument (Ojeda et al. 2018). The maximum number of sub-arrays operating at one time, N_{sub}, must be at least ten to fulfill the CSP technical requirements.

Sub-arrays assigned to pulsar or VLBI observations are trivially compliant, as the CSP already features the ability to direct individual beams to the appropriate hardware. However, some consideration is required to efficiently support multiple simultaneous synthesis imaging sub-arrays.

As inspection of Equations 3 and 4 shows, reducing the number of receivers N_{ant} proportionally reduces the input data bandwidth, but reduces the amount of computation required much more rapidly – correlation of sub-arrays is much less demanding than the entire array. It is helpful to note that when a receiver is moved from one sub-array to another, potentially with a different frequency configuration, the correlator's total input data rate is unchanged and the computation required is either the same or decreases. The resources required are therefore necessarily within the correlator performance envelope, depending on the efficiency of their co-ordination.

If the X-Engine nodes are supplied with excess input data bandwidth relative to the full-array case, it should be possible to increase the signal bandwidth processed by each node for smaller array sizes. This will require significant flexibility in the routing of data from the B&C nodes, but allows the resulting 'excess' nodes to be re-tasked or powered down to reduce overall energy costs.

A difficulty arises in dealing with the 'fractional nodes' which are divided between multiple sub-arrays. It is not clear that the resources within one node may be easily and efficiently shared, particularly when the frequency-channelization parameters (and therefore packet timing) differ. The deployment of $\approx N_{sub}$ additional nodes allows each sub-array to fully occupy all of its hardware, eliminating this consideration entirely. This additionally enables arbitrary sub-array configurations using different firmware, which is desirable for later development and testing (Selina et al. 2020b).

2.2. *Corrections and Calibration*

As per the preliminary technical requirements (Ojeda et al. 2019), the X-Engine may be required to apply calibration factors to the data. Per-receiver and/or per-frequency instrumental gains may be applied prior to correlation; similarly, corrections for quantization and lost data may be applied either in either the B&C or XE.

2.3. *RFI Excision*

RFI excision is a formal requirement of the correlator system (Ojeda et al. 2018) and current plans have identified a set of potential excision methods which could be implemented in the X-Engine (Rau et al. 2019; Selina et al. 2020a). These involve a range of potential computational loads; statistical tests on high-time-resolution data would require a significant increase in the computational load on the X-Engine, and associated data flags would similarly increase the node I/O requirements. As there is uncertainty as to the decision to excise RFI in the X-Engine as well as the nature of any excision, incorporation of RFI excision in the X-Engine has been deferred for the time being.

2.4. *I/O Handling*

Once received, the input data packets will be distributed to a set of complex-valued signal buffers. The Station ID, Thread ID, and data frame timing information are sufficient to route each set of receiver data to its correct location in memory. If buffers are zeroed after readout, the attempted read of an all-zero buffer would (almost certainly) indicate that a given packet was not received prior to its being required for a computation; otherwise, a 'valid/current data' flag may be required to avoid the CMAC units ingesting old or invalid packets.

A separate buffer for each receiver and frequency channel entails

$$\left(\frac{N_{ant}}{263}\right)\left(\frac{N_{pol}}{2}\right)\left(\frac{N_{chan}}{16k}\right) * 8.6 \cdot 10^6 \tag{7}$$

buffers; if each contains N_{pkt} whole packets, the total buffer memory required in the XE is

$$\left(\frac{N_{ant}}{263}\right)\left(\frac{N_{pol}}{2}\right)\left(\frac{N_{chan}}{16k}\right)\left(\frac{N_{bits}}{16}\right)\left(\frac{N_{pkt}}{1}\right) * 16.4\,\mathrm{MiB} \tag{8}$$

The value of N_{pkt} required will depend on how reliably the packets arrive in chronological order, as well as on the size of a dot product or tensor core (if one is employed, see §2.6).

2.5. *Argument Broadcasting*

For each time-step and frequency channel, correlation requires N_{ant} input data points but produces $\frac{1}{2}N_{ant}(N_{ant}+1)$ output values. The ability to broadcast input data within the XE would vastly reduce internal bandwidth requirements, especially as N_{ant} grows.

2.6. *Acceleration via Dot Product*

One of the simplest performance improvements involves replacing the individual-value CMAC with a dot product along the time axis; a dot product with vector length L_{dp} reduces the rate at which computations are required by the same factor. However, dot product modules are generally not built for complex number support; I have not yet found a pre-existing general complex integer dot product

IP. If we were able to construct a complex dot product which supported argument broadcasting it would offer a considerable advantage in data transfer and performance. The associated costs are primarily in the system's complexity and the requirement to buffer at least L_{dp} time-samples' additional data.

A related case is the use of 'Tensor Cores', which provide high-rate matrix-vector and matrix-matrix multiply-and-accumulate operations for data of specified sizes and bit widths; these have similar implications for accelerating correlation but to a much greater extent.

3. POTENTIAL X-ENGINE ARCHITECTURES

The computation of the complete set of visibilities for the entire array while operating at maximum instrumental bandwidth is the most computationally expensive of the processing tasks required; as such, this document evaluates system properties relative to their ability to complete this task.

The task of correlation is too large for any single currently-available computational device; it must therefore be divided along one or more axes. A few options for the division of this task are described below.

3.1. *Time Division*

Division along the time axis requires some number of separate fractionally-sized correlators, each of which receives a corresponding fraction of the post-B&C packets. This results in a multiplication of infrastructure requirements and a huge additional networking cost, while limiting possible operational modes. As such, it is not considered further.

3.2. *Receiver Division*

Division of the correlation task on a receiver-group basis, as presented in earlier SCREAM designs, scales well with a gradually-expanding array but introduces significant network complexity. Previous CDL-internal SCREAM X-Engine documents explored this option in significant detail; it is not under consideration at this time.

3.3. *Frequency Division*

Division along the frequency axis scales less gradually with array size but has a much simpler implementation; it may be seen as an extension of the sub-band division which occurs in the antennas' digital back-end (DBE) (Ojeda 2020a,b). The remainder of this document considers this option; each node of the X-Engine will complete the correlation for all receiver-polarizations for some fraction of the instrument's overall bandwidth.

4. POTENTIAL X-ENGINE HARDWARE

4.1. *CPUs*

General-purpose Central Processing Units (CPUs) provide a powerful and flexible platform for many types of scientific computing, but are not well-optimized for the types of extremely large parallel computations the X-Engine requires. If selected, CPUs would require host machines and external Network Interface Cards (NICs).

4.2. *GPUs*

Modern Graphics Processing Units (GPUs) are heavily optimized for large-scale parallel operations on limited-bit-depth input. A number of types, sizes, and formats of GPU are available; two of the most promising options are described below.

4.2.1. *NVIDIA Ampere*

The NVIDIA Ampere[2] GPU architecture features a large number of Streaming Microprocessors, each of which combines caches, schedulers, parallel computing modules, and tensor cores (NVIDIA Corporation 2020a, 2021). These GPUs are capable of extremely high rates of low-bit-depth processing, particularly a matrix multiply–accumulate operation which is extremely suitable for correlation. Their power consumption on a per-operation basis is extremely low, with recent correlator implementations (Romein 2021) approaching 1 picojoule per operation.

The primary difficulty with a GPU-based X-Engine is with data transport to and from the GPU. For an ngVLA-like telescope, with a large instantaneous observing band relative to the number of receivers, the operational intensity[3] of the correlation task is relatively low. GPUs are almost all limited to PCIe interfaces with relatively low bandwidth; for the correlation required, they have difficulty receiving data quickly enough to occupy their processing capability.

An additional consideration for GPU hardware selection is the choice of specific model and form factor. They typically take the form of individual GPUs hosted in servers and connected by PCIe to separate NICs. One alternative is multi-GPU self-hosting units like the NVIDIA HGX A100 (NVIDIA Corporation 2020c) or DGX A100 (NVIDIA Corporation 2020b) which support high-speed internal networking via the NVLINK protocol. PCIe GPUs offer increased flexibility in networking configuration and power-bandwidth-cost choices, while the latter make inter-GPU data transport an order of magnitude faster.

4.2.2. *Jetson AGX Xavier*

A low-power GPU optimized for mobile and autonomous computing, the NVIDIA Jetson AGX Xavier module (NVIDIA Corporation 2020d,f) hosts eight Volta-architecture Streaming Multiprocessors (NVIDIA Corporation 2017) equipped with tensor cores capable of up to approximately 23 TOPs of 8-bit integer computation.

Although the Jetson AGX Xavier is described in marketing materials as having an x8 PCIe Gen4 connection, in practice only one lane is accessible as an endpoint, limiting its data transfer bandwidth severely (NVIDIA Corporation 2020e).

4.3. *FPGAs*

With a range of configurations and options, Field-Programmable Gate Arrays (FPGAs) offer a number of possible solutions; a few are elaborated upon below.

4.3.1. *General-Purpose FPGA*

Current top-of-the-line all-purpose FPGAs such as the Xilinx Virtex UltraScale+ (Xilinx Inc 2021a,b) and Intel Stratix 10 (Intel Corporation 2020a, 2021) families combine large numbers of high-bandwidth interfaces and a wealth of programmable logic resources.

[2] Although named in honor of physicist André-Marie Ampère, the architecture's name is unaccented.
[3] Computations required per byte of input, see Williams et al. (2009).

These FPGAs are available in a number of hosts; these include pre-built standalone development boards like the Xilinx VCU 118 (Xilinx Inc 2018) and PCIe cards with integrated network interfaces like the Bittware XUP-VV8 and 520N (Bittware Inc 2021a,c). If selected, PCIe FPGAs would require host machines.

Many pre-built FPGA host boards feature one or more ANSI/VITA 57.1 FPGA Mezzanine Connector (FMC) or ANSI/VITA 57.4 FMC+ interfaces, which enable high-speed data transport via a standard for interchangeable modules. Currently-available Commercial Off-The-Shelf (COTS) adapters provide up to 6x 100GbE-capable QSFP28 connectors per FMC+ array (HiTech Global LLC 2021).

4.3.2. *AI-Optimized FPGA*

Following the same industry trends which gave rise to GPU tensor cores, FPGA manufacturers have released models with additional support for tasks commonly associated with AI implementation. These have data transport and programmable logic resources similar to the general-purpose FPGAs, with additional dedicated resources for limited-bit-depth matrix-matrix and/or vector-matrix computations.

The Xilinx Versal AI Core line of 'Adaptive Compute Acceleration Platforms' (ACAPs) (Xilinx Inc 2021c,d) feature a set of 'AI cores' for large-scale vector processing alongside more traditional programmable logic resources. Alternatively, the Versal Premium line of ACAPs provides an extremely large number of DSP cores capable of many types of computation, integrated alongside the programmable logic.

The Intel Stratix 10 NX family (Intel Corporation 2020b,c) features purpose-built 'AI Tensor Blocks' optimized for 4- and 8-bit dot products, alongside on-die HBM and high-speed transceivers featured on other Stratix 10 FPGAs.

4.3.3. *Achronix Speedster7t*

A special case of the AI-optimized FPGA, the Achronix Speedster7t series of FPGAs (Achronix Semiconductor Corporation 2019a) features a high-speed 'Network on Chip' (NoC) and a set of 'Machine Learning Processors' (MLPs) which perform a heavily-optimized vector dot product (Achronix Semiconductor Corporation 2019b). Arbitrarily programmable logic resources are reduced relative to other AI FPGAs; the hardware is streamlined and heavily focused on high-speed data transport and vector product computation. Speedster7t FPGAs are currently available as a COTS PCIe card (Bittware Inc 2021b), or may be integrated into a custom-designed host board.

4.3.4. *TALON-DX*

Featured in the ngVLA Reference Design (Ojeda 2018) and in previous minimally-modified derived designs (Ojeda 2020a,b) the TALON-DX board hosts one of a selection of Intel Stratix 10 SX-variant FPGAs along with large quantities of DDR4 memory and a number of high-speed interfaces (Pleasance et al. 2017; Carlson & Pleasance 2018).

The TALON-DX has a great deal of similarity to notional NRAO-designed ngVLA-specific FPGA host boards, save for its use of Leap On-Board Transceivers[4] (OBTs) to interface with a planned passive optical interconnect. For the purposes of this discussion, it is considered an instance of the General-Purpose FPGA design described in §4.3.1 which has committed to a specific FPGA package and optical interface.

[4] Referred to in Carlson & Pleasance (2018) as the FCI Leap Mid-Board Optic (MBO).

4.4. *ASIC*

4.4.1. *NRAO Custom*

An Application-Specific Integrated Circuit (ASIC) designed explicitly for the ngVLA correlator would require an initial design process comparable to that for a dedicated FPGA solution on an NRAO-designed board, followed by an in-depth verification and fabrication process. Performance could be precisely as required, and power consumption would be substantially reduced relative to an FPGA implementation.

D'Addario & Wang (2016) describe a 32 nm correlation ASIC which consumes 1.8 pJ per CMAC, less than even the latest 7 nm NVIDIA Ampere devices. Their overall system design focuses on an ASIC which performs only the outer product and integration, with some number of ASICs connected to a host FPGA which provides external data interfaces, handles buffering and data routing, and controls the ASICs' operation.

The potential advantages of this solution are primarily the power and space savings, which must be balanced against the cost of implementing and manufacturing an ASIC.

4.4.2. *Cerebras Wafer-Scale Engine*

The Cerebras Wafer-Scale Engine (WSE) (Cerebras Systems Inc 2021; Rocki et al. 2020) is a large monolithic ASIC designed for bulk AI inference calculations, which involve vector products similar to those required in the X-Engine. Applying the WSE (or the later WSE-2) to radio correlation was explored in earlier design stages, and is mentioned here for completeness; it is no longer under active consideration.

5. EVALUATION

5.1. *Metrics Used*

The metrics used to assess the choice of hardware employed in the conceptual design description (Denman et al. 2021a) are presented and briefly described in Table 1.

Due to the high level of parallelism present in the correlation task, any of the hardware options considered would be capable of satisfying the key requirements if applied in sufficient quantity. The evaluation of potential hardware choices is therefore an optimization; the capabilities of each hardware option inform the required quantity of nodes, and therefore the Hardware Cost and Power Consumption. The Hardware Cost metric represents the cost of enough of a given choice of hardware to satisfy the requirements of the X-Engine, rather than a per-unit price. This evaluation models the lifecycle cost as predominantly NRE plus Hardware Cost plus Power Consumption. The maturity of a proposed solution is included as a consideration to represent our confidence in (and the probability of) deploying a stable and effective correlator based upon a specific choice of hardware.

Potentially also a consideration, but not currently included due to high uncertainty, is the difficulty of sourcing a given hardware choice independent of price.

For the purposes of balancing hardware cost and NRE against power consumption, we have used the value of \$0.15 per kWh or \$1.31 per watt-year from Ojeda (2018) and a twenty-year hardware refreshment cycle; the results in an additional cost-of-power of ≈\$26 per node per watt of additional power consumption. This is intended to capture both the direct cost of additional power and the cost of cooling required to dissipate the additional waste heat generated.

5.2. *Hardware Option Evaluation*

(Hardware Cost)	Estimated cost of purchasing the hardware itself
(Non-Recurring Engineering)	Design effort required to implement correlator
(Power Consumption)	Cost of power consumed during operation
Maturity	Previous use of this hardware for a similar purpose

Table 1. Metrics used to evaluate hardware options. For those in parentheses, lower/less is better; for others, higher/more is better.

Table 2 presents a qualitative evaluation of the hardware options described in §4 against the metrics in Table 1. A more detailed description of each option's evaluation follows, with quantitative comparisons where appropriate.

Hardware		Ref	(HW Cost)	(NRE)	(Power)	Maturity
CPU	Generic	§4.1	DQ	Low	DQ	High
GPU	Ampere	§4.2.1	Med	Med	High	Med
GPU	Jetson	§4.2.2	High	Med	High	Med
FPGA	General	§4.3.1	Med	Med	Med	High
FPGA	AI	§4.3.2	Low	Med	Med	Med
FPGA	Speedster	§4.3.3	Med	Med	Med	Low
FPGA	TALON	§4.3.4	Med	Med	Med	Med
ASIC	Custom	§4.4.1	?	High	Low	Low
ASIC	WSE	§4.4.2	DQ	High	DQ	Low

Table 2. Qualitative evaluation of hardware options from §4 against criteria from §5.1 For those metrics in parentheses, lower/less is better; for others, higher/more is better. 'DQ' indicates that performance regarding a particular metric is considered disqualifyingly bad; '?' indicates that evaluation of a metric is ongoing.

CPUs: Intended for flexible general-purpose computing, CPUs are poorly suited for the X-Engine's large-scale outer product. The vast number of CPUs required and corresponding power consumption remove this option from consideration.

NVIDIA Ampere GPUs: Although capable of low-bit-depth processing at extremely high rates and with low per-operation power consumption, GPUs have relatively restricted input data bandwidth; this results in extreme inefficiency when correlating an ngVLA-like array. The initial hardware cost is therefore several times that of comparable solutions (and the power consumption no less), significantly disfavouring this choice of hardware.

Future developments in the GPU field may alter this conclusion; the introduction of user-accessible high-speed external interfaces should prompt a re-evaluation of this option.

NVIDIA Jetson AGX Xavier GPUs: As with other GPUs, I/O limitations require an extremely large number of nodes to receive the incoming voltage data. This, in turn, results in increased initial hardware costs and power consumption an order of magnitude greater than other solutions, removing

this option from consideration.

General-Purpose FPGAs: Modern general-purpose FPGAs are, at the chip level, well-supplied with transceivers supporting extremely high input and output data rates. They are not, however, capable of computational densities on the level of GPUs; specific models differ, but are generally much less capable of high-rate low-bit-depth processing.

These remain a good fall-back solution, particularly given the parameters of the ngVLA's correlation task when operating in sub-array modes. The most powerful general-purpose FPGAs currently available offer approximately one third of a GPU's INT8 processing power but more than sixteen times the I/O bandwidth, a better match to the system requirements.

Beyond the chip-level discussion, most commercial FPGA host boards do not expose a significant fraction of the available data bandwidth; if an FPGA-based correlator is to succeed, it will require additional data transfer resources. These may be obtained through either onboard FMC+ expansion modules or through the construction of a dedicated host board, and evaluation of these options' relative cost is ongoing.

AI-Optimized FPGAs: The accelerator cores ('AI engines', 'AI cores', etc.) present in AI-optimized FPGAs permits a significant increase in their computational power and efficiency relative to general-purpose FPGAs, while retaining the high-speed transceivers which GPUs lack. It is essential that the accelerator cores be efficiently integrated into the design; at present, some manufacturers only support them through dedicated AI-optimized compilers (Conly 2021), which may limit their effectiveness. Close consultation with the manufacturers of potential hardware is vital to ensuring it can be effectively employed in the correlator system.

Given this, however, AI-optimized FPGAs offer a combination of the advantages of GPUs and general-purpose FPGAs for a low overall cost; their primary drawback relative to general-purpose FPGAs is a lack of publicly-documented deployments in a digital signal processing role, and corresponding uncertainty about their practical performance.

Achronix Speedster7t FPGAs: These devices, a form of AI-optimized FPGA which emphasize internal data transfer bandwidth and processing, have a low initial hardware cost relative to other devices in the category. This is balanced by their relative novelty and a dearth of flexible resources alongside their 'Machine Learning Processors'. Although they offer the possibility of a very low price-to-performance ratio, this comes with significant technical risk in ensuring the design meets all its requirements. This hardware is therefore not presently favoured as a basis for the X-Engine design.

TALON-DX: The TALON-DX board features an Intel Stratix 10 SX FPGA with significant programmable logic and data transfer resources; it is effectively a general-purpose FPGA host board, although designed for a large-scale radio beamforming task with different parameters than ngVLA. The total available I/O is similar to that envisioned for an ngVLA-custom FPGA host, but is divided between multiple form factors. Due to the incorporation of a passive optical mesh into the TRIDENT design, the TALON board features only two 100 GbE QSFP28 connectors, with the majority of its I/O taking the form of specialized mid-board optical connections. The total hardware cost of a non-TRIDENT X-Engine built using TALON hardware is significantly higher than that of a

non-TALON, non-TRIDENT solution but is still far less than the TRIDENT case; as such, this option appears strictly inferior to a purpose-designed FPGA host board.

NRAO Custom ASIC: As detailed in D'Addario & Wang (2016), the power consumption of an ASIC may be significantly less than that of comparable general-purpose hardware. This comes with extremely high development and verification costs, the precise extent of which is presently unclear. Implementation would likely also require the ASICs to be hosted by FPGA data transfer and control units, requiring their own hardware selection and development. At present, it is not believed that the power savings relative to an FPGA-based solution would justify the additional costs.

Cerebras WSE: The Cerebras WSE is sub-optimal for the X-Engine's computations, and has a high unit cost and very high power consumption; these are sufficient to remove it from further consideration.

5.3. *Rationale for Selection of Hardware for Conceptual Design*

The design option currently under exploration (Denman et al. 2021a) is a board hosting an FPGA with additional network interfaces hosted either directly or via FMC+ modules. This exposes the extremely high data transfer bandwidth which is required for the ngVLA, particularly in heavily-sub-arrayed modes, while minimizing the quantity of hardware required to complete the outer product for the full array.

The FPGA device itself may be either an 'AI-optimized' model equipped with accelerator cores or a large general-purpose model with significant programmable DSP resources; the former offers more cost-efficient computational power, but only if the technical risks inherent in employing a novel architecture can be mitigated. Both varieties share a design path until a relatively late stage, differing primarily in the within-chip extent and distribution of computational resources. Prior to the detailed signal-processing design stage, the design is agnostic on the specific FPGA model, permitting optimization on a cost and efficiency basis once the performance of each option is well-determined.

An additional benefit is that when combined with the recommendations of the accompanying Pulsar Engine trade study (an HBM-equipped FPGA on a similar host board, Denman et al. (2021c)) it may be feasible to select package-compatible HBM and AI FPGAs which permit a unified host board, resulting in a significant reduction in design costs. The release of future FPGAs featuring both HBM and accelerator cores (likely, given current industry trends) would potentially permit convergence on a single hardware platform for all sub-band processing,

REFERENCES

Achronix Semiconductor Corporation. 2019a, Speedster7t FPGA Datasheet, Tech. Rep. DS015, Achronix Semiconductor Corporation

—. 2019b, Speedster7t Machine Learning Processing User Guide, Tech. Rep. UG088, Achronix Semiconductor Corporation

Bittware Inc. 2021a, 520N PCIe FPGA Board Datasheet, Tech. rep., Bittware Inc

—. 2021b, S7t-VG6 VectorPath Accelerator Card Datasheet, Tech. rep., Bittware Inc

—. 2021c, XUP-VV8 PCIe FPGA Board Datasheet, Tech. rep., Bittware Inc

Carlson, B., & Pleasance, M. 2018, TRIDENT Correlator-Beamformer for the ngVLA: Preliminary Design Specification, Tech. Rep. TR-DS-000001, National Research Council Canada

Cerebras Systems Inc. 2021, Achieving Industry Best AI Performance Through A Systems Approach, Tech. rep., Cerebras Systems Inc

Conly, M.-A. 2021, Private Communication with ND

D'Addario, L. R., & Wang, D. 2016, Journal of Astronomical Instrumentation, 5, 1650002

Denman, N., et al. 2021a, A SCREAM-Compatible ngVLA Cross-Correlation Engine: Design Description, Tech. Rep. TBA, National Radio Astronomy Observatory

—. 2021b, A SCREAM-Compatible ngVLA Pulsar Engine: Design Description, Tech. Rep. TBA, National Radio Astronomy Observatory

—. 2021c, ngVLA Electronics Memo #11: A SCREAM-Compatible ngVLA Pulsar Engine: Key Requirements Review and Option Trade-Off Study, Tech. Rep. E11, National Radio Astronomy Observatory

HiTech Global LLC. 2021, 6-Port QSFP28 (6x100G) FMC+ Module, http://www.hitechglobal.com/FMCModules/x6QSFP28.htm

Intel Corporation. 2020a, Intel Stratix 10 GX/SX Device Overview, Tech. Rep. S10-OVERVIEW, Intel Corporation

—. 2020b, Intel Stratix 10 NX FPGA Product Brief, Tech. Rep. ss-1121, Intel Corporation

—. 2020c, Intel Stratix 10 NX FPGA Product Table, Tech. Rep. Gen-1070, Intel Corporation

—. 2021, Intel Stratix 10 Device Datasheet, Tech. Rep. S10-DATASHEET, Intel Corporation

Koski, W., Baca, J., & Durand, S. 2019, ngVLA Monitor & Control Hardware Interface Layer: Reference Design Description, Tech. Rep. 020.30.45.00.00-0004-DSN-A, National Radio Astronomy Observatory

Küsel, T. 2021a, ngVLA Subsystem CoDR Scope Definition, Tech. rep., National Radio Astronomy Observatory

—. 2021b, ngVLA Systems Engineering Management Plan, Tech. Rep. 020.10.00.00.00-0001-PLA, National Radio Astronomy Observatory

NVIDIA Corporation. 2017, NVIDIA Tesla V100 GPU Architecture, Tech. Rep. WP-08608-001, NVIDIA Corporation

—. 2020a, NVIDIA A100 Tensor Core GPU Architecture, Tech. rep., NVIDIA Corporation

—. 2020b, NVIDIA DGX A100 Datasheet, Tech. rep., NVIDIA Corporation

—. 2020c, NVIDIA HGX A100 Datasheet, Tech. rep., NVIDIA Corporation

—. 2020d, NVIDIA Jetson AGX Xavier Developer Kit Carrier Board Specification, Tech. Rep. SP-09778-001, NVIDIA Corporation

—. 2020e, NVIDIA Jetson AGX Xavier Series OEM Product Design Guide, Tech. Rep. DG-09840-001, NVIDIA Corporation

—. 2020f, NVIDIA Jetson AGX Xavier Series System-on-Module Data Sheet, Tech. Rep. DS-09654-001, NVIDIA Corporation

—. 2021, NVIDIA Ampere GA102 GPU Architecture, Tech. rep., NVIDIA Corporation

Ojeda, O. 2018, ngVLA Central Signal Processor: Preliminary Reference Design, Tech. Rep. 020.40.00.00.00-0002-DSN-03, National Radio Astronomy Observatory

—. 2020a, Trident 2.0 Concept: A Minimum Delta Update to the Central Signal Processor Reference Design, Tech. Rep. E04, National Radio Astronomy Observatory

—. 2020b, Trident 2.1 Concept: Updates to the CSP Reference, Tech. Rep. E05, National Radio Astronomy Observatory

Ojeda, O., Lacasse, R., Selina, R., et al. 2018, ngVLA Central Signal Processor: Preliminary Technical Specifications, Tech. Rep. 020.40.00.00.00-0001-SPE-07, National Radio Astronomy Observatory

—. 2019, ngVLA Central Signal Processor: Preliminary Technical Requirements, Tech. Rep. 020.40.00.00.00-0001-REQ-A, National Radio Astronomy Observatory

Pleasance, M., Zhang, H., Carlson, B., et al. 2017, in 2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS), 1–4

Rau, U., Selina, R., & Erickson, A. 2019, RFI Mitigation for the ngVLA: A Cost-Benefit Analysis, Tech. Rep. 70, National Radio Astronomy Observatory

Rocki, K., Van Essendelft, D., Sharapov, I., et al. 2020, arXiv e-prints, arXiv:2010.03660

Romein, J. 2021, The Tensor-Core Correlator, Presentation at CASPER Workshop 2021

Selina, R., Rau, U., Hiriart, R., & Erickson, A. 2020a, RFI Mitigation for the ngVLA: A Cost-Benefit Analysis, Tech. Rep. 71, National Radio Astronomy Observatory

Selina, R., et al. 2020b, ngVLA: CSP System Concept Options & Trade-Offs, Tech. Rep. 020.10.25.00.00-0005-REP, National Radio Astronomy Observatory
—. 2020c, ngVLA: System Requirements, Tech. Rep. 020.10.15.10.00-0003-REQ, National Radio Astronomy Observatory
Williams, S., Waterman, A., & Patterson, D. 2009, Commun. ACM, 52, 65–76
Xilinx Inc. 2018, VCU118 Evaluation Board User Guide, Tech. Rep. UG1224, Xilinx Inc
—. 2021a, UltraScale Architecture and Product Data Sheet: Overview, Tech. Rep. DS890, Xilinx Inc
—. 2021b, UltraScale+ FPGAs: Product Tables and Product Selection Guide, Tech. Rep. XMP103, Xilinx Inc
—. 2021c, Versal ACAP Design Guide, Tech. Rep. UG1273, Xilinx Inc
—. 2021d, Versal Architecture and Product Data Sheet: Overview, Tech. Rep. DS950, Xilinx Inc