

# ngVLA Electronics Memo #20

## XE and PSE Hardware Class Down-Selection

Nolan Denman

May 22, 2025

### Contents

<b>1</b>	<b>Overview</b>	<b>1</b>
1.1	Representative Hardware . . . . .	2
<b>2</b>	<b>Considerations</b>	<b>4</b>
2.1	Processing . . . . .	4
2.2	Data Transport . . . . .	4
2.3	Power and Cooling . . . . .	4
2.4	Lifespan . . . . .	5
2.5	Development Effort . . . . .	5
2.6	Availability . . . . .	5
<b>3</b>	<b>Discussion</b>	<b>6</b>
3.1	Hardware Cost . . . . .	6
3.2	Operational Cost . . . . .	6
3.3	Development Schedule . . . . .	7
<b>4</b>	<b>Conclusion</b>	<b>7</b>

### 1 Overview

The decision as to what processing hardware will be used on the ngVLA CSP X-Engine (XE) and Pulsar Engine (PSE) must be made in the near future, to permit further development of their designs. It is strongly desired that the XE and PSE share their hardware family, if not their specific choice of hardware, in order to limit the duplication of development effort. In ngVLA Electronics Memos #10 and #11, a variety of options were presented; development following the publication of those memos has informed the narrowing of those options to one of the following:

- An FPGA-based system with a set of independently-operating PCIe-hosted nodes, each of which hosts a large FPGA along with a set of network interfaces, memory, and management components. The LRU server primarily supplies power and serves monitoring and control functions.
- A GPU-based system, where a set of PCIe- or SXM-hosted GPUs are paired with PCIe Network Interface Cards (NICs) and are managed and operated by the LRU server.

This document will summarize the ways in which the two options may differ in relation to XE and PSE performance considerations, and the effect this has on the resources required to design and construct the ngVLA CSP. The current perspective of the XE and PSE development teams will be offered, along with a conditional down-selection for further development.

## 1.1 Representative Hardware

For the purposes of discussion, representative hardware has been identified: the NVIDIA L40 GPU and the Bittware IA-860m, which hosts an Agilix AGM039 FPGA. Both are powerful devices with broad present-day availability, and each represents the most efficient selection within its hardware category. This choice is not meant to suggest even a tentative selection for actual construction, but to provide concrete values for comparison. A summary of relevant characteristics of each hardware option is provided in Table 1. Figures 1 and 2 show the overall structure of LRUs based on each of the hardware options.

Previous discussion of a GPU-based X-Engine considered pre-integrated multi-GPU solutions such as the NVIDIA DGX platform; these are aimed at more compute-intensive problems, and their relative cost is higher as a result. The XE or PSE would not make good use of their capabilities, being limited by I/O and unable to make use of more than a tiny fraction of the available computing power.

	NVIDIA L40	Bittware IA-860m
Form Factor	Dual-Slot PCIe	Dual-Slot PCIe
Max. Power Draw	300 W	250 W
TFLOPS	90.5	18.4
INT8 TOPS	362	88.6
I/O	256 Gbps PCIe	3x 400GbE
Memory	48 GiB GDDR6	32 GiB HBM2e
Retail Price	~\$8k (+ ~\$2k NIC)	~\$15k

Table 1: Representative hardware chosen for this discussion.

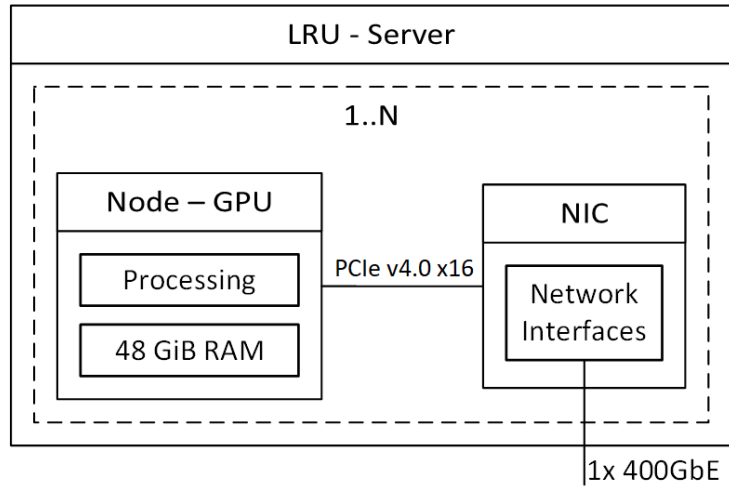


Figure 1: A schematic of the GPU-based LRU structure; each server LRU contains one or more pairs of L40 GPUs and 400GbE NICs, connected via PCIe.

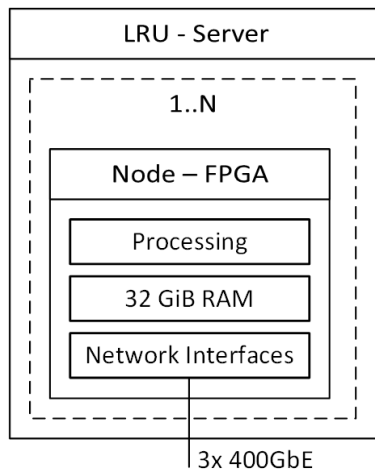


Figure 2: A schematic of the FPGA-based LRU structure; each server LRU contains one or more FPGAs, each of which hosts its own network interfaces.

## 2 Considerations

### 2.1 Processing

The fundamental processing task of the X-Engine, an outer product of all the receivers within each frequency channel, requires approximately  $2.7 \times 10^{15}$  complex multiply-and-accumulate operations per second (CMAC/s). At 8 integer operations per CMAC, this is the equivalent of  $2.2 \times 10^{16}$  integer operations per second (OPS).

The Pulsar Engine has comparatively relaxed requirements for processing; these are dominated by the large FFTs required for coherent dedispersion when operating at the lowest available frequencies and highest time resolution. The most extreme scenario would require about  $4.8 \times 10^{11}$  operations per second for the lowest-frequency subband (of  $\sim 50$ ), with the processing load dropping rapidly at higher frequencies.

The FPGA and GPU hardware options differ significantly in their overall processing capacity; in addition to an across-the-board advantage in processing capacity, many modern GPUs (including the L40 discussed here) are equipped with ‘tensor core’ systems, which greatly accelerate certain categories of matrix multiplication operations, including those involved in correlation. As such, many fewer GPU nodes are required to achieve the same overall processing capacity relative to those based on FPGAs.

### 2.2 Data Transport

Data transport requirements, specifically the high rate of input data to the XE and PSE, are one of the key factors informing the scale of the systems required. In full utilization, the XE must handle  $\sim 168$  Tbps of input data while the PSE receives  $\sim 5.7$  Tbps (both exclusive of metadata).

An FPGA-based CSP system would have each PCIe node provide its own network interfaces, located on the same card as the FPGA itself. This enables extremely high-bandwidth connections directly to the node (3x 400GbE for the IA-860m), minimizing the limitations placed on the system by data transport.

A GPU-based CSP system would instead pair each GPU with a separate NIC, connecting the two via PCIe. This is the primary bottleneck in the resulting system, particularly as the L40 example hardware only supports 16 lanes of PCIe v4.0, for a total of  $\sim 256$  Gbps. A system using PCIe v5.0 supports a nominal data transfer rate of  $\sim 63$  GB/s (504 Gbps) for a 16-lane connection, and PCIe v6.0 approximately doubles this to  $\sim 121$  GB/s (968 Gbps).

### 2.3 Power and Cooling

Although previous discussions of CSP hardware choices identified power consumption as an area in which FPGAs had a significant advantage over GPUs, the latest generations of GPU have narrowed this gap substantially. Based upon the available information, overall power consumption is expected to be similar between the two options.

Cooling requirements are largely set by the devices' power consumption, and both options have similar cooling profiles – dual-slot PCIe passive cooling via forced air, with liquid cooling available as a separate upgrade.

## 2.4 Lifespan

At present, the GPU sales lifecycle is short; performance improvements generation-to-generation are significant, but supplies of any one model are very time-limited. This complicates the ability to procure drop-in replacements for failures which occur after the initial stock of spares is exhausted. FPGAs have a longer commercial lifecycle, and any specific FPGA (the chip, rather than the integrated board) is available for a longer period of time. However, the integrated FPGA units do not have particularly long lifecycle, so when evaluating the availability of drop-in replacements there is little additional grace compared to the GPUs.

There is little information publicly available about the relative service lifespans and failure rates of the selected hardware; what information there is does not suggest a significant difference between the two options.

## 2.5 Development Effort

The overall scale of development effort for GPU and FPGA hardware is not incomparable; both would require significant work to achieve the required performance. Previous discussions categorized both as 'medium' effort, with CPUs ('low') and ASICs ('high') at the extremes. However, FPGA development requires an extremely specific skillset and is highly labor-intensive; GPU software development has a relatively lower barrier to entry and many of the interfaces and software packages would be general-purpose rather than specialized.

Software packages for digital signal processing in general, and radio-telescope correlation in particular, are available for both FPGAs and GPUs, but correlator development has emphasized GPUs in recent years. Initial testing suggests that existing packages (e.g. John Romein's Tensor-Core Correlator<sup>1</sup>) may be adapted for use on the ngVLA with little modification. Additionally, the level of effort required to move software from one generation of GPU to another while maintaining high resource utilization is considered to be lower than in the case of FPGAs.

## 2.6 Availability

At present, acute supply chain issues from the COVID-19 pandemic appear to be largely settled. The short- to medium-term availability of hardware is not expected to be a significant obstacle to CSP construction. There is additionally no reason to believe that any particular hardware choice would be more or less affected by uncertainties in the international trade and tariff situation.

---

<sup>1</sup><https://git.astron.nl/RD/tensor-core-correlator>

## 3 Discussion

It appears entirely reasonable to construct a correlator system based upon either FPGA or GPU processing nodes; there are currently-operating and design-stage systems using either technology – FPGAs (e.g. ALMA ATAC, present-day MeerKAT) and GPUs (e.g. ALMA TPGS, CHORD, planned MeerKAT upgrade). The question is therefore one of resource-efficiency, including the initial hardware cost, operational costs, and the development effort required to meet the ngVLA schedule.

### 3.1 Hardware Cost

There is necessarily a significant uncertainty in the initial hardware cost, which is determined by the number of nodes required and their unit cost. For the purposes of this document, though, it is sufficient to compare the relative number and relative cost of the options, without requiring the specific totals be accurately predicted – prices are constantly evolving, and hardware purchases for CSP construction are multiple years in the future.

Table 2 shows the relative number of nodes required to meet the processing and data transportation needs of the XE and PSE. The total node count is dominated by the XE, driven by I/O for the GPU option and processing for the FPGAs; the PSE is entirely I/O limited.

For the reference hardware described in Subsection 1.1, the GPU XE and PSE would require about 2.7 times as many nodes as the FPGA system. At a per-node cost of roughly 1.5 times as much for FPGAs as GPUs, this would indicate that FPGAs are substantially lower-cost. With a PCIe v5.0-equipped GPU (and larger NIC), however, this drops to about 1.3 times as many nodes – making FPGAs and GPUs similar in total hardware cost. Currently, only NVIDIA’s top-of-the-line GPUs (the H100 and RTX 5090) are offered with PCIe v5.0, but there are strong indications that future generations will adopt this as standard.

This comparison relies centrally on the assumption that the actual performance of each system relative to its nominal capabilities for the most restrictive aspect (i.e. I/O for GPUs and processing for FPGAs) is in approximately the same proportion. That is, there is not a reason to expect that the (actually achieved) efficiency of GPU I/O is very different from the efficiency of FPGA processing. The comparison would become more favorable to FPGAs only if their use of processing resources was much more efficient than the GPUs’ use of I/O. Initial tests of GPU PCIe data transfer have shown performance at about 75% of nominal can be achieved with the supplied high-level libraries; if FPGA performance was 100%, this would change the relative node counts from  $\sim 1.3$  to  $\sim 1.7$ , still not substantially different from the per-node cost ratio of  $\sim 1.5$ . As such, the total hardware cost of each option appears the same, within the uncertainty inherent in these estimates.

### 3.2 Operational Cost

The operational cost, although similarly uncertain to the initial hardware cost, is largely set by the power consumption and service lifespans of the hardware in question. As discussed in

	XE Proc.	XE I/O	PSE Proc.	PSE I/O
NVIDIA L40	62	<u>658</u>	<1	<u>23</u>
with PCIe v5.0	62	<u>334</u>	<1	<u>12</u>
IA-860m	<u>250</u>	140	<1	<u>5</u>

Table 2: Relative measures of the number of nodes required for the XE and PSE based on data transfer rates (‘I/O’) and processing (‘Proc.’) The underlined values are those which are most restrictive in each case; the XE is I/O-limited for GPUs but processing-limited for FPGAs, while the PSE is always I/O limited. These should *not* be treated as actual or potential node counts, but are expected to be roughly proportional to the number required and so can be used to inform the relative hardware cost of each option.

Subsections 2.3 and 2.4, these are not believed to differ significantly between the two options.

### 3.3 Development Schedule

The development schedule for ngVLA CSP systems has been modified due to resource constraints, with implications for the choice of CSP processing hardware. Previous discussions (ngVLA Electronics Memos #10 and #11) assumed steady development over the entire period; in contrast, the expectation is now that there will be limited resources until late in the development period, after which there will be more resources but much less time to complete the final design and system construction.

This suggests that the specificity and labor-intensive nature of FPGA development (per Subsection 2.5) would pose an obstacle to reduced-timeframe development. Additionally, GPU development may be ‘scaled up’ more easily in response to a go-ahead signal on system construction; the interfaces and software packages are relatively standard, and the overall complexity of implementations is generally lower.

## 4 Conclusion

The ngVLA CSP X-Engine and Pulsar Engine, when considered jointly, would be more efficiently constructed using GPUs rather than FPGAs. The initial hardware costs and operational costs do not significantly differ between the two options. GPUs, however, permit a shorter and less resource-intensive development cycle which is a much better fit to the evolving schedule of ngVLA design and construction.

This is contingent upon the belief that a GPU with price and performance comparable to the L40 but a PCIe v5.0 x16 interface will be commercially available by the time hardware is required for the ngVLA CSP. Although an acknowledged risk, this is well-supported by current trends in hardware manufacture; the announcement of next-generation hardware specifications in the moderate future would largely retire this concern.

*The National Radio Astronomy Observatory and Green Bank Observatory are facilities of the U.S. National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This work was supported by awards AST-2034328 (MSIP Prototype Antenna) and AST-2334267 (ngVLA Design Activities); NRAO related activities are funded under award AST-1647378 (NRAO Operations/Development).*

