#### **EDHPC 2025**





European Data Handling & Data Processing Conference

13 - 17 October 2025 | Elche | Spain



# Accelerating Neural Network Inference in space by offloading heavy operations on systolic arrays

Constantin Papadas<sup>2</sup>, Ioannis Katelouzos<sup>2</sup>, Agamemnonas Kyriazis<sup>2</sup>, Tilemachos

Tsiapras<sup>2</sup>, Roland Brochard<sup>1</sup>, Jérémy Lebreton<sup>1</sup>, Lucas Marti<sup>1</sup>, Evangelos Kioulos<sup>2</sup>, Tomasz

Grzegory<sup>2</sup>, Tim Helfers<sup>4</sup> and Laurent Hili<sup>3</sup>,

<sup>1</sup>Airbus Defence & Space, Toulouse Toulouse, France

<sup>2</sup>ISD S.A.Athens GREECE

<sup>3</sup>ESA-ESTEC Microelectronics Section (TEC-EDM), THE NETHERLANDS

<sup>4</sup>AIRBUS DEFENCE & SPACE GmbH, Taufkirchen, GERMANY

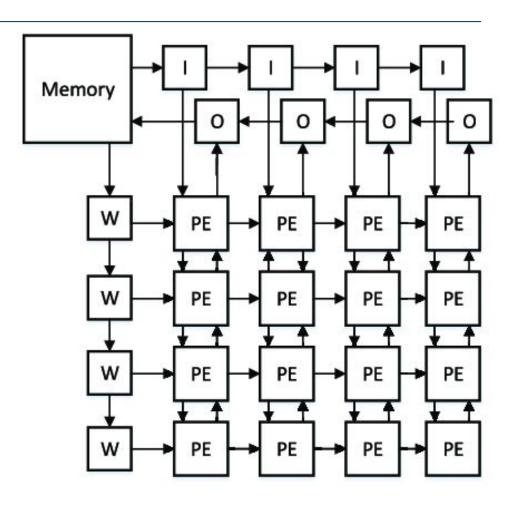
## **Outline**

- Systolic Arrays and HPDP40
- Use cases
- Methodology
- Results
- Conclusion



## **Systolic Arrays**

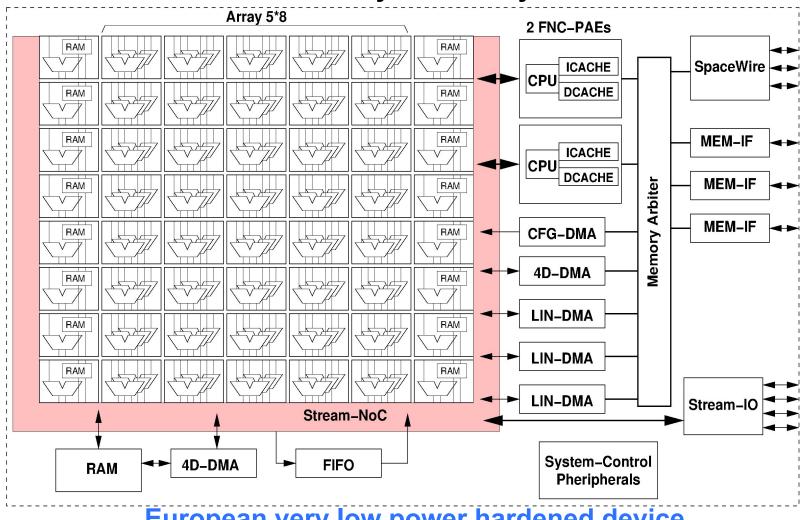
- Parallel processing Architecture
  - A grid of processing elements (PEs) that compute and pass data between them.
- Continuous Data flows
  - Inputs pass through the array in a synchronized pattern, allowing continuous computation without waiting for global memory
- Optimized for Matrix operations.
- Direct local communication
  - Each PE only talks to its neighbors, reducing the need for complex interconnects or global buses
- Deterministic Data Flow
  - Predictable timing and control make them suitable for real-time systems and fault-tolerant designs
- Energy Efficient
  - Optimized for multiply-accumulate operations, which dominate many space AI workloads





## HPDP40

#### **HPDP40** is a systolic array SoC



European very low power hardened device



## **HPDP40 - features**

- 40Gops of arithmetic operations (floating point operations can be emulated)
- 1.65W power consumption
- Re-configuration on the fly (1us)
- fast boot time (few ms)
- 4x 1.4Gbps Streaming Ports
- >4 MB on-chip SRAM
- **256MB SDRAM**
- Very low BoM solution
- TID > 300Krad
- SEL free for LET up to 72.2 MeV-cm2/mg @ 90°C junction temperature and max supply voltage
- Worst case SEFI under radiations: 1.25E-4 errors/device/day

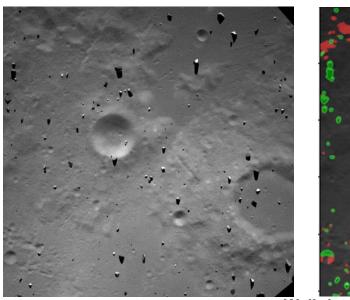
#### European very low power hardened device

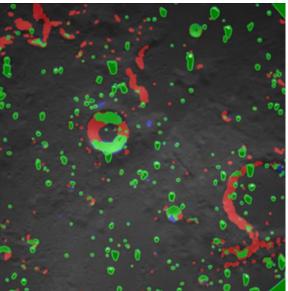


## LiDAR-Free HDA

VBN is an enabler for complex missions requiring onboard autonomy for example Lunar landers

- → Terrain relative navigation (AbsNav, RelNav)
- → Hazard Detection and Avoidance
  - Camera-based hazard detection algorithm
  - U-net Convolutional Neural Network
  - LiDAR-based solution are costly and heavy
  - So far only **very small neural networks** can be used with existing platforms (1000 parameters)
  - Works very well on rocks/roughness, less on slopes





Well-detected hazards

Undetected hazards

Accelerating Neural Network Inference in space by offloading heavy operations on systolic arrays – EDHPC – October 2025

## Vision Based Navigation

Planetary Landing

In-orbit rendezvous

Debris removal

Space Situational Awareness

Rovers

## **LiDAR-Free HDA**

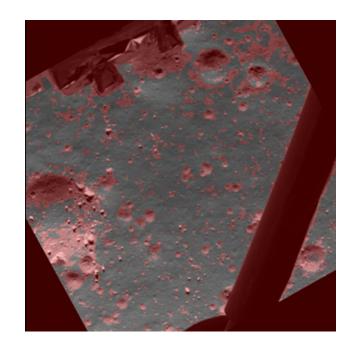
**Current targeted platforms can only run small networks** 

- CPUs have limited processing capabilities
- Tensor convolution is heavy

#### New hardware acceleration could bring

- Potentially bigger networks
  - More robust => More challenging scenarios
- Enable online 3D estimation to help with the detection of slopes
  - No LiDAR => Less costs, less weight
- More complicated architectures with more operations
- Earlier Detection to improve Avoidance strategy
- Running the algorithm multiple times

Chang'E 3 Descent image from NavCam and Hazard detection



## Vision Based Navigation

Planetary Landing

## Software is important

Hardware is not the end of the story!

- Acceleration is better than dedicated hardware
  - HW accelerated common algorithmic primitives can be reused
  - Dedicated implementations are costly
- We need good software along with the hardware!
  - We want to target an API, not a platform
  - Focus on the added algorithmic value, not the implementation details close to the hardware
- Already validated software blocks accelerate industrialisation of solutions
  - O Why validate a 2D convolution every time ?

## **GIGA**

A new API for porting Neural Networks

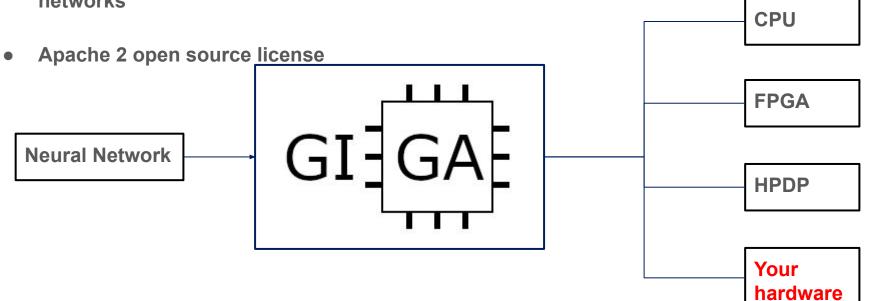
## **Generic Interface Generic Accelerator**

- Our solution : GIGA
  - Platform independent API for porting neural networks
- Make any accelerator easy to use
  - An accelerator implementing the API is (almost) ready to use as is

Already three ongoing projects (GENEVIS2, HPDP, NEURAVIS) use GIGA for porting neural networks

CPU

CPU



## **GIGA**

A new API for porting Neural Networks



## **Generic Interface Generic Accelerator**

#### Operations

- 2D convolution (Backbone of many classical networks)
- Dense Layers
- 0 ...

#### Memory

- Memory layout and allocation is carefully considered when transcoding the network into C
- The memory usage is solved and explicit before flight

#### Validation

- A network is a C-code with no dependencies
  - No proprietary code
  - The entire network in one C file
- C-code is generated from the very explicit and human readable NNEF format
  - Non-proprietary format
  - Explicit and fully specified operations



## **GIGA**

A new API for porting Neural Networks

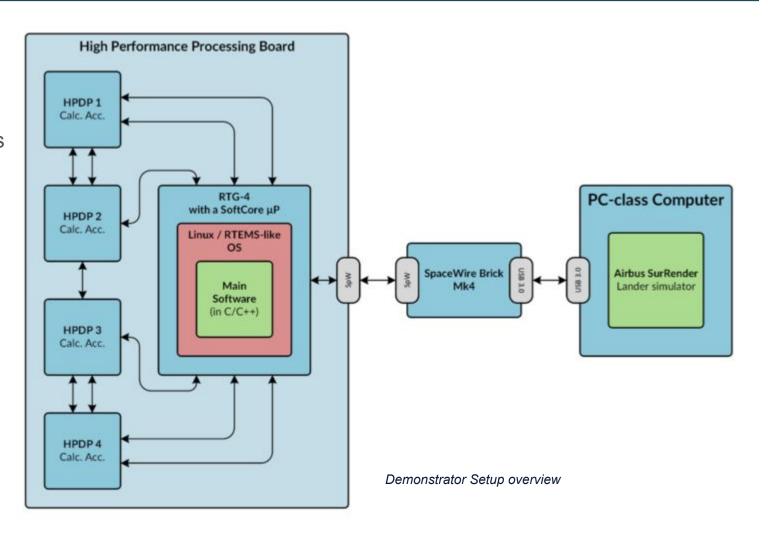
## **Methodology - Demonstrator**

#### **Demonstrator Components:**

- The HPPB board.
  - A board comprising of 4 HPDP chips and one FPGA
- Host computer
- SurRender image simulator
- Host Demonstrator Application Software



HPPB Board

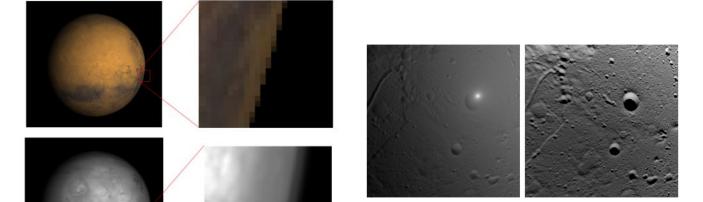


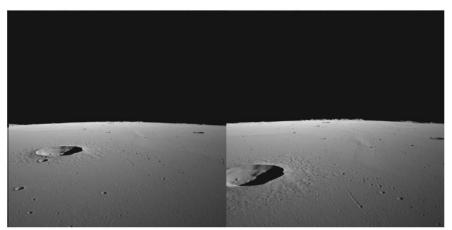


## **Methodology - Simulator**

**SURRENDER software**: Airbus high performance image simulator specialized for space applications (MSR ERO, JUICE, EL3/Argonaut, in-orbit rendezvous, SSA, etc.)

- ▶ Physical rendering in raytracing (illumination, shadows, brightness, ...)
- **✓** Smart raysampling, RAM mgt, double prec., giant datasets rendered continuously
- ✓ Tradeoff high quality images vs real-time





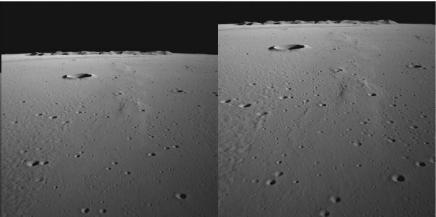


Figure 1: Examples of Moon images rendered with the SurRender software (dataset NAT-DATA-S5-slow) Here the site is the one of Chang'e-3 lander, the Lunar surface model is a fusion of a low resolution DEM and a high-resolution one in the central region, with constant albedo and a Hapke BRDF model.

Figure 2: Rastering (OpenGL, DirectX) vs high quality raytracing (SurRender)



## Methodology - Simulator design

#### Lunar terrain model:

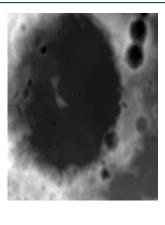
- 260 × 209 km DEM¹ of Von Karman crater at 1 m (MADNET 2.0²)
- Textures: 370 monocular LROC NAC<sup>3</sup> images
- Test / train / valid datasets to avoid bias

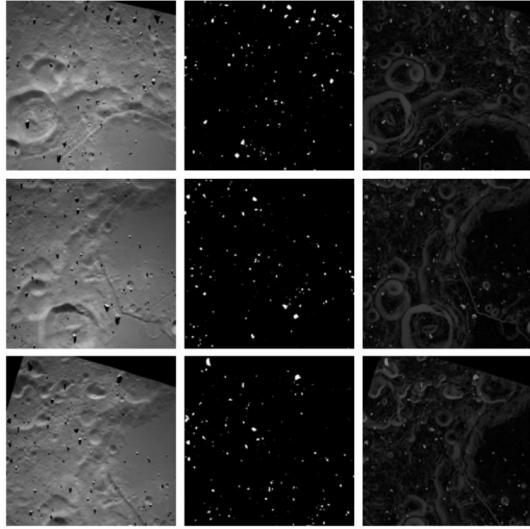
#### Input:

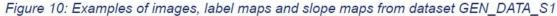
- Camera pose and Sun position (SPICE kernels)
- 70° FoV, 1024x1024 pixel resolution, PSF
- Sample altitudes [500–1000 km] (500m => GSD<sup>4</sup> ~0.5 m/pixel)
- Slopes and craters from the DEM
- Synthetic boulders are distributed on the surface with models for size, density, albedo and shapes from the scientific literature

#### **Output:**

- 11,000 images
- Label maps: slopes, craters, boulders (⇒ hazards)
- Depths maps
- <sup>1</sup> DEM: Digital Elevation Model
- <sup>2</sup> UCL-MSSL\_Moon\_von\_Karman\_V1.0, https://doi.org/10.57780/esa-fb921t3
- <sup>3</sup> LROC NAC: Lunar Reconnaissance Orbiter Narrow Angle Camera
- <sup>4</sup> Ground Sample Distance









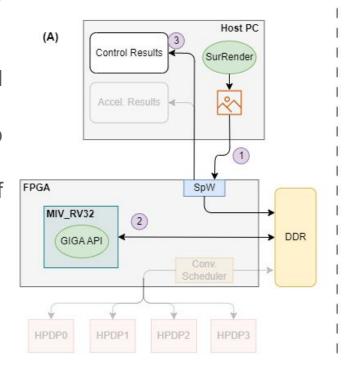
## **Methodology - Procedure**

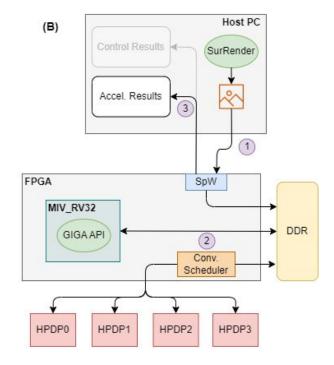
#### **Procedure:**

- Divided into two phases. A) control B) accelerated
- Test images common for both phases
- Execute all processing on the Risc-V and measure performance (Phase A)
- Repeat processing by offloading heavy duty to HPDP and measure performance (Phase B)
- Provide detailed report and comparison of results.

#### **Algorithms:**

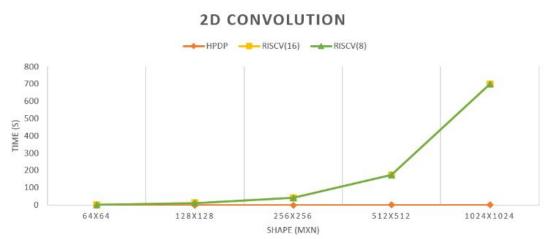
- Convolution
- Dense layer
- Bilinear Upscale
- Nearest Neighbour Upscale
- Addition







## Results – RiscV vs HPDP (1/2)



HPDP (s)	RISC-V (s)	speedup
0.18	2.56	14.22
0.19	10.56	55.58
0.23	43.24	188.00
0.36	173.9	483.06
0.79	700.3	886.46

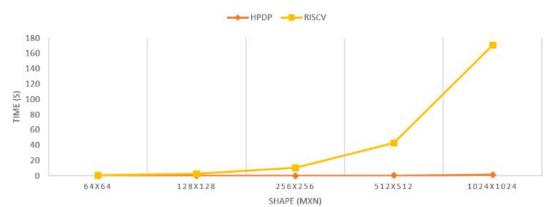


HPDP (s)	RISC-V (s)	speedup
0.14	1.84	13.14
0.14	3.69	26.36
0.14	7.38	52.71
0.15	14.75	98.33
0.29	29.51	101.76



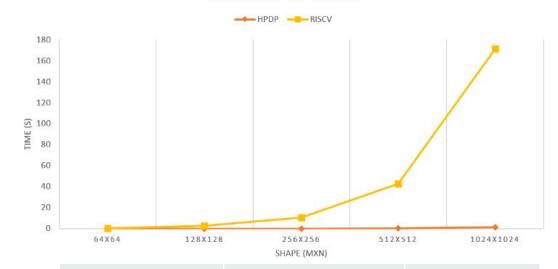
## Results – RiscV vs HPDP (2/2)

#### **NEAREST NEIGHBOUR UPSCALE**



HPDP (s)	RISC-V (s)	speedup
0.14	0.67	4.79
0.16	2.67	16.69
0.21	10.71	51.00
0.41	42.82	104.44
1.28	171.26	133.80

#### **BILINEAR UPSCALE**

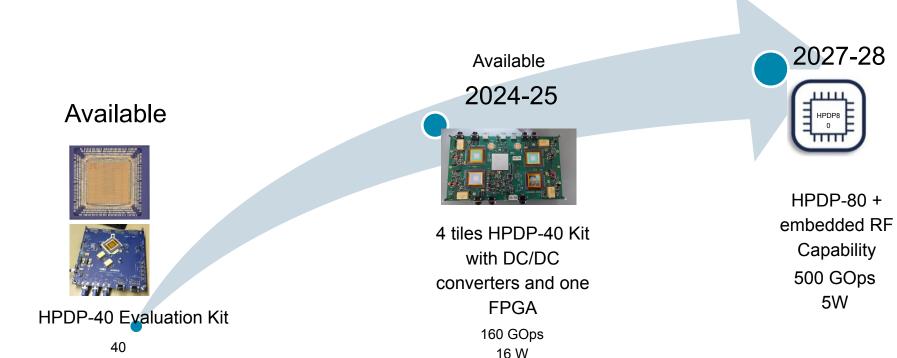


HPDP (s)	RISC-V (s)	speedup
0.13	3.23	24.85
0.15	13.23	88.20
0.2	53.95	269.75
0.4	216.72	541.80
1.22	871.56	714.39



## **Conclusion**

- □ Such technology seems to be quite promising for autonomous navigation and landers
- □ Over the next years with the arrival of HPDP80 these performances are expected to be further increased by a factor of 10
- Over the next years ISD will reinforce the cooperation with Klepsydra in order to get the maximum benefit from their innovative framework.





GOps 1.65W

## Thank you for your attention!

Questions?

