www.bsc.es

GPU4S (GPUs for Space):

Are we there yet?

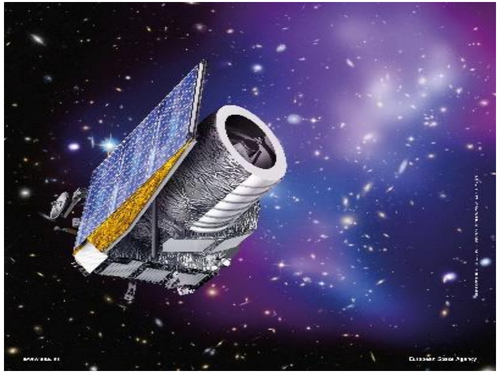
Leonidas Kosmidis, Iván Rodriguez, Alvaro Jover-Alvarez, Guillem Cabo, Sergi Alcaide, Jérôme Lachaize, Olivier Notebaert, Antoine Certain, David Steenari

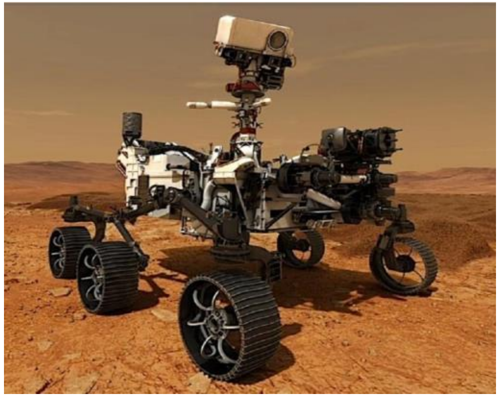
Barcelona Supercomputing Center
Centro Nacional de Supercomputación
Universitat Politècnica de Catalunya
Airbus Defence & Space
esa

15/06/2021 OBDP 2021

Introduction

- Increased need for computing power in on-board computers for satellites, rovers and spacecraft
- Bigger sensors sizes and more complex tasks
  - Vision-based Navigation
  - Autonomous Operation
- Radhard processors are limited in their processing performance
- Dedicated FPGA designs are used when additional compute performance is needed

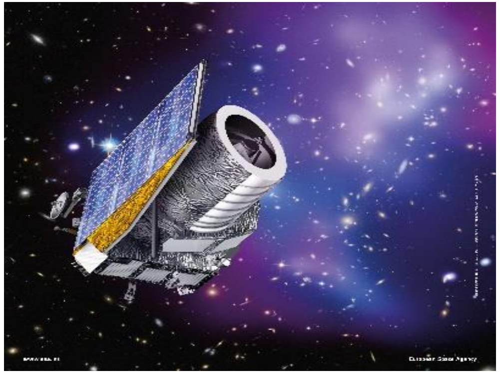


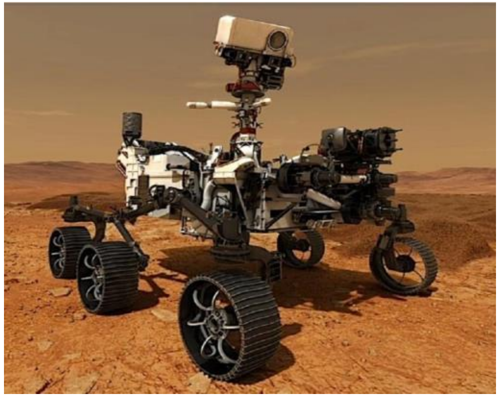


Images courtesy of ESA and NASA

Introduction

- Embedded GPUs can provide high performance, with low power consumption
  - Effectively allowing the use of software for processing tasks on-board
  - More flexibility, easier reconfiguration and can support several different processing tasks through reuse of compute resources
  - Easy access to specialised developers, widely used programming models
  - Overall lower development cost





Images courtesy of ESA and NASA

GPU4S (GPU for Space) ESA AO/1- 9010/17/NL/AF Overview

- Goal:
  - Evaluation of GPU IP for possible future space processors
  - Evaluation of COTS GPUs
- Tasks:
  - Perform a survey of the state of the art in
    - Existing embedded GPU, mainly European and major US (Nvidia, AMD)
    - Existing and future space algorithms amenable to GPGPU acceleration
  - Select promising embedded GPUs
    - benchmark and compare them with existing on-board technologies
  - Build a demo of a space application on the most appropriate candidate
  - Define the roadmap for the adoption of embedded GPUs in space

GPU4S (GPU for Space) CCN1 Overview

- Extend GPU4S Benchmarking to AMD Devices (Embedded Ryzen, Unibap DDX-i5)
  - originally not available during GPU4S
  - Promising radiation results
- Extend the demonstrator
  - Implement image processing 1.1 OBPMark in CUDA and OpenCL
  - Port Euclid NIR in OpenCL
  - Evaluate both of them on all platforms, including the new ones
- Perform detailed power measurements of all boards
  - During GPU4S we conservatively used the manufacturer's TDP for computing energy efficiency

GPGPU Development Frameworks Overview

Programming Framework	Type of API	Proprietary/ Vendor	Programmability/ Easy to use for Compute	Performance	Safety Certified/ Certifiable	Remarks
CUDA	Compute	Yes (NVIDIA)	++++	++++	No	
OpenCL	Compute	No	+++	++++	No	Limited on NVIDIA
OpenMP	Compute	No	+++++	+	No but	Limited support on embedded GPUs
HIP	Comput					Only on few AMD, Unibap
OpenACC	Comput					No support on embedded GPUs
Vulkan	Compute/Graphics	No	++	++++	No	
Brook Auto/SC	Compute	No	++++	++	Yes	Academic, open source
OpenGL SC 2	Graphics but solutions exist for compute too	No	+	++	Yes	Few sw vendors/devices
Vulkan SC	Compute/Graphics	No	++	++++	Yes	Standard not available yet
ComputeCore	library	Yes (CoreAVI)	++++++	++++	Yes	

Only CUDA/OpenCL explored in this study, due to their universal availability

Space Software Survey

- Theoretical analysis of algorithms found in several space domains [1][2]
- Most of existing space algorithms are a good fit for the GPU programming model
  - Especially on-board image-processing
- Some algorithms with several dependencies have been initially identified as not good candidates
  - Experimental evaluation later has shown speedup

Lesson 1: Modern GPUs can accelerate a wide range of existing and future space algorithms

Lesson 2: Porting an algorithm to a GPU is the only reliable way to find out whether it gets speedup

[1] GPU4S: Embedded GPUs for Space, DSD 2019
[2] GPU4S: Embedded GPUs for Space – Latest Project Updates, MICPRO 2020

Embedded GPUs Hardware Survey

- GPU Taxonomy defined [1][2]: Low and High End GPUs, COTS, Soft-GPUs, Many Cores and High-Level Synthesis
- Covered extensively the embedded GPU market
- NDA agreements signed with several European IP vendors
  - Challenge No 1: no vendor shares the price or any non-public information (area, power, number of gates) about their product before an upfront commitment
  - P… w…

Lesson3: There is no easy way to cost-effectively select the most appropriate GPU IP from several vendors without detailed tests.

[1] GPU4S: Embedded GPUs for Space, DSD 2019
[2] GPU4S: Embedded GPUs for Space – Latest Project Updates, MICPRO 2020

Embedded Hardware Survey

- IP vendors
  - Challenge No 2: No FPGA prototype without considerable costs
  - Vendor development/integration costs
  - Expensive equipment: special FPGAs with cost ~$50K
  - Reduced configurations, slow simulations

Lesson 4: Similar publicly funded projects in the future require considerable budget provisioned for FPGA prototypes of commercial IP designs

[1] GPU4S: Embedded GPUs for Space, DSD 2019
[2] GPU4S: Embedded GPUs for Space – Latest Project Updates, MICPRO 2020

Embedded Hardware Survey

- Open Source IP GPUs
  - Limiting capabilities, only GPU subsets, e.g. only integer operations
  - Mostly non-commercially friendly licensing e.g. GPLv3
- Situation has changed recently with RISC-V and Open Hardware movements
  - More capable and commercially viable hardware designs and licenses
  - Fully open source EDA options: OpenLane, SkyWater 130nm PDK
  - Identified and started evaluating options

Lesson 5: Existing open-source GPU designs cannot be used for our purpose.

Lesson 6: The RISC-V and Open Hardware movement can create opportunities for a commercially-friendly open source space GPU

[1] GPU4S: Embedded GPUs for Space, DSD 2019
[2] GPU4S: Embedded GPUs for Space – Latest Project Updates, MICPRO 2020

Embedded Hardware Survey

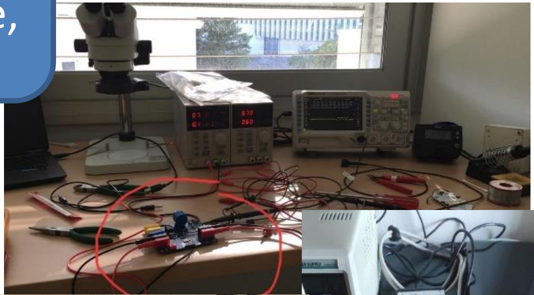
- COTS
  - NVIDIA
    - Higher performance
    - Used in rugged products
    - CUDA: Popular/Proprietary
    - Shorter product market availability window (7y)
  - AMD
    - Used in rugged products
    - Longer market availability (10y)
    - Better radiation properties
    - More open ecosystem (hardware and software)
    - Lower performance

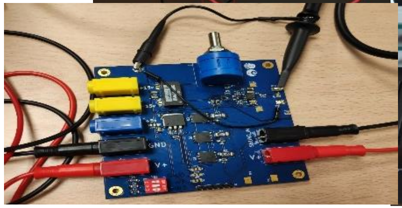
Lesson 8: NVIDIA GPUs offer higher performance but AMD better properties for space

Embedded GPU Benchmarking: HW

- Several selected candidate devices
  - NVIDIA Xavier, TX2
  - ARM Mali-G72 (HiKey 970)
  - AMD Embedd
- Difficulties to setu
  - OS, GPU driv
  - Delays in procurement
- Manufactured a PCB for the board measurements (CCN1)
- In total used 3 different ways to measure power

Lesson 9: NVIDIA's vertical integration results in tighter control of product releases in both hardware and software, but hw/sw is closed





Embedded GPU Benchmarking: SW selection

- Lack of benchmarks for Space
  - Proprietary code, export restrictions
- Lack of GPU benchmarks for critical systems
- Definition of an open source GPU Benchmark suite: GPU4S Bench[1]
  - Building blocks from many domains identified in the space survey
  - ESA GPL-3 compatible license, released together with OBPMark [2]

Lesson 10: Complex Space software is subject to restrictions

Lesson 11: Open Source benchmarks are required to maximise benefit from public funding

[1] GPU4S Bench: Design and Implementation of an Open GPU Benchmarking Suite for Space On-board Processing: https://www.ac.upc.edu/app/research-reports/public/html/research_center_index-CAP-2019,en.html
[2] OBPMark (On-Board Processing Benchmarks) – Open Source Computational Performance Benchmarks for Space Applications, OBDP 2021, http://OBPMark.org

Embedded GPU Benchmarking: GPU4S Bench Overview

- Identified building blocks and the domains they represent

Domains / Building Block	Compression	Vision Based Navigation	Image Processing	Neural Network Processing	Signal Processing
Fast Fourier Transform			GENEVIS		ADS-B, NGDSP
Finite Impulse Response Filter					ADS-B, NGDSP
Integer Wavelet Transform	CCSDS 122				
Pairwise Orthogonal Transform	CCSDS 122				
Predictor	CCSDS 123				
Matrix computation		GENEVIS (Solver)		Image classification	
Convolution Kernel		OpenCV	GO3S,GENEVIS	Image classification	
Correlation		OpenCV	GO3S,GENEVIS		ADS-B
Max detection			GO3S	Image classification	ADS-B
Synchronization mechanism		GENEVIS	EUCLID NIR, GO3S	TensorFlow	ADS-B, NGDSP
Memory Allocation		CERES Solver , OpenCV	EUCLID NIR, GO3S	TensorFLow	ADS-B, NGDSP

CCN1

- Complex application: Image recognition pipeline, based on CIFAR-10
- Euclid NIR (Near InfraRed)

[1] GPU4S Bench: Design and Implementation of an Open GPU Benchmarking Suite for Space On-board Processing: https://www.ac.upc.edu/app/research-reports/public/html/research_center_index-CAP-2019,en.html

OBPMark Overview and implementation status

src/1.1-image/	Done in CCN1, including OpenMP
src/1.2-radar/	To be defined (CCN2)
src/2.1-data_compression/	Done (+omp), to be part of CCN2
src/2.2-image_compression/	Done (+omp), to be part of CCN2
src/2.3-hyperspectral_compression/	Partially done (+omp), part of CCN2
src/3.1-aes_compression/	To be defined (CCN2)
src/4.1-fir_filter/	Done (+omp)
src/4.2-fft/	Done (+omp)
src/5.1-object_detection/	To be defined (CCN2)
src/5.2-cloud_screening/	To be defined (CCN2)

[1] OBPMark (On-Board Processing Benchmarks) – Open Source Computational Performance Benchmarks for Space Applications, OBDP 2021, http://OBPMark.org

Embedded GPU Benchmarking: GPU programmability

- Several benchmark versions: naïve, optimized, libraries
  - CUDA, HIP, OpenCL
    - CUDA/HIP versions faster implementation
    - CUDA better support in terms of development tools and libraries
  - Run into issues with OpenCL
- Power consumption: TDP <= 15W confirmed experimentally

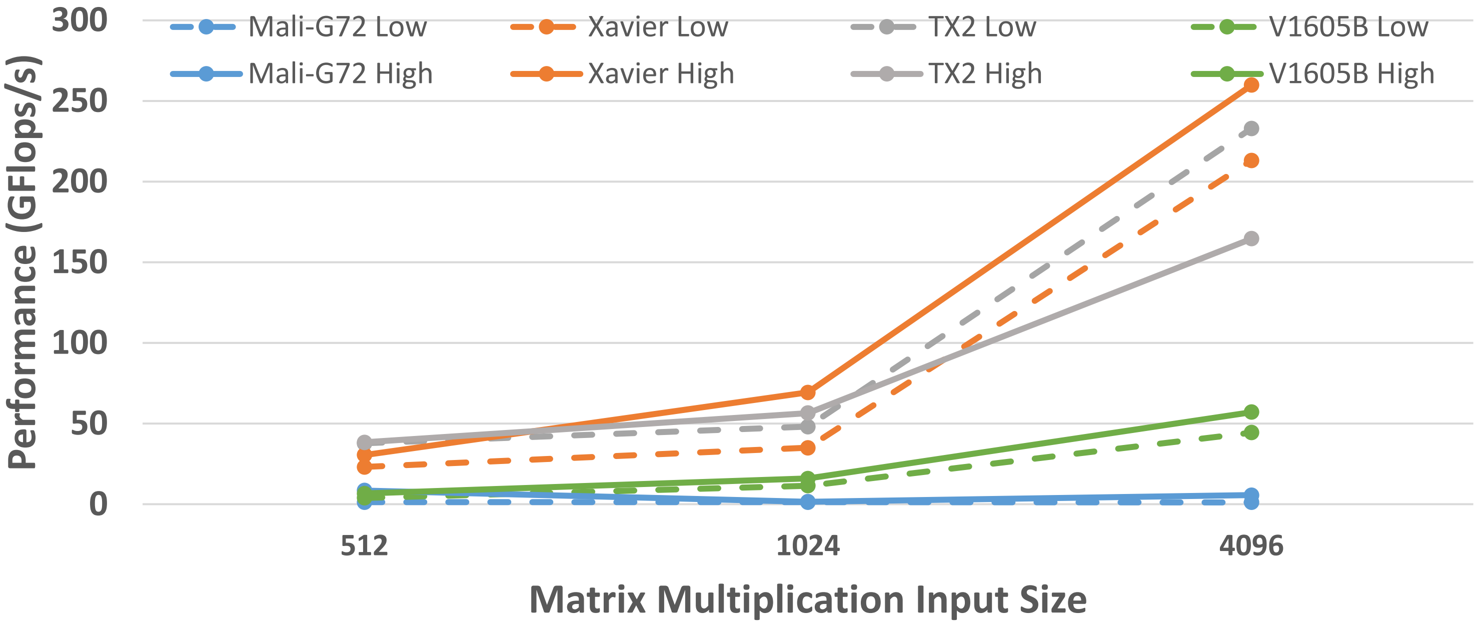
Lesson 12: Embedded GPUs comply with on Board power requirements

Lesson 13: CUDA and HIP offer easier programmability than OpenCL

[1] GPU4S Bench: Design and Implementation of an Open GPU Benchmarking Suite for Space On-board Processing: https://www.ac.upc.edu/app/research-reports/public/html/research_center_index-CAP-2019,en.html
[2] OBPMark (On-Board Processing Benchmarks) – Open Source Computational Performance Benchmarks for Space Applications, OBDP 2021, http://OBPMark.org

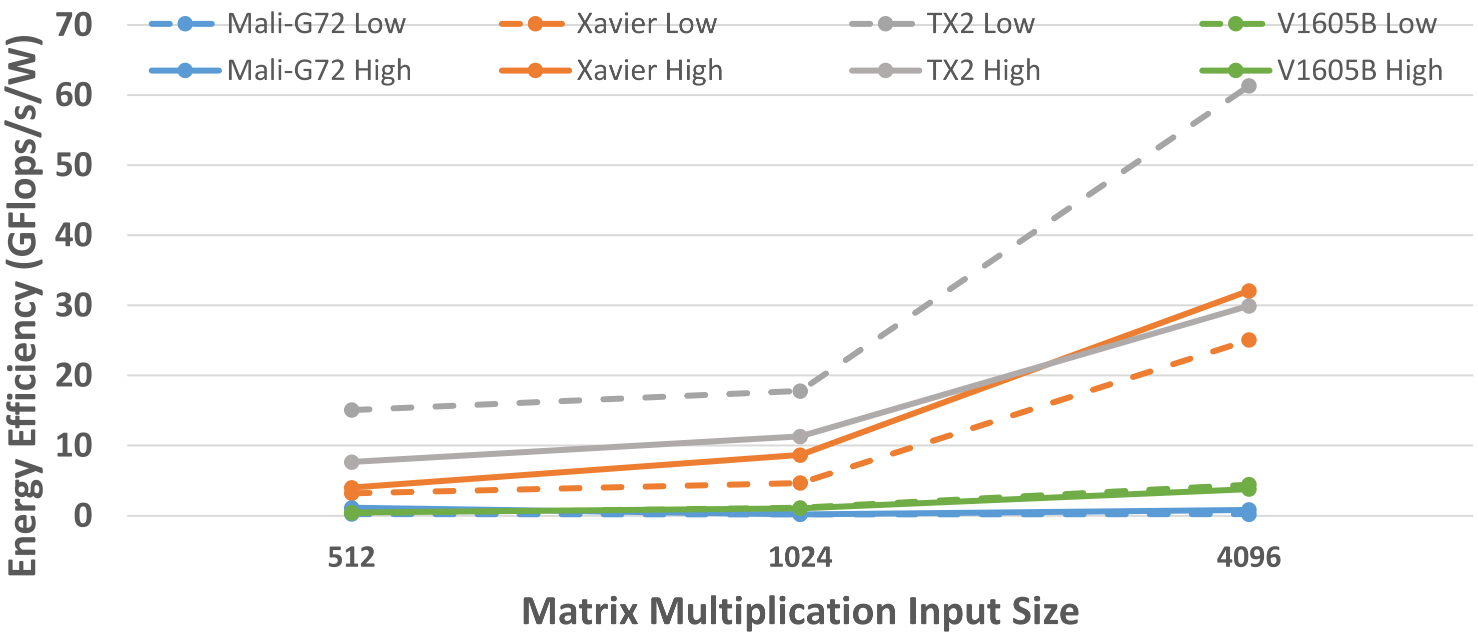
Embedded GPU Benchmarking: Results (Performance)

- NVIDIA platforms dominated in terms of performance and energy efficiency
  - Results depend on input size, and benchmark
- Library versions are not always faster



Embedded GPU Benchmarking: Results (Energy Efficiency)

- NVIDIA platforms dominated in terms of performance and energy efficiency
  - Results depend on input size, and benchmark
- Library versions are not always faster



Embedded GPU Benchmarking: Results

- NVIDIA platforms dominated in terms of performance and energy efficiency
  - Results depend on input size, and benchmark
- Library versions are not always faster

Lesson 14: Vendor optimised libraries have a large initialisation cost, so they are not always the best choice, but it depends on the application scenario

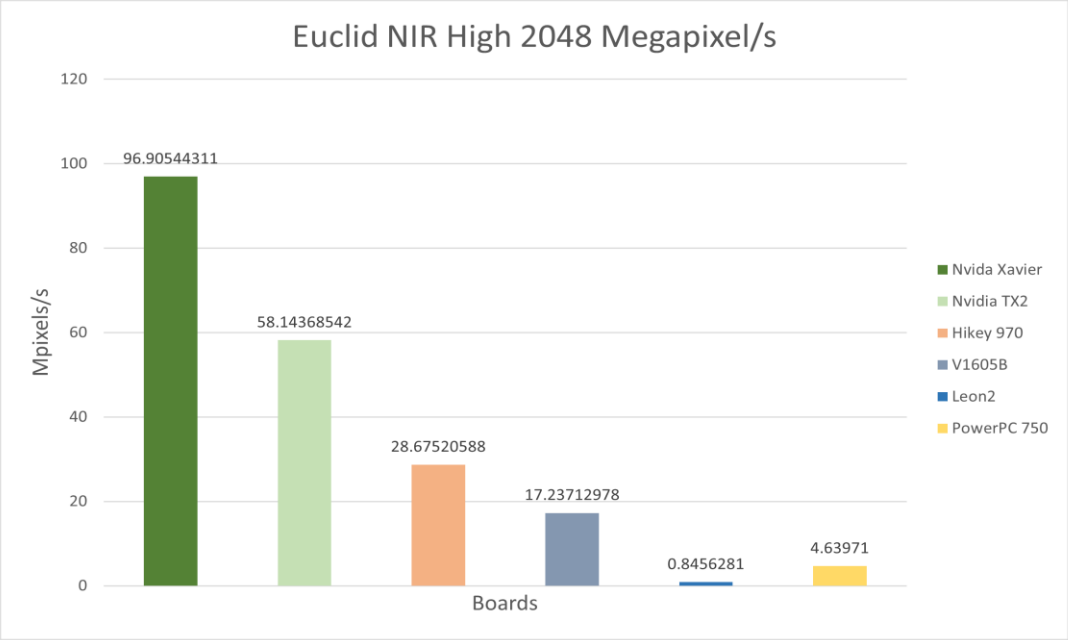
Lesson 15: It is possible to obtain high performance with reasonable GPU development effort

Lesson 16: Performance can only be assessed by actual implementation

[1] GPU4S Bench: Design and Implementation of an Open GPU Benchmarking Suite for Space On-board Processing: https://www.ac.upc.edu/app/research-reports/public/html/research_center_index-CAP-2019,en.html
[2] OBPMark (On-Board Processing Benchmarks) – Open Source Computational Performance Benchmarks for Space Applications, OBDP 2021, http://OBPMark.org

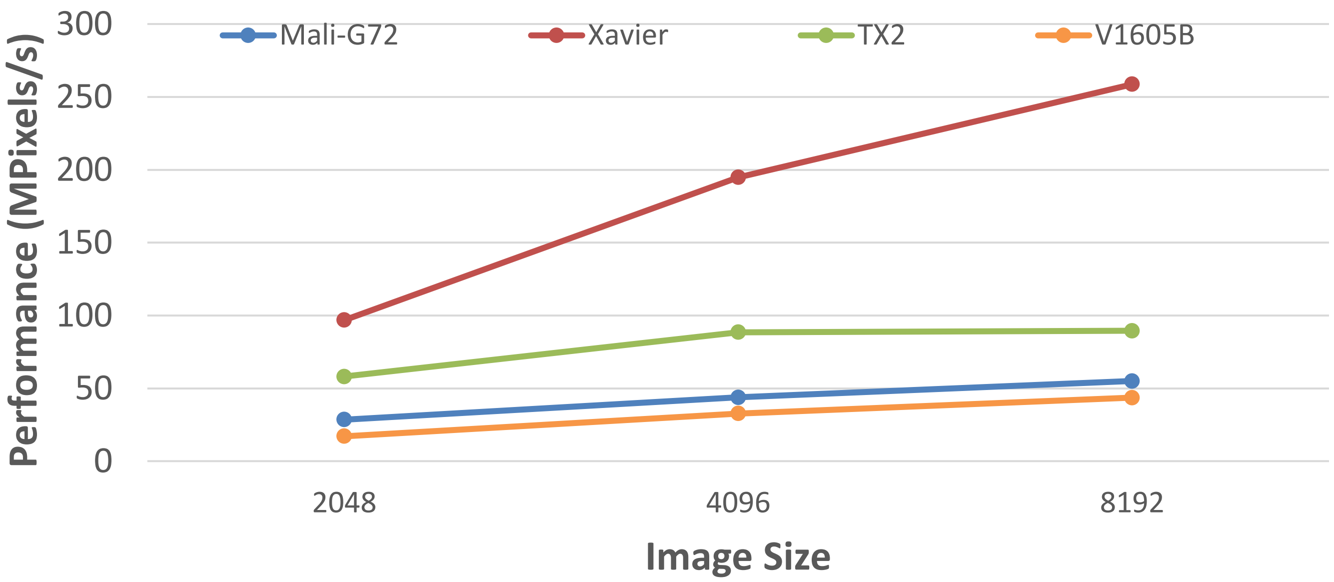
Embedded GPU Benchmarking: Complex applications / Demonstrator

- GPU4S: Inference application designed for the CIFAR-10 data set
- CCSDS Space compression standards [1]
- Ported Euclid NIR space case study to GPU [2][3]



Embedded GPU Benchmarking: Complex applications / Demonstrator

- GPU4S: Inference application designed for the CIFAR-10 data set
- CCSDS Space compression standards [1]
- Ported Euclid NIR space case study to GPU [2][3]



Embedded GPU Benchmarking: Complex applications / Demonstrator

- GPU4S: Inference application designed for the CIFAR-10 data set
- CCSDS Space compression standards [1]
- Ported Euclid NIR space case study to GPU [2][3]
  - Significant speedups compared to existing space processors
    - 2 orders of magnitude in small sizes (2048x2048)
  - Benefit increases with input size

Lesson 17: GPUs can significantly accelerate complex space processing compared to other technologies

[1] An On-board Algorithm Implementation on an Embedded GPU: A Space Case Study, DATE 2020

[2] I. Rodriguez, Master Thesis, https://upcommons.upc.edu/handle/2117/344892

[3] Parallelisation of On-Board CCSDS Compressors: a Benchmarking Approach, ESA OBPDC 2020

Radiation Effects and Future Adoption

- Radiation studies are required before GPU adoption in space
- Preliminary results available from other projects
- Nanosatellites and technology demonstration missions
  - Mars Ingenuity (still working!), HyTI Cubesat, Φ-Sat-1, OPS-SAT 1+2 etc.
  - Promising results for radiation tolerance of COTS devices
- Similar issues faced in the automotive domain
  - Built-in reliability required for ASIL-D
  - Hardware [1] or software solutions [2]

Lesson 18: GPU reliability solutions from the automotive domain can be adopted for use in space

[1] High-Integrity GPU Designs for Critical Real-Time Automotive Systems, DATE'19
[2] Software-only Diverse Redundancy on GPUs for Autonomous Driving Platforms, IOLTS '19

Overview of Tested GPUs

Company	Component	Performance	Software Support	ECC protection	Radiation Results Available	Flight board available
NVIDIA	Xavier	Best	CUDA/Vulkan	Yes	Yes	No
NVIDIA	TX2	High	CUDA/Vulkan	Yes	Yes	Yes
AMD	V1605B	Medium/High	OpenCL/HIP/ Vulkan	Yes	Yes	Yes
ARM Mali-G72	HiKey 970 HiSilicon	Medium	OpenCL	No	No	No
AMD/Unibap	DDX-i5	Medium	OpenCL	Yes	Yes	Yes
AMD/Unibap	DDX-i10	Medium/High	OpenCL/HIP	Yes	Yes	Yes

Conclusions

- GPUs in general are a good fit for highly flexibly high-performance on-board processing
- Embedded COTS GPU SoCs have high-performance and are suited for New Space applications where component qualification is not critical
  - GPUs can offer better time-to-market than FPGAs for New Space
- Reliability of COTS GPUs can be increased through HW/SW approaches
  - Radiation testing of COTS GPUs is required
  - FDIR software is recommended
- GPUs could be introduced in institutional missions, but would require significant investment
  - A strong industrial consortium, including the GPU IP provider is necessary
  - Open source fault-tolerant GPU IPs with commercial-friendly licensing could be a way forward

Acknowledgements

- ESA AO/1- 9010/17/NL/AF
- Spanish Ministry of Economy and Competitiveness (MINECO) under grants PID2019-107255GB and FJCI-2017-34095
- HiPEAC Network of Excellence, Technology Transfer Award 2019
- RISC-V Foundation Educator of the Year Award 2019
- Airbus Defence and Space, Getafe, Spain, TANIA-GPU ADS (E/200)
- CoreAVI
- Unibap
- UP2DATE H2020, European Community's Horizon 2020 Programme (H2020-ICT-2019-2) grant agreement 871465

www.bsc.es

GPU4S (GPUs for Space):

Are we there yet?

Leonidas Kosmidis, Iván Rodriguez, Alvaro Jover-Alvarez, Guillem Cabo, Sergi Alcaide, Jérôme Lachaize, Olivier Notebaert, Antoine Certain, David Steenari

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
Universitat Politècnica de Catalunya
Airbus Defence & Space
esa

15/06/2021 OBDP 2021