UNIVERSIDADE D
COIMBRA

Pedro Fonseca da Cunha Roque Póvoa

Combining HPC and low-power systems for data-intensive-acquisition sensors in space missions

Dissertação no âmbito do Mestrado Integrado em Engenharia Eletrotécnica e de Computadores, na especialização de Área da Robótica, Controlo e Inteligência Artificial, orientada pelo Professor Doutor Gabriel Falcão Paiva Fernandes e pelo Professor Doutor Rui Miguel Curado Silva e apresentada ao Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade de Ciência e Tecnologias da Universidade de Coimbra.

Julho de 2023

1 2  9 0

UNIVERSIDADE D
COIMBRA

Combining HPC and low-power systems for data-intensive-acquisition sensors in space missions

Pedro Fonseca da Cunha Roque Póvoa

Coimbra, July of 2023

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Combining HPC and low-power systems for data-intensive-acquisition sensors in space missions

Dissertation supervised by Professor Doctor Gabriel Falcão Paiva Fernandes and Professor Doctor Rui Miguel Curado Silva and submitted to the Electrical and Computer Engineering Department of the Faculty of Science and Technology of the University of Coimbra, in partial fulfilment of the requirements for the Degree of Master in Electrical and Computer Engineering.

Supervisors:
Prof. Dr. Gabriel Falcão Paiva Fernandes
Prof. Dr. Rui Miguel Curado Silva

Jury:
Prof. Dr. Vítor Manuel Mendes da Silva
Prof. Dr. Jorge Nuno de Almeida e Sousa Almada Lobo
Prof. Dr. Gabriel Falcão Paiva Fernandes

Coimbra, July of 2023

Agradecimentos

Gostaria expressar os meus sinceros agradecimentos a todos aqueles que contribuíram de forma significativa para a realização desta dissertação. O meu percurso académico não teria sido possível sem o apoio e incentivo de várias pessoas e gostaria de lhes expressar os meus agradecimentos.

Em primeiro lugar, um grande obrigado aos meus orientadores, o Professor Doutor Gabriel Falcão e o Professor Doutor Rui Curado, pela orientação, ajuda e paciência ao longo de todo o processo. As suas experiências e conhecimento foram fundamentais para o desenvolvimento deste trabalho, e estou extremamente grato por todo o apoio prestado.

Agradeço também à minha família pelo seu amor, apoio incondicional e compreensão durante esta jornada. As suas palavras de encorajamento e incentivo foram essenciais para a minha motivação e sucesso.

Não poderia deixar de mencionar os meus amigos e colegas, que me apoiaram ao longo deste caminho. As nossas discussões, troca de ideias e colaboração foram fundamentais para o desenvolvimento das minhas ideias e crescimento pessoal. Obrigado por estarem sempre ao meu lado e tornarem esta jornada mais especial.

Por fim, gostaria de agradecer a todos aqueles que me acompanharam nesta caminhada. Cada palavra de encorajamento, cada conselho e cada gesto de apoio foram importantes para a minha conquista. A todos, o meu mais profundo obrigado.

Que este trabalho possa contribuir para o avanço do conhecimento na área e inspirar futuros estudantes e investigadores.

Um obrigado a todos!

Esta dissertação foi desenvolvida em colaboração com:

1 2 9 0 UNIVERSIDADE D COIMBRA
deec.uc DEPARTAMENTO DE ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES
df.uc DEPARTAMENTO DE FÍSICA FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE DE COIMBRA
it instituto de telecomunicações
LIP LABORATÓRIO DE INSTRUMENTAÇÃO E FÍSICA EXPERIMENTAL DE PARTÍCULAS partículas e tecnologia
THOR TGF and High energy astrophysics Observatory for gamma Rays

Abstract

Data-intensive acquisition sensors involved in space missions produce a substantial number of data samples constrained in time and require high-performance computing (HPC) systems to process information in real-time. They present some challenges like intensive calculations and tight timings. The use of multi-core systems can overcome those obstacles by performing several tasks in parallel and dynamically handling workloads. The objective of this study is to develop a programming framework for a high-performance and low-power multicore system capable of efficiently handling the vast amount of data obtained from the data acquisition of the gamma-ray detector under investigation. This system must also adhere to various constraints imposed by the space environment, like limited power consumption, the inability to access the system during the experiment, and limitations on heat dissipation. Another goal of this research is to deploy the designed system in the THOR-SR experiment, which will be conducted aboard the Space Rider, a new recoverable space vehicle developed by the European Space Agency (ESA). The Space Rider is scheduled to fly for a period of two months. During this mission, the high-performance and low-power multicore system will handle all the data originating from the data acquisition of the gamma-ray detector. The system's design will enable it to meet tight space environment constraints, including limited power consumption, the need for autonomous operation during the experiment, and the ability to dissipate heat effectively.

Keywords : High-performance computing (HPC), multi-core, real-time, Graphics Processing Unit (GPU).

Resumo

Os sensores de alta aquisição de dados envolvidos em missões espaciais produzem um número substancial de amostras de dados, as quais estão frequentemente sujeitas a restrições temporais e que requerem processamento em tempo real. Para enfrentar esses desafios apresentados, tais como cálculos intensivos e prazos rigorosos, os sistemas de computação de alto desempenho (HPC - high-performance computing) são essenciais. Uma abordagem para lidar com esses obstáculos é a utilização de sistemas multicore, que podem executar várias tarefas em paralelo e lidar dinamicamente com as cargas de trabalho. O objetivo deste estudo é desenvolver um programa para um sistema multicore de alto desempenho e baixo consumo de energia, capaz de lidar de forma eficiente todos os dados provenientes da aquisição de dados do detetor de raios gama em estudo. Esse sistema também deve obedecer a diversas restrições impostas pelo ambiente espacial. Essas restrições incluem consumo de energia limitado, ausência de acesso ao sistema durante a experiência e limitações de dissipação de calor. A outra meta desta pesquisa é utilizar o sistema projetado na experiência THOR-SR, que será realizado a bordo do Space Rider, um novo veículo espacial reutilizável desenvolvido pela Agência Espacial Europeia (ESA). O Space Rider está programado para voar durante um período de dois meses. Durante essa missão, o sistema multicore de alto desempenho e baixo consumo de energia será responsável por lidar com todos os dados provenientes da aquisição de dados do detetor de raios gama. A experiência THOR-SR tem como objetivo estudar radiação orbital, bem como flashes de raios gama terrestres e emissões de fontes de raios gama para futuros telescópios espaciais de astrofísica de alta energia.

Palavras − chave : Computação de Alto Desempenho (HPC), multicore, tempo real, Unidade de Processamento Gráfico (GPU)

Contents

List of Acronyms

ASIC	Application-specific integrated circuit
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture
ESA	European Space Agency
FLOPS	FLoating-point Operations Per Second
FPGA	Field-Programmable Gate Array
FToA	Fast Time of Arrival
GPU	Graphics Processing Unit
GUI	Graphical User Interface
HLS	High-Level Synthesis
HPC	High Performance Computing
LIP	Laboratório de Instrumentação e Física Experimental de Partículas
OBC	OnBoard Computer
PRODEX	Programme de Dévéloppement d'Éxperiences Scientifiques
RAM	Random Access Memory
RTL	Register Transfer Level
SAXPY	Single-precision A.X Plus Y
SIMT	Single Instruction Multiple Thread
SM	Streaming Multiprocessor

TGF Terrestrial gamma-ray flashes

THOR-SR TGF and High-energy astrophysics Observatory for gamma-Rays on board the Space Rider

ToA Time of Arrival

ToT Time over Threshold

VHDL VHSIC Hardware Description Language

VHSIC Very High Speed Integrated Circuits

List of Figures

List of Tables

1

Introduction

As the technology evolved, the data became larger, faster and more sophisticated. As a result, existing platforms either became slow or were unable to keep up with necessary the speed for processing information. New processing platforms had to emerge that could keep up with this evolution. Thus high-performance computing (HPC) systems appeared. They are capable of processing data faster and more efficiently from an energetic perspective, which is timely and important at this stage of development. However, the use of such systems proposes requirements that must be met, such as the development of parallel programming skills, latency, power and temperature limits [1].

It is important to use this kind of platform to follow up on the evolution of the data coming from the sensors, to also manage quickly the complexity of that type of information and to overpass all the constraints that space environments imply.

Initially, Application-Specific Integrated Circuits (ASIC) and later Field-Programmable Gate Array (FPGA) have been used in space missions for processing all the information. Nowadays, Graphics Processing Units (GPUs) have become common platforms because they can perform tasks in parallel [2]. As these systems can perform several tasks quickly, they come with the cost of great energy demands [3]. For that reason, this work aims to develop HPC and low-power systems able to monitor and process the information coming from a detector in a space environment, i.e., being exposed to space operation constraints.

Space missions' planning defines a list of requirements such as the limitation of the weight and size of the components for the flight, the existence of maximum power consumption and power supply and restrictions of heat dissipation.

This study will evaluate the performance and low-power consumption of the platform *Jetson AGX Xavier* from *NVIDIA corporation* with the use of the roofline model and benchmarks to achieve high-performance workloads.

The system will be part of the LIP's (Laboratório de Instrumentação e Física Experimental de Partículas) THOR-SR experiment, which will be operated onboard the Space Rider [4], a new ESA recoverable space vehicle. This flight will take a couple of months in space. This experiment aims to study orbital radiation as well as terrestrial gamma-ray flash

radiation. After the flight, the results obtained will provide important conclusions for future high-energy astrophysics space telescopes.

Figure 1.1 shows an idea of the focus of the dissertation and an overview of the THOR-SR space experiment. The figure represents the space vehicle exposed to celestial gamma-ray radiation during the experiment. The sensors will capture that radiation and the onboard computer will process that information and store it. Subsequently, it will have the possibility to send it to the ground station or hold it until the vehicle return to the earth.

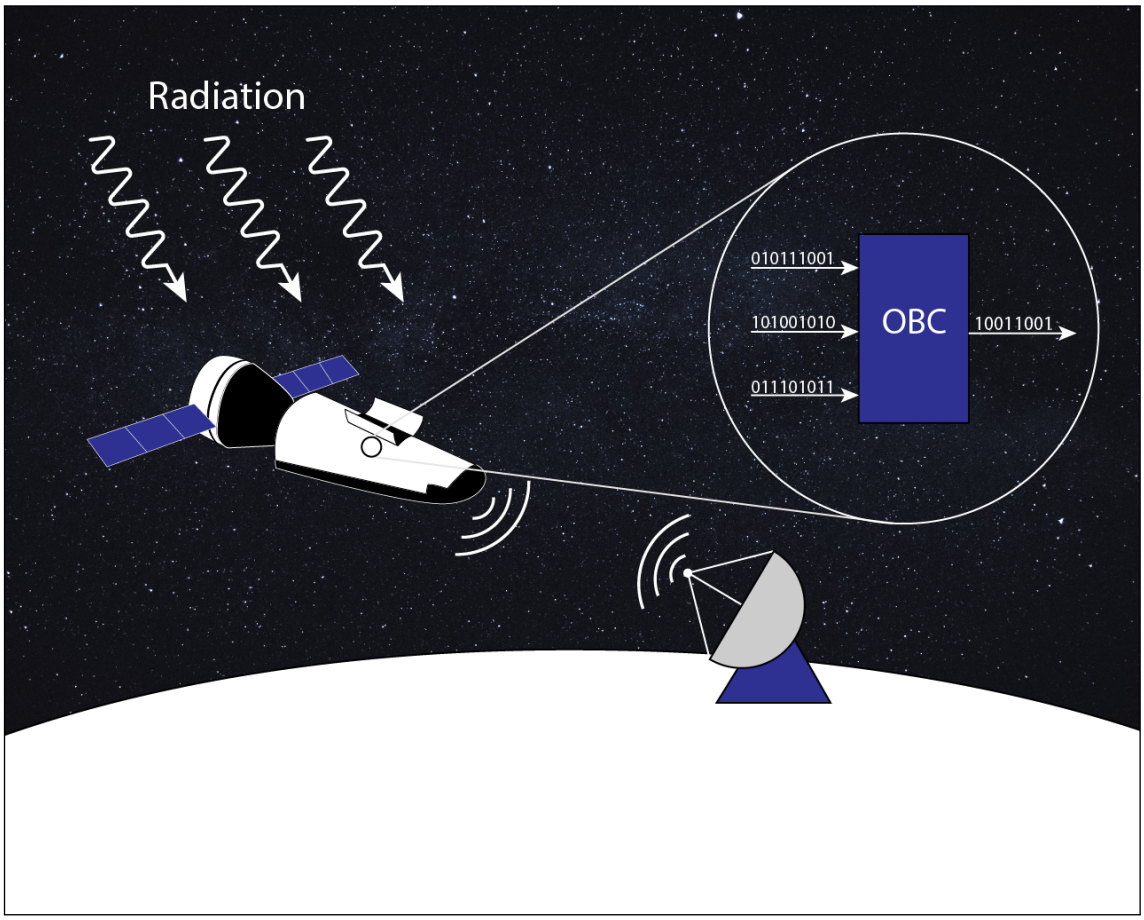

Figure 1.1: Simbolic representation of dissertation study.

1.1 Objectives and Contributions

This thesis proposes:

- To address the challenges of real-time data processing from data-intensive acquisition sensors in space missions.
- To use multi-core systems to overcome these challenges by performing multiple tasks in parallel and dynamically handling workloads.

- To design and implement a system that combines high-performance computing (HPC) and low-power systems to handle all the data from the data acquisition sensor.

A posterior, the following contributions are expected from this work:

- The final system is intended to be operated onboard the Space Rider program, a new recoverable space vehicle developed by the European Space Agency (ESA) [4].
- The data received from orbital radiation and terrestrial gamma-ray flashes radiation and gamma-ray sources' emissions will be part of a study to contribute to the advancement of high-energy astrophysics space telescopes.
- A paper entitled *TGF and High-energy astrophysics Observatory for gamma-Rays on board the Space Rider Mission* will be submitted to the *Advances in Space Research.*
- PRODEX project approved by ESA.

2

Background and state-of-the-art

In today's world, managing large amounts of data and executing complex tasks in real-time has become increasingly important. However, many systems are becoming more limited in terms of power consumption, as in the case of Space Rider, and there is a need to provide maximum computing power in such power-constrained context. This chapter discusses two examples of HPC for power-limited systems, FPGA and GPU, and compares their advantages and disadvantages. While FPGA is more scalable and occupies less area, GPU is more cost-efficient and powerful. Despite the debate over which system is better, both are becoming popular choices for space applications due to their ability to adapt to changing requirements, high performance, and ability to withstand harsh environmental conditions.

2.1 High-Performance Computing

The field of High-performance computing (HPC) combines various elements such as hardware technologies, architecture, operating systems, programming tools, software, and end-user problems and algorithms. Its aim is to achieve the highest computing capability possible with current technology. This interdisciplinary field makes use of electronic digital devices called "supercomputers" to carry out various computational tasks or "applications" as quickly as feasible. The term "supercomputing" is often used interchangeably with HPC and refers to the process of executing an application on a supercomputer. [5]

HPC processes enormous amounts of information (big data) and solves complex problems at incredibly fast rates using parallel processing in powerful computers. Compared to the fastest commercial desktop, laptop, or server computers, HPC systems often operate at speeds that are more than one million times faster [6].

HPC systems are important in a wide range of scientific and industrial fields because they enable researchers to perform complex simulations and calculations that would otherwise be infeasible or take prohibitively long to complete using traditional computing resources.

These systems are characterized by their ability to process large amounts of data quickly and efficiently, making them well-suited for tasks such as protein folding simulations,

climate and financial modelling.

HPC systems also have a wide range of practical applications in industry, such as in the design and optimization of new products, the simulation of industrial processes, and the analysis of large datasets. This can help companies to make more informed decisions, improve their products and services, and increase their competitiveness in the marketplace.

New issues in systemic resilience, energy efficiency, and software complexity have emerged as scale and complexity in HPC have increased [7].

Two unique architectural ideas are emerging, both of which are motivated by the difficulties of energy efficiency [7]. The first is characterized by a few heterogeneous nodes with multicore CPUs and accelerators, while the second is characterized by a much higher number of homogeneous nodes.

In space missions there are limitations in energy and power consumption and HPC systems have to fulfil those requirements with the highest computing performance possible.

2.2 HPC for power-limited systems

Power-limited systems are becoming increasingly common in present times. These systems have a constraint to consume less power than traditional systems while requiring them to provide the necessary functionality to achieve the objective. These systems can be found in a wide range of applications, including mobile devices, wearables, HPC platforms, and the Internet of Things (IoT) [8].

One of the key technologies used in power-limited systems is energy-efficient hardware design. This can be achieved through the use of specialized low-power integrated circuits and microprocessors [9]. Additionally, the use of energy-efficient algorithms and software can also help reduce power consumption.

Another technique used in power-limited systems is dynamically scaling the voltage and frequency to adjust the power consumption of the system in response to changes in workload. Power management can also involve the use of sleep modes, where the system enters a low-power state when not in use [10].

Two of the most widely used HPC platforms are FPGAs and GPU architectures. FPGAs are commonly used for power-limited real-time systems and projects where hardware configuration is subject to change and a circuit that can be adjusted to these changes is required [11]. The GPU, meanwhile, is most commonly used on computers for graphics and video rendering and is also known for its capabilities in gaming [12]. More recently, it has been adopted for performing scientific computing as well.

In areas such as image classification and bioinformatics which is computationally costly and have large amounts of data, FPGA and GPU can be used to increase the performance

of an implementation of that kind [13].

2.2.1 FPGA

FPGAs are power efficient due to a large number of transistors that have a small amount of leakage current. By customizing data paths to the needs of a particular algorithm or application, FPGAs provide chances for low-level fine-grained parallelism [14]. They are also flexible because it is possible to change the hardware configuration to suit the new circuit. This equipment is even, to some extent, scalable because it is possible to connect more FPGA chips to the system to increase the available computational resources [15][16]. The common programming method used in FPGAs is RTL, namely through VHDL and Verilog. It is also possible to use high-level synthesis (HLS) to program these devices. This makes them well-suited for use in applications such as scientific simulations and modelling, where high performance is required.

2.2.2 GPU

GPUs are more cost-efficient and energy-efficient than FPGAs because they are built to perform a specific task and most likely cannot be reusable for other purposes. This makes them powerful but in terms of hardware not very flexible. While in terms of programming the GPUs have more tools than the FPGAs.

The GPU is a component that is commonly used to create images in a frame buffer intended for output to a display device [17].

Unlike CPUs, GPUs have a highly threaded architecture and are more efficient at processing graphics data and complex algorithms. Between the GPU and CPU structures, most of the CPU area is in the controller and register, while GPU utilizes ALU (Arithmetic Logic Unit) for intensive and parallel processing of data [17].

Figure 2.1 represents the difference between CPU and GPU architectures. On the GPU, meanwhile, there are more areas dedicated to the ALU and fewer areas saved for control and cache than on the CPU.

Overall GPUs have a high-power consumption which involves that they cannot be installed in systems with constrained power, space or temperature requirements [15][16]. The programming language most used is C/C++, Java and Python.

2.2.2.1 Generic low-power GPU platforms

This section shows shortly three specifications examples of low power and HPC GPU boards, table 2.1.

All platforms are capable of working with different power consumption, which is a good point for projects that have low power requirements and also have a high computational performance which is great to perform several operations quickly. In this study, the chosen

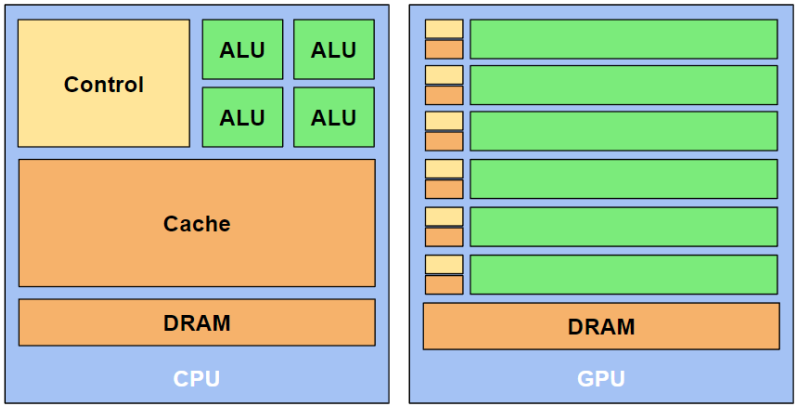

Figure 2.1: Differences between CPU and GPU architectures [17].

Table 2.1: Generic specifications of platforms. [18][19][20]

	Jetson AGX Xavier	Jetson TX2 NX	Jetson Nano
Year of production	2018	2017	2019
CPU	Octal-core NVIDIA Carmel ARMv8,2 2,26GHz	Dual-core Denver 2 64-bit CPU and Quad-core Arm Cortex-A57 MPCore processor 2GHz	Quad-core ARM Cortex-A57 MPCore processor 1,43GHz
GPU	512-core Volta 1377MHz and 64 Tensor cores	256-core NVIDIA Pascal GPU 1,3GHz	128-core NVIDIA Maxwell architecture GPU 921MHZ
Memory	32GB 256-bit LPDDR4x 136,5GB/s	4GB 128-bit LPDDR4 51,2GB/s	4GB 64-bit LPDDR4 25,6GB/s
Power	10/15/30W	7,5/15W	5/10W
Performance	32 TeraOPS	1,33 TeraOPS	472 GFLOPs

platform is the Jetson AGX Xavier because it has more cores on GPU and CPU, with 512 CUDA cores and 8 CPU cores, and faster clock frequencies.

2.2.3 GPU vs FPGA Timeline

The following images were taken from [16], where it can be seen in figure 2.2 the processing efficiency for power and cost and in figure 2.3 the comparison in different areas of both platforms. The data from both images come from platforms up to 2016.

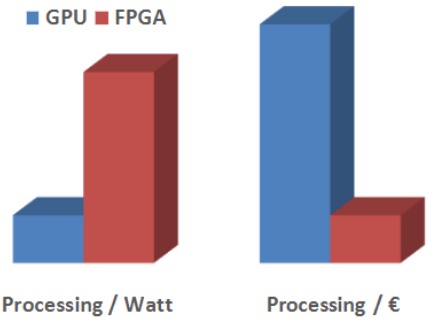

Figure 2.2: Overall Processing Efficiency [16].

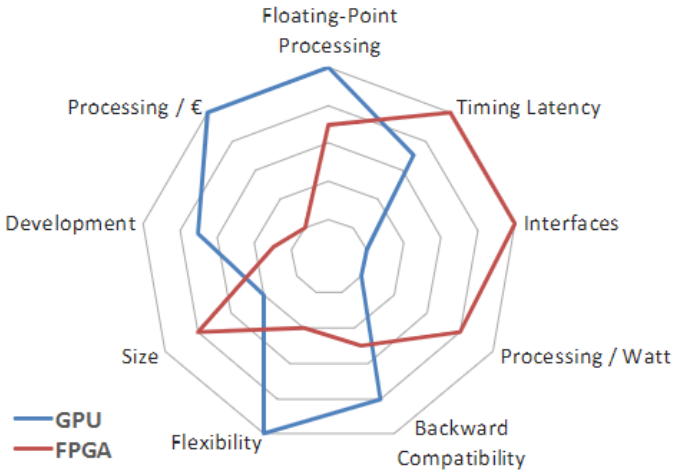

Figure 2.3: GPU vs FPGA Overall Qualitative Comparison [16].

Nowadays, the difference between FPGA and GPU was starting to narrow. These platforms race each other to reach better performance on target workloads. Both platforms can be used for example for image processing [21], deep learning [22] and artificial intelligence [17][23]. They can achieve high performance with low power consumption for any of these tasks.

There is no common opinion about what is the better platform and articles usually take one side.

2.2.4 Systems in Space context

The system most used was the ASIC which is a customized system which has only one particular use. Experiments like POLCA [24], VZLUSAT-1 [25] and Fermi Gamma-ray Space Telescope [26] utilized ASICs for handling data.

Either FPGAs or GPUs have been increasingly used in space contexts for example New Horizons mission [27] that uses FPGAs for telecommunications, due to their ability to adapt to changing requirements and their high performance. This makes them well-suited for use in space applications, where the ability to adapt to changing needs is important.

The Space Rider planned to recover their space vehicle and it is important that THOR-SR's system is capable to change and adapt for future projects, one more reason to use these systems.

Both FPGAs and GPUs are capable of performing high-speed data processing and analysis, which is crucial for space missions that require real-time data processing [28] [29].

Another advantage of using FPGAs and GPUs in space context is their ability to withstand harsh environmental conditions. Both platforms are radiation-hardened, meaning they can withstand high levels of radiation, which is a common problem in space environments [30].

2.3 THOR-SR project

The THOR-SR project is part of the LIP's involvement in the development of future missions in high-energy astrophysics, focusing on the concept of all-sky monitoring instruments. These instruments, utilizing CdTe technology, detect and analyze gamma-ray emissions from the most intense celestial sources in the sky. They perform various measurements such as spectroscopy, imaging, polarimetry, and time variability analysis. Within this framework, THOR-SR serves as a pathfinder mission, aiming to validate the capabilities of a CdTe-based instrument for all-sky mode measurements of gamma-ray bursts and bright gamma-ray emitters, including the Crab Nebula pulsar. Beyond that, it monitors Earth's terrestrial gamma-ray flashes (TGFs) emissions and the space radiation environment in low Earth orbit.

The Space Rider offers a valuable opportunity to test scientific instruments' operations in space, reaching high technology readiness levels of over 7. The project focuses on addressing technological and scientific objectives:

- High-energy astrophysics: The project aims to study the spectroscopy, imaging, polarimetry, and time variability of the most intense gamma-ray sources in the sky, such as the Crab Nebula and gamma-ray bursts. Notably, the all-sky operational mode of the instrument as a polarimeter represents a pioneering achievement in a space mission.
- Terrestrial gamma-ray flashes (TGFs) science: By recording TGF emissions using spectroscopy, imaging, and time variability analysis, the project aims to assess the potential of CdTe pixelated detectors as TGF monitors. The long-term objective is to develop a commercial product for aviation safety, providing alerts and risk assessments related to TGF emissions for passengers and crew members.
- Space radiation in low Earth orbit and space weather: The project's focus includes the recording of proton and electron fluxes around the Van Allen belts, examining particle fluxes in relation to energy and particle fluxes concerning time variability in low Earth orbit.

2.4 Summary

To handle big information and perform several tasks quickly HPC is one solution. It provides high parallel processing power. Power-limited systems are becoming increasingly more common as we strive to use this system in constrained environments. These systems are designed to consume less power than traditional systems while still providing the necessary functionality.

FPGA and low-power GPUs are examples of HPC used in power-limited systems that can be applied to real-time projects. FPGA is reconfigurable and scalable because it is possible

to increase the resources available by connecting more FPGA chips. On the other hand, GPU is more cost-efficient and energy-efficient but in terms of hardware not very flexible. Back in 2016 GPU and FPGA were more distinct. FPGA had more processing/watt and less processing/euro. Nowadays these differences have narrowed.

Because of the ability to adapt to changing requirements, high performance, and ability to withstand harsh environmental conditions, FPGAs and GPUs are becoming increasingly popular choices for space applications. They are well-suited for use in space missions, where power consumption and data processing are critical considerations.

THOR-SR is a high-energy astrophysics pathfinder mission with a scientific payload based on CdTe detectors on board the Space Rider. The mission addresses gamma-ray astrophysics, space weather, and TGF monitoring through the performance of spectroscopy, imaging, time variability, and polarimetry measurements.

3

GPU system

This chapter describes the GPU system that will be used in this study. It focuses on the platform that will be the object of study and will delve into the methods and practices of programming this type of device. More specifically, it addresses the CUDA platform, memory hierarchy, the roofline model and the various benchmarks to extract the maximum computational performance from the platform.

3.1 NVIDIA GPU Platforms

NVIDIA is a multinational technology company that specializes in the design and development of graphics processing units (GPUs) and system-on-a-chip (SoC) units for the gaming, professional visualization, data centre, and automotive industries. NVIDIA is well-known for its GeForce line of gaming GPUs, which are widely used by gamers worldwide. These GPUs are designed to deliver high-quality graphics and fast performance to enhance the gaming experience. NVIDIA also offers a range of professional GPUs, including the Quadro and Tesla series, which are used in industries such as architecture, engineering, and scientific research. In addition to the hardware systems offerings, NVIDIA has also developed a range of software platforms and tools that are designed to enhance the performance and capabilities of its GPUs. These platforms include CUDA, which is a parallel computing framework and programming model that enables developers to harness the power of GPUs for a range of applications, and TensorRT, which is an inference optimizer and runtime engine for deep learning applications. NVIDIA's products and technologies have been widely recognized for their innovation and performance.

3.1.1 CUDA Execution Model

NVIDIA released CUDA (Compute Unified Device Architecture) [31] in November 2006, a general-purpose parallel computing and a programming framework that takes advantage of the parallel computing engine in NVIDIA GPUs. CUDA supports different programming languages such as C, C++ and Fortran and provides interoperability with OpenGL.

Unlike the CPU, the GPU is a highly threaded and programmable accelerator, that excels in performing data-parallel tasks, which involve the execution of the same instruction on

several data elements simultaneously. The GPU execution is controlled by the host (CPU) which is in charge of GPU kernel invocations and memory transfers to and from the GPU device memory (global memory). The GPU kernel represents the program's core, which is executed simultaneously by a number of threads. These threads are organised in warps and thread blocks (CUDA blocks), each of which can hold up to 1024 threads. The blocks of threads are arranged in a grid representing all the threads running a certain kernel. The CUDA programming model can be visualized in Figure 3.1.

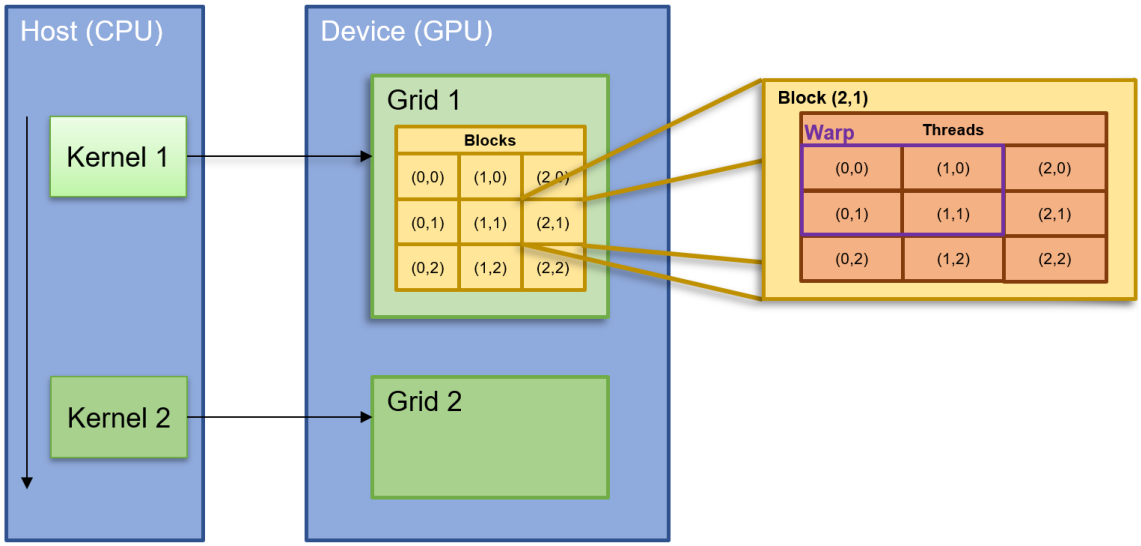

Figure 3.1: CUDA programming model [32].

CUDA follows a Single Instruction, Multiple Thread (SIMT) [31] execution model, where a set of threads in the same block, commonly 32 threads, known as warps, executes the same instruction and may share and exchange data. CUDA blocks are usually launched in multiples of 32 threads (the same size as wraps) to better utilize warps' execution properties which will get the most out of GPU performance.

Each CUDA block is executed by one streaming multiprocessor (SM) and cannot be migrated to other SMs in GPU. Several blocks can run concurrently on the same SM depending on the resources needed. Each kernel is executed on one device and CUDA supports running multiple kernels on a device at one time.

3.1.2 Memory hierarchy in GPU architecture

The GPU architecture has a memory hierarchy [31] constituted by :

- **Registers** - It is a private section of the memory, which means each register within a thread cannot be visible to other threads and can only be used by the compiler.
- **L1/Shared memory** - Every SM has a fast, on-chip scratchpad memory that has access to the L1 cache and shared memory. All threads share shared memory within a CUDA block, and all CUDA blocks running on a given SM can share the physical

memory (RAM) resource provided by the SM.

- **Read-only memory** - Section of memory that is read-only by the kernel code, where each SM has an instruction cache, constant memory, texture memory and RO (read-only) cache.
- **L2 cache** - This cache is shared by all SMs, unlike the L1 cache which is shared by the threads within a single SM. This cache usually is bigger but slower on average than the L1 cache to access.
- **Global memory** - Main memory section where DRAM is located and can be accessed by the host (CPU) and all threads in GPU. It has the largest size but it is also the slowest to access. The latency can be up to hundreds of cycles.

The image 3.2 shows an example of a memory hierarchy in a GPU architecture.

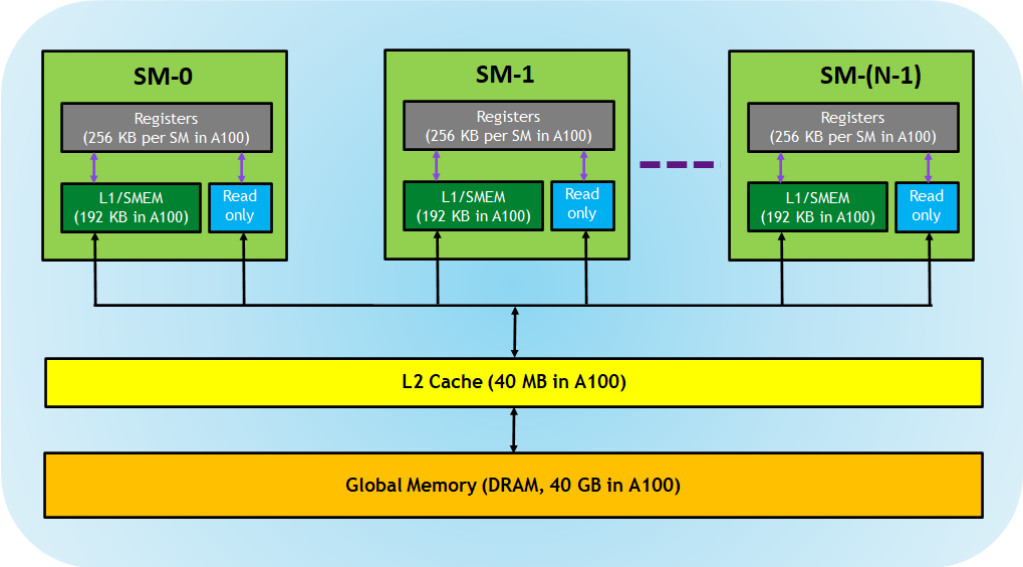

Figure 3.2: Memory hierarchy in GPUs [31].

3.1.3 NVIDIA Jetson AGX Xavier

The Jetson AGX Xavier is a high-performance artificial intelligence (AI) computing system developed by NVIDIA Corporation. It is designed to meet the demands of complex AI applications, including autonomous robots, drones and intelligent video analytics. The system is built on the NVIDIA Xavier architecture, which includes a custom 64-bit eight-core ARMv8 CPU with 2,26GHz of frequency, a 512-core Volta GPU with 1377MHz of frequency, and a high-speed memory subsystem.

The Jetson AGX Xavier is highly energy-efficient and can perform 30 Tera operations per second (TOPS) with a maximum power consumption of only 30 watts. This makes it an ideal platform for developing and deploying intelligent AI systems in a wide range

of environments, including industrial automation, healthcare, transportation, and defence [33].

An important key feature of the Jetson AGX Xavier is its support for multiple AI frameworks, including TensorFlow, PyTorch, and Caffe. This makes it easy for developers to create and deploy AI applications using their preferred frameworks and tools. The system also includes NVIDIA's DeepStream SDK, which provides tools and libraries for building intelligent video analytics applications [34].

The Jetson AGX Xavier also includes a range of connectivity options, including Gigabit Ethernet, USB 3.1, and PCIe, as well as support for multiple cameras and sensors. This makes it possible to integrate the system with a wide range of peripherals and devices, enabling developers to build highly customized and flexible AI systems [18].

The NVIDIA Jetson AGX Xavier board has the following specifications:

Table 3.1: Specifications of Jetson AGX Xavier [18].

Jetson AGX Xavier	
CPU	NVIDIA Carmel ARMv8.2 8 Cores 2,26 GHz 4 x 2MB L2 4MB L3
GPU	512 Cores Volta 1377 MHz 64 Tensor Cores
Memory	16 GB 256-bit LPDDR4x 2133 MHz 137 GB/s
Power	10/15/30W
Performance	32 Tera operations per sec

3.1.3.1 The CPU and GPU Architecture

The Jetson AGX Xavier is equipped with an 8-core ARM v8.2 64-bit CPU, which is capable of running complex computing tasks. This CPU includes 8MB of L2 and 4MB of L3 cache, which helps to improve performance by reducing the time it takes to access frequently used data. Figure 3.3 shows the CPU architecture based on the Carmel microarchitecture and features a combination of high-performance and low-power cores with an L2 cache sharing every two cores and a common cache L3.

The board also features a 512-core NVIDIA Volta GPU. This GPU is optimized for deep learning applications and provides a significant boost in performance for tasks such as image and video processing. Figure 3.4 represents the GPU architecture of Volta which includes eight SMs with 64 CUDA cores and 8 Tensor Cores. Each Volta SM has a 128KB

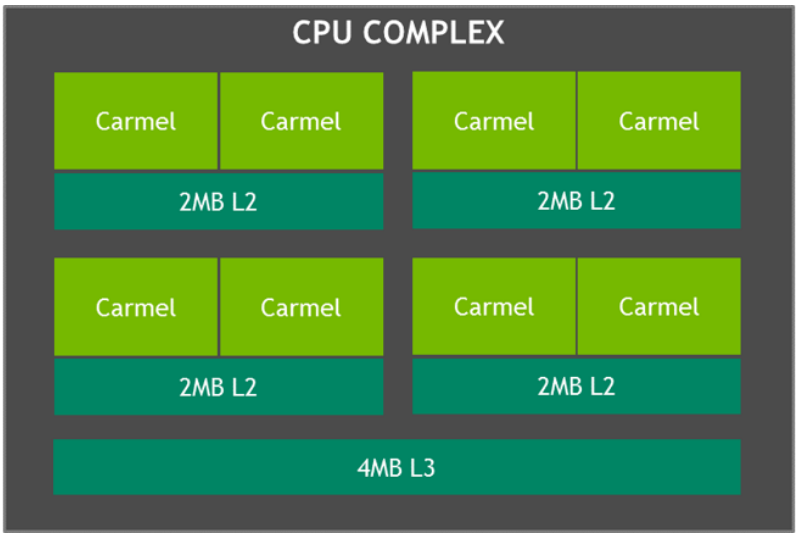

Figure 3.3: CPU architecture [33].

L1 cache, 8x larger than previous generations, and a shared cache L2 with 512 KB, which offers 4x faster access than previous generations [33]. Each SM can run up to 2048 threads concurrently. With 8 SMs the platform can use 16384 threads concurrently.

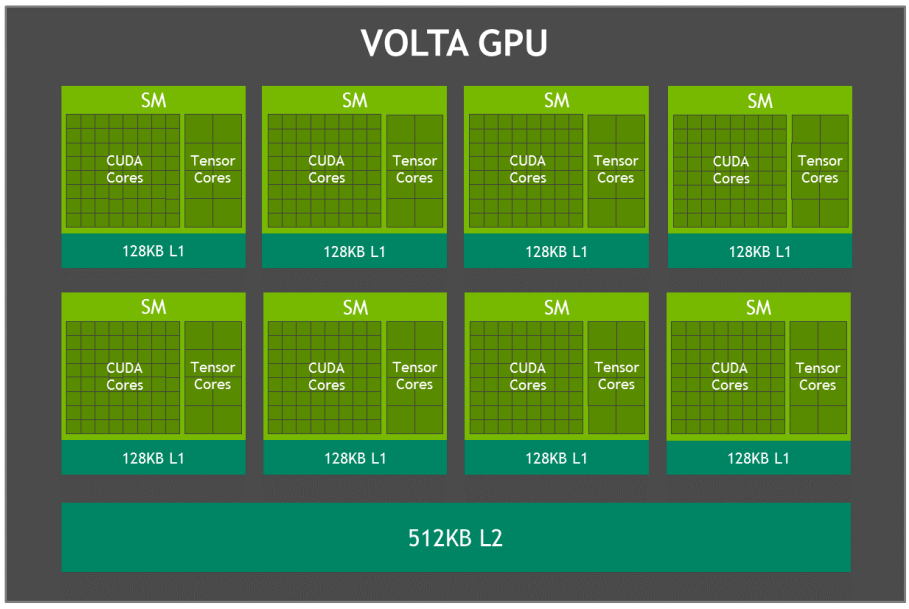

Figure 3.4: GPU architecture [33].

There is a zone of memory that is shared by both architectures, called global memory, which represents most of the system memory. This memory area is the slowest access.

3.2 Roofline Model

To achieve good performance from the program of the HPC and power-limited platform the roofline model was used as a guideline. The roofline model [35] [36] [37] combines floating-

point performance, arithmetic intensity and memory performance in a two-dimensional graph. The performance, the y-axis of the plot, is measured in Floating-point operations per second (Flops) and is theoretically calculated by the hardware specifications or through micro-benchmarking.

The arithmetic intensity or arithmetic intensity, in the x-axis, is expressed in Flops/Byte which is the operations per byte of memory traffic. It should be mentioned that both axes are logarithmic scales.

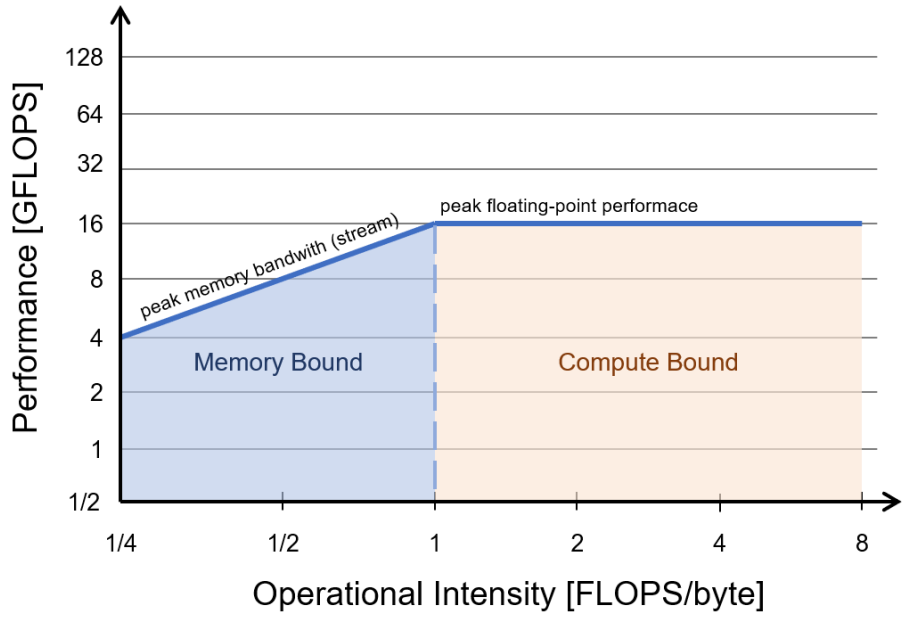

Figure 3.5: Roofline model with two optimizations regions [36].

In Figure 3.5 the slope of the graph is dictated by the memory bandwidth which can be obtained only through specifications. If a vertical line is drawn from the kernel's arithmetic intensity, the kernel can be memory or compute bound. If the vertical line hits the slope, which corresponds to memory bandwidth then it is memory bound. If it intersects the horizontal line then it is compute bound.

The two bounding lines intersect at a point, i.e., where the slope and horizontal lines meet representing the minimum arithmetic intensity required to reach the peak performance. Therefore if the arithmetic intensity is low then it is challenging to achieve maximum performance. But if the arithmetic intensity is high the kernel can potentially get the peak performance.

The roofline model is a simple and visual performance model, which can help to identify which changes should be made or which hardware should run the program.

3.3 Optimization Strategies

In order to develop a system capable of processing all the information quickly it is important to study the performance of a system and understand how to get the most out of it. Based on that, this section shows the algorithms used to extract the computational performance and arithmetic intensity of a platform followed by the experimental results of that strategy.

The computational performance is measured in FLOPS followed by the formula 3.1 and the arithmetic intensity is calculated in FLOPs/byte by the expression 3.2.

$$Perf = \frac{\text{Number of Operations}}{\text{Time of Execution (in seconds)}} \tag{3.1}$$

$$AI = \frac{\text{Number of Operations}}{\text{Bytes of memory transferred}} \tag{3.2}$$

According to table 3.1 the maximum performance is 32 TFLOPS. This value is only theoretical and for perfect working conditions. This means in practical terms this value will never be reached. As a result of this, the goal of this study is to find the highest possible performance for more realistic working scenarios.

3.3.1 Setup

The majority of the practice results have as base the SAXPY algorithm [38]. The following code (3.1) is an example of the SAXPY algorithm in C/C++. The variable n corresponds to the dimension of the arrays.

Listing 3.1: SAXPY code in C [38].

```
1 void saxpy_cpu(float *a, float *b, float *c, int n)
2 {
3     for (int i = 0; i < n; ++i)
4         c[i] = 2.1 * a[i] + b[i];
5 }
```

In each cycle, the kernel makes two operations, one addition and one multiplication, which means the program makes $2 \times n$ operations. Varying the size of the arrays will give different times of execution and numbers of operations. Fig 3.6 represents the variation in the size of the arrays according to their performance (see Appendix A.1).

As it is possible to observe in Figure 3.6 the peak performance stabilises for longer array dimensions and it is still low, compared with the maximum theoretical performance. Consequently, maximum performance has not yet been achieved with this configuration.

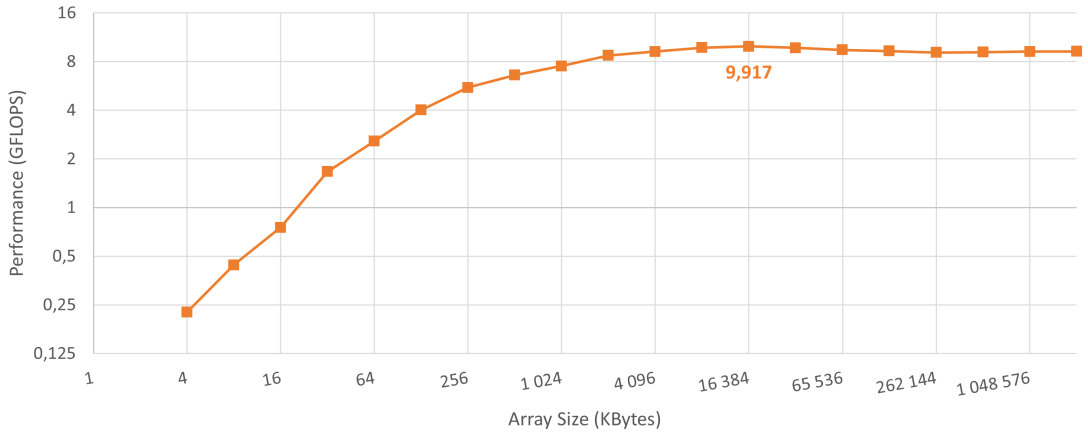

Figure 3.6: Graph of array size according to performance on the GPU (see Appendix A.1).

It is possible to expand the operations per cycle using arithmetic operations to raise the arithmetic intensity even more. The code in 3.2 uses the example of the SAXPY algorithm (3.1) and adds 2 additional operations per cycle increasing the arithmetic intensity.

Listing 3.2: SAXPY code in C with 2 additional operations per cycle.

```
1 void saxpy_cpu(float *a, float *b, float *c, int n)
2 {
3     for (int i = 0; i < n; ++i)
4         c[i] = ((2.1*a[i] + b[i]) * (-0.2) + 0.3);
5 }
```

Figure 3.7 uses the SAXPY algorithm with 64 and 128 operations per cycle (see Appendices A.1, A.2 and A.3).

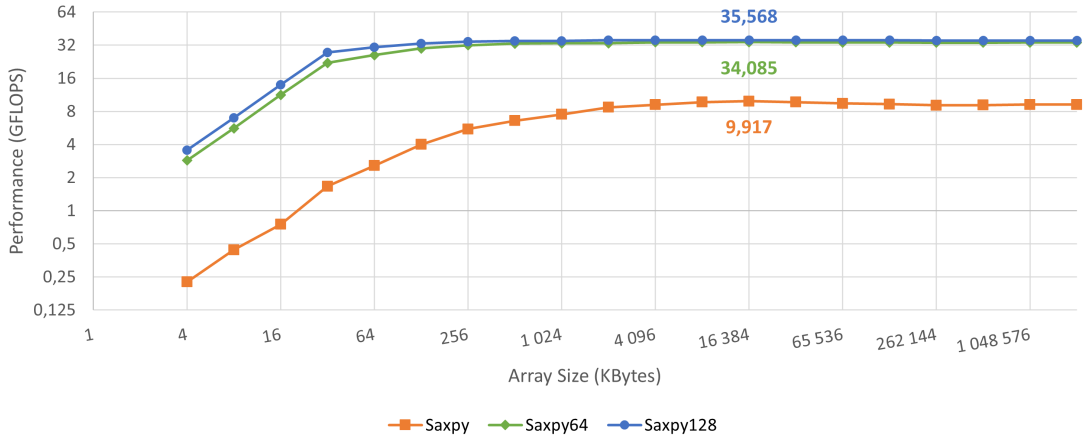

Figure 3.7: Graph of array size according to performance with 64 and 128 operations per cycle (see Appendices A.1, A.2 and A.3).

It is observable in Figure 3.7 that the simple saxpy configuration has not yet saturated. With 64 operations per cycle, the performance is rising by 25 GFLOPS. By comparing the 64 and 128 operations per cycle is viable to conclude that it is not probable to increase the performance further with the increase of operations per cycle.

3.3.2 Built-in vectors

Built-in vectors in CUDA are data types that allow for the manipulation of multiple data elements in a single instruction. These vectors are built into the CUDA programming model and can be used to improve the performance of parallel computations on CUDA-enabled devices.

According to NVIDIA [32], built-in vectors in CUDA are available in two forms:

- A scalar data type, such as float or int, with a built-in vector type, such as float4 or int4, that contains four elements of the scalar data type.
- A new data type, such as float16 or int8, that represents a vector of a specific number of elements of the corresponding data type.

Built-in vectors in CUDA can be used in a variety of ways, including:

- Performing mathematical operations on multiple data elements at once, such as vector addition or dot product.
- Loading and storing multiple data elements to and from memory in a single instruction.
- Performing operations on multiple data elements in a single instruction, such as comparing or selecting elements.

With the utilisation of the built-in vector feature, the number of operations per second will increase during the same period. This allows, for example, an operation with two vectors of four elements to process four operations in the same time it would take to perform a single non-vector operation. The listing 3.3 is a structure example of a built-in vector with four float elements called *float4*.

Listing 3.3: float4 structure

```
1 struct __device_builtin__ __builtin_align__(16) float4
2 {
3     float x, y, z, w;
4 };
```

This feature is a significant boost for the performance and can be observed in figure 3.8 (see Appendices A.2 and A.4). The blue line represents the vector operations and the orange

one is the previous result of the SAXPY algorithm. Taking into account the results of the previous section both graphs make use of 64 operations per cycle.

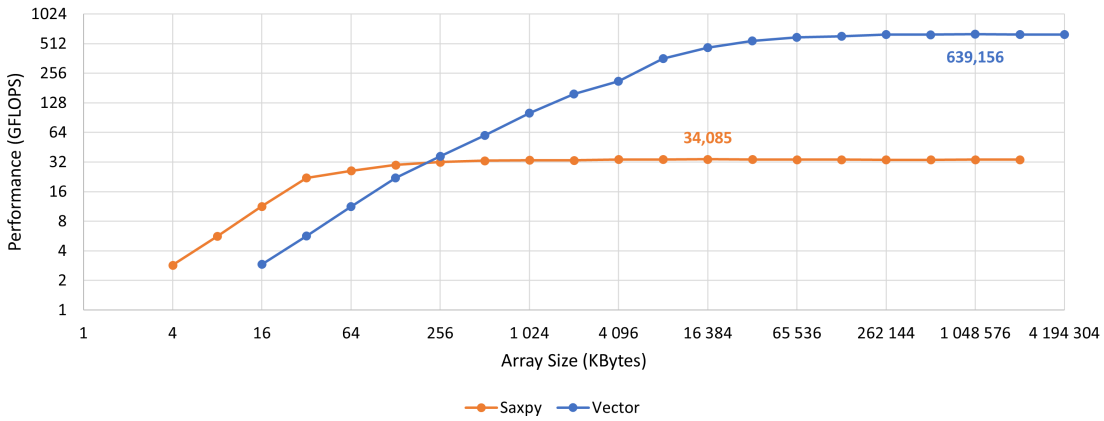

Figure 3.8: Graph of array size according to performance for SAXPY algorithm and vectorial operations (see Appendices A.2 and A.4).

Previously, in Figure 3.7, there was no significant difference between 64 or 128 operations per cycle in terms of compute performance. If built-in vectors are added, both methods produce completely different outcomes. Figure 3.9 shows these differences (see Appendices A.4 and A.5).

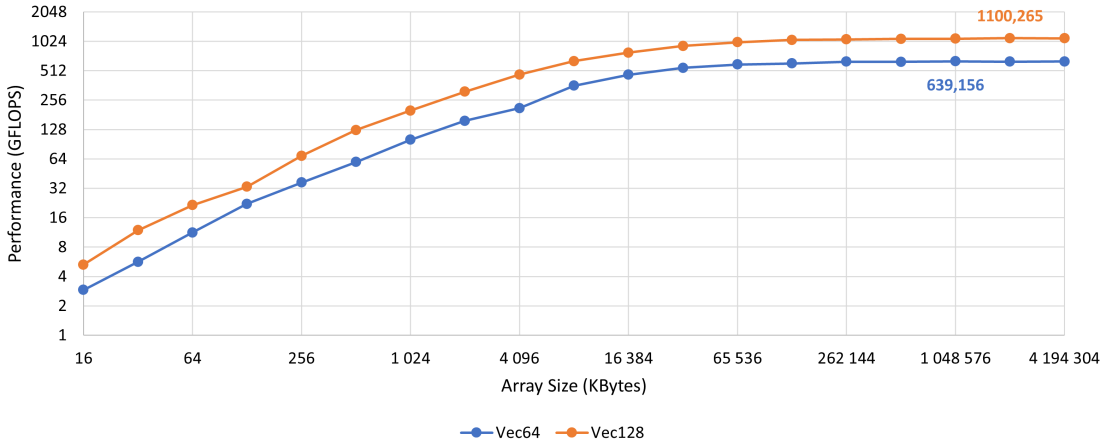

Figure 3.9: Graph of array size according to performance with 64 and 128 vector operations per cycle (see Appendices A.4 and A.5).

With enough operations, the one with 128 per cycle has about more than 500 GFLOPS than the 64 one. This demonstrates that GPU is not even close to reaching its full potential.

3.3.3 Memory Bandwidth

Memory bandwidth can limit the performance of CUDA kernels, as it is a key factor in determining the amount of data that can be transferred between the GPU and the system

memory. The GPU's memory bandwidth is the rate at which data can be transferred from the GPU memory to the GPU's processing units.

When executing a CUDA kernel, it often requires data to be transferred from the system memory to the GPU memory, and the result to be transferred back to the system memory. If the amount of data required by the kernel exceeds the available memory bandwidth, this can cause a bottleneck in the kernel's performance.

In addition, the GPU's memory hierarchy can also limit the performance of CUDA kernels. GPU memory is divided into multiple levels, each with its own bandwidth and latency characteristics. If a kernel requires data that is not in the highest level of the GPU memory hierarchy, it has to be transferred from a lower level, which can add additional latency and limit the performance of the kernel.

To minimize the impact of memory bandwidth limitations on CUDA kernel performance, developers can use techniques such as memory coalescence, which improves memory access patterns and increases the GPU's memory bandwidth utilization. Also, minimizing data transfers between the GPU and the memory system by keeping data on the GPU for as long as possible, and using shared memory to reduce the number of global memory accesses are good strategies for improving performance.

To further extend the performance of the system it is necessary to avoid being limited by the slow memory system. To get around this restriction we can use shared memory to speed up memory transfers.

Figure 3.10 shows the difference between using shared memory access and not, with 64 and 128 vector configuration (see Appendices A.4, A.5, A.6 and A.7).

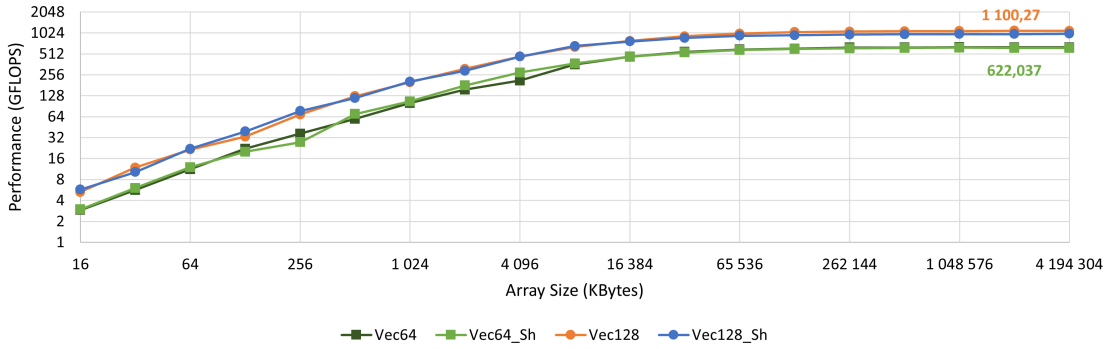

Figure 3.10: Impact of the compute performance of shared memory with 64 and 128 vector operations per cycle (see Appendices A.4, A.5, A.6 and A.7).

The graph shows that there is no difference in avoiding access to global memory because that configuration does not have enough access to memory to justify the use of shared memory. There are even some occasions when shared memory has poorer performance.

To increase the number of accesses it is possible to use the stencil algorithm [39] [40]. The

stencil algorithm is an optimization strategy that could be described as a set of weighted neighbour cells that should be loaded and combined. The radius is the distance from the centre of a stencil to its neighbouring elements. Figure 3.11 shows a representation of a 1D stencil algorithm.

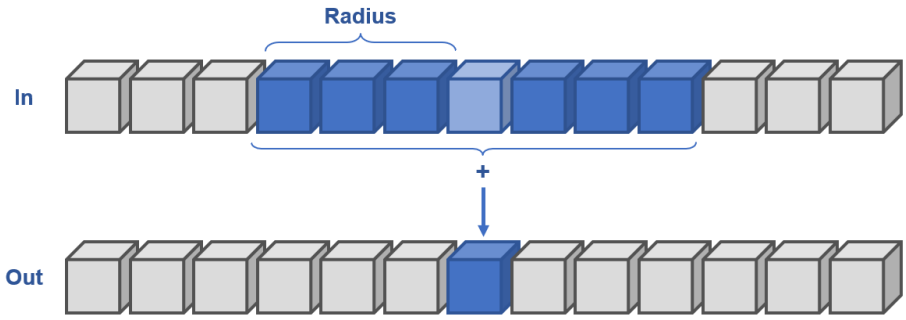

Figure 3.11: Representation of the sum of elements using 1D stencil algorithm with radius of 3.

Figure 3.12 explores the relationship between stencil radius and performance. Also shows the difference between 64 and 128 vector operations per cycle configuration (see Appendices A.8 and A.9).

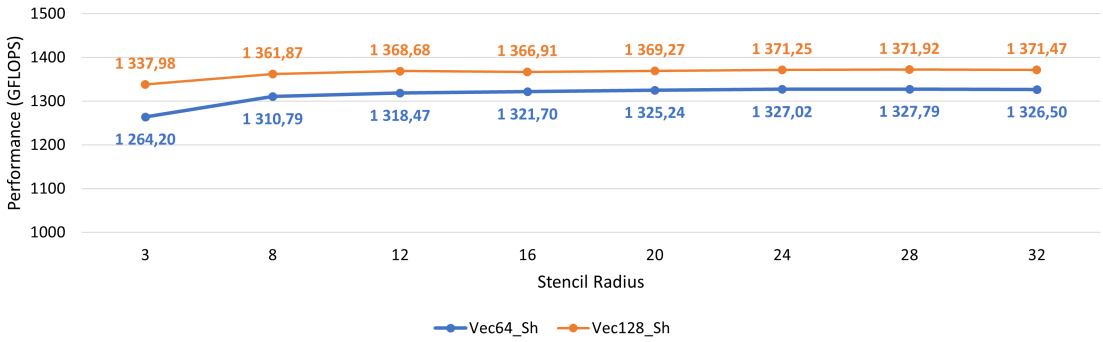

Figure 3.12: The relationship between stencil size and performance with 64 and 128 vector operations per cycle (see Appendices A.8 and A.9).

The performance for radius 3 is lower than the peak in Figure 3.9. The reason behind it is that shared memory consumes some time to transfer from global to shared memory because of that it is only possible to take advantage of this level of memory if there are enough memory accesses to be suppressed from global memory. It is possible to prove this by observing the curve of the graph in Figure 3.12. The more memory transferred with shared memory the more the performance will be until it stabilises.

3.3.4 Streams

One key feature of CUDA [32] is the ability to perform operations with streams, which allows multiple tasks to be executed concurrently on the GPU, which allows developers to use the power of NVIDIA GPUs for a variety of applications, including deep learning,

scientific computing, and video and image processing.

Streaming in CUDA is accomplished using streams, which are sequences of operations that are executed in order on the GPU. Each stream is associated with a specific CUDA device and can have its own set of CUDA contexts and memory spaces. This allows multiple tasks to be executed concurrently on the same GPU, with each task executing in its own stream.

In practice, streaming in CUDA is implemented by creating multiple CUDA streams and launching CUDA kernels (i.e., functions executed on the GPU) in each stream. The CUDA runtime automatically schedules the execution of kernels across the available streams and manages the dependencies between them. This allows for efficient use of the GPU and can significantly improve performance for applications that have a high degree of parallelism.

With the feature of stream operations, it is achievable to reach even higher performances. The ability to perform parallel tasks between transferring data between host and device and kernel execution is given by copy engines. The board Jetson AGX Xavier has only one copy engine. Consequently, it can only run 2 tasks in parallel, like transferring data from device to host and executing the kernel at the same time. Figure 3.13 illustrates this case.

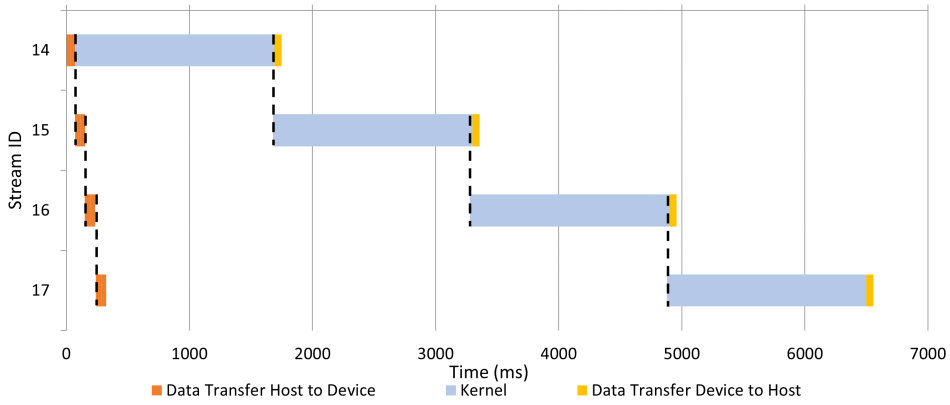

Figure 3.13: Profiling of one execution of streaming with 4 streams.

The data transfer from the host to the device (orange) starts at the same time as the execution of the first kernel (blue). Another example is when the data transfer from the device to the host (yellow) starts at the same time as the execution of the next kernel.

This board has a maximum of 16 streams. Therefore it is possible to study the influence of different numbers of streams on the performance. Figure 3.14 shows the relation between the number of streams and the compute performance (see Appendix A.10).

In this configuration using streams does not increase the compute performance. This occurs because the execution time of transferring data between host-device or/and device host is small and it takes time to set up the streams. It is possible to observe from the

graph that the more streams the better the results. This happens because there are more tasks being executed concurrently on the GPU. The execution time of stream operations will be more noticeable when the transfer of data between host-device or/and device-host is larger.

3.3.5 Tensor Cores

Tensor Cores are specialized hardware units that are built into the NVIDIA GPU architecture [41] and they are designed to accelerate the computation of deep learning workloads they provide a significant performance boost for tasks such as image classification, natural language processing, and video analytics. Tensor Cores are capable of performing mixed-precision matrix operations, such as the multiplication of a 4x4 matrix of 8-bit integers or 16-bit floating-point numbers with a 4x4 matrix of 16-bit floating-point numbers, and they can speed up deep learning workloads by a factor of up to 12x compared to traditional GPU cores.

They can perform mixed-precision matrix operations at much higher speeds than traditional GPU cores, which allows them to process more data in less time. This makes them ideal for deep learning workloads that require large amounts of data to be processed quickly.

In addition, Tensor Cores are also designed to support half-precision floating-point format, which uses 16-bit floating-point numbers instead of 32-bit floating-point numbers. This reduces the amount of memory required to store the data and can also reduce the bandwidth required to transfer the data between the GPU and the system memory. This can lead to significant performance improvements for deep learning workloads that require large amounts of data.

Through matrix multiplication, a study was made of the performance of the Tensor cores. These results are independent, which means the configuration of the SAXPY algorithm

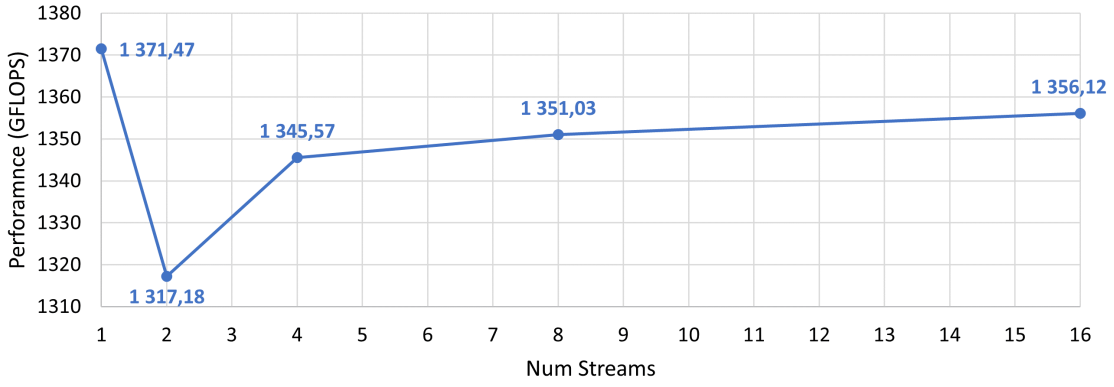

Figure 3.14: The relationship between a number of streams and performance with 128 vector operations per cycle using shared memory (see Appendix A.10).

within a loop was not tested. Figure 3.15 shows the computational performance matrix multiplication for different sizes with Tensor cores (see Appendix A.11).

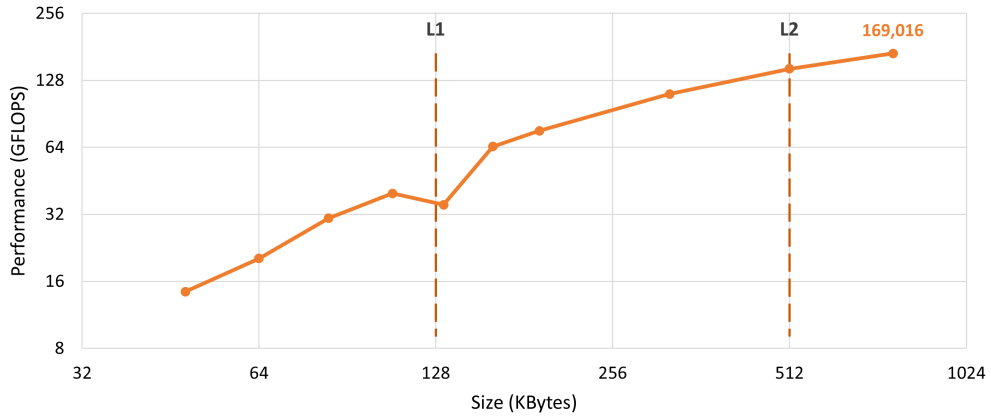

Figure 3.15: The relationship between the total size of the matrices and performance with the representation of cache L1 and L2 sizes (see Appendix A.11).

Observing Figure 3.15 when the size is close to the border of both the L1 and L2 cache the performance tends to drop.

3.3.6 Power Modes

Some platforms from NVIDIA support different power modes. Power modes are a set of configurations that can be used to optimize the power consumption of embedded systems by limiting the resources available, like changing the CPU clock frequency, the number of CPU cores available, etc. Some platforms of the Jetson family are designed, for example, for AI and computer vision applications, and as such, they require a significant amount of processing power to run efficiently. However, running at full capacity can also result in high power consumption and heat generation, which may not fulfil the requirements needed for the experiment.

To address this, NVIDIA has developed several power modes. For example, the Jetson AGX Xavier, has MAXN, 15W and 10W. Each power mode is designed to balance performance and power consumption based on the specific use case. For example, the MAXN mode maximizes the performance of the device, while the 10W mode is designed for low-power applications that require minimal processing power.

To control the power consumption of a Jetson device, users can switch between the different power modes depending on their use case or they can create their own custom power modes by changing any parameter of their choice.

The table 3.2 show some power modes there are available in the NVIDIA Jetson AGX Xavier.

The NVIDIA devices offer several power modes that can be used to optimize power con-

Table 3.2: Power modes of Jetson AGX Xavier [18].

Power Modes - Jetson AGX Xavier			
Mode	**MAXN**	**10W**	**15W**
Power budget	n/a	10W	15W
Mode ID	0	1	7
Online CPU cores	8	2	4
CPU maximal frequency (MHz)	2265,6	1200	2188
GPU TPC	4	2	4
GPU maximal frequency (MHz)	1377	520	670
DLA cores	2	2	2
DLA maximal frequency (MHz)	1395,2	550	115,2
PVA cores	2	0	1
PVA maximal frequency (MHz)	1088	0	115,2
CVNAS maximal frequency (MHz)	1356,8	601,6	115,2
Memory maximal frequency (MHz)	2133	1066	1333

sumption and performance for specific use cases. Users can switch between power modes depending on their needs. These power modes are essential for managing the power consumption of NVIDIA devices and for space, missions are critical to meet the requirements.

Power modes mentioned earlier, limit the available resources of the onboard computer then it is expected that changing the power mode from MAXN to, for example, 10W will naturally have an impact on the computational performance. Figure 3.16 indicates the impact of the performance for different power modes with the configuration of 28 vector operations per cycle using shared memory (see Appendices A.9, A.12 and A.13).

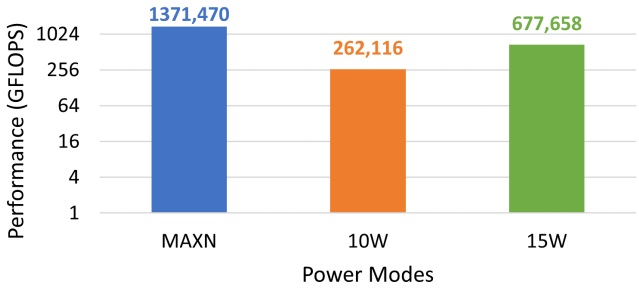

Figure 3.16: The impact of power modes on the computational performance with the configuration of 28 vector operations per cycle using shared memory (see Appendices A.9, A.12 and A.13).

It is clear that with the reduction of some resources of the platform, the computational performance decreased. The MAXN mode is not limited therefore the performance goes to 1.38 tera-operations per second. The 10W and 15W were some resource limitations and it is expected the observed performance of 262 and 678 GFLOPS, respectively. It is important to keep in mind that to meet the requirements of the project it will be

necessary to sacrifice some performance. As a result, this study is important to evaluate the resources available versus the compute performance achieved.

3.3.7 Analysis of the optimization strategies results

This section shows the analysis of computational performance and arithmetic intensity for each previous optimization strategy using its best result. Each strategy using the *saxpy* or *stencil* algorithms extracted its best result to build the Figure 3.17 and those values are in the Appendix A.14. Table 3.3 contains a description for all the abbreviations that were used in Figure 3.17.

By looking at Figure 3.17 it is possible to observe that the simple *saxpy* algorithm has the least arithmetic intensity and subsequently the worst performance. With the increase in the number of arithmetic operations to 64 and 128, *saxpy64* and *saxpy128* respectively, the arithmetic intensity rose very quickly away from the *saxpy*. At this moment both strategies have low computational performance and to increase it the use of built-in vectors, *Vec64* and *Vec128*, helps and also can maintain the arithmetic intensity of the algorithm.

The use of shared memory, *Vec64_Sh* and *Vec128_Sh*, allows to prevent the strategy to be limited by the memory bandwidth, thus accelerating its performance. But in this case, the small number of memory accesses performed does not justify the use of shared memory. To take benefit of this, it was necessary to use the *stencil* algorithm which increases the number of memory accesses without compromising arithmetic intensity. The *Vec64_Sh_R28* and *Vec128_Sh_R28* use the previous setup (*Vec64_Sh* and *Vec128_Sh* respectively) but now with the stencil algorithm of a radius of 28. Finally, the use of

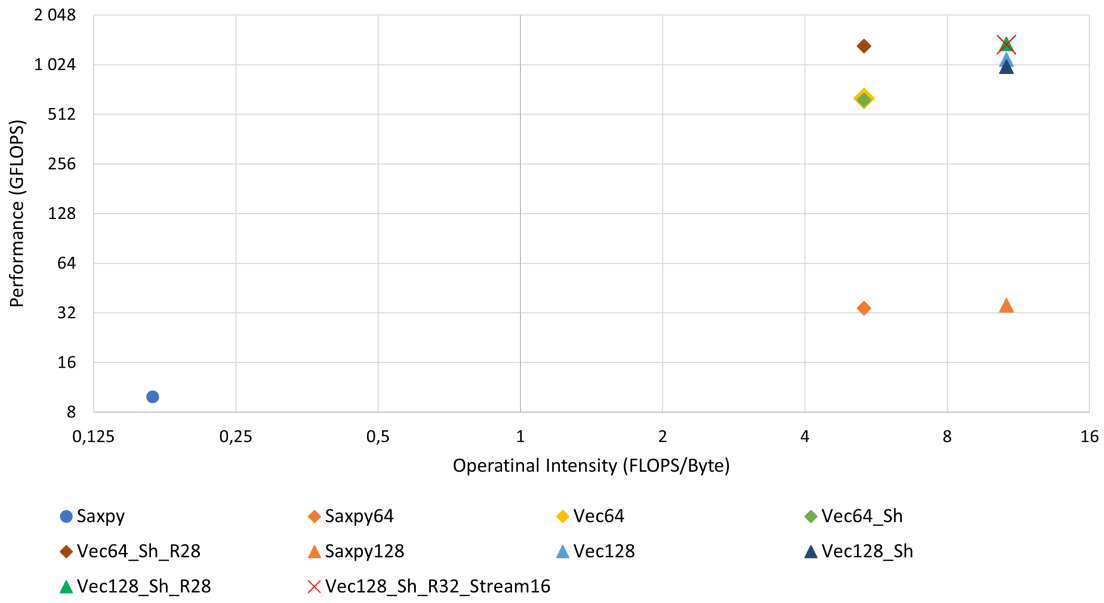

Figure 3.17: Analysis of the best results of each optimization strategy that uses the *saxpy* or *stencil* algorithms as a baseline (see Appendix A.14).

Table 3.3: Table with descriptions of abbreviations used in the Figure 3.17.

Abbreviations	**Description**
Saxpy	Original saxpy algorithm
Saxpy64	Saxpy with 64 operations
Saxpy128	Saxpy with 128 operations
Vec64	Saxpy64 with built-in vector
Vec128	Saxpy128 with built-in vector
Vec64_Sh	Vec64 with shared memory
Vec128_Sh	Vec128 with shared memory
Vec64_Sh_R28	Stencil algorithm with radius of 28 with 64 operations with shared memory
Vec128_Sh_R28	Stencil algorithm with radius of 28 with 128 operations with shared memory
Vec128_Sh_R32_Stream16	Vec128_Sh_R32 with 16 stream channels

streams for this algorithm does not help with the performance because, as mentioned in section 3.3.4, the execution time of transferring data between host-device or/and device host is small and it takes time to set up the streams. It is hard to see the difference in Figure 3.17 but by looking at those values in the Appendix A.14 is possible to observe that the computational performance decreased.

3.4 Summary

NVIDIA is a multinational technology company that specializes in the design and development of graphics processing units (GPUs) and system-on-a-chip (SoC) units for various industries, including gaming, professional visualization, data centre, and automotive. It offers a range of software platforms and tools to enhance the performance and capabilities of its GPUs. NVIDIA's products and technologies have been widely recognized for their innovation and performance.

This section also references the algorithm used to extract the computing performance of an NVIDIA platform, the built-in vector, the use of streams and the usage and understanding of memory bandwidth and tensor cores.

Built-in vectors in CUDA are data types that allow for the manipulation of multiple data elements in a single instruction, providing a way to improve the performance of parallel computations on CUDA-enabled devices.

CUDA allows for concurrent execution of tasks, known as streams, which are sequences of operations that are executed in order on the GPU. This allows multiple tasks to be executed concurrently on the same GPU and can significantly improve performance for applications with a high degree of parallelism.

Memory bandwidth is a key factor that can limit the performance of CUDA kernels. The

GPU's memory bandwidth is the rate at which data can be transferred from the GPU memory to the GPU's processing units. When a CUDA kernel requires more data than the available memory bandwidth, this can cause a bottleneck in the kernel's performance. Developers can reduce this impact by using techniques such as memory coalescing, minimizing data transfers between the GPU and system memory, and using shared memory to reduce the number of global memory accesses.

Tensor Cores are specialized hardware units that are built into NVIDIA GPU architecture, they are designed to accelerate the computation of deep learning workloads. They provide a significant performance boost for tasks such as image classification, natural language processing, and video analytics. Tensor Cores can perform mixed-precision matrix operations at much higher speeds than traditional GPU cores. They support a half-precision floating-point format which can lead to significant performance improvements for deep learning workloads that require large amounts of data.

NVIDIA devices can be used, for example, for AI and computer vision applications and they require a significant amount of processing power to run efficiently. To solve this, these devices offer several power modes that can be used to optimize power consumption and performance to further meet the requirements needed.

By analyzing the computational performance and arithmetic intensity of different optimization strategies. The simple *saxpy* algorithm has the lowest arithmetic intensity and the worst performance. As the number of arithmetic operations increases, *saxpy64* and *saxpy128* show an improvement in arithmetic intensity. Using built-in vectors (*Vec64* and *Vec128*) enhances computational performance while maintaining arithmetic intensity.

Shared memory (*Vec64_Sh* and *Vec128_Sh*) accelerates performance by overcoming memory bandwidth limitations, but it is not justified for low memory accesses. The *stencil* algorithm increases memory accesses without compromising arithmetic intensity. Strategies *Vec64_Sh_R28* and *Vec128_Sh_R28* incorporate a radius of 28 for the stencil algorithm. Streams for this algorithm do not improve performance.

4

Case Study on the gamma-ray detector

The main components of the experiment are a detector unit, an OBC (onboard computer), and a storage unit. The detector unit is formed by 16 sensors which are responsible to generate data from the radiation picked up from the open sky. OBC is where the low-power and HPC system is located which is in charge of processing data from the sensors and sending it to the storage unit. The storage unit stores all the information received. Image 4.1 is an illustration of the overview of the data flow of the experiment.

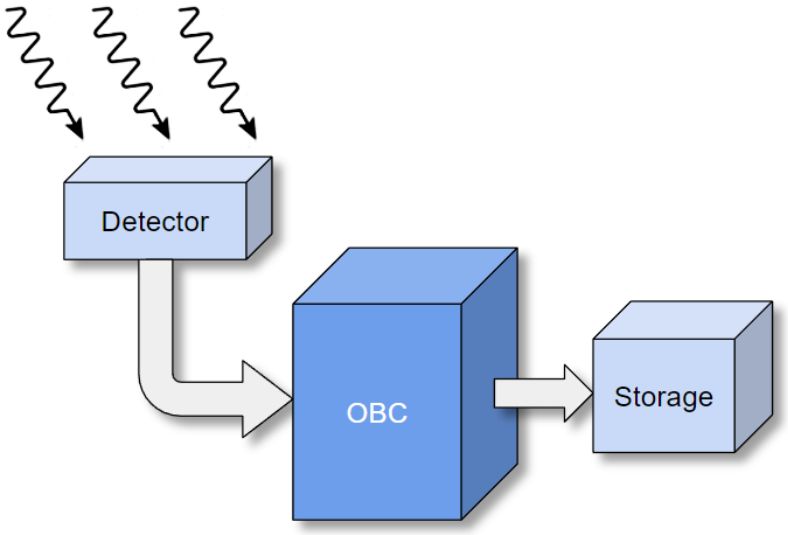

Figure 4.1: Overview of data flow of the experiment.

The detector unit is responsible to capture the radiation from outside. It is composed of 4 detection planes and each plane with 4 side-by-side CdTe sensors, called MiniPix Tx3, individually with a 256x256 matrix of pixels, which means each plane has a 256x1024 matrix of pixels. In total, the detector unit has around 10^6 pixels. When particles or photons interact with each pixel, it outputs the corresponding intensity. The information that the OBC receives is the coordinates of the pixel where there was an interaction and the corresponding intensity. After that, the OBC processes that data and sends the output

to the storage unit where it is stored for later analysis post-flight and can communicate with the ground station.

Since the detector unit has a matrix structure, working with a GPU, which is excellent at working with images, is a good idea and it is possible to associate this work with a 2D structure problem.

4.1 Detector Unit

The sensor, MiniPix Tx3, is produced by ADVACAM, a company that focuses on sensor manufacturing, water solder bumping, flip chip bonding and other services [42].

ADVACAM provides GUI (Graphical User Interface) software compatible with the detector, called *PIXet Pro*, that is used for testing and validation [42].

MiniPix CdTe sensor comes from a family called Timepix which detects ions, muons, protons, alpha particles, electrons, gamma and X-rays [43][44][45], by measuring their position, kinetic energy, and time of arrival.

Figure 4.2 shows all the basic particle track types that the detector can capture.

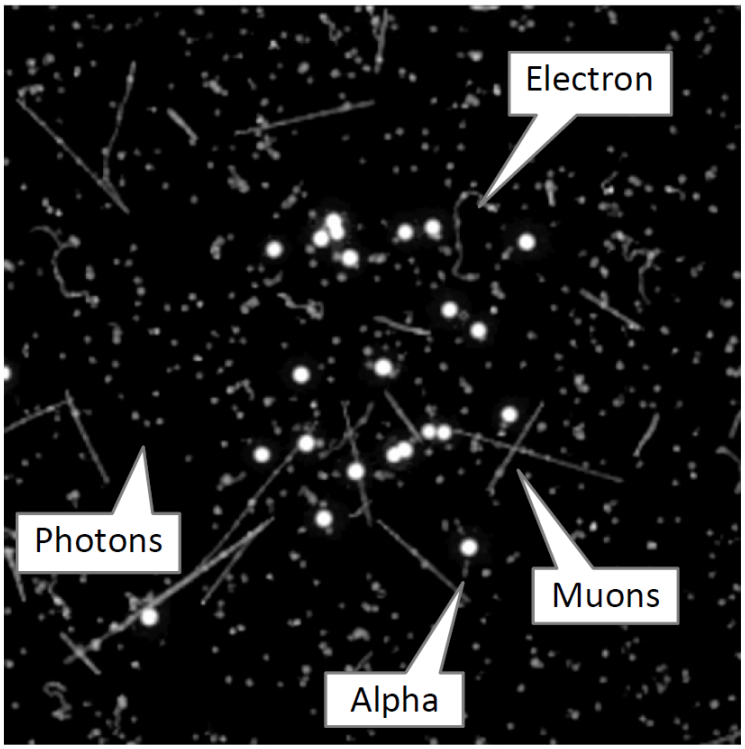

Figure 4.2: Illustration of particle tracking capability of Timepix3 device: The tracks of different particles of radiation background were recorded for 10 minutes in ADVACAM's office space in Prague. Brightness corresponds to energy intensity. No noise (clean zero) is seen in dark regions. All basic particle track types are seen nicely: muons = straight lines, alpha particles = bright balls, electrons = curving lines, gamma and X-rays = dots and blobs [46].

In the *PIXet Pro*, it is possible to connect a simulator to produce random values for testing. Figure 4.3 show an example of data produced by the simulation. Each dot inside the big square is the value that the sensor receives which represents an iteration between the particles and the pixel. The gradient of the figure illustrates the intensity of the interaction ranging, in this case, between 0 and 4×10^5.

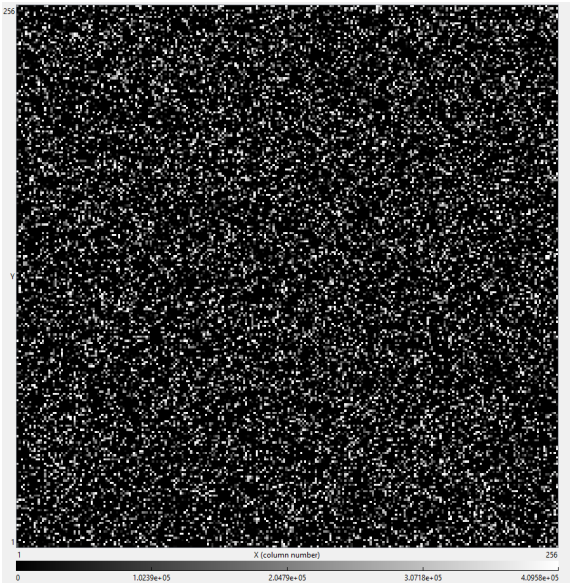

Figure 4.3: Example of random data generated by the simulator on the GUI software.

4.1.1 Operation modes

The detector offers different operation modes: Frame mode, Integral mode and Data-Driven mode.

- **Frame mode:** reads all pixels after the end of the exposure
- **Integral mode:** similar to Frame mode but compacts all the previous frames into one image.
- **Data-Driven:** reads only hit pixels continuously during exposure during a period of time.

Within all the operation modes, the experiment will work in the data-driven mode. Because it receives only the information about the hit pixels instead of the entire matrix and reads out immediately and continuously during the exposure time.

This software provides a download file of the captures. Also has addons for clustering and spectral imaging. The clustering addon has the purpose of identifying clusters of interactions. It is possible for the same particle to hit several adjacent pixels forming an area of hit pixels. Figure 4.4 shows an example of many clusters of particles hitting the sensor.

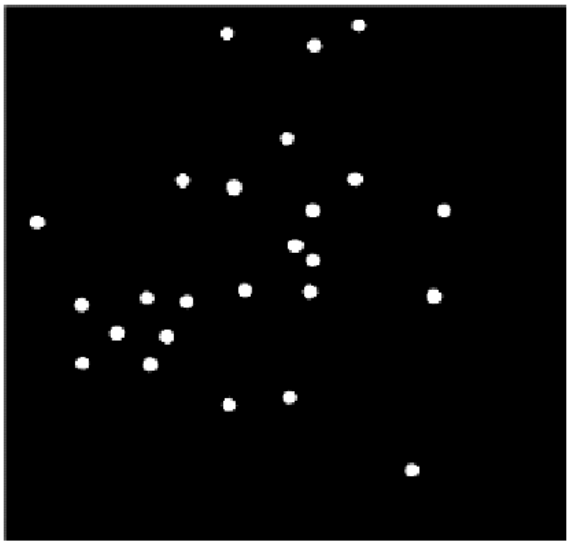

Figure 4.4: Example of several clusters of particles using the GUI software.

Spectral imaging is a tool capable of creating a histogram of a spectrum of the radiation to be captured. Figure 4.5 shows an example of a spectrum of Barium (Ba-133).

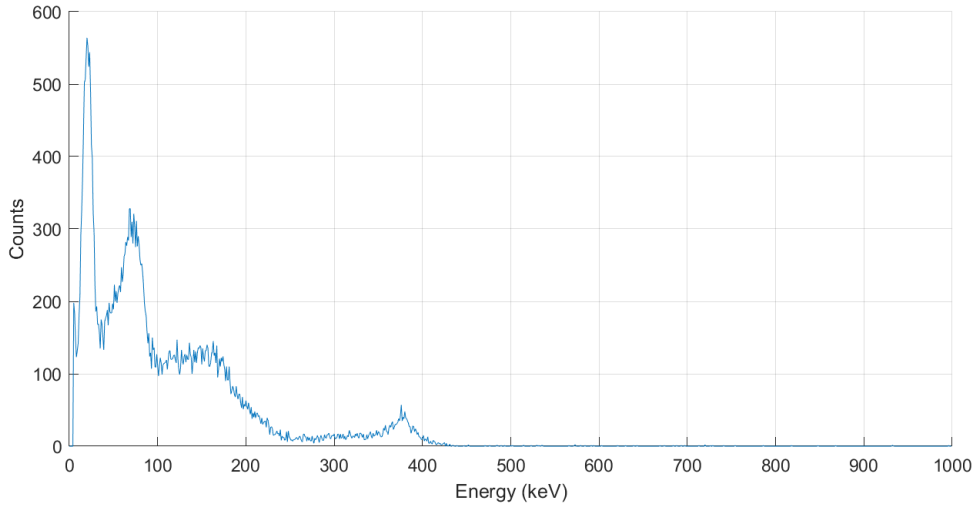

Figure 4.5: Example of a Barium (Ba-133) spectrum of a 10 minute acquisition in the LIP laboratory.

4.1.2 Example output data

The detector can provide a variety of information. It can output the coordinate of the active pixel, the timestamp of the acquisition, and for example the energy of the particle in that pixel. Figure 4.6 shows an example of the output data in a text file in data-driven mode.

Figure 4.6 displays several columns of data:

- The ***Index*** column represents the ID of the event.

Index	Matrix Index	ToA	ToT	FToA	Overflow
0	5561	43112	1	10	0
1	6075	43111	5	18	0
2	5566	43111	3	8	0
3	6076	43110	8	5	0
4	5050	43112	1	5	0
5	5053	43111	6	13	0
6	5306	43110	17	5	0
7	5309	43110	15	8	0
8	5051	43110	9	6	0
9	5052	43110	11	8	0
10	5562	43110	21	9	0
11	5565	43110	24	10	0
12	5818	43110	12	7	0
13	5821	43110	14	6	0
14	5307	43110	111	12	0
15	5308	43110	146	13	0
16	5563	43110	311	13	0
17	5820	43110	60	14	0
18	5819	43110	43	13	0
19	5564	43110	338	14	0
20	33490	109842	10	17	0
21	33235	109840	16	19	0
22	33491	109841	11	9	0

Figure 4.6: Example of an output file from the detector.

- The ***Matrix Index*** defines the location of the pixel in the matrix.
- The ***ToA*** denotes *Time of Arrival* which is the timestamp of the event in nanoseconds.
- The ***ToT*** signifies *Time Overthreshold* represents the energy of the interaction of the particle in the pixel.
- The ***FToA*** is the *Fast Time of Arrival* which is used to improve time resolution.
- Lastly, the ***Overflow*** shows if there is an overflow of data. This is unusual with the MiniPix sensor because of the high readout. There are two possible outputs:
  - index = 0x74: start of lost data
  - index = 0x75: end of lost data, ToA is the length of the missing time.

Figure 4.6 is an example in a text file, but is possible to produce the same file in binary format with the following structure (table 4.1).

Table 4.1: Table of the contents of a binary file with the raw data.

Data	Header Type	Header Description
Matrix Index	u32	unsigned int of 32bits
ToA	u64	unsigned int of 64bits
Overflow	byte	byte or 8bits
FToA	byte	byte or 8bits
ToT	u16	unsigned int of 16bits

From the *Matrix Index* it is possible to extract the coordinates of the pixel (x and y) by following the equations 4.1 and 4.2. The % symbolises the remainder after division and the *Int* is a function to round to the closest integer.

$$X = MatrixIndex\%256 \tag{4.1}$$

$$Y = Int(MatrixIndex/256) \tag{4.2}$$

ToA is the Time of Arrival in steps and to calculate the time of the interaction in nanoseconds follow the formula 4.3. The FToA operation ($FToA * 1.5625$) is not mandatory, only if further precision is required.

$$Time(ns) = ToA * 25 - FToA * 1.5625 \tag{4.3}$$

4.2 Operations modes of the on-board computer

Upon reaching orbit, the payload autonomously boots when the space rider turns on the power supply. The payload boots up and the OBC performs the *WakeUp* procedure by checking the operationality of the equipment. At the end of this procedure, the OBC enters the *Housekeeping mode* and produces a report of the payload status (housekeeping data). If the results of the report are as expected, the OBC autonomously enters the *Observational Mode*; if not, the OBC remains in *Housekeeping Mode* waiting for ground intervention.

Upon entering the *Observational Mode*, the OBC fully turns on the detector unit and begins gathering scientific data. For each event detected, the OBC time-stamps it and saves it to the internal memory. The payload continuously records scientific data regardless of the space rider's orientation. With the attitude data given by the space rider and the cumulative gamma-ray detection, the OBC can identify the gamma-ray sources and perform the required scientific calculations. During the *Observation Mode*, the OBC updates the space rider mass memory unit with housekeeping data to be sent to the ground station. Upon performing scientific calculations, the OBC updates the space rider mass memory unit with scientific data every 24 hours. The OBC operates in this mode the majority of the time.

In case the OBC identifies an anomaly in the normal functioning of the payload subsystems (e.g., overheating, over-current consumption, noisy pixels in the detector), it autonomously switches to the *Debug sub-Mode* where it solves the issue at hand while still taking scientific data. If it is an unknown error, the OBC switches to *Housekeeping Mode* and includes in the housekeeping data an error message explaining the problem to the LIP-GS. The OBC

waits for ground intervention. When the ground message arrives, the OBC enters the *Test Sub-mode* where it tests the new software patch. Before testing the new software version, the OBC saves the current version in the internal memory. If the test is successful, the OBC enters the *Firmware Update sub-Mode* to update the OBC.

The payload is ready, at any given time, for an emergency power cut-off. When it is time to prepare for the space rider's entry the OBC starts to shut down the payload. All the operation modes can be followed by Figure 4.7.

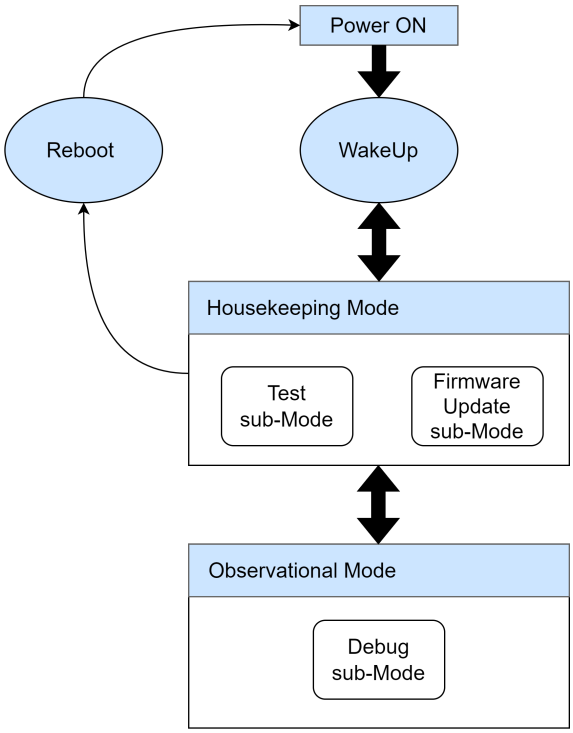

Figure 4.7: In-flight operation modes of the OBC.

In this thesis, the focus is on the *Observational Mode*. Which collects and handles the information from the detector unit.

4.3 Functions to be processed on the on-board computer

An onboard computer is a crucial component of many technological projects, as it allows for the automated processing and analysis of data. The focus of this section is to show some functions that must run on the computer, such as scientific data collection, converting raw data to energy, and spectroscopy.

- **Scientific Data Collection** - The onboard computer is responsible for collecting data from the detector. Its function is to record physics events that enable the mission to achieve its scientific objectives, specifically the detection of photons and particles. The computer stores this data, and it performs basic data processing and analysis tasks, such as data conversion and spectroscopy.

- **Raw Data to Energy Conversion** - Another important function of the onboard computer is the conversion of raw data into energy. In this project, the computer is programmed to convert the data collected from the detector, ToT into energy (keV), that can be used for example to make spectroscopy of the incoming gamma-ray photons or particles.

  The conversion from ToT data into energy in keV is given by Figure 4.8 according to the equation 4.4.

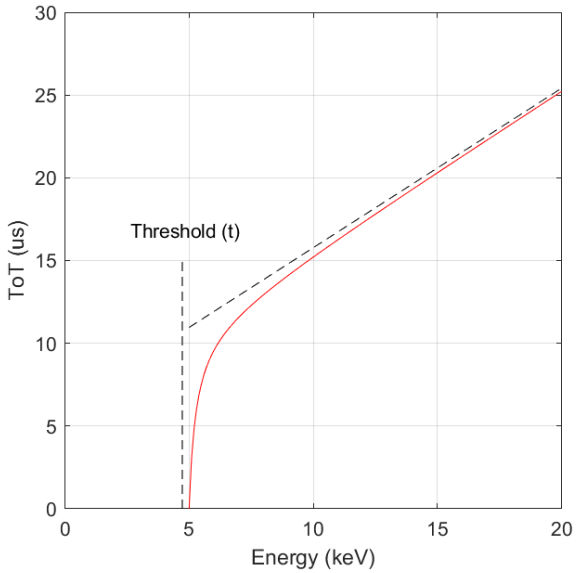

Figure 4.8: Graph of the conversion of ToT to keV.

$$f(x) = ax + b - \frac{c}{x - t} \tag{4.4}$$

Parameters of the equation 4.4 are:

- f(x) is the ToT value;
- x is energy in keV;
- a, b, c, and t are constants. Each pixel has its own values in the configurations of the sensors.

The detector gives ToT data and in order to get the corresponding energy is necessary to rewrite the equation 4.4 in order to x. The equation becomes:

$$x = \frac{at - b + f(x) \pm \sqrt{a^2t^2 + 2abt - 2af(x)t + b^2 + 4ac + f(x)^2 - 2bf(x)}}{2a} \tag{4.5}$$

- **Spectroscopy** - This function, which is constructing a histogram of energies, is responsible for analyzing the energy distribution of the incoming detected gamma-

ray photons or particles. Spectroscopy plays a crucial role in understanding various phenomena, such as the emission mechanisms, source composition, and physical processes occurring in astrophysical objects.

4.4 Summary

This experiment consists of a detector unit with 4 detection planes of MiniPix Tx3 sensors, an onboard computer (OBC), and a storage unit. The sensors collect radiation data from the open sky. OBC handles the sensor data, while the storage unit stores it. The detector has 10^6 pixels that output intensity in response to particle interaction. OBC receives pixel coordinates and intensity data, processes it, and stores the output for post-flight analysis. Using a GPU with the detector's matrix structure for image processing is ideal for a 2D problem.

The detector has three modes: Frame, Integral, and Data-Driven. It also has additional features for clustering and spectral imaging. Cluster addon identifies clusters of interactions. Spectral imaging is a tool capable of creating a histogram of a spectrum of the radiation to be captured.

The detector can provide a variety of information. It can output the coordinate of the active pixel, the timestamp of the acquisition, and the energy of the particle in that pixel. The output file can be in a text file or in binary format.

The OBC has some operations mode to execute. The *Housekeeping Mode* to monitor the operational status of the payload equipment. If everything is normal, it enters the *Observational Mode* and collects scientific data continuously. It can identify gamma-ray sources and perform calculations. Housekeeping data is periodically sent to the ground station, while scientific data is stored every 24 hours. If anomalies occur, it switches to *Debug sub-Mode* to solve the issue, or *Housekeeping Mode* if it's unknown.

This section covers essential functions like data collection, energy conversion, and spectroscopy. The onboard computer collects data from the detector to enable the detection of photons and particles and achieve scientific objectives. The computer analyzes and processes data, and converts it to energy in keV. Spectroscopy analyzes the energy distribution of detected gamma-ray photons or particles.

5

Experimental Data Analysis Results

This chapter covers some results performed on the functions of chapter 4. The function *Scientific Data Collection* shows the time required to read each file created and for *Raw Data to Energy Conversion* and *Spectroscopy* functions, it displays the compute computational and runtime.

5.1 Scientific Data Collection

The data from the detector is stored directly in the memory of the system and only then the OBC can read it to finally began the main data processing. The data can be stored both in a text file or binary file. Figure 5.1 shows the differences between reading a text file and a binary one with raw data from a detector for various power modes. The text file is a recollection of 25 minutes of raw data and 60 minutes for the binary file. Both acquisitions have the Barium (Ba113) source nearby of the detector and were acquired at the LIP laboratory.

Even though the text file is easier for a human to read but for a machine it is more complex and time-consuming. By observing the differences between the bars from the two format files it is obvious that reading the text file is slower than reading the binary one, even though the binary file is bigger. This happens because when reading a text file, the program needs to read each character one at a time and then convert it to its corresponding ASCII value. This process can be slow and time-consuming, especially for large text files. In contrast, binary files store data in a format optimized for quick reading and writing by the computer. Comparing now the heights of the bars in the same file format it is possible to observe the slowest is the 10W power mode because of the lower CPU maximal frequency (table 3.2). There is little difference between MAXN and 15W power modes in reading time because the maximal CPU frequency of both is close.

The files have different sizes not only because of the particular recollection time but also of the format of the file. Figure 5.1 shows contrasts in the sizes of the files.

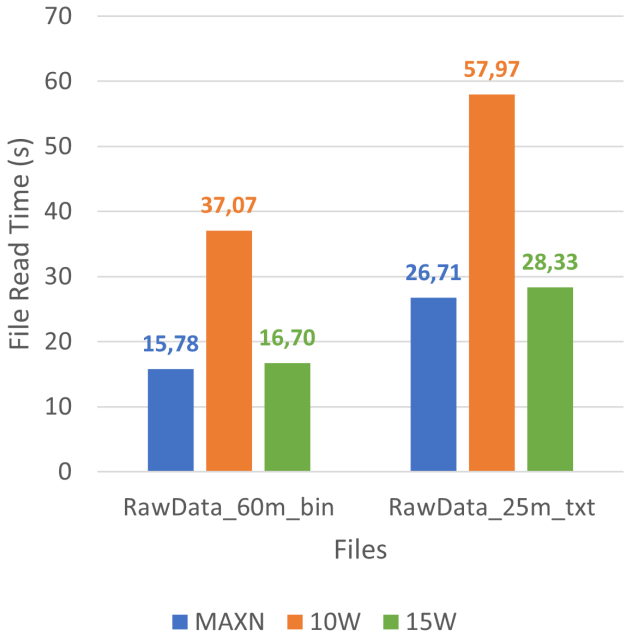

Figure 5.1: Reading times of raw data in seconds for the 60-minute binary data collection file and the 25-minute text file.

Table 5.1: Table of the sizes and number of pixels stored for different formats of files.

File Name	Size (MBytes)	num Pixels	Bytes/Pixel
RawData_60m	1 707,027	111 871 713	16,000
RawData_25m_txt	1 772,223	54 931 729	33,829

According to the table 5.1 the text file size is bigger even with less time for gathering data and consequently it has more bytes per pixel capture which is not desired. So collecting raw data from the detector, which gathers information for several days, in a text file is not practical because it will generate bigger files to store. Also opening an enormous file with an external application with several Gigabytes is not viable and probably the program could not even open because the file is too big.

5.2 Raw Data to Energy Conversion function

This section addresses the runtime and performance of the function that converts the raw data (ToT column in Figure 4.6) to energy data using the approach in section 4.3.

Figure 5.2 represents the time of the execution of the function to raw data transform energy data for different power modes and acquisition times (see table 5.3). From the results of the chapter 3 the first version of this function utilizes only the properties of CUDA cores. As this function uses low memory accesses it is not advantageous to use shared memory. The use of streams is also not feasible for the same reasons discussed in chapter 3. Due to the non-use of matrix operations, using tensor cores is not applicable.

The only possible optimization that is likely to improve the performance is the use of built-in vectors but at the moment, as the THOR project is still susceptible to change, this optimisation has not been implemented.

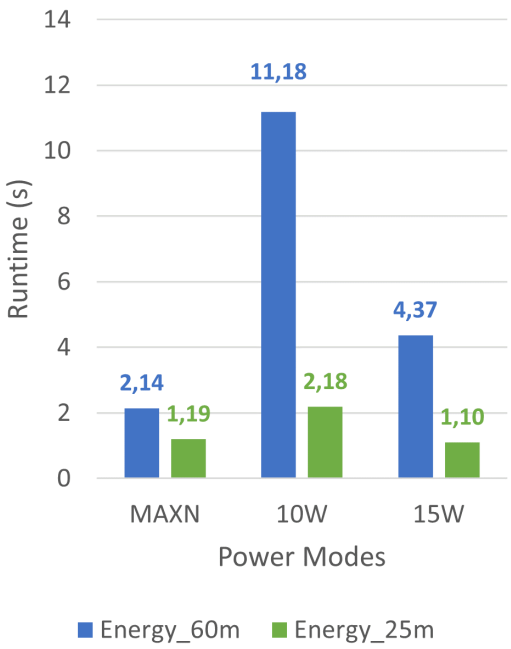

Figure 5.2: Duration in seconds to run the energy conversion function for different power modes and acquisition times (see table 5.3).

As it is expected that the file with 60 minutes of gathering data will take more time than the other one because it has around 57 000 000 more pixels (table 5.1). Therefore, the more pixels it has stored the more it takes to process all the information. Regarding information on power modes, it is predicted that the 10W mode is slower than the other ones because of the lower CPU and GPU maximum frequency.

Figure 5.3 represents the number of operations per second, in GFLOPS, of the energy conversion function for various power modes and raw data files (see table 5.3).

The computational performance is equal for both files. This means it already reaches the peak performance of that function. So the computational performance will not increase if the number of operations increases too. Table 5.2 indicates the number of operations it takes to complete the energy conversion function of each file. Since both files have a very high number of operations the compute performance in Figure 5.3 is stabilized. The MAXN mode has better performance because it has the highest GPU and CPU frequency.

Table 5.2: Table of the number of operations it takes in the energy function.

File Name	Operations
Energy_60m	5 369 842 224
Energy_25m	2 636 722 992

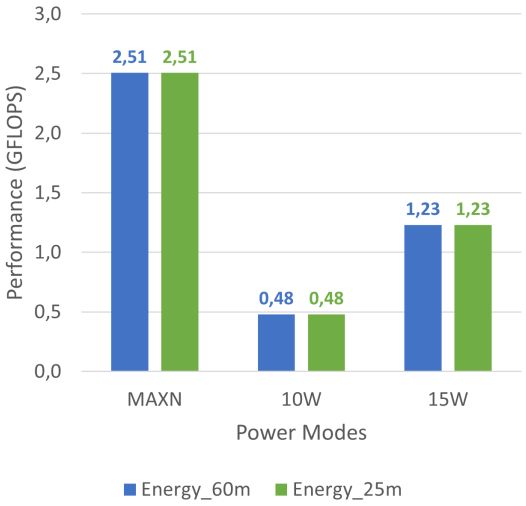

Figure 5.3: Number of operations per second in GFLOPS of the energy conversion function for various power modes and acquisition times (see table 5.3).

Table 5.3: Table of compute performances of conversion energy function with different files and for various power modes.

Name	Energy_60m	Energy_25m
Operations	5 369 842 224	2 636 722 992
MAXN Time (s)	2,143	1,053
MAXN Perf. (GFLOPS)	2,506	2,505
10W Time (s)	11,183	5,491
10W Perf. (GFLOPS)	0,480	0,480
15W Time (s)	4,367	2,144
15W Perf. (GFLOPS)	1,230	1,230

5.3 Spectroscopy function

This section is about the runtime and performance of the spectroscopy function for different files and power modes. Spectroscopy is a histogram of the distribution of the energy capture. This function uses the output data from the energy conversion function to create the histogram.

Figure 5.4 represents the runtime of the spectroscopy function for different power modes and acquisition times (see table 5.4).

The results of the runtime of this function are similar to the energy conversion function. Bigger files bigger execution time. The best result is for MAXN power mode.

Figure 5.5 shows the performance of the function (see table 5.4).

The spectroscopy function mainly runs on the CPU to avoid complex algorithms using CUDA cores. As the execution time is low for the largest file, around 3.91 seconds for 60 minutes of data, there is no guarantee that using these complex algorithms is worth the

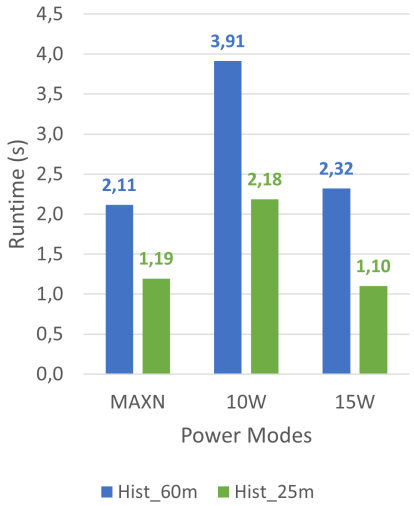

Figure 5.4: Duration in seconds to run the histogram function for different power modes and acquisition times (see table 5.4).

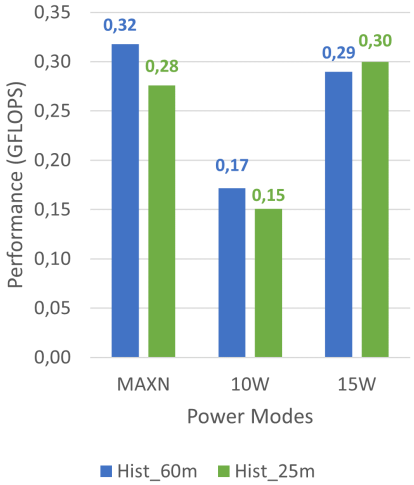

Figure 5.5: Number of operations per second, in GFLOPS, of the spectroscopy function for various power modes and acquisition times (see table 5.4).

Table 5.4: Table of compute performances of spectroscopy function with different files and for various power modes.

Name	Hist_60m	Hist_25m
Operations	671 230 278	329 590 374
MAXN Time (s)	2,112	1,194
MAXN Perf. (GFLOPS)	0,318	0,276
10W Time (s)	3,911	2,184
10W Perf. (GFLOPS)	0,172	0,151
15W Time (s)	2,318	1,099
15W Perf. (GFLOPS)	0,290	0,300

time investment. Thus, the bigger the data lower the performance. In any power mode, this is visible and due to CPU frequency being lower in the 10W mode, the differences between files are bigger.

With the help of *Matlab*, Figures 5.6 and 5.7 show the results of a spectroscopy function for 60 minutes of exposure file. The difference between both is that Figure 5.6 has noisy pixels that return wrong values constantly. In the future, it is important to create a function to remove that kind of noisy information. Not only because the data is wrong, but also because it takes up too much memory.

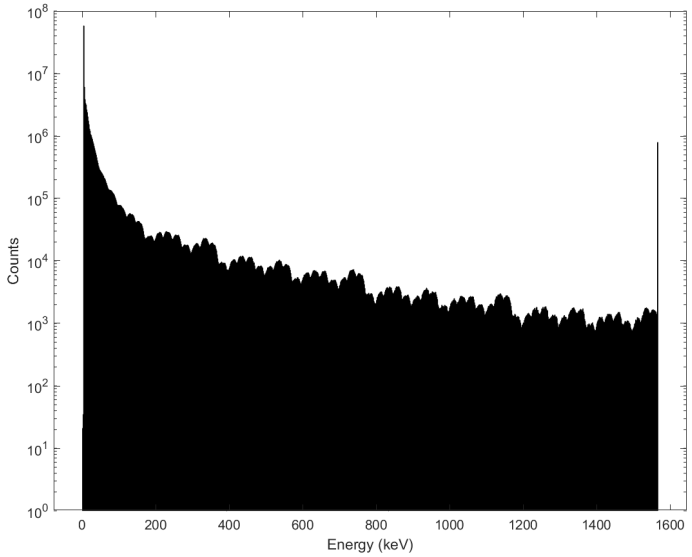

Figure 5.6: Histogram of noisy pixel energies of the 60-minute acquisition file of the Barium (Ba133) source.

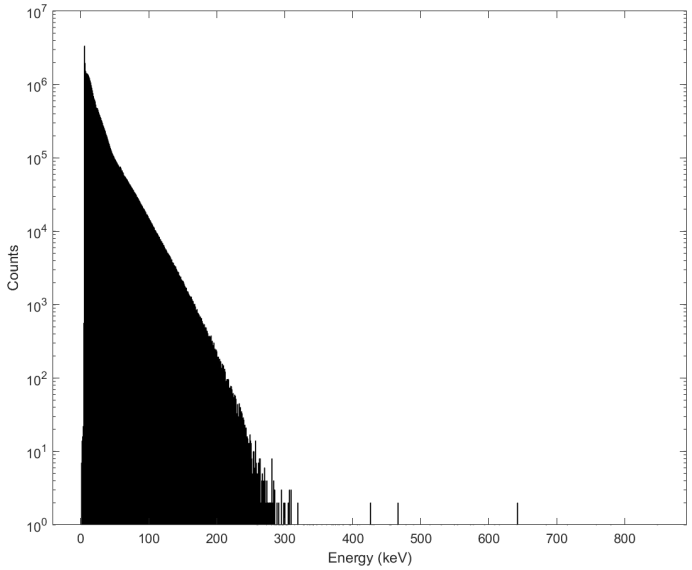

Figure 5.7: Histogram of noisy pixel energies of the 60-minute acquisition file of the Barium (Ba133) source without 3 most noisy pixels.

The removal of the noisiest pixels was also done with the help of *Matlab*, looking for the most hit pixels in the file and removing the top 3.

The number of interactions that were removed for Figure 5.7 was about 75 300 000, which is about 1,12 GB of data. That noisy information represents around 67% of the data. It is crucial to remove that information right from the detector's data. The malfunctioning of certain pixels can be derived from the temperature.

5.4 Summary

This chapter reveals some results performed on the functions, discussed in chapter 4. The function *Scientific Data Collection* displays the time required to read each file created and for *Raw Data to Energy Conversion function* and *Spectroscopy function*, it shows the compute computational and runtime.

The data collected by the detector is stored directly in the system's memory before being read by the OBC for processing. The file size is also affected by the file format. It is not practical to store raw data in text files for long periods as the files can become too large for external applications to handle.

The larger file takes longer to process due to its higher number of pixels. The 10W power mode is slower than the others due to lower CPU and GPU frequencies. The MAXN power mode has the highest GPU and CPU frequency, resulting in better performance.

The spectroscopy function for larger files takes longer to process. The function mainly runs on the CPU, consequently larger data results in lower performance. The noisy information takes excessive memory and represents around 67% of the data, indicating the need to remove it from the detector's data.

6

Conclusions and Future Work

In this Thesis, it becomes evident that high-performance computing (HPC) systems are crucial for processing and managing the large, complex data generated by modern sensor technologies, particularly in space environments. However, these systems come with the challenge of meeting requirements such as power, latency, and temperature limitations. The use of multi-core systems, such as FPGA and GPU, is common in space missions due to their parallel processing capabilities, but they come with high energy demands, which is a critical constraint in space environments.

The study and experiments that came out in this dissertation, allow us to conclude that the Jetson AGX Xavier platform is an effective solution for processing and monitoring data in a resource-restricted space environment. The platform's high-performance capabilities and low power consumption make it suitable for use in space missions with strict constraints on weight, size, power consumption, and heat dissipation. The use of benchmarks and the roofline model provides an effective method of evaluating the platform's performance and power consumption, enabling the development of high-performance workloads. The results obtained from the THOR-SR experiment will be useful in the development of future high-energy astrophysics space telescopes.

6.1 Future Work

In order to increase the computational performances it is worthwhile to investigate the artificial intelligence and deep learning domains in which the GPU specialises. This approach can yield significant improvements in noise detection processing speed and efficiency.

For the THOR-SR project, we will aim at:

- Monitoring the temperature of the system, because it has an impact on the noisy pixels;
- Handling data from *Housekeeping mode*;
- Develop an algorithm to detect and remove efficiently noisy pixels from the detector unit, to minimize valueless information.

References

[1] Joshua Lant, JN Palma, Mikel Lujan, and John Goodacre. Making the case for fpga based hpc. *IEEE Micro*, 40:25–34, 2020.

[2] Claude Tadonki. Conceptual and technical challenges for high performance computing. *arXiv preprint arXiv:2010.02769*, 2020.

[3] Shoaib Kamil, John Shalf, and Erich Strohmaier. Power efficiency in high performance computing. In *2008 IEEE international symposium on parallel and distributed processing*, pages 1–8. IEEE, 2008.

[4] European Space Agency. ESA - Space Rider. `https://www.esa.int/Enabling_Support/Space_Transportation/Space_Rider`.

[5] Thomas Sterling, Maciej Brodowicz, and Matthew Anderson. *High performance computing: modern systems and practices*. Morgan Kaufmann, 2017.

[6] W Hwu Wen-Mei, David B Kirk, and Izzat El Hajj. *Programming Massively Parallel Processors: A Hands-on Approach*. Morgan Kaufmann, 2022.

[7] Al Geist and Daniel A Reed. A survey of high-performance computing scaling challenges. *The International Journal of High Performance Computing Applications*, 31(1):104–113, 2017.

[8] Lech Jóźwiak. Advanced mobile and wearable systems. *Microprocessors and microsystems*, 50:202–221, 2017.

[9] S Madhura. A review on low power vlsi design models in various circuits. *J. Electron*, 4:74–81, 2022.

[10] Cheng Zhuo, Shaoheng Luo, Houle Gan, Jiang Hu, and Zhiguo Shi. Noise-aware dvfs for efficient transitions on battery-powered iot devices. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 39(7):1498–1510, 2019.

[11] Maciej Trochimiuk. FPGA programming - what is it, how it works and where it can be used - codilime, 4 2021. Accessed: 2 Febuary 2022.

[12] Intel. What is a gpu? graphics processing units defined. `https://www.intel.com/content/www/us/en/products/docs/processors/what-is-a-gpu.html`. Accessed: 3 Febuary 2022.

[13] Felipe F Lopes, João Canas Ferreira, and Marcelo AC Fernandes. Parallel implementation on fpga of support vector machines using stochastic gradient descent. *Electronics*, 8(6):631, 2019.

[14] Murad Qasaimeh, Kristof Denolf, Jack Lo, Kees Vissers, Joseph Zambreno, and Phillip H Jones. Comparing energy efficiency of cpu, gpu and fpga implementations for vision kernels. In *2019 IEEE international conference on embedded software and systems (ICESS)*, pages 1–8. IEEE, 2019.

[15] Rainer Keller, David Kramer, and Jan-Philipp Weiss. *Facing the Multicore-Challenge: Aspects of New Paradigms and Technologies in Parallel Computing*, volume 6310. Springer, 2010 edition, 2010.

[16] BERTEN DSP S.L. Gpu vs fpga performance comparison white paper 2. 2016.

[17] Bingzhen Li, Jiaojiao Gu, and Wenzhi Jiang. Artificial intelligence (ai) chip technology review. In *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, pages 114–117. IEEE, 2019.

[18] NVIDIA Corporation. NVIDIA Jetson AGX Xavier, 2018. `https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-agx-xavier/`.

[19] NVIDIA Corporation. NVIDIA Jetson TX2, 2017. `https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-tx2/`.

[20] NVIDIA Corporation. NVIDIA Jetson Nano, 2019. `https://developer.nvidia.com/embedded/jetson-nano-developer-kit`.

[21] Shuichi Asano, Tsutomu Maruyama, and Yoshiki Yamaguchi. Performance comparison of fpga, gpu and cpu in image processing. In *2009 international conference on field programmable logic and applications*, pages 126–131. IEEE, 2009.

[22] Yuexuan Tu, Saad Sadiq, Yudong Tao, Mei-Ling Shyu, and Shu-Ching Chen. A power efficient neural network implementation on heterogeneous fpga and gpu devices. In *2019 IEEE 20th international conference on information reuse and integration for data science (IRI)*, pages 193–199. IEEE, 2019.

[23] Andrew Boutros, Eriko Nurvitadhi, Rui Ma, Sergey Gribok, Zhipeng Zhao, James C Hoe, Vaughn Betz, and Martin Langhammer. Beyond peak performance: Comparing the real performance of ai-optimized fpgas and gpus. In *2020 International Conference on Field-Programmable Technology (ICFPT)*, pages 10–19. IEEE, 2020.

[24] RM Curado Da Silva, E Caroli, JB Stephen, and F Frontera. The polca experiments. *Proceedings of Polarimetry days in Rome: Crab status, theory and prospects*, 2008.

[25] Martin Urban, Ondrej Nentvich, Veronika Stehlikova, Tomas Baca, Vladimir Daniel, and Rene Hudec. Vzlusat-1: Nanosatellite with miniature lobster eye x-ray telescope and qualification of the radiation shielding composite for space application. *Acta Astronautica*, 140:96–104, 2017.

[26] W B Atwood et. al. The large area telescope on the fermi gamma-ray space telescope mission. *The Astrophysical Journal*, 697:1071, 5 2009.

[27] C.C. DeBoy, C.B. Haskins, T.A. Brown, R.C. Schulze, M.A. Bernacik, J.R. Jensen, W. Millard, D. Duven, and S. Hill. The rf telecommunications system for the new horizons mission to pluto. In *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720)*, volume 3, page 1478 Vol.3, March 2004.

[28] A Elshazly, A Elliethy, and MA Elshafey. Tactics overview for implementing high-performance computing on embedded platforms. In *IOP Conference Series: Materials Science and Engineering*, volume 1172, page 012034. IOP Publishing, 2021.

[29] Stefano Esposito, Cristian Albanese, Monica Alderighi, Fabio Casini, Luca Giganti, Maria Livia Esposti, Claudio Monteleone, and Massimo Violante. Cots-based high-performance computing for space applications. *IEEE Transactions on Nuclear Science*, 62(6):2687–2694, 2015.

[30] Heather Quinn, William H Robinson, Paolo Rech, Miguel Aguirre, Arno Barnard, Marco Desogus, Luis Entrena, Mario Garcia-Valderas, Steven M Guertin, David Kaeli, et al. Using benchmarks for radiation testing of microprocessors and fpgas. *IEEE Transactions on Nuclear Science*, 62(6):2547–2554, 2015.

[31] Pradeep Gupta. CUDA refresher: The CUDA programming model | nvidia technical blog. `https://developer.nvidia.com/blog/cuda-refresher-cuda-programming-model/`, 6 2020. Accessed: 17 December 2022.

[32] NVIDIA Corporation. CUDA C++ Programming Guide, 2022. `https://docs.nvidia.com/cuda/cuda-c-programming-guide/`.

[33] Dustin Franklin. Nvidia jetson agx xavier delivers 32 teraops for new era of ai in robotics | nvidia technical blog, 2018. `https://developer.nvidia.com/blog/nvidia-jetson-agx-xavier-32-teraops-ai-robotics/`.

[34] Dustin Franklin. Jetson agx xavier and the new era of autonomous machines, 2018. `https://info.nvidia.com/rs/156-OFN-742/images/Jetson_AGX_Xavier_New_Era_Autonomous_Machines.pdf`.

[35] Samuel Williams. Roofline: An insightful visual performance model for floating-point programs and multicore. *ACM Communications*, page 16, 2009.

[36] Samuel Williams, Andrew Waterman, and David Patterson. Roofline: an insightful visual performance model for multicore architectures. *Communications of the ACM*, 52(4):65–76, 2009.

[37] Aleksandar Ilic, Frederico Pratas, and Leonel Sousa. Cache-aware roofline model: Upgrading the loft. *IEEE Computer Architecture Letters*, 13(1):21–24, 2013.

[38] Dhruv Singal. N ways to saxpy: Demonstrating the breadth of gpu programming options | nvidia technical blog, 4 2021. `https://developer.nvidia.com/blog/n-ways-to-saxpy-demonstrating-the-breadth-of-gpu-programming-options/`.

[39] Bastian Hagedorn, Larisa Stoltzfus, Michel Steuwer, Sergei Gorlatch, and Christophe Dubach. High performance stencil code generation with lift. In *Proceedings of the 2018 International Symposium on Code Generation and Optimization*, pages 100–112, 2018.

[40] Gauthier Sornet, Fabrice Dupros, and Sylvain Jubertie. A multi-level optimization strategy to improve the performance of stencil computation. *Procedia Computer Science*, 108:1083–1092, 2017.

[41] NVIDIA Corporation. NVIDIA Tensor Cores, 2022. `https://developer.nvidia.com/tensor-cores`.

[42] ADVACAM s.r.o. ADVACAM official website, 2023. `https://advacam.com/`.

[43] Xavier Llopart, Rafael Ballabriga, Michael Campbell, Lukas Tlustos, and Winnie Wong. Timepix, a 65k programmable pixel readout chip for arrival time, energy and/or photon counting measurements. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 581(1-2):485–494, 2007.

[44] Carlos Granja, Jan Jakubek, Ulli Köster, Michal Platkevic, and Stanislav Pospisil. Response of the pixel detector timepix to heavy ions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 633:S198–S202, 2011.

[45] Jan Jakubek, Andrea Cejnarova, Tomas Holy, Stanislav Pospisil, Josef Uher, and Zdenek Vykydal. Pixel detectors for imaging with heavy charged particles. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 591(1):155–158, 2008.

[46] ADVACAM s.r.o. *MNXT3S-Xxx220520 - MiniPIX TPX3 Standard Datasheet*, 2022.

Appendices

A

GPU benchmarking for different optimization levels

This appendix contains all the tables used to create the graphs throughout the chapter 3.

The tables contain a few abbreviations in order to keep the overview concise. [Perf] refers to computational performance and [DIM] to dimension. The best performance is highlighted.

Table A.1: Table for original SAXPY algorithm configuration.

Name	DIM	Num Blocks	Threads/Block	Array Size (KBytes)	Time (ms)	Perf (GFLOPs)
saxpy	1 024	1	1 024	4	0,009	0,225
saxpy	2 048	2	1 024	8	0,009	0,441
saxpy	4 096	4	1 024	16	0,011	0,753
saxpy	8 192	8	1 024	32	0,010	1,668
saxpy	16 384	16	1 024	64	0,013	2,573
saxpy	32 768	32	1 024	128	0,016	4,008
saxpy	65 536	64	1 024	256	0,024	5,520
saxpy	131 072	128	1 024	512	0,040	6,591
saxpy	262 144	256	1 024	1 024	0,070	7,509
saxpy	524 288	512	1 024	2 048	0,120	8,713
saxpy	1 048 576	1 024	1 024	4 096	0,228	9,214
saxpy	2 097 152	2 048	1 024	8 192	0,431	9,734
saxpy	4 194 304	4 096	1 024	16 384	0,846	9,917
saxpy	8 388 608	8 192	1 024	32 768	1,728	9,710
saxpy	16 777 216	16 384	1 024	65 536	3,552	9,447
saxpy	33 554 432	32 768	1 024	131 072	7,221	9,294
saxpy	67 108 864	65 536	1 024	262 144	14,768	9,089
saxpy	134 217 728	131 072	1 024	524 288	29,358	9,144
saxpy	268 435 456	262 144	1 024	1 048 576	58,222	9,221
saxpy	536 870 912	524 288	1 024	2 097 152	116,242	9,237

Table A.2: Table for SAXPY algorithm with 64 operations per cycle configuration.

Name	Dim	Num Blocks	Threads/Block	Array Size (KBytes)	Time (ms)	Perf (GFLOPs)
saxpy64	1 024	1	1 024	4	0,023	2,868
saxpy64	2 048	2	1 024	8	0,023	5,619
saxpy64	4 096	4	1 024	16	0,023	11,315
saxpy64	8 192	8	1 024	32	0,024	22,081
saxpy64	16 384	16	1 024	64	0,040	26,027
saxpy64	32 768	32	1 024	128	0,070	29,871
saxpy64	65 536	64	1 024	256	0,131	31,961
saxpy64	131 072	128	1 024	512	0,254	33,016
saxpy64	262 144	256	1 024	1 024	0,502	33,392
saxpy64	524 288	512	1 024	2 048	1,007	33,316
saxpy64	1 048 576	1 024	1 024	4 096	1,976	33,954
saxpy64	2 097 152	2 048	1 024	8 192	3,944	34,028
saxpy64	4 194 304	4 096	1 024	16 384	7,876	34,085
saxpy64	8 388 608	8 192	1 024	32 768	15,782	34,019
saxpy64	16 777 216	16 384	1 024	65 536	31,685	33,888
saxpy64	33 554 432	32 768	1 024	131 072	63,450	33,845
saxpy64	67 108 864	65 536	1 024	262 144	127,250	33,752
saxpy64	134 217 728	131 072	1 024	524 288	254,353	33,772
saxpy64	268 435 456	262 144	1 024	1 048 576	508,418	33,791
saxpy64	536 870 912	524 288	1 024	2 097 152	1016,687	33,796

Table A.3: Table for SAXPY algorithm with 128 operations per cycle configuration.

Name	Dim	Num Blocks	Threads/Block	Array Size (KBytes)	Time (ms)	Perf (GFLOPs)
saxpy128	1 024	1	1 024	4	0,037	3,549
saxpy128	2 048	2	1 024	8	0,037	7,002
saxpy128	4 096	4	1 024	16	0,038	13,956
saxpy128	8 192	8	1 024	32	0,038	27,421
saxpy128	16 384	16	1 024	64	0,069	30,610
saxpy128	32 768	32	1 024	128	0,127	33,158
saxpy128	65 536	64	1 024	256	0,244	34,389
saxpy128	131 072	128	1 024	512	0,481	34,885
saxpy128	262 144	256	1 024	1 024	0,956	35,088
saxpy128	524 288	512	1 024	2 048	1,895	35,413
saxpy128	1 048 576	1 024	1 024	4 096	3,780	35,504
saxpy128	2 097 152	2 048	1 024	8 192	7,554	35,535
saxpy128	4 194 304	4 096	1 024	16 384	15,094	35,568
saxpy128	8 388 608	8 192	1 024	32 768	30,239	35,508
saxpy128	16 777 216	16 384	1 024	65 536	60,574	35,452
saxpy128	33 554 432	32 768	1 024	131 072	121,244	35,424
saxpy128	67 108 864	65 536	1 024	262 144	242,853	35,371
saxpy128	134 217 728	131 072	1 024	524 288	485,571	35,381
saxpy128	268 435 456	262 144	1 024	1 048 576	971,018	35,385
saxpy128	536 870 912	524 288	1 024	2 097 152	1941,874	35,388

Table A.4: Table for SAXPY algorithm with 64 vector operations per cycle configuration.

Name	Dim	Num Blocks	Threads/Block	Array Size (KBytes)	Time (ms)	Perf (GFLOPs)
Vec64	1 024	1	1 024	16	0,090	2,916
Vec64	2 048	2	1 024	32	0,093	5,654
Vec64	4 096	4	1 024	64	0,093	11,268
Vec64	8 192	8	1 024	128	0,095	22,088
Vec64	16 384	16	1 024	256	0,115	36,633
Vec64	32 768	32	1 024	512	0,141	59,470
Vec64	65 536	64	1 024	1 024	0,167	100,285
Vec64	131 072	128	1 024	2 048	0,214	156,644
Vec64	262 144	256	1 024	4 096	0,317	211,983
Vec64	524 288	512	1 024	8 192	0,373	359,779
Vec64	1 048 576	1 024	1 024	16 384	0,578	464,408
Vec64	2 097 152	2 048	1 024	32 768	0,982	546,952
Vec64	4 194 304	4 096	1 024	65 536	1,815	591,549
Vec64	8 388 608	8 192	1 024	131 072	3,537	607,221
Vec64	16 777 216	16 384	1 024	262 144	6,797	631,889
Vec64	33 554 432	32 768	1 024	524 288	13,625	630,441
Vec64	67 108 864	65 536	1 024	1 048 576	26,879	639,156
Vec64	134 217 728	131 072	1 024	2 097 152	54,184	634,126
Vec64	268 435 456	262 144	1 024	4 194 304	108,321	634,406

Table A.5: Table for SAXPY algorithm with 128 vector operations per cycle configuration.

Name	Dim	Num Blocks	Threads/Block	Array Size (KBytes)	Time (ms)	Perf (GFLOPs)
Vec128	1 024	1	1 024	16	0,099	5,277
Vec128	2 048	2	1 024	32	0,088	11,881
Vec128	4 096	4	1 024	64	0,097	21,523
Vec128	8 192	8	1 024	128	0,127	33,099
Vec128	16 384	16	1 024	256	0,122	68,822
Vec128	32 768	32	1 024	512	0,133	126,334
Vec128	65 536	64	1 024	1 024	0,168	199,501
Vec128	131 072	128	1 024	2 048	0,215	312,076
Vec128	262 144	256	1 024	4 096	0,287	467,697
Vec128	524 288	512	1 024	8 192	0,418	641,674
Vec128	1 048 576	1 024	1 024	16 384	0,687	781,899
Vec128	2 097 152	2 048	1 024	32 768	1,172	916,437
Vec128	4 194 304	4 096	1 024	65 536	2,146	1000,729
Vec128	8 388 608	8 192	1 024	131 072	4,073	1054,615
Vec128	16 777 216	16 384	1 024	262 144	8,023	1070,710
Vec128	33 554 432	32 768	1 024	524 288	15,880	1081,884
Vec128	67 108 864	65 536	1 024	1 048 576	31,628	1086,374
Vec128	134 217 728	131 072	1 024	2 097 152	62,457	1100,265
Vec128	268 435 456	262 144	1 024	4 194 304	125,255	1097,269

Table A.6: Table for SAXPY algorithm with 64 vector operations per cycle configuration using shared memory.

Name	Dim	Num Blocks	Threads/Block	Array Size (KBytes)	Time (ms)	Perf (GFLOPs)
Vec64_Sh	1 024	1	1 024	16	0,088	2,974
Vec64_Sh	2 048	2	1 024	32	0,087	6,052
Vec64_Sh	4 096	4	1 024	64	0,087	12,012
Vec64_Sh	8 192	8	1 024	128	0,104	20,097
Vec64_Sh	16 384	16	1 024	256	0,152	27,530
Vec64_Sh	32 768	32	1 024	512	0,120	69,998
Vec64_Sh	65 536	64	1 024	1 024	0,159	105,831
Vec64_Sh	131 072	128	1 024	2 048	0,187	179,797
Vec64_Sh	262 144	256	1 024	4 096	0,243	276,341
Vec64_Sh	524 288	512	1 024	8 192	0,359	373,458
Vec64_Sh	1 048 576	1 024	1 024	16 384	0,574	467,931
Vec64_Sh	2 097 152	2 048	1 024	32 768	1,006	533,830
Vec64_Sh	4 194 304	4 096	1 024	65 536	1,847	581,210
Vec64_Sh	8 388 608	8 192	1 024	131 072	3,580	599,872
Vec64_Sh	16 777 216	16 384	1 024	262 144	7,000	613,572
Vec64_Sh	33 554 432	32 768	1 024	524 288	13,826	621,273
Vec64_Sh	67 108 864	65 536	1 024	1 048 576	27,398	627,055
Vec64_Sh	134 217 728	131 072	1 024	2 097 152	55,237	622,037
Vec64_Sh	268 435 456	262 144	1 024	4 194 304	110,828	620,053

Table A.7: Table for SAXPY algorithm with 128 vector operations per cycle configuration using shared memory.

Name	Dim	Num Blocks	Threads/Block	Array Size (KBytes)	Time (ms)	Perf (GFLOPs)
Vec128_Sh	1 024	1	1 024	16	0,091	5,749
Vec128_Sh	2 048	2	1 024	32	0,102	10,250
Vec128_Sh	4 096	4	1 024	64	0,095	22,186
Vec128_Sh	8 192	8	1 024	128	0,107	39,196
Vec128_Sh	16 384	16	1 024	256	0,109	77,215
Vec128_Sh	32 768	32	1 024	512	0,142	118,537
Vec128_Sh	65 536	64	1 024	1 024	0,165	203,963
Vec128_Sh	131 072	128	1 024	2 048	0,231	291,069
Vec128_Sh	262 144	256	1 024	4 096	0,287	468,062
Vec128_Sh	524 288	512	1 024	8 192	0,405	662,660
Vec128_Sh	1 048 576	1 024	1 024	16 384	0,696	771,153
Vec128_Sh	2 097 152	2 048	1 024	32 768	1,242	864,448
Vec128_Sh	4 194 304	4 096	1 024	65 536	2,308	930,530
Vec128_Sh	8 388 608	8 192	1 024	131 072	4,519	950,476
Vec128_Sh	16 777 216	16 384	1 024	262 144	8,892	966,050
Vec128_Sh	33 554 432	32 768	1 024	524 288	17,592	976,566
Vec128_Sh	67 108 864	65 536	1 024	1 048 576	34,866	985,480
Vec128_Sh	134 217 728	131 072	1 024	2 097 152	69,739	985,374
Vec128_Sh	268 435 456	262 144	1 024	4 194 304	138,251	994,130

Table A.8: Table for Stencil algorithm with 64 vector operations per cycle configuration using shared memory.

GPU Perf		Radius							
Stencil Vec64_Sh	Dim Array	3	8	12	16	20	24	28	32
	1 024	118,725	182,522	210,700	223,234	234,057	239,218	247,584	252,241
	2 048	230,297	349,910	413,320	441,365	457,280	472,523	488,947	499,747
	4 096	411,806	647,740	794,568	823,567	887,963	915,412	939,525	981,982
	8 192	581,435	844,024	920,449	1015,347	1043,081	1076,883	1111,771	1156,308
	16 384	805,535	1020,716	1101,445	1159,929	1179,015	1202,045	1194,039	1207,438
	32 768	960,737	1153,922	1201,393	1230,723	1239,094	1259,072	1258,821	1274,161
	65 536	1096,509	1232,254	1197,551	1208,037	1287,174	1216,849	1301,643	1227,354
	131 072	1180,449	1235,414	1288,620	1298,229	1307,730	1312,763	1313,197	1279,856
	262 144	1216,647	1289,668	1301,673	1294,974	1314,668	1318,811	1320,654	1301,211
	524 288	1238,093	1298,593	1311,327	1315,441	1312,010	1323,422	1315,096	1313,992
	1 048 576	1253,528	1302,400	1314,516	1314,554	1322,478	1324,950	1326,104	1320,426
	2 097 152	1246,834	1303,394	1308,589	1313,403	1318,519	1321,187	1323,430	1320,127
	4 194 304	1221,233	1287,892	1299,979	1307,536	1313,972	1317,637	1319,992	1317,463
	8 388 608	1210,549	1285,747	1297,179	1305,431	1311,878	1316,028	1317,827	1317,160
	16 777 216	1195,647	1278,692	1292,204	1301,380	1308,951	1313,488	1316,151	1315,665
	33 554 432	1197,098	1280,188	1293,149	1302,280	1309,591	1313,727	1316,400	1315,966
	67 108 864	1198,323	1280,928	1293,639	1302,550	1309,864	1313,895	1316,540	1316,166
	134 217 728	1199,929	1281,231	1294,063	1303,135	1309,899	1314,297	1316,789	1316,607
	268 435 456	1264,204	1310,789	1318,466	1321,699	1325,236	1327,022	1327,793	1326,504

Table A.9: Table for Stencil algorithm with 128 vector operations per cycle configuration using shared memory.

GPU Perf Stencil Vec128_Sh		Radius							
		3	8	12	16	20	24	28	32
Dim Array	1 024	172,983	236,241	254,726	270,471	276,780	284,183	288,059	290,022
	2 048	329,090	461,904	507,874	533,470	547,692	558,481	567,541	574,257
	4 096	653,493	878,637	967,178	1034,779	1064,149	1097,493	1118,094	1130,531
	8 192	797,830	1050,058	1144,534	1197,833	1215,002	1232,731	1248,513	1270,456
	16 384	1029,169	1205,749	1239,100	1247,765	1284,712	1295,517	1303,516	1298,138
	32 768	1137,990	1263,524	1309,018	1309,430	1326,165	1330,198	1341,191	1342,316
	65 536	1250,431	1311,684	1339,931	1269,137	1343,068	1353,144	1279,901	1279,447
	131 072	1283,559	1338,524	1353,805	1316,079	1355,518	1362,672	1323,344	1325,608
	262 144	1309,200	1349,289	1360,074	1340,487	1345,517	1346,442	1366,816	1366,593
	524 288	1324,228	1355,470	1364,667	1355,068	1356,994	1359,090	1360,056	1359,194
	1 048 576	1330,123	1354,691	1366,396	1361,077	1362,518	1365,045	1365,797	1365,167
	2 097 152	1320,640	1352,046	1361,882	1360,494	1363,506	1367,665	1366,631	1365,165
	4 194 304	1304,500	1345,543	1358,437	1357,554	1361,308	1365,829	1365,863	1366,171
	8 388 608	1279,629	1344,088	1350,748	1353,058	1357,874	1365,019	1365,907	1364,080
	16 777 216	1288,795	1339,966	1353,699	1355,496	1359,679	1363,284	1364,914	1365,434
	33 554 432	1290,073	1341,173	1354,041	1356,089	1360,336	1363,715	1365,335	1365,795
	67 108 864	1291,027	1341,315	1354,396	1356,198	1360,611	1363,880	1365,341	1365,965
	134 217 728	1294,753	1342,683	1354,716	1357,289	1361,207	1364,156	1365,695	1366,235
	268 435 456	1337,980	1361,869	1368,677	1366,914	1369,271	1371,248	1371,915	1371,470

Table A.10: Table of streaming with 128 vector operations per cycle configuration using shared memory

GPU Perf Vec128_Sh		Num Streams				
Dim Array	1	2	4	8	16
1 024	290,022	147,277	251,704	547,537	142,641
2 048	574,257	273,137	292,853	677,887	288,022
4 096	1130,531	573,175	528,516	452,597	377,311
8 192	1270,456	785,152	817,628	480,442	385,471
16 384	1298,138	1018,735	1022,893	890,016	453,102
32 768	1342,316	1144,196	1145,080	1153,569	1100,414
65 536	1279,447	1240,060	1184,317	1186,730	1188,074
131 072	1325,608	1248,751	1285,092	1302,778	1268,057
262 144	1366,593	1312,272	1326,711	1333,003	1321,521
524 288	1359,194	1316,175	1328,783	1344,196	1338,064
1 048 576	1365,167	1314,030	1345,567	1346,701	1348,713
2 097 152	1365,165	1314,748	1340,066	1350,243	1353,398
4 194 304	1366,171	1313,863	1337,885	1350,651	1355,516
8 388 608	1364,080	1313,725	1337,982	1349,111	1356,116
16 777 216	1365,434	1315,276	1339,615	1349,410	1355,211
33 554 432	1365,795	1316,066	1340,289	1350,045	1355,260
67 108 864	1365,965	1316,675	1340,803	1350,801	1355,612
134 217 728	1366,235	1317,098	1341,025	1350,921	1355,942
268 435 456	1371,470	1317,180	1341,161	1351,033	1355,989

Table A.11: Table A.11 shows the stats of the multiplication of matrices using tensor cores. Matrices have MxN and NxK dimensions. The table displays matrix size in kilobytes, the number of operations of that multiplication, the time to execute the task and the corresponding performance in GFLOPS.

M	N	K	Total Mat Size (KBytes)	Op	Time (ms)	Perf (GFLOPs)
64	64	64	48	524288	0,036	14,385
64	64	96	64	786432	0,039	20,277
64	96	96	84	1179648	0,038	30,720
96	96	96	108	1769472	0,045	39,753
96	96	128	132	2359296	0,067	35,310
96	128	128	160	3145728	0,049	64,504
128	128	128	192	4194304	0,055	75,896
128	128	256	320	8388608	0,075	111,125
128	256	256	512	16777216	0,116	144,075
256	256	256	768	33554432	0,199	169,016

Table A.12: Table of Stencil with 128 vector operations per cycle configuration using shared memory with power mode 10W.

GPU Perf 10W Vec128_Sh		Radius							
		3	8	12	16	20	24	28	32
Dim Array	1 024	26,703	102,062	109,695	115,110	117,110	119,024	120,424	122,114
	2 048	53,368	199,304	214,226	224,485	230,642	234,604	238,085	241,023
	4 096	72,957	226,124	237,312	241,237	244,849	248,127	250,809	251,066
	8 192	106,377	242,673	249,243	249,215	253,536	254,822	255,650	256,555
	16 384	149,091	249,312	255,421	228,228	256,104	258,577	259,103	258,953
	32 768	239,745	253,676	258,897	243,625	259,671	260,356	260,708	260,650
	65 536	245,642	256,708	260,373	259,440	259,112	260,960	260,843	261,507
	131 072	248,929	257,868	261,085	260,384	261,012	261,690	261,770	261,776
	262 144	250,693	258,832	261,133	258,887	261,210	261,863	261,956	259,858
	524 288	251,554	259,108	261,287	259,775	260,484	261,960	262,042	261,016
	1 048 576	252,345	259,269	261,338	260,829	261,536	262,024	262,080	261,584
	2 097 152	252,517	259,446	261,380	260,728	261,132	262,038	262,074	261,799
	4 194 304	251,812	259,080	261,368	260,849	261,396	261,923	261,948	261,850
	8 388 608	251,543	258,915	261,169	260,844	261,374	261,805	261,955	261,950
	16 777 216	250,861	258,936	260,949	260,734	261,238	261,823	261,866	261,930
	33 554 432	250,726	258,793	261,024	260,809	261,321	261,828	261,899	261,718
	67 108 864	250,965	258,821	261,053	260,805	261,297	261,825	261,905	261,918
	134 217 728	251,047	258,884	261,086	260,817	261,312	261,829	261,910	261,928
	268 435 456	252,731	259,524	261,569	261,186	261,605	262,077	262,125	262,116

Table A.13: Table of Stencil with 128 vector operations per cycle configuration using shared memory with power mode 15W.

GPU Perf 15W Vec128_Sh		Radius							
		3	8	12	16	20	24	28	32
Dim Array	1 024	42,242	130,642	140,659	147,725	151,090	153,590	155,726	157,143
	2 048	79,534	253,207	276,944	290,138	298,155	303,350	307,605	313,224
	4 096	143,136	502,305	534,726	574,267	581,345	596,446	606,815	612,927
	8 192	449,537	570,900	597,956	614,138	619,404	629,536	637,793	643,578
	16 384	517,487	616,553	632,099	637,773	644,436	652,431	656,541	660,825
	32 768	577,773	634,460	654,738	651,756	657,043	666,099	667,853	665,353
	65 536	580,240	619,532	666,152	628,506	630,718	671,690	672,936	633,069
	131 072	627,353	660,618	650,926	673,761	651,768	674,711	675,552	675,338
	262 144	639,766	665,663	672,268	664,554	663,779	676,111	676,621	666,137
	524 288	647,091	668,534	674,341	668,349	669,958	676,847	671,930	671,927
	1 048 576	649,747	669,295	674,688	671,887	675,609	677,168	674,802	674,761
	2 097 152	651,919	667,635	674,284	673,851	674,654	676,043	677,555	675,618
	4 194 304	645,709	667,660	673,683	672,808	674,834	676,387	676,817	676,193
	8 388 608	642,915	666,724	673,159	672,868	674,462	676,140	676,517	676,243
	16 777 216	637,875	665,260	672,175	672,250	673,763	675,637	675,808	676,053
	33 554 432	639,107	665,406	672,638	672,409	674,018	675,674	676,104	676,178
	67 108 864	638,900	665,563	672,694	672,532	674,053	675,745	676,094	676,254
	134 217 728	639,117	665,616	672,722	672,553	674,085	675,779	676,152	676,292
	268 435 456	650,169	670,642	676,242	675,192	676,218	677,549	677,677	677,658

Table A.14: Roofline of the best results of each optimization strategy that uses the *saxpy* or *stencil* algorithms as a baseline.

Name	Dim	Array Size (KBytes)	Perf (GFLOPs)	AI (FLOPs/Byte)
Saxpy	4 194 304	16 384	9,917	0,167
Saxpy64	4 194 304	16 384	34,085	5,333
Vec64	67 108 864	1 048 576	639,156	5,333
Vec64_Sh	67 108 864	1 048 576	627,055	5,333
Vec64_Sh_R28	268 435 456	4 194 304	1327,793	5,333
Saxpy128	4 194 304	16 384	35,568	10,667
Vec128	134 217 728	2 097 152	1100,265	10,667
Vec128_Sh	268 435 456	4 194 304	994,130	10,667
Vec128_Sh_R28	268 435 456	4 194 304	1371,915	10,667
Vec128_Sh_R32_Stream16	8 388 608	131 072	1356,116	10,667

B

PRODEX project approved by ESA

Programme de Dévéloppement d'Éxperiences Scientifiques (PRODEX) is a program of the ESA in which participating member countries that do not have a significant space agency are involved. ESA member countries that subscribe to PRODEX pay an annual fee and then receive that money in ESA projects in which they participate.

GUIDELINES TO ESTABLISHING A PRODEX PROJECT PROPOSAL

A. Identification of the proposal

1. Title/Acronym: **T**GF and **H**igh-energy astrophysics **O**bservatory for gamma-**R**ays on board the **S**pace **R**ider (**THOR**-**SR**)
2. Abstract (max. 5 lines):

   THOR-SR is a high-energy astrophysics pathfinder mission whose scientific payload is based on CdTe and Si detectors to be launched on board the Space Rider. It will address gamma-ray astrophysics, space weather and TGF monitoring, by performing spectroscopy, imaging, time variability and polarimetry. It is a large field of view instrument (up to 2π sr) provided by a 4 layers CdTe stack detector, operating in the 0.1–10 MeV energy band.
3. Identification of the overall project selected/endorsed by ESA (Announcement of Opportunity or none):

   Announcement of Opportunity: Workshop "From Science to Business – How to Contribute to and Profit from Space Rider Payloads", February 3rd, 2022, Portugal;

   Original identification in March 29th 2021 Notice of Intent: CdTe Detector for High-Energy Astrophysics and TGF Aviation Safety and Science**;**

   Current Identification: TGF and High-energy astrophysics Observatory for gamma-Rays on board the Space Rider.
4. Role of the PI and of each co-I in the overall project:

   Rui Miguel Curado da Silva (PI): Project Management and Work Packages (WPs) supervision.

   Jorge Maia-Pereira (Co-I): Supervision of experimental and data analysis WPs.
5. Satellite(s) or flight opportunity(ies) (+ date(s) selected/endorsed by ESA):

   Space Rider maiden flight selected by ESA in may 2021.
6. Starting and ending dates of:
   a. the overall project selected/endorsed by ESA: 01/01/2022
   b. the requested PRODEX Experiment Arrangement (PEA): 01/10/2022
7. Previous PEA's (reference, starting and ending dates, budget) related to the project (If applicable): None.
8. Field of research: Physics - Astrophysics.

B. Identification of the Participating State's Principal Investigator (PI)

1. Title, name, surname: Prof. Rui Curado da Silva
2. Institute: LIP–Laboratório de Instrumentação e Física Experimental de Partículas
3. Address: LIP-Coimbra, Departamento de Física, Universidade de Coimbra, Rua Larga

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

3004-516 Coimbra, Portugal

4. Tel.: (secretariat: + 351 23 941 06 55)
5. E-mail: rui.silva@coimbra.lip.pt
6. Website: https://www.lip.pt/?section=about&poles&page=coimbra
7. Institute Head, endorsing this Project Proposal: Profª Isabel Lopes

C. Identification of the Participating State's Co-Investigator(s) (if applicable)

1. Title, name, surname: Prof. Jorge Maia-Pereira
2. Institute: University of Beira Interior, Physics Department
3. Research unit: LIP–Laboratório de Instrumentação e Física Experimental de Partículas
4. Address: Departamento de Física, Universidade da Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal
5. Tel.: (secretariat: + 351 275329132)
6. E-mail: jmaia@ubi.pt
7. Website: http://www.ubi.pt/
8. Institute Head, endorsing this Project Proposal: Profª Sílvia Socorro

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

D. Proposal description and motivation (max. 3 pages)

1. Objectives of the activity/Description of the problem;

The SR (Space Rider) programme provides a new opportunity to test the operation of scientific instruments in space at the highest TRLs (> 7). In this experiment proposal, we address technological and scientific objectives within the following research lines of our group:

i) High-energy astrophysics: spectroscopy, imaging, polarimetry and time variability of the most intense gamma-ray sources in the sky (Crab Nebula and GRBs). An instrument operating in all-sky mode as a polarimeter will be a premiere in a space mission;

ii) Terrestrial gamma-ray flashes (TGFs) science: record TGFs emissions (spectroscopy, imaging and time variability) to evaluate the potential of CdTe pixelated detector as a TGF monitor. The ultimate goal is the development of a commercial product for aviation safety to alert and assess the risks associated with TGF emissions for passengers and crew members;

iii) Space radiation in low earth orbit (LEO) and space weather: record protons and electrons around the the Van Allen belts (particle flux vs. energy and particle flux vs. orbit time); record the protons from solar proton events (proton flux vs. energy and proton flux vs. time).

2. Rationale.

LIP i-Astro group has been developing CdTe detectors in the framework of high-energy astrophysics space telescopes' mission proposals [1]-[6] and stratospheric balloon experiments (e.g., STRATOSPOLCA [7]) to study the Gamma-ray Universe. We have been performing simulation and experimental development of the main scientific instruments of these missions as well as radiation hardness analysis with proton beams [8]-[16]. Several gamma-ray space observatories (NASA, ESA) in operation [17]-[18] or in design, with our group contribution [3, 6], include the CdTe technology. Solutions such as Laue lens focusing systems [6] or all-sky mode instruments based on celestial sources photon trajectory reconstruction [3, 19] are being proposed to improve the soft to medium gamma-ray domain. To complement the orbital radiation hardness studies [20]-[23], several tests were carried out by our group with CdTe detectors under proton beams with energy in the range of 3-14 MeV produced at UC ICNAS cyclotron, for proton equivalent fluences up to ~20 years in LEO [13], [14]. Furthermore, we are implementing the project GLOSS (Gamma-ray Laue Optics and Solid-State detectors) selected in the framework of the Euro Material Ageing Facility (ESA/CNES) programme [24]. In this project, a set of passive samples of CdZnTe detectors will be exposed to the ISS orbital environment on Bartolomeo platform during 1 year, providing an unprecedented estimation of materials ageing in LEO. Recently, we initiated a new line of research on TGFs. These gamma-ray ms-long bursts with energies up to 100 MeV produced in cumulonimbus clouds at altitudes between 10 and 20 km, mainly in tropical and the equatorial regions are observed by gamma-ray observatories in space [25]-[30]. The risks associated with TFGs remains an open issue [31]-[32] and further research is needed to assess TFGs' exposure risks for crews and passengers of commercial flights. A master thesis performed in our group provided a preliminary but potentially significant result concerning TGF risks for aviation. The development of a sensor for aeroplanes to monitor TGF emissions is envisaged [33].

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

3. Justification of the relevance of the methods and approaches in the light of the state of the art of the research. Describe the innovation contained in the proposal;

This experiment addresses three scientific domains: i) high-energy astrophysics, ii) TGFs science and monitoring, and iii) space radiation in LEO and space weather. For these objectives, we are proposing a CdTe monitor with all-sky operational capabilities to perform spectroscopy, imaging, time variability celestial source observation, and polarimetry of the most intense gamma-ray sources in the sky (e.g., Crab Nebula and GRBs). So far, limited polarimetric measurements have been performed in space. The CdTe monitor allows the implementation of deep space and Earth observations in the 100 keV to 10 MeV energy band. The detector consists of a tracker composed by 4 detection planes, each with 4 CdTe pixelated matrices in a 4x1 configuration, each 1.4x1.4 cm^2 matrix with 256 x 256 square pixels (each 55 x 55 μm^2) and 2 mm thickness, providing an area of 31.4 cm^2 and 10^6 pixels (Fig. 1 and Fig. 2). The instrument tracker configuration, its high pixel granularity with per-pixel spectrometry, in coincidence regime (1µs) and its ability to track charged particles will provide the sensitivity level required to observe the referred celestial objects and perform polarimetric measurements. It allows implementing techniques for rejecting charged particles from the orbital background, namely protons and electrons, through reconstructing their tracks in the detection medium. The CdTe monitor will be operated in two complementary configurations: – a) Wide field of view Compton camera to observe: i) terrestrial gamma-ray flashes (TGFs) with the instrument pointed at Earth (Nadir); ii) gamma-rays from deep space sources with the instrument pointed at the deep sky (Crab Nebula and GRBs); – b) Particle tracking detector to observe: i) space radiation in low earth orbits, namely of protons and electrons around the Van Allen belts (characterization of the flux of protons and electrons); ii) solar activity, through the detection of solar flares and solar proton events, i.e., the study of space weather. Preliminary simulations performed with the MEGAlib toolkit indicate that the most critical observational parameter, the emissions' polarization status, will be potentially measured for the most intense gamma-ray sources (e.g. Crab Nebula and GRBs). These measurements will validate the all-sky operation mode at the highest TRLs (> 7), providing a very relevant conclusion for the future design of high-energy telescopes [3]. Furthermore, this experiment will leverage our research on the effects of the orbital environment on scientific instruments since it will operate exposed to particle radiation in the LEO and to solar proton events. This study should bring new data to compare with other experiments and models used to calculate the radiation environment in LEOs. Moreover, this experiment will allow studying the CdTe detector modules and the electronic system's response degradation in a real orbital environment while fully operating. Our scientific payload will be exposed to the orbital environment since it will be installed in the SR Field of View Locker that will be open to space when the SR reaches the cruise orbit. During the two months of flight, the response of our payload will be recorded to evaluate the detector's degradation during the time of flight. Further characterization of the CdTe detector modules will be performed after recovering the experiment. Finally, this experiment will pave the way to develop a CdTe monitor for aviation safety applications, addressing the potential dosimetric risks of TGFs to human health. This CdTe monitor will record the TGF emissions. The best conditions for TGF detection will be available when the SR Field of View Locker will be oriented towards Earth during circa one month. TGF's energy and time spectra and localization emission sites will be analysed and

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

compared with the data recorded by other astrophysical missions, e.g., AGILE and ASIM. We expect to be able to evaluate the potential of our device as a TGF monitor. The conclusions of this experiment will be essential to scale this concept to smaller monitor suitable to operate on board commercial flight aeroplanes.

4. Justification of the timelines and urgency of the project: strategic importance of the proposed research with respect to the objectives of the mission or overall project the proposal ties in with and with respect to international competition on this specific research subject;
TGF science is a relatively new research field, especially the knowledge of effects of TGFs on the health of passengers and crews of commercial aviation require further research. Each scientific work published on this domain has a large potential to have a relevant impact on TGF science and/or aviation safety. The SR flight will help to accelerate this research and maximise its impact. Additionally the SR orbit (~5° low-inclination LEO, altitude ~400 km) will cover the region of the sky above the typical latitudes of production of TGFs in the atmosphere. Testing the all-sky operation mode of a CdTe tracker in the LEO environment will be relevant to the design and operation modes of AMEGO mission [3]. Moreover, in its maiden 2-month flight, the SR will orbit Earth during a period when the Sun reaches its maximum of activity, after January 2024, almost the ideal moment to probe solar phenomena and solar proton events.

5. *Justification for the need of space-based compared to ground-based data;*
The SR provides an ideal platform since the Gamma-ray Universe can only be observed from high-altitude balloons, but with limited life-time, or space platforms. The SR flight platform provides space exposure allowing us to monitor our experiment in operation in the LEO environment. Furthermore, SR experiments are recoverable, which allows post-flight analysis of our scientific payload, a relevant output for future gamma-ray telescopes.

6. Work to be done at each partner in the Participating State, showing work to be done by institutes and work to be done at industry. General work logic;
LIP – Lead the proposal and coordinate the project at all stages of its development to ensure the fulfilment of the technical-scientific objectives: from the concept and design, the estimated scientific payload performance with simulation tools, the development of the scientific data processing software, and of the on-board computer, the development and integration of the monitor, the experimental performance tests and the validation tests.
AST – Company responsible for developing and building the experiment's enclosure. Also responsible for carrying out the validation tests (TVAC and vibration tests, see Fig. 4).

7. Interfaces, if any, with other partners outside of the Participating State, work carried out outside of the Participating State (summary only).
ADVACAM – SME company responsible for supplying the 16 CdTe sensor matrices and the electronic system of the scientific payload. Also, it will participate with LIP in developing the integration software between the detection unit and the on-board computer (see Fig. 4).

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

References

[1] J. Knödlseder et al., Exp. Astron., vol. 23, pp. 121–138, 2009.
[2] D. Angelis et al., Exp. Astron. vol. 44, pp. 25–82, 2017.
[3] A. Moiseev et al., PoS(ICRC2017) 798, 2017.
[4] E. Virgilli et al., Memorie Della Soc. Astron. Italiana, vol. 90, nos. 1–2, pp. 252–255, 2019.
[5] F. Frontera et al., Memorie Della Soc. Astron. Italiana, vol. 90, nos. 1–2, pp. 247–251, 2019.
[6] F. Frontera et al., ESA Voyage 2050 White Paper, 2021.
http://www.cosmos.esa.int/FronteraFWhitePaperESA-voyage2050.pdf.
[7] https://rexusbexus.net/experiments/scientific-research/astrophysics/.
[8] R. M. Curado da Silva et al., J. Appl. Phys., vol. 104, no. 8, pp. 084903-1–-7, 2008.
[9] R. M. Curado da Silva et al., IEEE Trans. Nucl. Sci., vol. 58, no.4, pp. 2118–2123, 2011.
[10]R. M. Curado da Silva et al., IEEE Trans. Nucl. Sci., vol. 59, no. 4, pp. 1628–1635, 2012.
[11] J. M. Maia et al., IEEE Trans. Nucl. Sci., vol. 62, no. 2, pp. 577–587, 2015.
[12] M. Pinto et al., Nucl. Instrum. Methods. Phys. Res. A, vol. 840, pp. 69–76, 2016.
[13] N. Simões et al., Nucl. Instrum. Methods. Phys. Res. A, vol. 877, pp. 183–191, 2018.
[14] M. P. Páscoa et al., IEEE Trans. Nucl. Sci., vol. 66, no. 9, pp. 2063–2071, 2019.
[15] A. F. V. Cortez et al., Exp. Astron., vol. 48, pp. 65–76, 2019.
[16] M. Moita et al., IEEE Trans. Nucl. Sci., vol. 68, no. 11, pp. 2655–2660, 2021.
[17] N. Gehrels et al., *The Astrophys. J.*, vol. 611, pp. 1005–1020, 2004.
[18] C. Winkler et al., *Astron. Astrophys.*, vol. 411, pp. L1–L6, 2003.
[19] A. Zoglauer et al., New Astron. Rev., vol. 50, nos. 7–8, pp. 629–632, 2006.
[20] Y. Eisen et al., Nucl. Instrum. Methods. Phys. Res. A, vol. 491, pp. 176–180, 2002.
[21] M. Zanarini et al., Nucl. Instr. and Meth. Phys. Res. B, vol. 213, pp. 315–320, 2004.
[22] M. M. Murakami et al., IEEE Trans. Nucl. Sci., vol. 50, no. 4, pp. 1013–1019, 2003.
[23] O. Limousin et al., Nucl. Instrum. Methods. Phys. Res. A, vol. 787, pp. 328–335, 2015.
[24] https://www.esa.int/ESA_Multimedia/Images/2020/06/Euro_Material_Ageing_facility.
[25] J. R. Dwyer et al., Space Sci. Rev., vol. 173, pp. 133–196, 2012.
[26] M. Marisaldi et al., J. Geophys. Res. Space Physics, vol. 119, pp. 1337–1355, 2014.
[27] M. Tavani et al., Phys. Rev. Lett., vol. 106, 018501, 2011.
[28] J. R. Dwyer et al., J. Geophys. Res., 117, A10303, 2012.
[29] J. R. Dwyer et. al., J. Geophys. Res., vol. 115, D09206, 2010.
[30] D. Sarria et al., Geophys. Res. Letters, vol. 48, e2021GL093152, 2021.
[31] M. Tavani et al., Nat. Hazards Earth Syst. Sci., vol. 13, pp. 1127–1133, 2013.
[32] M. Pallu et al., J. Geophys. Res. Atmosph., vol. 126, e2020JD033907, 2021.
[33] J. Mingacho, Departamento de Física, FCTUC, Universidade de Coimbra, 2020.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

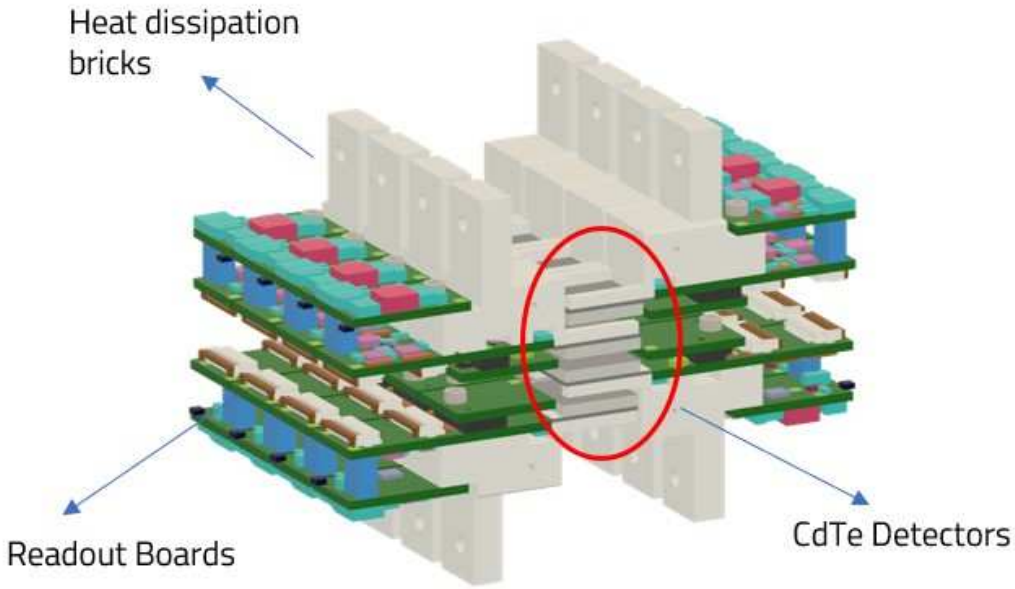

Fig. 1 - View of THOR-SR scientific instrument showing the CdTe stack composed by 4 layers of 4× *1* detector matrices, the electronic readout boards and the heat dissipation bricks.

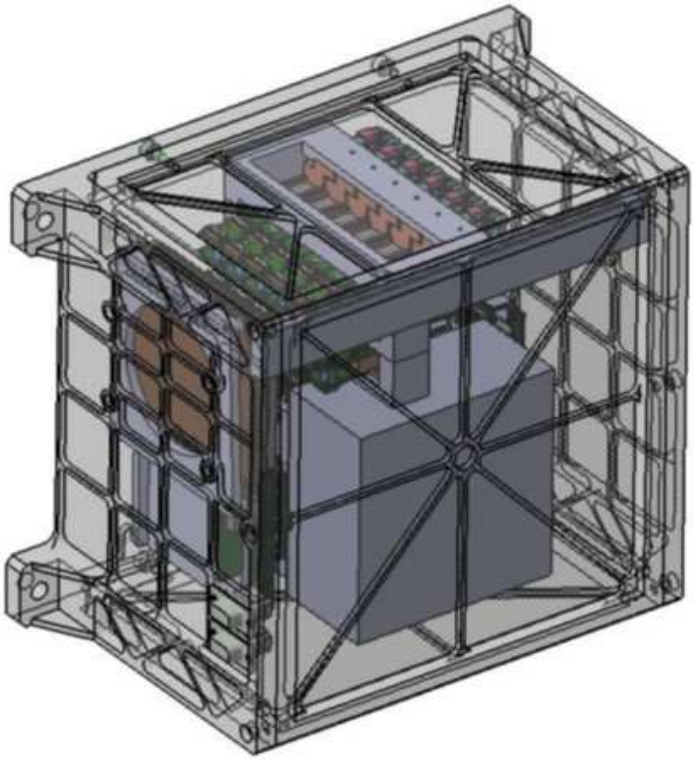

Fig. 2 - Transparent view of the enclosure containing THOR-SR scientific payload. It will be mounted on top of the fixation plate of the Space Rider Field of View Locker.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

E. Description of the scientific team(s)

LIP team members and short CVs [b,c,d,e,f,g,h]

Rui Miguel Curado da Silva, PhD (LIP/UC): Principal Investigator (PI) - Overall project management, WPs supervision, coordination of activities with external partners and stakeholders, project outreach and dissemination and project financial management. Has extensive background on instrumentation for high-energy astrophysics [4]-[16].

Jorge Manuel Maia-Pereira, PhD, Hab. (LIP/UBI): Co-Investigator (Co-I) - Co-management of overall project. Coordinator of WPs 2 to 5, supervision of the scientific payload development and post flight data analysis. Has solid background in development of radiation detectors for nuclear and astrophysics Instrumentation [4], [5], [9]-[16].

Gabriel Falcão, PhD (IT/UC): Coordinator of onboard computer development and testing. Has extensive experience in parallel computer architectures and compute-intensive signal processing applications.

Joana Diogo Mingacho, MSc (LIP): TGF events' simulations as well as measured TGF data analysis. Master thesis on TGF science [33].

José Carlos Ferreira Sousa, MSc student (LIP/UC): CdTe, readout board and power unit development and implementation. Flight measurements survey and data analysis. Experience on hardware & design of STRATOSPOLCA experiment [7].

Pedro Fonseca da Cunha Roque Póvoa, MSc student (LIP/UC): Onboard computer development and testing. Experience on STRATOSPOLCA FPGA development [7].

Margarida Rodrigues (LIP): Project financial manager. Administrative staff of LIP.

Research projects in the area under consideration (last 5 years):

-Terrestrial Gamma-ray Flash Science and Monitoring for Aviation Safety, ref.: EXPL/FIS-PAR/0333/2021; 2022-01 to 2023-06, Fundação para a Ciência e a Tecnologia.

- GLOSS-Gamma-ray Laue Optics and Solid-State Detectors, PEA: 4000136945, 2021-07 to 2024-06, ESA PRODEX Program.

- AHEAD2020: Integrated Activities for the High Energy Astrophysics Domain, ref.: 871158; 2020-03 to 2024-03, Research Executive Agency.

Most important relevant publications in peer reviewed international (last 5 years):

-O. Ferraz, S. Subramaniyan, (...), **G. Falcao**, *IEEE Communi. Surveys and Tutorials*, Vol. 24, no. 1, FIRST QUARTER, pp. 524–556, 2022.

- A. De Angelis, V. Tatischeff, (…), **R. M. Curado da Silva**, *et al*., *Exp Astron.,* vol. 51, 1225–1254, 2021.

- M. Moita, **R. M. Curado da Silva**, **J. M. Maia**, *et al*., *IEEE Trans. Nucl. Sci.*, vol. 68, no. 11, pp. 2655-2660, 2021.

- M.P. Páscoa, **J.M. Maia**, N. Auricchio, **R.M. Curado da Silva**, *et al.*, *IEEE Trans. Nucl. Sci.*, vol. 66, no. 9, pp. 2063-2071, 2019.

- M. Moita, E. Caroli, **J.M. Maia**, **R. M. Curado da Silva**, *et al*., *Nucl. Instrum. Methods Phys. Res. A*, vol. 918, pp. 93-98, 2019.

- N. Simões, **J.M. Maia**, **R. M. Curado da Silva**, *et al*., *Nucl. Instrum. Methods Phys. Res. A,* vol. 877, pp. 183-191, 2018.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Relevant international organizations, networks, working groups, ... the team belongs to;

- ESA M7 Call pre-selected ASTROGAM mission proposal [2];

- NASA AMEGO (All-sky Medium Energy Gamma-ray Observatory) mission proposal [3];

-Peer review: IEEE Transactions on Nuclear Science, Institute of Electrical and Electronics Engineers: Piscataway, New Jersey, US.

Mid-term R&D strategy with regard to this project.

The THOR-SR project is a pathfinder experiment for high-energy astrophysics future missions (e.g. AMEGO and ASTENA [3, 6]) in particular for an all-sky gamma-ray dedicated polarimeter based on tracker instrument (S to M ESA class standards) with high observational sensitivity that may operate in a future Space Rider flight or in a satellite platform.

The outputs of this project will contribute to leverage the mid-term strategy of our group to develop a small monitor for TGFs to fly on board commercial aeroplanes.

Lastly, the fine evaluation of the CdTe material degradation in LEOs environments, the effect of charged particles (protons) using the THOR-SR collected data on proton fluxes, including the solar proton events will contribute to develop radiation hardness solutions for future missions and to develop our group research on space weather.

ADVACAM team [b,c,f]

ADVACAM SME company core business is to design and produce detector cameras for imaging with ionising radiation. ADVACAM cameras scientific applications include particle accelerator, neutron radiography, Ion tracking and space science, with a strong partnership with CERN and with NASA on the Artemis Program. ADVACAM´s CEO and CSO collected the prestigious Czech EY Technology Entrepreneur of the Year 2021 award.

Company: ADVACAM s.r.o.
Address and Contacts: U Pergamenky 12, 17000 Praha 7, Czech Rep., https://advacam.com/

Carlos Granja, PhD: Co-coordinator of scientific payload detector system development. Development of detectors' readout software and development and testing of charged-particle detection systems. Dr. Granja has an extensive background in the development of instrumentation for space radiation research.
Jiří Šesták, PhD: Advacam Project Manager will co-coordinate joint activities between the company and LIP team and between ADVACAM and AST company on the experiment enclosure development. Dr. Šesták has an extensive background in the management of projects in the domain of radiation detectors for different applications: medical, nuclear safety, space science, etc.

Most important relevant publications in peer reviewed international (last 5 years):

- Marek Sommer, **Carlos Granja**, Satoshi Kodaira, Ondřej Ploc, *Nucl. Instrum. Methods Phys. Res. A*, vol. 1022, 165957, 2022.

- Anatoly Rosenfeld, Saree Alnaghy, (...), **Carlos Granja**, *et al.*, *Radiation Measurements*, vol. 130, 106211, 2020.

- **Carlos Granja**, Karel Kudela, Jan Jakubek, *et al.*, *Nucl. Instrum. Methods Phys. Res. A*, vol. 911, 142–152, 2018.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

F. Institute work under the present proposal PEA

Description of the work breakdown structure, work package description and production of deliverables per work package for the work done at the institute.

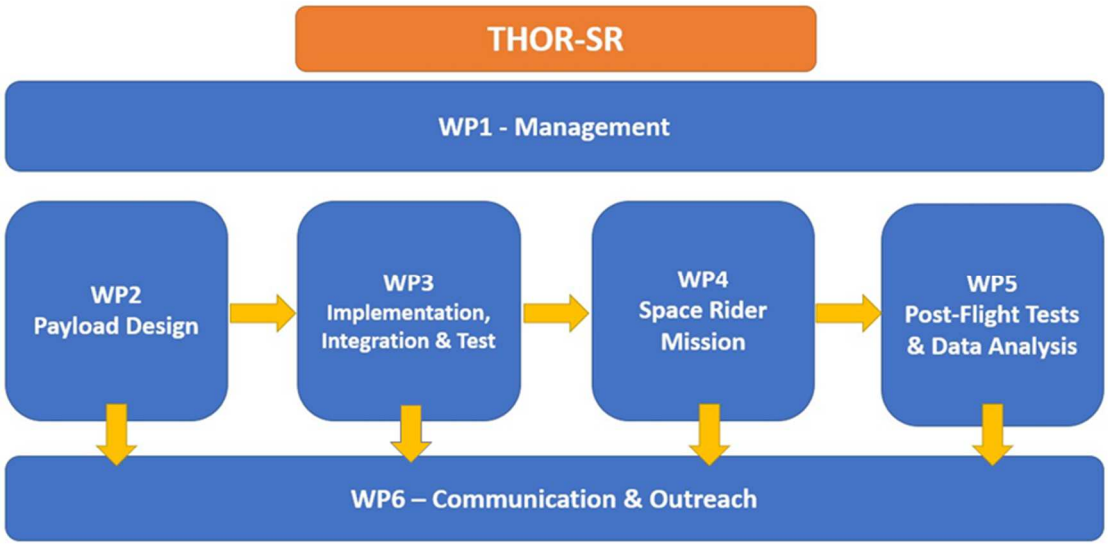

Fig. 3 - THOR-SR project work plan showing the workflow between each of 6 work packages.

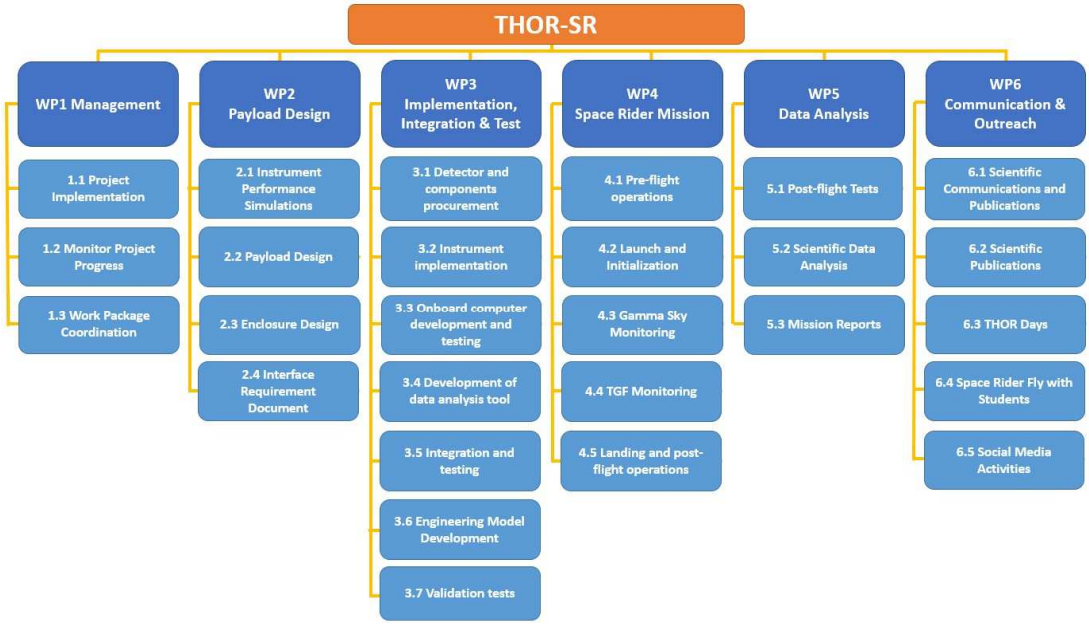

Fig. 4 - THOR-SR work breakdown structure per work package (zoomed in the next page)

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

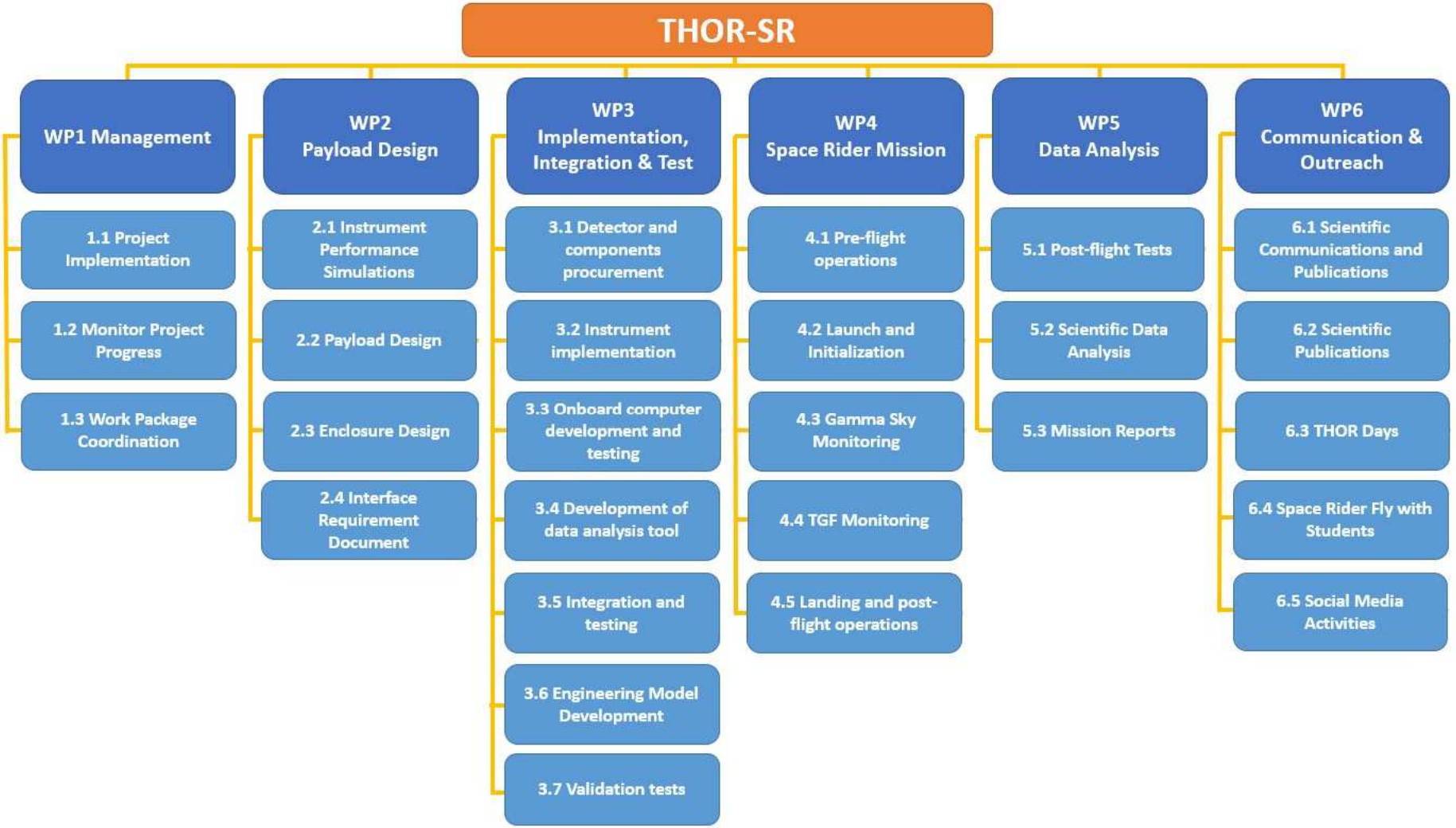

Fig. 4 - THOR-SR work breakdown structure per work package.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Project's WP detail

Work Package number:	WP1
Work Package Title:	Management
Responsible entity:	LIP
Local Managers:	Rui Silva (LIP/UC), Jorge Maia (LIP/UBI) and Margarida Rodrigues (LIP)
Beginning and End of WP	T_0 – T_{42} (01/2022 to 06/2025)

Objectives: Ensure that project is implemented as scheduled, monitor WP's tasks developments and that WP's objectives and overall project's objectives are attained.

Inputs:

- Project proposal document;
- Project organisation documents: gantt chart, work plan and timeline.

Description of work:

1.1 - **Project Implementation**: The members of the Management Board are: the Principal Investigator (PI) of the project, the Co-I and the LIP project manager (LIP Secretariat). The Management Board of the project will ensure the realisation of the project's objectives within the defined budget limits. It will ensure adequate timing and effective interaction between tasks, with the project partners, equipment and service suppliers. The Management Board is responsible for the budget distribution and to gather and integrate the information needed for the scheduled milestones of the project (interim task reports, financial statements, reviews-related documents and final report). The LIP project manager will ensure the day-to-day administration of the project, communication with ESA and support to the other elements in the Management Board;

1.2- **Monitor project progress**: The Principal Investigator (PI) with the Co-Investigator (Co-I) will also monitor progress and technical evaluation of the activities. Meetings will be held regularly during the project's lifetime (every 3 months), and when necessary, ad hoc meetings can be added during the project to discuss and resolve urgent issues. Meeting

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

minutes will be written to document the progress of the meeting and the decisions made and will be available to the project team. The PI with the Co-I will monitor the provision of the various project outcomes in the terms presented in the description of work. They will support internal and external partners to explore the results of the project. The participation in the international meetings foreseen in the project plan will be followed closely by the PI and/or the Co-I in order to foster the best practices for the dissemination of results so that relevant achievements can be efficiently shared with the stakeholders, both to the scientific and to the general public (here special emphasis will be placed to explain the promising achievements for a non-specialized public and its importance to tackle societal challenges);

1.3- **Work Package Coordination**: Within each work package (WP), the WP Coordinator Researcher will be responsible for the implementation of the action plan, by monitoring qualitative and quantitative scientific and technical results and by reviewing project task outcomes.

Deliverables:

- Annual project reports and financial reports;
- Project Final Report.

Work Package number:	WP2
Work Package Title:	Payload Design
Responsible entity:	LIP
Local Managers:	Rui Silva (LIP/UC) and Jorge Maia (UBI/LIP)
Beginning and End of WP	T_0 – T_{15} (01/2022 to 03/2023)
Objectives: Define the scientific payload design as well as the configuration of the external enclosure that will interface the Space Rider Locker.	

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Inputs:

- Space Rider general and operational specifications' documents;
- ADVACAM CdTe detector systems' datasheets.

Description of work:

In order to perform the scientific payload and external enclosure design the following tasks will be performed:

2.1 - **Scientific performance simulations with MEGAlib tool-kit**: continuum and line sensitivity, and polarisation level and angular sensitivity. Several possible detector configurations will be tested under different celestial high-energy astrophysical sources (GRBs and the Crab Nebula), typical TGF emissions and orbital background;

2.2 - **Enclosure design**: Active Space Technologies (AST) will design and manufacture the enclosure, which will take into account the thermal, electrical, mechanical, and data flow characteristics of the scientific payload components;

2.3 - **Interface Requirement Document (IRD) writing**: In cooperation with ESA Space Rider flight direction, the IRD development includes general payload description, project overview, mission list of requirements (mission, system, interface, operational and outreach) and interfaces between systems;

2.4 - **Scientific payload design**: a) sensitive detector configuration definition according to the scientific requirements and ADVACAM CdTe module characteristics; b) front-end and readout electronics design according to the scientific requirements and ADVACAM ASICS Timepix3 electronics offer; c) onboard computer (OBC) and power distribution unit (PDU) selection according to data processing and electrical power requirements. A preliminary design review (PDR) document will be submitted to ESA. Afterwards, in regular communication with the SR management team the scientific payload design will undergo several review interactions until the critical design review (CDR) where the final payload configuration will be freezed.

Deliverables:

- Payload IRD - Interface Requirements Document;
- Payload PDR - Preliminary Design Review;
- Payload CDR - Critical Design Review.

Work Package number:	WP3

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Work Package Title:	Implementation, Integration and Test
Responsible entity:	LIP
Local Managers:	Rui Silva (LIP/UC) and Jorge Maia (UBI/LIP)
Beginning and End of WP	T_{10} – T_{24} (10/2022 to 12/2023)

Objectives: Detector, electronics, onboard computer and power distribution unit procurement. Scientific payload integration; scientific payload performance tests; scientific payload and enclosure integration; engineering model TVAC tests. Development of scientific data processing/analysis software tools.

Inputs:

- Payload IRD;
- Payload PDR;
- Payload CDR.

Description of work:

3.1-**Detector and components procurement**: Detector Unit, electronic modules, onboard Computer (OBC) and power distribution unit (PDU) procurement;

3.2-**Scientific instrument implementation**: Firstly, a single CdTe module will be implemented with a readout board. Characterization and performance tests will be performed in this configuration. Performance tests will be carry-out either with laboratory gamma-ray radioactive sources, or performed in the LARIX beamline (continuum energy spectrum until 300 keV) at the University of Ferrara, Italy;

3.3-**Onboard computer development and testing**: Integration software tests between the Detector Unit and the onboard Computer (OBC) will be performed with the participation of ADVACAM. For the integration between the payload OBC and the SR OBC software will be developed to simulate the payload communication system. The payload OBC implementation will follow ECSS standards;

3.4-**Development of data analysis tool:** A tool based on MEGAlib tool-kit will be developed for the processing/analysis of the scientific data from the Detector Unit;

3.5-**Integration and testing**: The 16 CdTe and 2 Si modules (including the boards based on ASICS Timepix3) will be mounted in the final configuration (Fig. 1), together with the OBC

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

and the PDU. Functional tests between systems will be performed at LIP laboratory facilities and at the ESRF or LARIX beamlines (access will be provided via AHEAD2020 Horizon project). Finally, the scientific payload will be integrated with the enclosure build AST. This flight model will undergo final scientific performance tests;

3.6-**Engineering model development**: An engineering model will be implemented with a similar operating CdTe detector module coupled to similar operating readout electronics, while the remaining instrument being composed by mockup structure, dummy heat dissipators and non operating components with the same mass and structure as in the flight model;

3.7-**Validation tests:** Thermal-Vacuum (TVAC) and Vibration tests, with the engineering model, will be performed in order to validate the payload thermal behaviour, the structural integrity and the survivability of the components in extreme conditions. Outsourced tests, carried out by AST. Finally, a Qualification Review (QR) document will be provided to ESA.

Deliverables:

- Engineering model;
- Flight model;
- Qualification Review document;
- Scientific publication on-ground instrument performance.

Work Package number:	WP4
Work Package Title:	Space Rider Mission
Responsible entity:	LIP
Local Managers:	Rui Silva (LIP/UC) and Jorge Maia (UBI/LIP)
Beginning and End of WP	T_{25} – T_{33} (01/2024 to 09/2024)

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Objectives: Space Rider flight onboard measurements: deep-space observation (Crab Nebula, GRBs, others), terrestrial gamma-ray flashes (TGFs) observation, and particle field monitoring.

Inputs:

- Flight Model
- Instrument Operation Plan

Description of work:

4.1- **Pre-flight operations**: When ready the flight model will be delivered to the ESA ESTEC facilities. From ESTEC it will be shipped to Kourou launch site in the French Guiana in South America. Final verification operational tests shall be performed by Kourou staff as requested by us;

4.2- **Launch and Initialization**: The experiment will be launched on board the Space Rider Field-of-View Locker. The Space Rider will be placed in LEO by a Vega-C rocket. The initialization phase is set when the Space Rider reaches a stable LEO. The experiment power is switched-on and boots to a housekeeping mode. Power: 14 W; No TD; Com: 2 MByte/orbit;

4.3- **Gamma Sky Monitoring**: During the Space Rider flight, the the Detector Unit radiation window (jointly with the top of SR) should be oriented towards the deep sky (ideally to the Zenith), to preferentially record gamma-ray astrophysical emission sources (Crab Nebula, GRBs or other gamma-ray sources). During this phase the scientific data sent to the ground station will be analysed on the LIP ground station facilities in order to, if needed, adjust the Detector Unit's response. Power: 50 W; TD: 50 W; Com: 19 MByte/orbit;

4.4- **TGF Monitoring**. During the Space Rider flight, the Detector Unit radiation window (jointly with the top of SR) should be oriented towards Earth (ideally to the Nadir), in order to record the TGFs emissions. During this phase the scientific data sent to the ground station will be analysed on the LIP ground station facilities in order to, if needed, adjust the Detector Unit's response. Power: 50 W; TD: 50 W; Com: 19 MByte/orbit;

4.5- **Landing and post-flight operations**. One hour before the landing operations start, the power of the experiment will be switched off (Power OFF; No TD; No Com). The Space Rider will land at Kourou where our scientific payload will be recovered from the Field-of-View Locker by local staff.

Deliverables:

- Scientific data sets measured during SR (gamma sources, orbital radiation and TGF);
- Space Rider data sets of flight parameters (altitude, orientation, velocity, etc.) recorded during the 2 months flight.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Work Package number:	WP5
Work Package Title:	Post-Flight Tests and Data Analysis
Responsible entity:	LIP
Local Managers:	Rui Silva (LIP/UC), Jorge Maia (UBI/LIP) and Gabriel Falcão (IT/UC)
Beginning and End of WP	T_{34} – T_{42} (10/2024 to 06/2025)

Objectives: Experiment scientific data analysis and integration of Space Rider flight parameters with scientific data.

Inputs:

- Scientific data sets form the Gamma Sky and TGF observations;
- Space Rider set of flight parameters (altitude, orientation, speed, etc) recorded during the 2 months flight;
- Pre-flight calibration data;
- Flight model recovered from the Space Rider.

Description of work:

5.1-**Post-flight Tests**: Upon retrieval of the THOR-SR payload, the Detector Unit and the electronics boards will be carefully inspected. Operational and performance tests will be performed and compared with the results obtained before the flight. A scientific publication on the ageing/degradation at LEO of CdTe detector modules and electronic systems will be performed;

5.2-**Scientific Data Analysis**: The onboard recorded data as well as the data transmitted during the flight will be thoroughly processed and analysed. Crab nebula, GRBs, TGFs and orbital charged particle measurements data will be compared with simulated data in order to interpret the recorded data during the flight, as well as, to validate simulation models.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Reference data from the literature (mainly from non-transient sources), and from other space missions, in operation, will be used too;

5.3-**Mission reports**: The main scientific and technological outputs of the data analysis will be analysed and compiled in a set of mission reports and scientific publications.

Deliverables:

- Experiment final report;
- A report on the ageing of CdTe detector modules and electronics systems;
- Scientific publications on high-energy astrophysics' and TGFs' flight measurements and recorded orbital particle environment.

Work Package number:	WP6
Work Package Title:	Communication and Outreach
Responsible entity:	LIP
Local Managers:	Rui Silva (LIP/UC), Jorge Maia (UBI/LIP) and Joana Mingacho (LIP)
Beginning and End of WP	T_{10} – T_{42} (10/2022 to 06/2025)
Objectives: Communication of the main experiment conclusions at international symposia and on international peer-review scientific journals. Outreach activities for the general public, including university and high-school students.	

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Inputs:

- Project technical documents;
- Project scientific documents.

Description of work:

6.1-**Scientific communications**: The main scientific and technological outputs of the data analysis will be presented in international scientific symposia, in the fields of astrophysics, atmospheric physics and nuclear sciences (IEEE NSS, SPIE or other);

6.2-**Scientific publications**: The main scientific outputs will be published in international journals, in the fields referred above, with refereeing and open access;

6.3-**THOR Days**: Two events entitled THOR-SR Days will be organised at local Universities, UBI and UC, for students, academics and stakeholders. The objective is to promote the scientific potential of our experiment, its interest in the graduating students' training, as well as its potential benefits to society, in particular the aviation sector;

6.4-**Space Rider Fly with Students**: Outreach activities will be performed together with local high-schools, and UC and UBI students, at the time of SR launch into orbit and during the two months flight until its landing back to Earth in a French Guiana airport track. Life broadcasting of the flight will be performed for a students' audience, followed by questions and answers;

6.5-**Social Media Activities**: In order to promote proximity with the students from all levels and locations, as well as with the general public, several dissemination platforms will be used: social media; institutional networks (e.g., LIP, UC and UBI webpages); THOR-SR project website; etc. In particular we intend to promote the project and related scientific topics (astrophysics, space science, Earth observation, space weather, etc.) by organising live talks, debates and interviews on social media streaming platforms in collaboration with UC and UBI scientific students associations.

Deliverables:

- 8 Scientific publications on international journals with refereeing in open access and international symposium proceedings;
- 4 Master theses;
- Project website;
- Social media live sessions.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

H. Institute - Detailed milestones and justification of related costs

Detail the milestones (at least per half year) of all work to be performed and deliverables to be presented.

a. All costs presented in the financial plan of section I (salaries, travel, small and large equipment, industrial developments and services) must be clearly justified in line with this timetable.
b. Scientific and engineering work to be carried out at institute level during the term of the requested PEA (tables 1 and 2 of the financial plan) must be clearly separated from work outside the said term, as well as from developments and services to be carried out by industry (table 3 of the financial plan).
c. In case the proposal is part of an overall project, indicate the constraints imposed by it.

Describe the risks related to each milestone (*e.g.* dependency on other scientific and industrial partners, data quality, modelling, etc.).

Summary

Schedule recap:

Provide a schedule chart of the activity indicating expected beginning/end and duration of each work package.

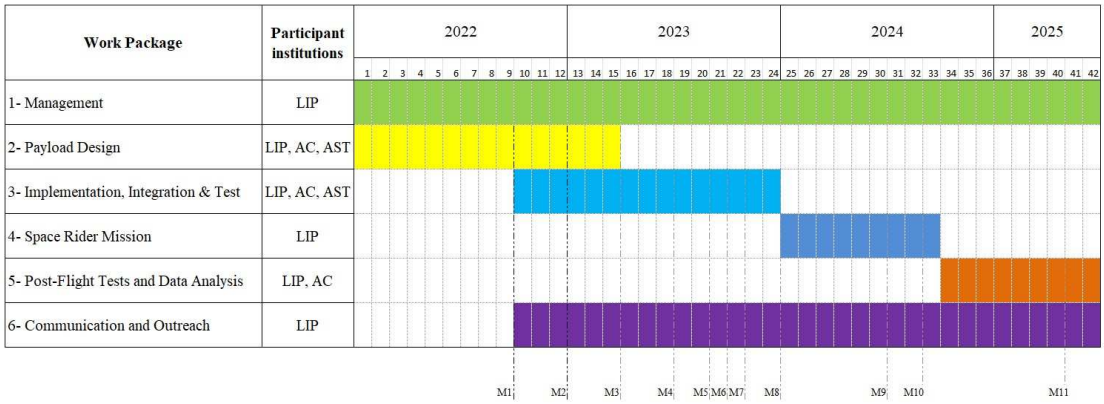

Fig. 5 - Gantt chart of THOR-SR project by work package with indication of the milestones (zoomed in the next page).

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

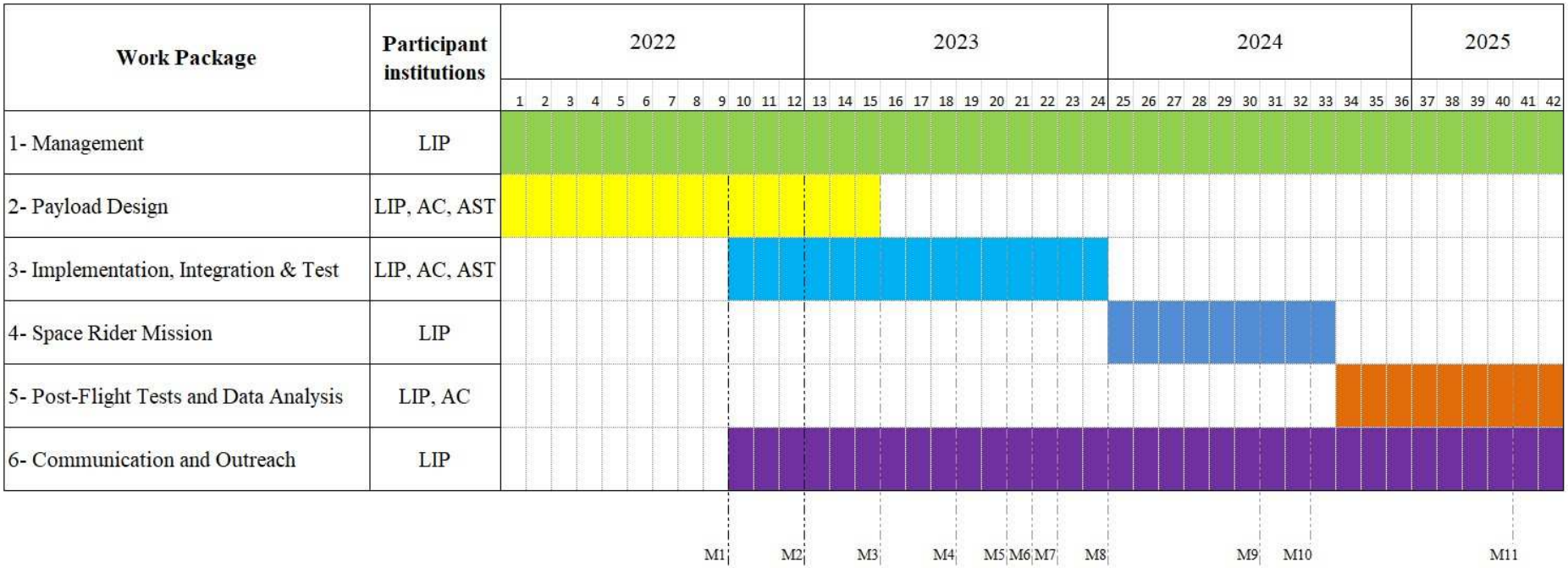

Fig. 5 - Gantt chart of THOR-SR project by work package with indication of the milestones.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Milestones and Deliverables recap:

Milestone	Project's Month	Description	Related risks
M1	9	Interface Requirement Document submitted to ESA	None (document submitted)
M2	12	Preliminary Design Review submitted to ESA	International shortage/price rise of components and materials, due to Covid+War, may require important changes in the payload design.
M3	15	Critical Design Review submitted to ESA	Same as M2 risk.
M4	18	Scientific payload enclosure production by AST	Price of aluminium with inorganic black anodization for the enclosure is increasing to worrying levels for the present project budget.
M5	20	End to end scientific payload system testing	Professional instability of human resources in Portuguese institutions, in particular change in specialised team members, may impact on the overall system coordination integration efficacy and consequently in the end to end system testing. Joint development tests between LIP and ADVACAM require fine coordination between both teams and efficient and safe transport of components from Prague to Coimbra.
M6	21	Engineering model completed	Professional instability referred to in M5 and international shortage/price rise of components and materials referred to in M2 might delay this milestone.
M7	22	Flight model completed	Same as M6 risk.
M8	24	Qualification Review	Same as M6 risk.
M9	30	Space Rider launch from Kourou, French Guiana	Meteorological weather conditions on the launch site in Kourou are a permanent risk of mission delay. International shortage/price rise of components and materials may impact

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

			on Kourou facilities capacity to be within SR maiden flight schedule. SR Vega C launcher depends on success of previous Vega launches, since presently only a number of engines for the second stage are available, since these were built in Ukraine before the war started.
M10	32	Space Rider landing at Kourou and scientific payload recovery	Although SR landing will be quite smooth with a parachute system at a conventional airport runway and that the Intermediate experimental Vehicle (test version of the SR) flew a flawless reentry in 2015, the reentry in the atmosphere is always a relatively risky operation.
M11	40	Presentation of first SR experiment results at an International Symposium	Anomalies during the SR flight in any of the subsystems of the SR or the scientific payload, affecting the SR data sets (altitude, position, time, etc.) or the scientific measurements may impact on data analysis delays or compromise certain scientific objectives. In particular polarimetric measurements are the most critical since to obtain the required sensitivity the observation time must be maximised.

Recapitulate the deliverables for each work package.

Work Package	Deliverables
WP1	● Annual project reports and financial reports; ● Project Final Report.
WP2	● Payload IRD - Interface Requirements Document; ● Payload PDR - Preliminary Design Review; ● Payload CDR - Critical Design Review.
WP3	● Engineering model; ● Flight model; ● Qualification Review document;

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

	• Scientific publication on-ground instrument performance.
WP4	• Scientific data sets measured during SR (gamma sources, orbital radiation and TGF); • Space Rider data sets of flight parameters (altitude, orientation, velocity, etc.) recorded during the 2 months flight.
WP5	• Experiment final report; • A report on the ageing of CdTe detector modules and electronics systems; • Scientific publications on high-energy astrophysics' and TGFs' flight measurements and recorded orbital particle environment.
WP6	• 8 Scientific publications on international journals with refereeing in open access and international symposium proceedings; • 4 Master theses; • Project website; • Social media live sessions.

Costs justification of the financial plan

Project Costs	Amount (€)	Costs Justification
Principal researcher salary complement	38 841	Complement of PI salary. This amount matches the salary difference between Principal Researcher level (its level in this project) and Junior Researcher level (the salary level granted by LIP) during 18 months.
Auxiliary researcher salary	85 058	18 months of salary for an auxiliary researcher to perform the scientific payload development: detector+readout electronics.
PhD student salary	23 094	18 months for a PhD Student that will perform the detector response simulations and measurements' data analysis.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Travel	10 000	- 2 Trips to ESA ESTEC Noordwijk, the Netherlands; - 8 Trips to ADVACAM Czech Rep.; - 2 Trips to Kourou, French Guiana for the Space Rider launch; - 2 International Meetings (IEEE Nuclear Science Symposium): 1 in Europe and 1 Canada.
Scientific payload enclosure	43 940	Enclosure development costs: 1. Model Philosophy; 2. Engineering Model (EM) + Flight Model (FM): 2.1- Interface system between experiment payload and Space Rider: 2.2- 4U volume structure for the experiment payload: Build 2 models EM +FM; 2.3- Documentation for Interface and 4U structure;
Materials: Scientific payload detectors + readout electronics	125 000	16 ADVACAM MiniPix TPX3 CdTe detectors; 2 Si detectors; Readout boards for CdTe and Si detectors.
Materials: Onboard computer and power distribution unit	25 000	- Onboard computer to process the detectors' output data; - Power unit to feed scientific payload: detectors, readout electronics and OBC.
Dissemination and Outreach	5 000	- 2 publications open access; - Multimedia services for recruitment and stages announcement, project logo, infographics for website and social media content.
Overheads	88 983	- Institutional overheads for LIP and UBI at 20% of the overall project's budget.

CVs:

Provide names/function/CV % of a full time equivalent involvement into the project of:

- study leader and person to be funded under PRODEX

Rui Miguel Curado da Silva, PhD (LIP/UC), 65% FTE (PRODEX will fund complement to salary): Principal Investigator (PI); Dr. Curado da Silva has an extensive background in instrumentation for high-energy astrophysics, in particular on gamma-ray polarimeter

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

development both by experimental and simulation research. He coordinated two stratospheric balloon experiments (NASA and ESA) and he is coordinating a material ageing experiment to be launched to the ISS. He is an Invited Auxiliary Professor at the UC, teaching in Master's degree in Astrophysics and Instrumentation for Space.
ORCID: 0000-0002-9961-965X

- Other team members

Jorge Manuel Maia-Pereira, PhD, Hab. (LIP/UBI): Co-Investigator (Co-I); Dr. Maia is Assistant Professor with Habilitation with Department of Physics, University of Beira Interior, and a researcher with LIP–Laboratório de Instrumentação e Física Experimental de Partículas. He has a solid background in development of radiation detectors for nuclear and astrophysics Instrumentation, mainly micropattern gas detectors, GEMs, THGEMs and MHSPs, for X-rays and compound semiconductor detectors, CdTe and CdZnTe, for gamma-rays (for polarimetry/imaging/spectroscopy). More recently, he focused his research on the effects on CdTe/CdZnTe detectors of the orbital radiation in LEOs, namely protons. ORCID: https://orcid.org/0000-0002-9314-1763

Gabriel Falcão, PhD (IT/UC): Dr. Falcão is Assistant Professor with the Department of Electrical and Computer Engineering, University of Coimbra, and a Researcher with Instituto de Telecomunicações. His research interests include parallel computer architectures, energy-efficient processing, and compute-intensive signal processing applications, namely in image and communications. More recently he focused his research on processing-in-memory architectures for mitigating the data movement bottleneck. He will be responsible for the development of the on-board computer. Gabriel is a Senior Member of the IEEE. ORCID: https://orcid.org/0000-0001-9805-6747

Joana Diogo Mingacho, MSc (LIP/UC): Researcher fellow; J.Mingacho develops research within the scope of TGF science and simulation studies with MEGALib toolkit related with dosimetry associated with TGFs and the risks for aviation. Recently, she finished a master thesis on TGF science. She will participate in TGFs related measurements and simulations.

José Carlos Ferreira Sousa, Engineering Physics MSc student (LIP/UC): Researcher fellow; J.Sousa develops work within the scope of systems engineering. Recently participated in hardware&design WP in the STRATOSPOLCA high-altitude balloon payload. He will participate in hardware related tasks (the implementation and integration of scientific payload) and in several tests of the instrument and its subsystems.

Pedro Fonseca da Cunha Roque Póvoa, Electrotechnical Engineering MSc student (LIP/UC): Researcher fellow; P.Póvoa develops work within the scope of software engineering. Recently participated in the software WP in the STRATOSPOLCA high-altitude balloon payload. He will participate in software related tasks, related with the communication between the Detector Unit and the onboard Computer (OBC), with close supervision of Prof. Falcão.

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Carlos Granja, PhD (ADVACAM s.r.o, Czech Republic): Researcher; Dr. Granja has an extensive background in nuclear instrumentation, namely the development of radiation detection systems based in Si and CdTe semiconductor materials, for several applications such as: particle tracking of energetic charged particles (electrons, protons, ions), radiation dosimetry, and space radiation detection in LEOs (including in ISS, spacecraft micro-satellites and CubSats). He will participate in the tests of the gamma-ray/charged-particle detection system, and in the development of the integration software between the Detection Unit and the OBC.

Jiří Šesták, PhD (ADVACAM s.r.o, Czech Republic): Project Manager; Dr. Šesták, has an extensive background in the management of projects within the company. He will be responsible for the interface within ADVACAM with the THOR-SR partners: the LIP research team and AST company.

I. Financial plan

Provide a financial plan – you may use own format of the tables suggested here after. Items to be costed (institute):

a. Manpower;
b. Equipment:
    - Small equipment (<5000K per unit price) to be purchased;
    - Equipment with single value > 5000 Euros;
    - Itemise (2 separate lists) the small equipment and the "bigger" equipment required for the project and provide a short description with rationale why equipment is required;
c. Travels. Provide a travel estimate per work package and (TBC delegation) a travel plan indicating: mission destination, reason, nb days, nb persons, travel costs / person, subsistence cost /person/day, total costs;
d. Funding received by third parties (e.g. universities or other) contributing to the project and not requested to PRODEX;
e. Industrial costs – provide a summary and the offers from industry consultations.

Example of a table (research activities). For hardware definition/development activity provide a financial plan per work package:

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

Funding requested from PRODEX:

INSTITUTE COSTS (€) - LIP	**2022**	**2023**	**2024**	**2025**	**Total**
Salaries:					
Principal researcher salary complement	6 474 €	25 894 €	6 474 €	0 €	38 841 €
Auxiliary researcher salary	14 176 €	56 705 €	14 176 €	0 €	85 058 €
PhD student salary					0 €
Travel	2 000 €	2 000 €	2 000 €	4 000 €	10 000 €
Small Equipment (<5 000 Euro) to be purchased by institute	0 €	0 €	0 €	0 €	0 €
External Services contracted by institute:					
- Scientific payload enclosure	43 940 €				43 940 €
- Materials (18 AdvaCAM MiniPix TPX3 CdTe + readout boards)	125 000 €				125 000 €
- Materials (Onboard computer and power unit)	25 000 €				25 000 €
- Dissemination and Outreach	0 €	1 250 €	1 250 €	2 500 €	5 000 €
Overheads	54 147 €	21 462 €	5 975 €	1 625 €	83 210 €
TOTAL Institute (PEA)	**270 737 €**	**107 312 €**	**29 875 €**	**8 125 €**	**416 049 €**
Larger Equipment (> 5000 Euro) to be purchased via ESA	**0 €**	**0 €**	**0 €**	**0 €**	**0 €**
Grand Total Institute	**270 737 €**	**107 312 €**	**29 875 €**	**8 125 €**	**416 049 €**
INDUSTRY COSTS (ESA industrial contracts)	**2022**	**2023**	**2024**	**2025**	**Total**
Industry 1					**0 €**
Total Industry	**0 €**	**0 €**	**0 €**	**0 €**	**0 €**
GRAND TOTAL PROJECT (Institute + Industry)	**270 737 €**	**107 312 €**	**29 875 €**	**8 125 €**	**416 049 €**

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

INSTITUTE COSTS (€) - UBI	2022	2023	2024	2025	Total
Salaries:					
Principal researcher salary complement					0 €
Auxiliary researcher salary					0 €
PhD student salary	3 849 €	15 396 €	3 849 €	0 €	23 094 €
Travel					0 €
Small Equipment (<5 000 Euro) to be purchased by institute					0 €
External Services contracted by institute:					
- Materials					0 €
- Dissemination and Outreach					0 €
Overheads	962 €	3 849 €	962 €	0 €	5 774 €
TOTAL Institute (PEA)	**4 811 €**	**19 245 €**	**4 811 €**	**0 €**	**28 868 €**
Larger Equipment (> 5000 Euro) to be purchased via ESA	**0 €**	**0 €**	**0 €**	**0 €**	**0 €**
Grand Total Institute	**4 811 €**	**19 245 €**	**4 811 €**	**0 €**	**28 868 €**
INDUSTRY COSTS (ESA industrial contracts)	**2022**	**2023**	**2024**	**2025**	**Total**
Industry 1					**0 €**
Total Industry	**0 €**	**0 €**	**0 €**	**0 €**	**0 €**
GRAND TOTAL PROJECT (Institute + Industry)	**4 811 €**	**19 245 €**	**4 811 €**	**0 €**	**28 868 €**

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt

INSTITUTE COSTS (€) - TOTAL PROJECT	2022	2023	2024	2025	Total
Salaries:					
Principal researcher salary complement	6 474 €	25 894 €	6 474 €	0 €	38 841 €
Auxiliary researcher salary	14 176 €	56 705 €	14 176 €	0 €	85 058 €
PhD student salary	3 849 €	15 396 €	3 849 €	0 €	23 094 €
Travel	2 000 €	2 000 €	2 000 €	4 000 €	10 000 €
Small Equipment (<5 000 Euro) to be purchased by institute	0 €	0 €	0 €	0 €	0 €
External Services contracted by institute:					
- Scientific payload enclosure	43 940 €				43 940 €
- Materials (18 AdvaCAM MiniPix TPX3 CdTe and Si detectors + readout boards)	125 000 €				125 000 €
- Materials (Onboard computer and power unit)	25 000 €				25 000 €
- Dissemination and Outreach	0 €	1 250 €	1 250 €	2 500 €	5 000 €
Overheads	55 110 €	25 311 €	6 937 €	1 625 €	88 983 €
TOTAL Institute (PEA)	**275 549 €**	**126 557 €**	**34 686 €**	**8 125 €**	**444 916 €**
Larger Equipment (> 5000 Euro) to be purchased via ESA	**0 €**	**0 €**	**0 €**	**0 €**	**0 €**
Grand Total Institute	**275 549 €**	**126 557 €**	**34 686 €**	**8 125 €**	**444 916 €**
INDUSTRY COSTS (ESA industrial contracts)	**2022**	**2023**	**2024**	**2025**	**Total**
Industry 1					**0 €**
Total Industry	**0 €**	**0 €**	**0 €**	**0 €**	**0 €**
GRAND TOTAL PROJECT (Institute + Industry)	**275 549 €**	**126 557 €**	**34 686 €**	**8 125 €**	**444 916 €**

Av. D. Carlos I, 126 1249-074 Lisboa Portugal
Tel.: +351 213 924 300 Fax: +351 213 907 481 NIF: 503904040 www.fct.pt