







RISC-V Processor Technologies for Aerospace Applications in the ISOLDE Project

William Fornaciari¹ , Federico Reghenzani¹ , Giovanni Agosta¹ ,
Davide Zoni¹ , Andrea Galimberti¹ , Francesco Conti² , Yvan Tortorella² ,
Emanuele Parisi² , Francesco Barchi² , Andrea Bartolini² ,
Andrea Acquaviva² , Daniele Gregori³ , Salvatore Cagnetta⁴ ,
Carlo Ciancarelli⁴, Antonio Leboffe⁴, Paolo Serri⁴, Alessio Burrello^{2,5} ,
Daniele Jahier Pagliari⁵ , Gianvito Urgese⁵ , Maurizio Martina⁵ ,
Guido Maserà⁵ , Rosario Di Carlo⁶ , and Antonio Sciarappa⁶ 

¹ Politecnico di Milano, Milan, Italy

{william.fornaciari,federico.reghenzani,giovanni.agosta,
davide.zoni,andra.galimberti}@polimi.it

² University of Bologna, Bologna, Italy

{francesco.conti,yvan.Tortorella,emanuele.parisi,
francesco.barchi,andra.bartolini,andra.acquaviva}@unibo.it

³ E4 Computer Engineering SpA, Scandiano, Italy

daniele.gregori@e4company.com

⁴ Thales Alenia Space Italia S.p.A., Turin, Italy

{salvatore.cagnetta,carlo.ciancarelli,
antonio.leboffe,paolo.serri}@thalesaleniaspace.com

⁵ Politecnico di Torino, Torino, Italy

{alessio.burrello,danieleJahier.pagliari,gianvito.urgese,
maurizio.martina,guido.masera}@polito.it

⁶ Leonardo SpA, Rome, Italy

{rosario.dicarlo.ext,antonio.sciarappa}@leonardo.com
<https://heaplab.deib.polimi.it>, <https://www.unibo.it>,

<https://www.e4company.com/>, <https://www.thalesaleniaspace.com/en>,
<https://www.polito.it>, <https://www.leonardo.com>

Abstract. Modern space applications impose significant challenges to the design of hardware and software platforms. Beyond traditional applications such as avionics, Attitude Orbit Control, and signal/telemetry processing, new developments increasingly leverage Machine Learning models to enhance the autonomy of spacecraft. Such AI-based functionalities promise significant advantages, but require computing power beyond what can be provided by current on-board platforms. At the same time, the challenge of technological sovereignty requires a move towards open hardware and software. To achieve these objectives, within the KDT ISOLDE project started in 2023, we propose the development of a new family of processors for AI-based applications to be deployed on board of satellites. In this paper, we showcase some examples of space applications with their requirements, and highlight the possible solutions as

well as the corresponding work that will be carried out in ISOLDE, and the expected results.

Keywords: High Performance Computing · RISC-V · Power Modeling and Control · Space Applications

1 Introduction

Europe has led the way in defining a common open software ecosystem from the cloud to the Internet-of-Things (IoT), with Linux being the de-facto standard Operating System (OS) for academic and industrial user communities alike. The European Chips Act considers the creation and expansion of a RISC-V open-source ecosystem to be a strategic investment that will enable Europe to reach the ambitions of doubling the value of design and production of semiconductors in Europe by 2030. RISC-V already has a European and worldwide momentum as can be seen from the 3100 members from 70 countries of RISC-V International. Driven by chip developments based on open-source RISC-V cores, industry starts making RISC-V based products, whereas these activities are rarely made public. One reason is that RISC-V is often used deeply embedded and not customer visible. Nevertheless, the existing RISC-V mostly open-source eco-system is often used to build such products.

However, these cores offer compute features at the lower end of the performance scale. The high-performance CPU, GPU and ASIC (Application Specific Integrated Circuits) development is still mostly proprietary with very few - mainly US-based - players. These players are meeting the exponential increase in demand for computational performance, but the associated development costs for design, and verification before sign-off -becomes less and less sustainable. In this international context, European companies are in a subordinate position: currently they have to either buy processors designed and manufactured outside Europe, or they have to rely on non-European IP providers for their processor chip designs.

Hence, RISC-V and open hardware/software have been seen as a major opportunity for Europe to leverage on the High-Performance Computing (HPC) as well as the embedded and IoT market segments. Expertise that has been built up over the years in strategic fields such as Automotive, Aerospace, communication and industrial, will push it from application to hardware development. To date, high performance adoption has largely been by academics, open-source enthusiasts and a selected number of industrial early adopters. However, general industry interest in high performance embedded computing is growing fast despite Intellectual Property (IP) restrictions or lack of industry strength CPU/GPU cores at low costs or even being royalty free.

1.1 The ISOLDE Project

ISOLDE is a new project within the Key Digital Technologies Joint Undertaking (KDT JU) that aims at improving the functional and non-functional properties

of European high-performance RISC-V-based CPUs within the next five years, to reach or surpass the established competitors and proprietary alternatives. This will be achieved by exploring and implementing advanced architectures and in addition by providing novel accelerators as well as IPs to complete the high-performance compute infrastructure based on inputs of partners that cover the entire value chain. Further, this goal will be supported by following and contributing to specifications from suitable industrial bodies and to Europe's long-term strategy for RISC-V-based ecosystem, including the creation of a repository of industry-grade building blocks to be used for SoC designs in different application domains, such as automotive, industrial, and aerospace. The ISOLDE approach includes all players along the value chain, covering, besides hardware designs, also electronic design automation tools (EDA), the full software stack, as well as a range of industry-strength use-case applications. The ISOLDE ecosystem will contribute to achieve a European sovereignty.

The broad, industry-led ISOLDE consortium includes the largest EU companies and globally operating semiconductor IDMs, thus enabling a large number of engineer to gain exposure to RISC-V technologies, as well as to bridge the confidence gap needed to persuade the industry to make investments needed to tapeouts through the development of prototype solutions employing well-verified, efficient, open-source RISC-V building blocks, as well as by supporting these building blocks with the appropriate software infrastructure, documentation, and benchmarks.

1.2 Objective of the Paper and Its Organization

The objective of this paper is to describe the aerospace scenario in terms of requirements and technological solutions, that are planned to be developed within the ISOLDE project. After a general introduction to the tight requirements of this critical class of applications in Sect. 2, the platform that will be considered is detailed in Sect. 3. The description of the main technologies that will be made available for the development of the platform and of the application is provided in Sect. 4 where both aspects related to specific functionalities as well as to the support of non functional properties are contained. Some conclusions are drawn in Sect. 5, showing also an outline of the work that will be carried out during the three years of the ISOLDE project.

2 Requirements of Modern Space Applications

The increment of computational power can permit to evolve the services offered both, on-board and ground side, for space applications likewise for other application domains. On the ground side, the HPC system are able to support the scientific data processing needs whereas for the on-board embedded applications, following the edge computer paradigm, the direct processing of data acquired can permit to shorten the overall system latency time as well as to improve the quality of the service selecting only significant data with the support of AI.

An example is the satellite Fault Detection Isolation and Recovery (FDIR) that is organized as a hierarchical architecture aiming at detect, isolate and recover faults at unit, subsystem or equipment level. It is a deterministic approach based on several predefined tables containing selected monitoring items and relative recoveries. These tables are identified leveraging the a-priori experience of the domain knowledge of space subsystems manufacturer and then implemented in the Avionic SW (ASW). The goal is to verify that the telemetry channels, computed by the ASW, does not exceed the operational thresholds: if a monitoring criterion is violated, a failure is detected and a recovery action will be performed. The a priori identification and design of both the thresholds and the confirmation time affect the flexibility and performance of the approach, without guaranteeing predictive capabilities and therefore preventive maintenance. For this reason, the use of ML-based algorithms (like AutoEncoder) can significantly enhance the capabilities of the on-board FDIR, especially in identifying the failures at channel and sub-system level, favoring equipment reuse and potentially extending the mission operation life.

Another example of space application, of particular interest for Leonardo, is hyperspectral imaging (HSI): this is an advanced technology that allows for the collection of a wide range of spectral data acquired by remote sensors, such as those present on satellites. This method has been shown to be useful for a wide range of applications, e.g. object detection, classification and material recognition; this is due to the fact that hyperspectral images provide unique material fingerprints, which can be used to identify different types of materials. These tasks are particularly suited for Deep Learning based methods, which have been become dominant for visual-related problems in the last few years. One example of DL model are CNN-3D, which are able to exploit both spatial and spectral features from the images. This kind of image analysis can have practical applications in fields such as remote sensing, geology, environmental monitoring and target detection.

The specific space application domain (telecommunication, earth observation, space exploration, ..) requires to embedded applications dedicated HW/SW Co-design customization to reach the desired performances taking also into account non functional constraints (power budget, radiation tolerance, form factor, costs, ..). A RISC-V based system can be an opportunity to apply the desired customization maintaining an HW/SW product that can be easily evolved following the mission specific requirements.

3 A RISC-V Processor Family for Onboard AI

ISOLDE targets an aggressively heterogeneous architecture to cope with the requirements of onboard processing and, in particular, Artificial Intelligence (AI). Space environment is hostile, and thus systems such as satellites and spacecraft need to tackle criticality by reacting fast. This requires a tight link between onboard processing, communication, sensing, and actuation elements [20], to complete tasks in a short enough amount of time to meet deadlines and ensure functionality. Onboard computing capabilities are crucial to reduce the latency

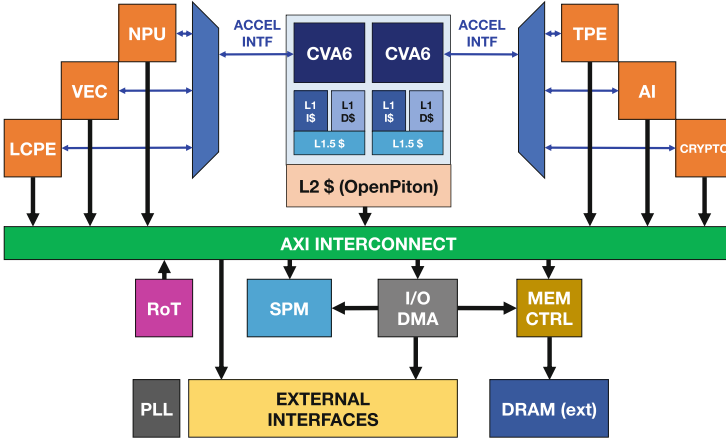


Fig. 1. ISOLDE heterogeneous architecture template for Onboard AI, with sample technology bricks.

overhead given by raw data transmission, allowing the data processing to be directly performed onboard a spacecraft (for example, for image processing or fault detection, isolation and recovery through machine-learning-based techniques), sharing only valuable information on the communication line.

A heterogeneous “RISC-V + accelerators” architecture tackles both the criticality of selected applications and the necessity for strong general-purpose computing capabilities. The ISOLDE space-centered use-case will be centered on a state-of-the-art multi-core Linux-capable RISC-V processor: CVA6 [35], originally developed as Ariane by ETH Zürich and more recently maintained by the *OpenHW Group* industry consortium. Figure 1 shows a simplified diagram of the ISOLDE heterogeneous architecture template, centered on a dual-core L2-coherent version of CVA6 that will be developed within the project starting from OpenPiton. The CVA6 are compounded by a set of tightly and loosely coupled accelerators, of which some examples are shown in the diagram: Neural Processing Units (NPU), Vector co-processors (VEC), loosely coupled Parallel Engines (LCPE), Tensor Processing Engines (TPE), AI accelerators, and crypto accelerators. These will be linked to the CVA6 cores by means of a common accelerator interface developed within the ISOLDE project, possibly inspired by the interface of the Ara vector co-processor [8]. The system is completed by a Root-of-Trust unit to enable secure boot and other trusted computing services.

4 Technology Bricks

In this section, we present the technical advances pursued by the ISOLDE project to enforce the requirements of aerospace applications, in terms of hardware accelerators and compiler and system software support for energy efficiency, real-time operation, and security, with a particular focus on the need of AI-based applications.

4.1 Hardware Accelerators

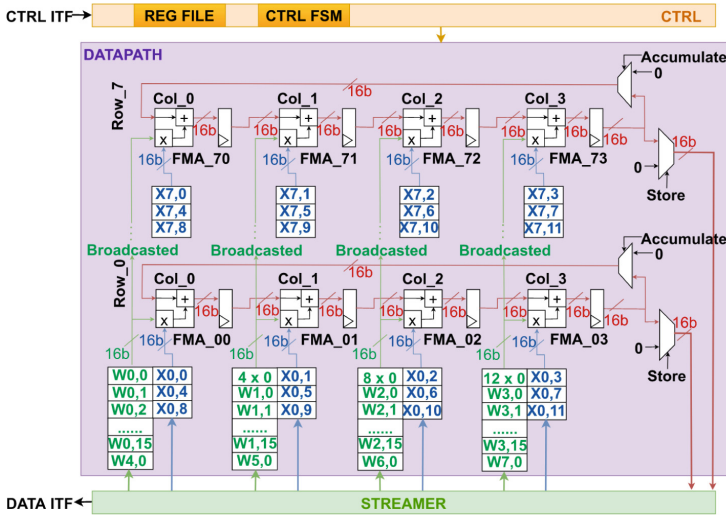


Fig. 2. Tensor Processing Engine (TPE) base architecture [18,33].

Tensor Processing Engine. Compute-intensive matrix multiplication operations are ubiquitous in Signal/Image Processing, Control, Machine Learning, and Deep Learning, ISOLDE will integrate a Tensor Processing Engine (TPE) with the RISC-V core of the ISOLDE platform (Fig. 2). TPEs focus on accelerating matrix multiplication of the kind $D = A \times B + C$, exploiting an internal high-efficiency systolic structure extending RedMule [18,33], an open-source systolic array with multi-precision Fused Multiply-Add Modules that achieves up to 920 GFLOPS/W when operating on FP8 inputs with FP16 accumulators and 775 GFLOPS/W on full FP16. ISOLDE aims at further extending the TPE capabilities in several directions: more internal and input/output formats; tight integration with the RISC-V CVA6 cores to enable TPE utilization within performance-critical software code; larger performance gains; and better integration with software.

Vector and Parallel Processing Accelerators. Complex digital signal processing algorithms in aerospace applications, such as SAR and hyperspectral imaging [6] or channel code decoders [12] and cryptography primitives [21], require significant processing capabilities. Moreover, such applications exhibit intrinsic data parallelism, which can be exploited to increase the throughput and reduce the power consumption. For these reasons, vector and parallel processing accelerators are an interesting and promising solution. During the project vector and parallel

processing approaches will be investigated in order to design loosely coupled engines connected to a RISC-V core. Such an approach requires to consider and investigate both effective interfaces to connect the accelerator to the processor and optimized computation units, able to reach the required processing speed while keeping the area limited. A relevant aspect which will be investigated in this context is flexibility, namely studying which classes of algorithms can share similar hardware structures to maximize the resources utilization.

Accelerators for AI. AI is known to be a computationally intensive task, which can benefit from dedicated accelerators [7]. Besides, different approaches have been proposed to implement AI systems based on artificial neural networks, ranging from Convolutional Neural Networks (CNNs), to Spiking Neural Networks and neuromorphic hardware. In this context, one of the objectives will be to design novel CNN accelerators able to trade an imperceptible accuracy reduction for energy savings. To achieve such a result, arithmetic operators, mainly multiplier and adder architectures, will be studied [31]. In particular, starting from the literature different alternatives will be studied and compared (e.g. [32]), with the aim of improving their structure or even to propose new solutions. On the other side, neuromorphic computational paradigms and hardware architectures have matured to a level where their ability to learn and adapt to changing conditions and tasks, while adhering to power constraints, make them well-suited for a wide range of applications from the edge to the HPC domains [34]. Through the encoding of signals from different types of standard digital sensors in the spike space [14], it becomes possible to fully utilize the theoretical underpinnings of neuromorphic computing paradigms when running on neuromorphic hardware. As a result, integrating neuromorphic computing systems into digital data analytic systems becomes a feasible possibility. In this project, our objective is to develop a chip-level integrated system that performs on-edge configuration of a neuromorphic platform, thereby eliminating the need for a host server for remote configuration [15]. Multiple open neuromorphic architectures will be evaluated for integration with the RISC-V technology supported in the project.

4.2 Energy Efficiency, Real-Time, and Power Monitoring

Mixed-Precision Computing. An emerging opportunity for improving the energy efficiency of edge application is represented by the possibility to trade off accuracy for performance, thereby reducing the need to rely on wide floating point units [22], and employing instead either fixed point arithmetics or narrow floating point representations (e.g., float16 or bfloat16, or even custom floating point sizes), as well as enabling opportunities for performance optimization, e.g. through vectorization. This opportunity is enabled by the combination of traditional digital signal processing and emerging AI applications, both of which can benefit from reduced precision. In the embedded systems domain, this is usually exploited through a manual redesign of the algorithm by an expert designer, but it is still a tedious and error prone task. Furthermore, the exploration of custom

floating point types is unfeasible without a degree of automation both at the compiler and at the unit design side. In the AI domain, mixed precision results from quantization, but is currently managed at a very coarse grain.

In ISOLDE, we aim at combining two existing methodologies: on the compiler side, TAFFO [10], a set of plugins for LLVM that support mixed precision computing; on the hardware design side, an FPU design template to develop custom mixed precision units within a RISC-V core [39]. TAFFO was first developed as part of the ANTAREX FETHPC project [30] to support precision tuning in high performance computing, and further extended in the TEXTAROSSA EuroHPC project [1] to support heterogeneous architectures. In ISOLDE, support for RISC-V platforms will be developed, together with the ability to support the design space exploration of customized architectures suitable for edge computing by means of the design template developed in [39, 40], which significantly widens the design space for RISC-V mixed precision architectures.

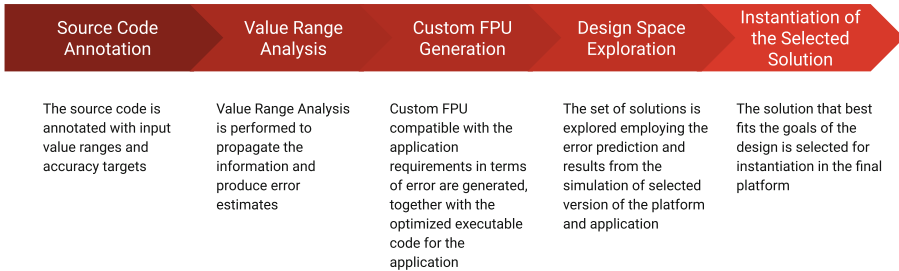


Fig. 3. The ISOLDE Mixed Precision Co-Design Flow

Figure 3 shows the proposed design flow, highlighting the main steps. The first step, *source code annotation*, is performed by the designer, leveraging his knowledge of the application domain. The second and third step entail the use of TAFFO, and the third step leverage the integration of TAFFO with the FPU design template. The fourth step, *design space exploration*, can be performed with any existing methodology – several of which come with the support of tools, which might be necessary if the design space is particularly large.

Real-time Properties Assessment and Enforcement. The problem of estimating a safe and tight Worst-Case Execution Time (WCET) in modern processors is an open problem and key issue of real-time embedded systems, and, for certain applications, of HPC computing nodes [26]. The increasing complexity of modern computing platforms (many cores, advanced pipelines, multi-level caches, heterogeneity, etc.) hinders the ability to estimate the exact WCET, forcing to introduce several approximations to make the computation of an overestimated WCET feasible. However, such overestimation can result in multiple orders of magnitude larger than the real WCET. Researchers in the last years explored

probabilistic techniques to estimate the so-called probabilistic-WCET, even if several fundamental open problems still exist [25].

The ISOLDE project aims to exploit the compiler (particularly LLVM), to improve the WCET estimation. The compiler can be the producer or the consumer of such information: on one hand, the compiler can provide information regarding the worst-case execution path, memory accesses information, and other compile-time analyses already available in LLVM, because exploited by LLVM passes for optimization reasons; on the other hand, the compiler can perform targeted optimizations to the WCET, instead of the usual average-case, including the previously described mixed-precision. To perform such optimizations, the information of the WCET must be brought up to the intermediate-level representation of LLVM, which is a challenging and open problem. An example of integration between compilers and timing information has been recently published by Cagnizi et al. [3], where the compiler inserts hooks into the code to improve the WCET estimation at run-time.

These novel and challenging research activities based on the integration of the compilation flow with the WCET estimations make the use of hybrid WCET analyses even more interesting. The compiler can provide the static part of the analysis, for example by detecting the worst-case execution path in the control-flow graph, while the probabilistic tool [27] can provide information on the single basic block execution time. The ISOLDE project aims to develop novel tools and compiler passes targeting any architecture, but with a special focus on the RISC-V architecture features that are subjects of the project.

Power Monitoring for RISC-V Accelerators. The possibility to perform comprehensive system-level power monitoring and optimization requires taking into account all the components of the overall power consumption [41]. One value added of ISOLDE is providing a design flow capable of automatically augmenting any hardware accelerator with an all-digital power monitor [13,38]. In such a way, reconfiguring the FPGA to accommodate different accelerators according to the evolution of the workload still maintains the possibility to easily perform power management, being the power monitors merged with the functionality of the accelerators.

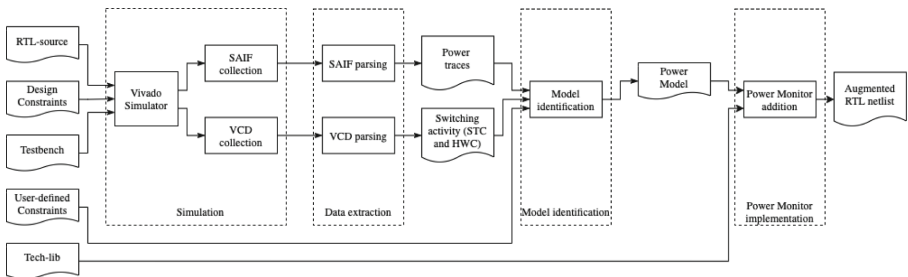


Fig. 4. Top level view of the flow to generate the on-line power monitors.

Figure 4 depicts the toolchain to generate a hardware-level on-line power monitor. The entry point is the description of the hardware to be monitored (`RTL_Source`), the set of design constraints (`Design Constraints`), a Testbench (`Testbench`), any user-defined constraints (`User-defined Constraints`), and the technology library files (`Tech-lib`). Design constraints are expressed in terms of timing and physical requirements e.g., respectively, operating frequency and pinout for the accelerators, while user-defined constraints allow the user to specify the maximum amount of resources to implement the power monitoring infrastructure, to keep under control the area overhead. At the end of the flow, the original RTL netlist is augmented with a run-time power monitoring infrastructure (`Augmented RTL-netlist`).

The realization of the power monitor can be viewed as the sequence of four different steps: (i) Simulation, (ii) Data extraction, (iii) Model Identification and (iv) Power Monitor Implementation. Initially the design is simulated using the Testbench and the provided constraints, generating in output two files containing power values (SAIF – Switching Activity Interchange Format) and switching activity information (VCD – Value Change Dump). The data extraction step parses the SAIF and VCD files preparing and filtering the data for the step (iii). During the Model Identification stage, a power model is identified minimizing the accuracy error within the resource budget. Finally, the RTL description of the computing platform is merged with a power monitoring infrastructure implementing the identified power model.

During the project, a proper template to map the implementation of the power models will be developed, in order to make possible the automatic generation of all-digital on-line power monitors.

4.3 Security

Post-quantum Cryptography Acceleration. In the upcoming decades, quantum computers are expected to break the currently employed public-key cryptography (PKC) schemes, which are fundamental to secure communication protocols. The USA’s National Institute of Standards and Technology (NIST) is conducting a standardization process for post-quantum cryptography (PQC) schemes, divided into key encapsulation mechanisms (KEM) and digital signature schemes, that can substitute the current PKC standards and government agencies are already planning their adoption and deployment. The worse performance and larger memory requirements compared to traditional PKC solutions and the need to deploy PQC schemes across the computing continuum, from data centers to low-power embedded devices, make it paramount to provide efficient support at the hardware level. The literature presents various solutions for KEMs and digital signature schemes, ranging from human-designed accelerators for the main arithmetic operations [17, 36, 37] and cores implementing whole cryptosystems [16] to HLS-generated modules executing portions of those schemes [23].

In ISOLDE, we aim to first identify PQC schemes to implement, among those being standardized in worldwide efforts, taking into account their ease of inte-

gration into existing secure communication protocols, their security properties, their performance in terms of latency and throughput, their suitability to hardware acceleration, and the applicability of protections against SCA attacks. The implementation of the accelerators for the identified cryptosystems will explore multiple design choices that include implementing accelerators for the whole schemes or only for the most computationally expensive operations, designing separate cores dedicated each to a different scheme or aiming for shared components supporting operations shared between multiple schemes, and realizing those accelerators through human-described RTL designs or HLS-generated components. Finally, additional efforts will be devoted to enhancing the hardware accelerators for PQC with cryptographically secure randomness sources and protections against physical attacks.

Trusted Execution Environment and Root of Trust. With the advancement in technology, cyber-attacks have become more refined, and security measures have advanced accordingly, at the cost, however, of higher consumption of resources. These measures include the integration of HW components, called Root of Trust, that can be trusted by design to implement features like secure boot and acceleration of crypto functions. Additionally, in critical domains, higher level of security requires to implement the concept of secure execution. It refers to the execution of programs in a controlled environment that isolates them from the underlying system and other programs, thereby preventing unauthorised access or tampering. In both industry and research, several solutions have been introduced to implement the concept of *secure execution* [19]. Amongst them, Global Platform (GP), a technical standards organisation, aims to define a common standard that describes a secure environment, named Trusted Execution Environment (TEE), operating alongside the regular operating system, the Real Execution Environment (REE), to provide a trusted execution environment for sensitive operations. TEE environments are designed to be tamper-resistant, isolated from the rest of the system, and cryptographically secured to protect against unauthorized access, malware, and other security threats.

The REE and the TEE can share the same computational resources (e.g. Trustzone, Keystone) or reside in dedicated ones [9, 24]. Hardware TEE employs specialized hardware components to create a secure and isolated environment. They typically use hardware-based isolation mechanisms, such as memory encryption, secure boot, and secure key storage, to protect against side-channel attacks, tampering, and reverse engineering. Although Hardware TEE offers high levels of security and performance, it can be expensive and less flexible than Software TEE. However, the added hardware can be optimised according to the cryptographic task desired.

In ISOLDE, we will explore how to integrate and efficiently interface RoT and TEE based on open hardware components, such as OpenTitan [11]. It has been designed by lowRISC in partnership with Google and other commercial and academic partners to act as silicon RoT and TPM (Trusted Platform Module). The objective is to enable a flexible but tamper-resistant solution providing RoT functionalities and secure execution as TEE. We will study and design software

and hardware interfaces to enable the wanted security features with minimum performance overhead.

Control Flow Integrity. Cyber-physical systems in safety-critical application domains are equipped with devices, such as TPM (Trusted Platform Modules), to support secure boot and firmware signature verification, preventing the execution of malicious code. Also, encryption and authentication protocols reduce the probability of shipping malicious code as part of payload in the network messages. To break these security defenses, Code-Reuse Attack (CRA) is a technique of exploitation which relies on executing the code which is already present in the memory (e.g. as part of the standard library). This significantly complicates the job of the attack mitigation software since the surface area of the attack shrinks and makes it much harder to detect and distinguish from the legitimate traffic. Code Reuse Attacks can force arbitrary, possibly Turing-Complete, behaviours. To achieve this target, techniques such as Return-oriented programming (ROP) are adopted. ROP allows an attacker to execute arbitrary code in the “.text” segment of a vulnerable process by chaining a set of attacker chosen gadgets. A gadget is a snippet of code placed in the execution memory of the vulnerable process, ending with a “ret” instruction. The attacker exploits a stack vulnerability to overwrite the return address of the current routing with the addresses of the sequence of gadget to be executed. Code Reuse Attacks (CRAs) poses serious challenges to computer security even when memory protection is enforced. State-of-the-Art literature works show that ROP attacks are Turing-Complete and can target also RISC-V architectures.

Control-Flow Integrity CFI is a general term for computer techniques that aim at preserving any attacker to redirect the control flow of a process. CFI can be enforced using either software or hardware approaches or both. Most software approaches are based on binary instrumentation, where a custom toolchain adds instructions to enforce security checks. The main advantage is that no dedicated hardware is required to protect vulnerable processes. However, typically, software approaches may impose very high runtime overhead, depending on the executed program. The HW alternative is to use external CFI coprocessors to dynamically check process execution. These solutions have lower execution overhead as the check is performed in parallel by the coprocessor. From the other side, a HW design overhead is required to implement the CFI coprocessor or to support the execution of custom CFI instructions.

In ISOLDE, we will explore fully open hardware solutions to implement CFI schemes on RISC-V processors. The objective is to ensure a good trade off between tight integration of the CFI monitor (needed to enable real-time detection of control-flow diversions) and ease of programmability of CFI policies and upgradability of the HW components. Along this direction, we will explore the adoption of OpenTitan [11] Root of Trust as CFI monitor to observe the instruction stream to enforce security policies. Moreover, crypto accelerators present in OpenTitan can be exploited to authenticate CFI data structure in main memory to avoid tampering.

4.4 AI Optimization Toolchain

Compilation, NAS, Mixed-precision for DNNs. Within the scope of the project, which includes the development of several new solutions in terms of computational hardware blocks, we will build on top of the current infrastructure in terms of facilities to deploy AI solutions on edge to further extend the boundaries of edge computation, unburdening the designers from long and tedious model design and optimization phases.

Specifically, during the five-year ISOLDE project, two key contributions will be made to advancing edge AI computation: The initial step will be to propose new neural architecture search (NAS) methods to optimize networks for the new hardware platforms being introduced in the project. We will construct on top of the lightweight NAS algorithm [28,29] to incorporate hardware models and customize the AI architectures to maximize the use of new accelerators, in contrast to state-of-the-art solutions that do not consider specific hardware platforms [4,5]. Additionally, we will “democratize” our algorithms to produce models that work with the new hardware platforms and maximize their memory utilization and computation capability *within a single search*.

As a second contribution in the same direction, ISOLDE will draw on the expertise of researchers who have developed advanced compilation toolchains for specific hardware platforms. Starting from existing open-source tools that target RISC-V platforms, such as DORY, Deployment Oriented to Memory, [2], focusing on optimizing memory access patterns for improved performance, or HTVM, an extension of TVM that enables deployment on heterogeneous RISC-V-based platforms equipped with multiple accelerators, we aim to develop more general and optimized tools for exploiting the full potential of accelerators and vector processing units in high-performance computing applications.

5 Concluding Remarks

The ISOLDE project is part of a family of European initiatives to spearhead the development of Open Source hardware, a critical need for the European technological sovereignty, since the European Union does not have major proprietary ISAs and micro-architectures. Within this effort, the role of ISOLDE is raising the Technology Readiness Level (TRL) of RISC-V-based solutions, providing demonstrators to key industrial sectors leveraging reusable technology bricks, including both hardware components (e.g., accelerators) and system software (e.g., compiler extensions), to hasten the uptake of RISC-V by the European industry. Within ISOLDE, the Italian cluster, composed of 3 companies and 3 universities, focuses on the aerospace domain, covering both traditional applications and novel AI-based ones. In this paper, we highlighted the requirements posed by the aerospace sector, presented the ISOLDE architecture template for onboard AI, and introduced the planned accelerator components and system software.

Acknowledgements. This work is partially supported by the European Commission and the Italian Ministry of Enterprises and Made in Italy (MIMIT) under the KDT ISOLDE project (G.A. 101112274). Web site: <https://www.isolde-project.eu/>.

References

1. Agosta, G., et al.: Towards extreme scale technologies and accelerators for eurohpc hw/sw supercomputing applications for exascale: the textarossa approach. *Microprocess. Microsyst.* **95**, 104679 (2022). <https://doi.org/10.1016/j.micpro.2022.104679>
2. Burrello, A., Garofalo, A., Bruschi, N., Tagliavini, G., Rossi, D., Conti, F.: Dory: automatic end-to-end deployment of real-world DNNs on low-cost IoT MCUs. *IEEE Trans. Comput.* **70**(8), 1253–1268 (2021)
3. Cagnizi, L., Reghenzani, F., Fornaciari, W.: Poster abstract: run-time dynamic WCET estimation. In: *Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation*, pp. 458–460. IoTDI 2023, Association for Computing Machinery, New York, NY, USA (2023). <https://doi.org/10.1145/3576842.3589168>
4. Cai, H., Gan, C., Wang, T., Zhang, Z., Han, S.: Once-for-all: train one network and specialize it for efficient deployment. *arXiv preprint arXiv:1908.09791* (2019)
5. Cai, H., Zhu, L., Han, S.: Proxylessnas: direct neural architecture search on target task and hardware. *arXiv preprint arXiv:1812.00332* (2018)
6. Caon, M., et al.: Very low latency architecture for earth observation satellite onboard data handling, compression, and encryption. In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp. 7791–7794 (2021). <https://doi.org/10.1109/IGARSS47720.2021.9554085>
7. Capra, M., Bussolino, B., Marchisio, A., Maserà, G., Martina, M., Shafique, M.: Hardware and software optimizations for accelerating deep neural networks: survey of current trends, challenges, and the road ahead. *IEEE Access* **8**, 225134–225180 (2020). <https://doi.org/10.1109/ACCESS.2020.3039858>
8. Cavalcante, M., Schuiki, F., Zaruba, F., Schaffner, M., Benini, L.: Ara: a 1-GHz+ scalable and energy-efficient RISC-V vector processor with multiprecision floating-point support in 22-nm FD-SOI. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **28**(2), 530–543 (2020). <https://doi.org/10.1109/TVLSI.2019.2950087>
9. Cerdeira, D., Santos, N., Fonseca, P., Pinto, S.: Sok: understanding the prevailing security vulnerabilities in trustzone-assisted TEE systems. In: *2020 IEEE Symposium on Security and Privacy (SP)*, pp. 1416–1432 (2020). <https://doi.org/10.1109/SP40000.2020.00061>
10. Cherubin, S., Cattaneo, D., Chiari, M., Agosta, G.: Dynamic precision autotuning with TAFFO. *ACM Trans. Archit. Code Optim.* **17**(2), 1–26 (2020). <https://doi.org/10.1145/3388785>
11. lowRISC CIC: Opentitan official documentation (2019). <https://opentitan.org/documentation/index.html>
12. Condo, C., Maserà, G.: Unified turbo/LDPC code decoder architecture for deep-space communications. *IEEE Trans. Aerosp. Electron. Syst.* **50**(4), 3115–3125 (2014). <https://doi.org/10.1109/TAES.2014.130384>
13. Cremona, L., Fornaciari, W., Zoni, D.: Automatic identification and hardware implementation of a resource-constrained power model for embedded systems. *Sustain. Comput. Inf. Syst.* **29**, 100467 (2021). <https://doi.org/10.1016/j.suscom.2020.100467>

14. Forno, E., Fra, V., Pignari, R., Macii, E., Urgese, G.: Spike encoding techniques for IoT time-varying signals benchmarked on a neuromorphic classification task. *Frontiers Neurosci.* **16**, 999029 (2022)
15. Forno, E., Spitale, A., Macii, E., Urgese, G.: Configuring an embedded neuromorphic coprocessor using a risc-v chip for enabling edge computing applications. In: 2021 IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc), pp. 328–332. IEEE (2021)
16. Galimberti, A., Galli, D., Montanaro, G., Fornaciari, W., Zoni, D.: FPGA implementation of bike for quantum-resistant TLS. In: 2022 25th Euromicro Conference on Digital System Design (DSD), pp. 539–547 (2022). <https://doi.org/10.1109/DSD57027.2022.00078>
17. Galimberti, A., Montanaro, G., Zoni, D.: Efficient and scalable FPGA design of GF(2m) inversion for post-quantum cryptosystems. *IEEE Trans. Comput.* **71**(12), 3295–3307 (2022). <https://doi.org/10.1109/TC.2022.3149422>
18. Garofalo, A., et al.: DARKSIDE: a heterogeneous RISC-V compute cluster for extreme-edge on-chip DNN inference and training. *IEEE Open J. Solid-State Circ. Soc.* **2**, 231–243 (2022). <https://doi.org/10.1109/OJSSCS.2022.3210082>
19. Jauernig, P., Sadeghi, A.R., Stapf, E.: Trusted execution environments: properties, applications, and challenges. *IEEE Secur. Priv.* **18**(2), 56–60 (2020)
20. Klesh, A.T., Cutler, J.W., Atkins, E.M.: Cyber-physical challenges for space systems. In: 2012 IEEE/ACM Third International Conference on Cyber-Physical Systems, pp. 45–52 (2012). <https://doi.org/10.1109/ICCPS.2012.13>
21. Koleci, K., Santini, P., Baldi, M., Chiaraluce, F., Martina, M., Masera, G.: Efficient hardware implementation of the LEDAcrypt decoder. *IEEE Access* **9**, 66223–66240 (2021). <https://doi.org/10.1109/ACCESS.2021.3076245>
22. Lasri, I., Cherubin, S., Agosta, G., Rohou, E., Sentieys, O.: Implications of reduced-precision computations in HPC: performance, energy and error. *Parallel Comput. Everywhere* **32**(2018), 297 (2018)
23. Montanaro, G., Galimberti, A., Colizzi, E., Zoni, D.: Hardware-software co-design of bike with HLS-generated accelerators. In: 2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS), pp. 1–4 (2022). <https://doi.org/10.1109/ICECS202256217.2022.9970992>
24. Pinto, S., Santos, N.: Demystifying arm trustzone: a comprehensive survey. *ACM Comput. Surv.* **51**(6), 1–36 (2019). <https://doi.org/10.1145/3291047>
25. Reghenzani, F., Massari, G., Fornaciari, W.: Probabilistic-WCET reliability: statistical testing of EVT hypotheses. *Microprocess. Microsyst.* **77**, 103135 (2020). <https://doi.org/10.1016/j.micpro.2020.103135>
26. Reghenzani, F., Massari, G., Fornaciari, W.: Timing predictability in high-performance computing with probabilistic real-time. *IEEE Access* **8**, 208566–208582 (2020). <https://doi.org/10.1109/ACCESS.2020.3038559>
27. Reghenzani, F., Massari, G., Fornaciari, W., et al.: chronovise: measurement-based probabilistic timing analysis framework. *J. Open Source Softw.* **3**, 711–713 (2018)
28. Risso, M., et al.: Lightweight neural architecture search for temporal convolutional networks at the edge. *IEEE Trans. Comput.* **72**, 744–758 (2022)
29. Risso, M., et al.: Pruning in time (PIT): a lightweight network architecture optimizer for temporal convolutional networks. In: 2021 58th ACM/IEEE Design Automation Conference (DAC), pp. 1015–1020. IEEE (2021)
30. Silvano, C., et al.: The ANTAREX tool flow for monitoring and autotuning energy efficient HPC systems. In: Internat. Conf. on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS), pp. 308–316 (2017). <https://doi.org/10.1109/SAMOS.2017.8344645>

31. Singh, R., Conroy, T., Schaumont, P.: Variable precision multiplication for software-based neural networks. In: 2020 IEEE High Performance Extreme Computing Conference (HPEC), pp. 1–7 (2020). <https://doi.org/10.1109/HPEC43674.2020.9286170>
32. Strollo, A.G.M., Napoli, E., De Caro, D., Petra, N., Meo, G.D.: Comparison and extension of approximate 4–2 compressors for low-power approximate multipliers. *IEEE Trans. Circuits Syst. I Regul. Pap.* **67**(9), 3021–3034 (2020). <https://doi.org/10.1109/TCSI.2020.2988353>
33. Tortorella, Y., Bertaccini, L., Rossi, D., Benini, L., Conti, F.: RedMule: a compact FP16 matrix-multiplication accelerator for adaptive deep learning on RISC-V-based ultra-low-power SoCs. In: Proceedings of the 2022 Conference & Exhibition on Design, Automation & Test in Europe, pp. 1099–1102. DATE 2022, European Design and Automation Association, Leuven, BEL (2022)
34. Urgese, G., Rios-Navarro, A., Linares-Barranco, A., Stewart, T.C., Michmizos, K.: Editorial: powering the next-generation IoT applications: new tools and emerging technologies for the development of neuromorphic system of systems. *Frontiers in Neuroscience* **17**, 1197918 (2023). <https://doi.org/10.3389/fnins.2023.1197918>
35. Zaruba, F., Benini, L.: The cost of application-class processing: energy and performance analysis of a linux-ready 1.7-GHz 64-Bit RISC-V core in 22-nm FDSOI technology. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **27**(11), 2629–2640 (2019). <https://doi.org/10.1109/TVLSI.2019.2926114>
36. Zoni, D., Galimberti, A., Fornaciari, W.: Efficient and scalable FPGA-oriented design of QC-LDPC bit-flipping decoders for post-quantum cryptography. *IEEE Access* **8**, 163419–163433 (2020). <https://doi.org/10.1109/ACCESS.2020.3020262>
37. Zoni, D., Galimberti, A., Fornaciari, W.: Flexible and scalable FPGA-oriented design of multipliers for large binary polynomials. *IEEE Access* **8**, 75809–75821 (2020). <https://doi.org/10.1109/ACCESS.2020.2989423>
38. Zoni, D., Cremona, L., Cilaro, A., Gagliardi, M., Fornaciari, W.: PowerTap: all-digital power meter modeling for run-time power monitoring. *Microprocess. Microsyst.* **63**, 128–139 (2018). <https://doi.org/10.1016/j.micpro.2018.07.007>
39. Zoni, D., Galimberti, A.: Cost-effective fixed-point hardware support for RISC-V embedded systems. *J. Syst. Architect.* **126**, 102476 (2022). <https://doi.org/10.1016/j.sysarc.2022.102476>
40. Zoni, D., Galimberti, A., Fornaciari, W.: An FPU design template to optimize the accuracy-efficiency-area trade-off. *Sustain. Comput. Inf. Syst.* **29**, 100450 (2021). <https://doi.org/10.1016/j.suscom.2020.100450>
41. Zoni, D., Galimberti, A., Fornaciari, W.: A survey on run-time power monitors at the edge. *ACM Comput. Surv.* **55**, 1–33 (2023). <https://doi.org/10.1145/3593044>