RADLER: A Permissive Open Source Hardware Platform for Increased Autonomy

Leonidas Kosmidis[1,2] Eric Rufart[1] Jannis Wolf[1] Guillermo Vidal[2,1] Marc Solé[2,1]
Javier Beiro[1,2] Nicolás Zhao[1,2] Beatriz Navidad[2] Juan Carlos Rodriguez[1] Matina Maria Trompouki[1]

[1]Barcelona Supercomputing Center (BSC)
[2]Universitat Politècnica de Catalunya (UPC)

***Abstract*—RADLER (RISC-V Autonomous Driving LEvel fouR Hardware/Software Co-design) is research project funded by Samsung Advanced Institute of Technology, which is focused on the development of an open source, RISC-V based autonomous driving platform for L4 autonomous vehicles, with a permissive license. The RADLER System-on-Chip platform is similar to existing MPSoC-GPU architectures for high performance automotive systems, consisting of Safety, Application and GPU IPs. Thanks to the built-in reliability features required for automotive functional safety compliance, the RADLER platform is a good candidate for space systems.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Autonomous Driving for SAE Level 4, i.e. high driving automation in which a vehicle is self-driving under most circumstances, requires an increased performance capability, which can only be provided by complex hardware architectures, involving multicores, Artificial Intelligence (AI) accelerators and Graphics Processing Units (GPUs) capable of general purpose processing. At the same time, non-functional properties such as reliability, functional safety certification according to ISO 26262, worst case execution time (WCET) computation and programmability are key challenges, which need to be addressed by their hardware and software.

Despite the availability of several high performance platforms which satisfy some of these features, there is no hardware platform that complies with all of these requirements. For example, NVIDIA's Xavier and Orin platforms are examples of ASIL-D (i.e. the highest integrity level in the automotive safety standard) certified hardware platforms which can provide very high performance, suitable for L4 autonomy needs. However, several of their features are either undocumented like RAS (Reliability and Serviceability), or not publicly available like how to use their lockstep processors and the internal details of their CPU and GPUs. Some of this information is only available through hard to obtain Non-Disclosure Agreements (NDAs) or kept as commercial secrets like the CPU and GPU internals.

On the software side, due to the closed source nature of the GPU specification and its software stack, it is not possible to use the GPU under a certified (real-time) operating system, since GPU drivers are only available for Linux and Android which are not possible to be used in ASIL-D software environments. Note that these challenges are shared among all GPU vendors not only NVIDIA, including AMD, ARM and Imagination Technologies to name a few.

Despite these challenges, it has been demonstrated that these architectures can provide much higher computational performance than existing space processors [7]. Moreover, as shown in the ESA ASIL2ECSS project [6], automotive solutions compliant with automotive certification standards are good candidates for adoption in space, since they offer similar requirements with space systems, and have much higher manufacturing quality with reproducible processes than regular COTS devices.

At the same time, the rise of RISC-V, an open specification Instruction Set Architecture, allows the possibility to overcome all these aforementioned challenges. In addition to the open specification, which allows to leverage software stacks developed for other compliant processors, there are also several open source hardware designs, which can be customised according to the particular needs of both automotive and space safety critical domains.

For this reason, in the RADLER project we perform the hardware/software co-design of a high performance, RISC-V based platform for L4 autonomous driving and its software stack. However, as already mentioned, given safety critical domains have similar requirements, the RADLER platform can be used in other markets too.

Both the hardware and software stacks are going to be released as open source with a permissive license, allowing not only further research on this domain, but also to be commercially adopted by industry.

The project started in January 2024, so this abstract focuses on the developments performed during the first year. The following research goals have been achieved:

- A functional prototype of the high performance RISC-V platform on an FPGA. The platform contains: a) a high performance general purpose application core enhanced with AI acceleration capabilities, b) a multithreaded, dual-lockstep safety core enhanced with AI acceleration capabilities and c) a configurable RISC-V based general purpose GPU. All hardware components are highly configurable, both types of CPUs support configurable

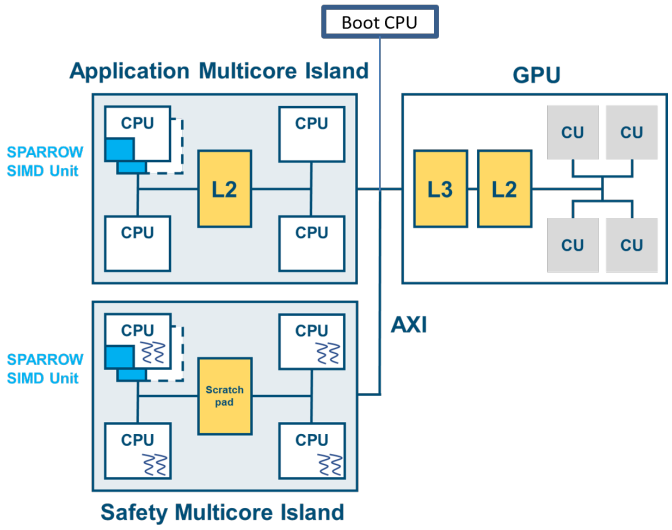

Fig. 1. RADLER Platform overview.

dual lockstep functionality and support for Worst Case Execution Time (WCET).

- A virtual platform of the designed architecture has been developed, to facilitate software development without the need of slow RTL simulations or the availability of several expensive FPGAs.
- A certifiable software stack. This includes the use of the GPU from a bare-metal and a qualifiable real-time operating system and a qualifiable GPU compiler and software stack.

Our work leverages extensive research work performed in our research team both targeting the automotive sector, as well as other certified safety critical domains like space and avionics.

II. Hardware and Software Overview

The RADLER hardware/software co-designed platform resembles safety critical MPSoCs like AMD/Xilinx Zynq Ultrascale+ or the NVIDIA Xavier or Orin GPU multicore platforms. It has the following features:

a) A secure boot core. RADLER uses a BSC-designed microcontroller-class RISC-V processor supporting the minimum RISC-V ISA, SafeTCo, which is responsible of the platform boot process and setup. It is similar to the hard Microblaze IP used in Ultrascale+ devices, or the secure boot ARM R5 cores in NVIDIA GPU platforms. It uses a read-only memory and data scratchpad.

b) A set of Functional Safety processors. These cores are simple, time predictable processors focused on safety, which can be used for the processing of time sensitive, critical tasks including AI, and they play a similar role with the ARM R5 cores used in the aforementioned platforms. The RADLER platform includes the high performance VeeR EH1 and dual threaded VeeR EH2 core (formerly known as SweRV) developed by Western Digital and currently managed by Chips Alliance. This processor is very similar to automotive ECUs like Infineon's Aurix TriCore microcontrollers, since it includes both scratchpads which enable time predictability as well as caches, and contains ECC protection, which makes it a good candidate for ISO 26262 certification. The core has been modified, including dual-lockstep functionality, support for worst case execution time, timing predictable branch prediction, and integration with the SPARROW AI accelerator [2]. Moreover, the RADLER EH2 variant was modified to include a predictable multithreading feature, which allows the execution time of one thread not to be negatively affected by the other thread. In this way, the WCET of software can be computed, without knowing the software executed in the other thread. Such features like robust partitioning and dual lockstep functionality are currently offered by NASA's HPSC [8].

b) A Set of Application Processors. RADLER uses BSC's Sargantana RISC-V Application Core, which is based on the Lagarto series of processors and it is licensed under a permissive open source license. Unlike functional safety processors which only support Memory Protection Units (MPU), Application processors feature a Memory Management Unit (MMU), which allows also the execution of feature rich operating systems if needed, such as Linux/Android, for less critical tasks. The RADLER-modified Sargantana processor is also integrated with the SPARROW AI accelerator [2] and supports the computation of Worst Case Execution Time, as well as configurable dual lockstep functionality, similar to the ARM A78AE application cores in NVIDIA Orin.

c) Integration of the Application and Functional Processors with the RISC-V Vortex [9] GPU. The GPU driver and software stack have been ported to both RISC-V types of processors, and supports both bare-metal and RTOS. This means that even the functional safety processors can offload GPU computations. This is a big difference with NVIDIA platforms, in which only the application cores can do so, and only under Linux. In RADLER, the Vortex GPU has been enhanced with a TensorCore, similar to the one included in NVIDIA Xavier and Orin. Our TensorCore follows the same design and programming API, facilitating code migration from NVIDIA platforms.

d) Qualifiable software stack. The RTEMS RTOS has been ported to VeeR EH1, EH2 and Sargantana cores, as well as the Ada SPARK Ravenscar and Jorvik profiles with SMP support. TensorFlow-micro has been ported to both types of processors with support for SPARROW as well as for Vortex GPU, and it can be used both in bare metal and under RTEMS. In addition, a GPU server has been implemented, which allows multiple RTEMS tasks to obtain exclusive access to the GPU, allowing GPU time sharing. Finally, LLVM support has been added for the TensorCore of the Vortex GPU.

e) FPGA prototype. Since the RADLER platform includes several hardware resources (i.e. dual lockstep functional safety processors), and requires a powerful configuration (high number of CPU and GPU cores, big caches etc.), a large FPGA is required. For this reason, the Xilinx Virtex Ultrascale+ VCU

118 is used which is currently among the largest Xilinx FPGAs available. All IP cores of the platform are clocked at 100 MHz.

f) Virtual platform. RADLER features a virtual simulation platform based on QEMU which facilitates the software development. In particular, VeeR EH1, EH2 and Sargantana models have been developed, which support all RADLER modifications in the cores.

III. Evaluation

A. Safety Cores

Early evaluation of the AI capabilities of RADLER VeeR EH1 shows speedups of 2-5× compared to the original core version, thanks to the inclusion of the SPARROW AI accelerator. The use of the scratchpad instead of the instruction cache offers higher performance in the examined AI workloads.

Moreover, a traffic sign detection inference use case based on the German Traffic Sign Recognition Benchmark Dataset achieves a 3× speedup compared to the scalar version, thanks to the SPARROW acceleration.

In terms of WCET estimation, the RADLER VeeR EH1 achieves at least 20% lower WCET using Measurement Based Probabilistic Timing Analysis (MBPTA) [3] compared to conventional measurement based WCET estimation using the EEMBC Autobench benchmarking suite. MBPTA is a probabilistic execution time analysis method, which simplifies significantly the WCET estimation, using only end-to-end execution time measurements. It is supported by BSC's open source probabilistic Execution time WCET tools [1].

The effectiveness of this method was demonstrated in the PROXIMA EU-funded project and in the PROARTIS4Space ESA-funded project, which resulted in the time-randomised version of LEON3 [4] and LEON4 processors respectively.

In addition to Worst Case Execution Time reduction in single threaded execution configuration, RADLER's dual-threaded VeeR EH2 version offers a significant improvement of WCET in a multithreaded configuration.

The original dual-threaded VeeR EH2 core shows up to 40% slowdown compared to the single threaded execution when all combinations of EEMBC benchmarks are explored. On the other hand, RADLER's EH2 predictable multithreading with robust partitioning shows less than 1% slowdown compared to when the second thread is idle. In principle, there should be no slowdown over this guaranteed performance, however we are still investigating the reason. The cost of partitioning is up to 20% compared to the execution in isolation with multithreading disabled, and comes mainly due to the reduced size of the cache available to each thread with the robust partitioning.

In addition, RADLER's EH2 core includes an optimised predictable multithreading implementation, which allows each thread to opportunistically execute faster than its guaranteed performance, without ever slowing down the other thread. This results to up to 33% performance improvement compared to the baseline robust partitioning.

B. Application Cores

Similar to the safety cores, SPARROW provides 2× speedup for inference thanks to the use of SPARROW and 20% WCET improvement in the Sargantana core, thanks to the support of MBPTA.

C. Vortex GPU

The TensorCore implementation of the VortexCore provides a speedup between 1.4× and 2.8×. It is worth noting that currently our focus has been to have a functional hardware and software platform, but not to optimise its performance. Therefore, the obtained results are going to be improved in the future, when both RADLER hardware and software is optimised.

IV. Conclusion and Future Work

In this abstract we have presented an overview of the RADLER hardware platform and its achievements over the first year of the project.

So far, we have a fully functional FPGA prototype with an advanced software stack, with promising results.

However, the work is not yet completed, and more work is expected, mainly in the Application Core and GPU IPs. In particular, multicore support for Sargantana and support for the hypervisor extension is under development. Reliability features similar to the ones available for the safety cores will be implemented too, both for the Sargantana and the GPU.

Moreover, the GPU implementation will be improved, optimising both the AXI interface between the CPU complex and the GPU, as well as adding DMA support. Finally, modifications for WCET estimation in the GPU will be implemented.

Acknowledgments

The research presented throughout this paper has received funding by the Samsung Global Research Outreach Program. Moreover, the development of the WCET features of the Functional Safety Cores has been partially supported by the ESA funded activity "Open source software randomisation framework for probabilistic WCET prediction and security on (multicore) CPUs, GPUs and Accelerators" [5].

References

[1] Mbpta-cv, tool for the estimation of probabilistic wcet based on execution time measurements (in the form of an r script).

[2] Marc Solé Bonet and Leonidas Kosmidis. SPARROW: A Low-Cost Hardware/Software Co-designed SIMD Microarchitecture for AI Operations in Space Processors. In *DATE*, 2022.

[3] Liliana Cucu-Grosjean, Luca Santinelli, Michael Houston, Code Lo, Tullio Vardanega, Leonidas Kosmidis, Jaume Abella, Enrico Mezzetti, Eduardo Quiñones, and Francisco J. Cazorla. Measurement-based probabilistic timing analysis for multi-path programs. In Robert Davis, editor, *24th Euromicro Conference on Real-Time Systems, ECRTS 2012, Pisa, Italy, July 11-13, 2012*, pages 91–101. IEEE Computer Society, 2012.

[4] Gaisler. Leon3 probabilistic platform. https://www.gaisler.com/index.php/products/processors/leon3.

[5] Leonidas Kosmidis. Open source software randomisation framework for probabilistic WCET prediction and security on (multicore) CPUs, GPUs and Accelerators. https://activities.esa.int/4000140111.

[6] Leonidas Kosmidis. Reusing Automotive Certification and Qualification Standards. https://nebula.esa.int/4000136128.

[7] Leonidas Kosmidis, Iván Rodriguez, Alvaro Jover-Alvarez, Sergi Alcaide, Jérôme Lachaize, Olivier Notebaert, Antoine Certain, and David Steenari. Gpu4s: Major project outcomes, lessons learnt and way forward. In *2021 Design, Automation Test in Europe Conference Exhibition (DATE)*, 2021.

[8] NASA. NASA's High Performance Spaceflight Computer White Paper. https://www.nasa.gov/wp-content/uploads/2024/07/hpsc-white-paper-tmg-26jun2024-final.pdf.

[9] Blaise Tine, Krishna Praveen Yalamarthy, Fares Elsabbagh, and Kim Hyesoon. Vortex: Extending the RISC-V ISA for GPGPU and 3D-Graphics. In *International Symposium on Microarchitecture (MICRO)*, 2021.