A Safe and Secure Platform for Autonomous Driving

Leonidas Kosmidis[1,2] Eric Rufart[1] Jannis Wolf[1] Guillermo Vidal[2,1] Marc Solé[2,1]
Juan Carlos Rodriguez[1] Matina Maria Trompouki[1] Sergi Alcaide[1] Jeremy Giesen[1]

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

Abstract

In the context of the SMARTY EU project, at BSC we develop a safe and secure RISC-V platform for autonomous driving. The developed System-on-Chip platform is similar to existing MPSoC-GPU architectures for high performance automotive systems, consisting of Safety, Application, GPU and security IPs.

Introduction

Autonomous Driving 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 security are key challenges, which need to be addressed by their hardware and software.

RISC-V 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 the automotive domain.

For this reason, in this project we design a RISC-V based platform for autonomous driving and its software stack. 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.

Currently, the following goals have been achieved:

- A functional prototype of the platform on an FPGA. It 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 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.

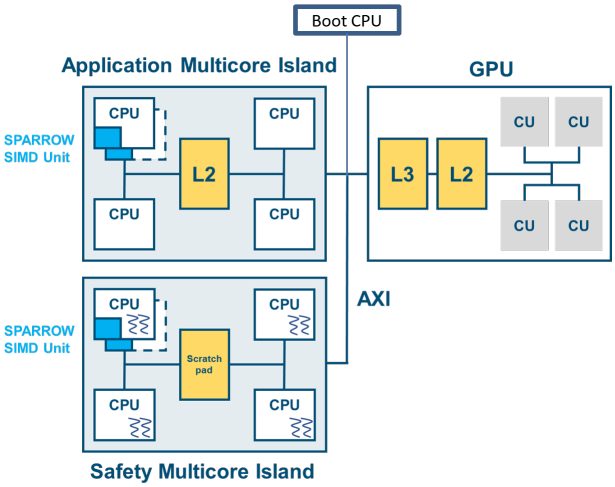

Figure 1: *Current Platform overview.*

Hardware and Software Overview

Our autonomous driving 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. A BSC-designed microcontroller-class RISC-V processor supporting the minimum RISC-V ISA, SafeTCo, 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, for processing of time sensitive, critical tasks

including AI, and they play a similar role with the ARM R5 cores used in the aforementioned platforms. Our 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. It 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 (WCET), timing predictable branch prediction, and integration with the SPARROW AI accelerator [1]. Moreover, our modified EH2 variant supports predictable multithreading, 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.

b) A Set of Application Processors. We use BSC's Sargantana RISC-V Application Core, which is based on the Lagarto series of processors. 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, such as Linux/Android, for less critical tasks. Our Sargantana processor is also integrated with the SPARROW AI accelerator [1] 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 Processor complex with the RISC-V Vortex [2] GPU. The GPU driver and software stack has been ported to both RISC-V types of processors, and supports both bare-metal and RTOS, so 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. 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. 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 our 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.

Evaluation

Early evaluation of the AI capabilities of the platform with a traffic sign detection inference use case achieves a 3× speedup compared to the scalar version, as well as at least 20% of WCET improvement.

The original dual-threaded VeeR EH2 core shows up to 40% slowdown compared to the single threaded execution when all combinations of EEMBC Autobench benchmarks are explored. Conversely, our modified EH2 predictable multithreading with robust partitioning shows less than 1% slowdown compared to when the second thread is idle. In addition, our modified EH2 core includes an optimised predictable multithreading mode, 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.

Conclusion and Future Work

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

Future work includes improvements in both in the hardware and software of the application core and the GPU, as well as the inclusion of additional safety and security features. This includes the integration of a Post Quantum Cryptography Accelerator, as well SafeSU, a statistic unit focused on the collection of SoC level statistics to increase the safety of the platform.

Acknowledgments

This work is supported by the Chips Joint Undertaking (JU), European Union (EU) HORIZON-JU-IA, under grant agreement No. 101140087 (SMARTY).

References

[1] 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.

[2] Blaise Tine et al. "Vortex: Extending the RISC-V ISA for GPGPU and 3D-Graphics". In: *International Symposium on Microarchitecture (MICRO)*. 2021.