

The SELENE Platform for Space Applications

Carles Hernández

Universitat Politècnica de València

RISC-V in Space Workshop



H2020 SELENE project grant No. 871467

14th December 2022

Outline

- Introduction
- Platform Architecture
 - Baseline SoC
- Deep Learning Toolchain
 - Hardware Acceleration
 - Programing Framework
- Safety Features
 - Mixed-criticality support
 - Contention Monitoring
 - Fault-tolerant features
- Demonstrators
 - Robotic Arm Control (ADS-DE)
 - Statellites use-case (ADS-FR)

- SELENE is a 3-year European Collaborative Research Project
 - Consortium (11 partners)
 - Academia and Research
 - UPV (coordinator), BSC, Ikerlan, Virtual Vehicles
 - Industry
 - Technology providers: SIEMENS (AT and DE), Cobham Gaisler, OpenTech
 - End Users: CAF, Airbus (FR and DE)
- From December 2019 to November 2022
- Supported by H2020 EU funding (grant nr 871467)
 ~5M€ Budget

SoC Overview

- General Purpose Processor (GPP)
- IO system
- Memory Interface system
- AI accelerators
- High performance Network On Chip(NoC) interconnect
- Debug over multiple interfaces
 - JTAG
 - UART
 - Ethernet
- Xilinx VCU118 Virtex Ultrascale Plus FPGA demonstrator

Overall SoC is <u>technology independent</u> and can be ported to other target technologies



SELENE SoC Block Diagram



SELENE SoC Block Diagram



H2020 SELENE

- Single GPP element with 6 NOEL-V cores in RV64GCH configuration with dedicated FPU and MMU
 - The <u>open-source Gaisler nanoFPU</u> is used in the open-source SELENE platform
 - The <u>commercial GRFPUnv</u> has been used for evaluation during the activity
- Pure AHB system inside GPP element
- AHBCTRL performs the main AHB bus arbitration
- APB UART, GP Timer and PLIC are connected on the APB bus
- APBCTRL connects them to the main AHB bus inside GPP element
- Debug support unit (DSU)

- Private L1 cache for each NOEL-V core
- Common L2 Cache for the GPP element
 - Supports <u>write-through</u> and <u>write-back</u> policies
 - Supports <u>cachable and non-cachable</u> accesses
 - <u>AXI backend</u> to support integration to NoC
 - IMiB size, 4-ways set associative

• GPL cores in the demonstrator design

- Megabit Ethernet interface
- General purpose IO
- RS485 UARTs

IO system

- DMA controller
- In addition to the GPL cores, the evaluation platform contains some commercial cores for use case evaluation

The IO system is tailored for the demonstrator design

Can be <u>easily modified</u> for other applications

- IOMMU to provide access protection
- Gigabit Ethernet interface.
- SpaceWire interface
- CAN-FD interface

- Xilinx Memory Interface Generator(MIG) to access the onboard DDR4 memory banks
 - The memory subsystem is tailored for the VCU118 demonstrator board when porting the SELENE platform for other targets, this should be adapted to the new technology

- An AXI based NoC to interface <u>GPP</u> elements, <u>IO system</u>, <u>accelerators</u> and <u>memory</u> interface system in SELENE SoC
- Current interconnect implements 2 crossbars
 - An **AXI crossbar** to interconnect <u>cores and accelerators to memory</u>
 - An AXI-lite crossbar to enable <u>configurations</u> and <u>cores to accelerator</u> communications
- The baseline interconnect has been extended with specific support for safety and security
 - Requests Owner ID is propagated
 - Monitoring units to control the usage of shared resources

SELENE Deep Learning Toolchain



Components

- HLSinf FPGA-based AI acceleration
- Runtime
- Machine Learning Library

HLSinf Accelerator

SELENE

https://github.com/PEAK-UPV/HLSinf

- Designed in High-level synthesis (HLS)
- Open-Source
- Customizable in functionality and performance
 - Layers to be implemented
 - Resources/Paralelism
 - Data type format
- Targets FPGA-based acceleration



SELENE

https://github.com/PEAK-UPV/HLSinf



- Modules connected with streams building a dataflow model.
- CPI = channels per input CPO = channels per output

- Uses Channel slicing.
- Module-base design allows pipelining.

SELENE Acceleration Runtime



- Objective of the runtime:
 - Offloading computations from CPU to generic accelerators using a common interface
- Capabilities
 - Kernel's control and parametric registers description from JSON file
 - Contiguous (physical) memory allocation for data input/output
 - Light OpenCL like C++ compatibility layer
 - Handles multiple kernels
- Current Limitations
 - Polling status register kernel execution (no async processes)
 - Support for 32-bit memory architecture
 - Not L2-cache coherent due to lack of HW support

SELENF

H2020 SELENE

EDDL

- EDDL (European Distributed Deep Learning Library)
 - General-purpose, open-source deep-learning library
 - Used for training and inference processes
 - Offloads heavy computations to the accelerators
 - Native support for FPGAs
 - SELENE platform has been implemented as computing target
 - Supports ONNX format



Enabling compatibility with other Frameworks (e.g tensorFlow)



Eddl



SELFNF

Using the SELENE AI framework

SELENE



H2020 SELENE

Robotic Arm Use-Case (ADS-DE)

Unit Tester Workstation Robotic Control Unit (RCU) SELENE SoC (Xilinx VCU118) Camera SpW Stimuli SpW Generator () (LIDAR. Stimuli SpW FMC SpW Generator CAN FMC I/FTest Bench Console Core 2 Core 3 Core 1 LOCARM LIDAR Nav Robot Control Pose Estimation Simulation Ethernet Power Dist. _ Unit (PDU) +

Main functionalities

- Robot Control (LOCARM)
- LIDAR Navigation
- AI Pose Estimation

- Unit Tester for Camera and LIDAR H/W
- Central user-interface
- Stimuli generators for images and LIDAR scans
- USB/SpW Interfaces
- Robotic arm simulation on Unit Tester
- Commanded via CAN Bus I/F
- Test bench manages Xilinx VCU118 via TCP/IP

Core 3: AI Pose Estimation (ADS-DE)

- Main Benchmarking application
- Neural Network uploaded on VCU118
- Camera images by stimuli generator via SpW
- Convolutional Neural Network detects key points via heatmaps, i.e. probability distributions
- Detected key points in Perspective-n-Point algorithm (OpenCV)
- Return of pose, key points and inference time to EGSE (host PC for testing purposes)
- Usage of AI acceleration, EDDL and OpenCV



H2020 SELENE



Key Point Prediction

Pose Estimate

SELENF



AI Performance (ADS-DE)

SELENE



H2020 SELENE

- HLSinf improves NOEL-V (single-core) Stacked HourGlass inference time by ~14X
- For the accelerated inference the bottleneck is on the CPU side (99% of the time)
 - Some computations/adaptations still need to be carried out by the CPU
 - 100MHz CPU prototyping frequency is obviously not helping
- Some other models tested achieve better speed-up

SELENE

Example: TinyYoloV4 for a CAF Railway Application
 NOEL-V single core Inference Time improved by ~2000X

10 images	tiny-yolov4						
	CPU	Floating point CPI/O 4		Fixed point 16 bits CPI/O 8		8 bit integer CPI/O 16	
Layers	Time (ms)	Time (ms)	%	Time (ms)	%	Time (ms)	%
HLSinf	0	17994	78,83%	15476	67,9%	7964	41,2%
Transform (CPU)	0	4793	21,00%	7260	31,9%	11347	58,6%
Others (CPU)	45685922	39	0,17%	41	0,2%	40	0,2%
Total	45685922	22825	1	22777	1	19351	-
ms x img	4568592	2283	-	2278	-	1935	-
speedw.r.t CPU		2002		2006		2361	

- Classic embedded architectures use dedicated microcontrollers in a distributed system
- Modern multi-core systems offer an abundance of processing power → consolidate system functionality onto a single HW platform
- Linux is a highly desirable cornerstone of a modern product
 ISAR and Buildroot-based Linux images are available for SELENE
- SELENE Mixed-criticality support goals
 - Isolation of critical from non-critical functions
 - At the functional level (memory protection)
 - At the performance level (timing guarantees)
 - And in the presence of faults (fault-isolation)



Hypervisor-based Safety Architecture

- We use Jailhouse hypervisor (RISC-V port)
 - Statically partitions system into multiple *cells*
 - Each cell assigned CPU cores, memory, and devices <u>exclusively</u>
 - Spatial isolation
 - Limited temporal isolation no scheduling
 - # cells ≤ # cores
 - Cells support guest OSs, including Linux, and bare metal code (*inmates*)
 - Certifiability
 - Prefer simplicity over features
 - Focus on device pass-through, not virtual devices
 - # lines of code < 10000</p>
 - RTEMS RTOS guest-capable
- Xtratum (XNG) RISC-V has also been tested successfully

H2020 SELENE





- At software level with hardware support
 - Shared L2 Cache Partitioining
 - Memory Bank Allocation
- Other sources of interference in the SoC
 - AHB bus
 - All writes go to the L2 cache (L1 is write-trough)
 - AXI NoC
 - L2 cache limited paralelism limits core-to-core AXI contention
 - Accelerators are memory intensive



End to End Contention Control

- Contention information is propagated from AHB and AXI interconnects to a hardware monitor (SafeSU)
 - Contention can be accurately attributed to specific initiators
 - Requests are extended to include an Owner ID
 - Usage quotas can be established to guarantee a target performance
 - Core exceeding quotas are stalled



AHB Contention Monitoring



AXI contention monitoring





Platform Mixed-Criticality Evaluation

 Highly Integrated Satellite Control and Data Management use-case from Airbus Defense and Space



- We have evaluated the capabilities of the SafeSU on top of RTEMS
 - SafeSU is able to limit performance slowdown due to contending applications
 - Further investigations are need to better characterize the behaviour of applications using the SafeSU on top of RTEMS (on-going)

SELENE platform Roadmap

FPGA fault-injection tool Adaptation
 ECSEL FRACTAL Project



- Implementing additional fault-tolerant support
 - IFAC Project funded by GVA (CISEJI/2022/30)
- Porting HLSinf to embedded FPGA technology
 - NimbleAI Horizon Europe Project
- Extending Cache Coherence and Interconnect Support
 - Potentially in a KDT RISC-V proposal





Platform Availability

- Integrated Gitlab repository SELENE GPL platform
 - <u>https://gitlab.com/selene-riscv-platform</u>
- Other Related Repositories
 - ISAR with ROS2 support for NOEL-V
 - <u>https://github.com/siemens/isar-riscv</u>
 - EDDL with support for the SELENE platform
 - https://github.com/deephealthproject/eddl
 - HLSINF
 - https://github.com/PEAK-UPV/HLSinf





Contact

SELENE

www.selene-project.eu





ikerlan

MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE













HLSinf Accelerator Performance



CPU: Intel Core i7-7800X(12 threads, 6 cores) FPGA: AlveoU200 H2020 SELENE

Inference time of 10 images model StackedHG (model by Airbus)



SELFRE