PULP Lab - Energy efficient Computing Circuits and Systems

Luca Benini

DEI/ARCES + ETH Zürich (luca.benini@unibo.it, lbenini@ethz.ch)

PULP Lab Overview

1 Full Professor, 1 Associate Professor,
3 Assistant Professors, 4 Post Docs,
22 Ph.D Students, 1 Research Fellow

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

What is PULP = Parallel Ultra Low Power

PULP
Parallel Ultra Low Power

- **Project started in 2013**
- **A collaboration between University of Bologna and ETH Zürich**
  - Large team. In total more than **100** people, not all are always working on PULP
  - We wanted to start with a clean slate, open architecture, based on an open ISA no dependencies with any commercial IP – Cofunded RV in 2014

- **Academic/Research goal**
  - Design Energy-efficient computing systems from IoT to HPC, addressing «the twilight of Moore's law»
  - Create an open compute platform and an ecosystem used for research on digital circuits and computing systems by our team as well as other groups in the World
  - Demonstrate innovation in silicon; more than **40** PULP SoCs prototyped in 10 years
  - Leverage emerging technologies (2.5D, 3D, NVMs) whenever possible

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Hardware Ecosystem & Capabilities of the Lab

PULP
Parallel Ultra Low Power

RISC-V Cores

RI5CY	Micro riscy	Zero riscy	Ariane
32b	32b	32b	64b

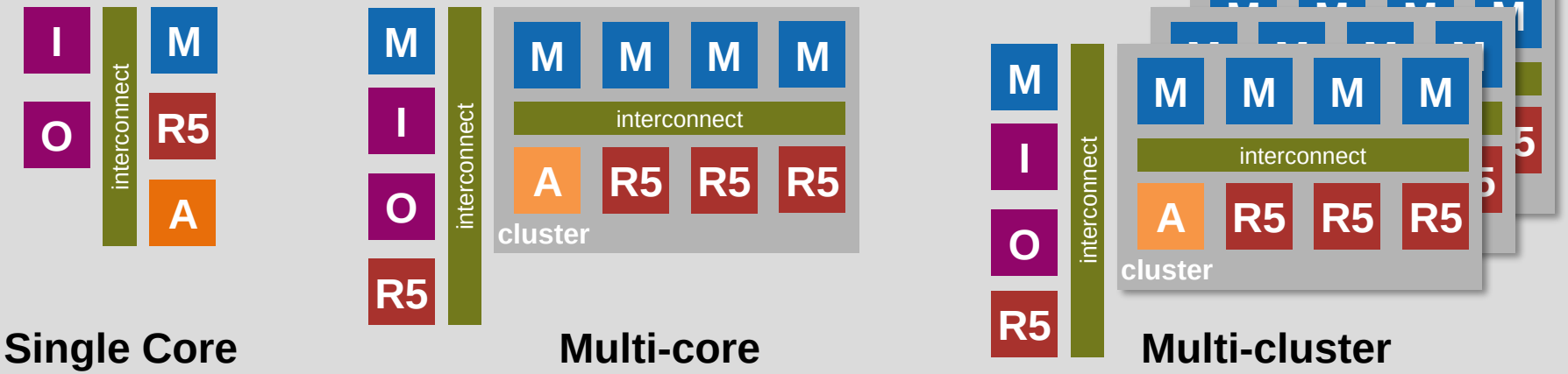
Peripherals

JTAG	SPI
UART	I2S
DMA	GPIO

Interconnect

- Logarithmic interconnect
- APB – Peripheral Bus
- AXI4 – Interconnect

Platforms



Single Core
- PULPino
- PULPissimo

Multi-core
- Mr. Wolf
- Kraken

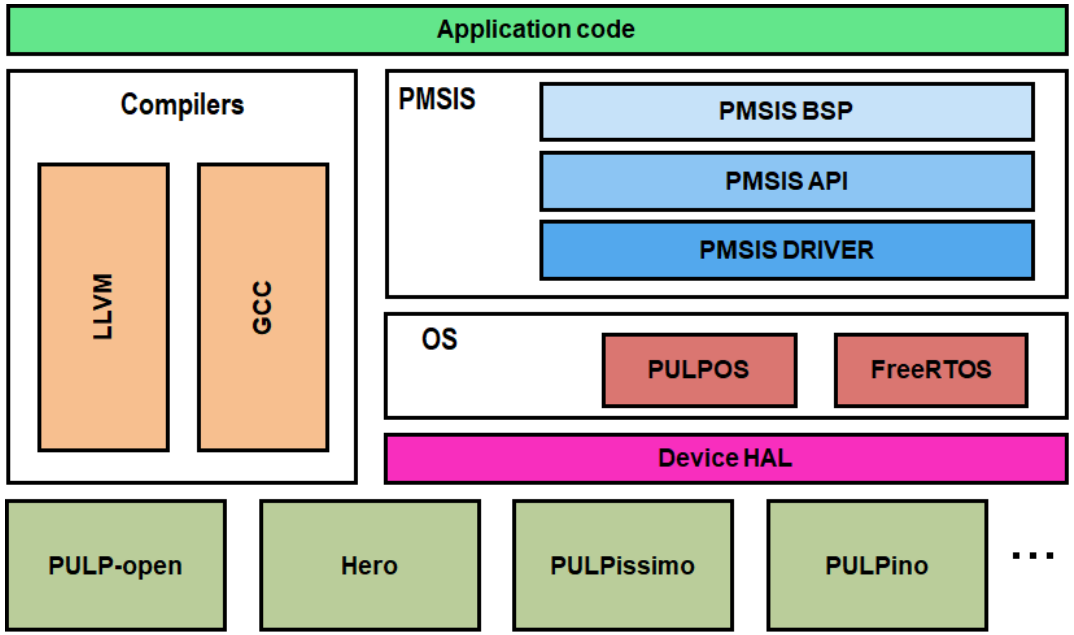
Multi-cluster
- Hero
- Occamy

IoT → HPC

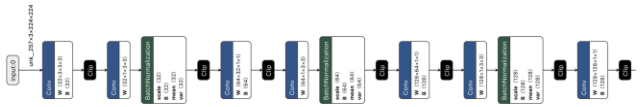
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

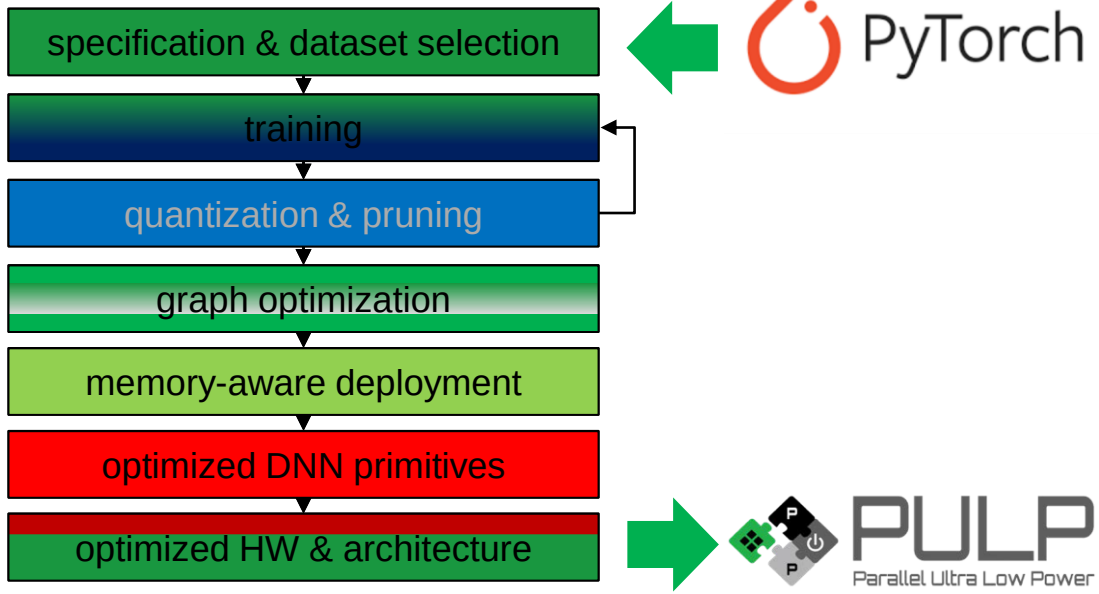
SW Ecosystem & Capabilities of the Lab

General Purpose SW Stack



DNN Mapping flow

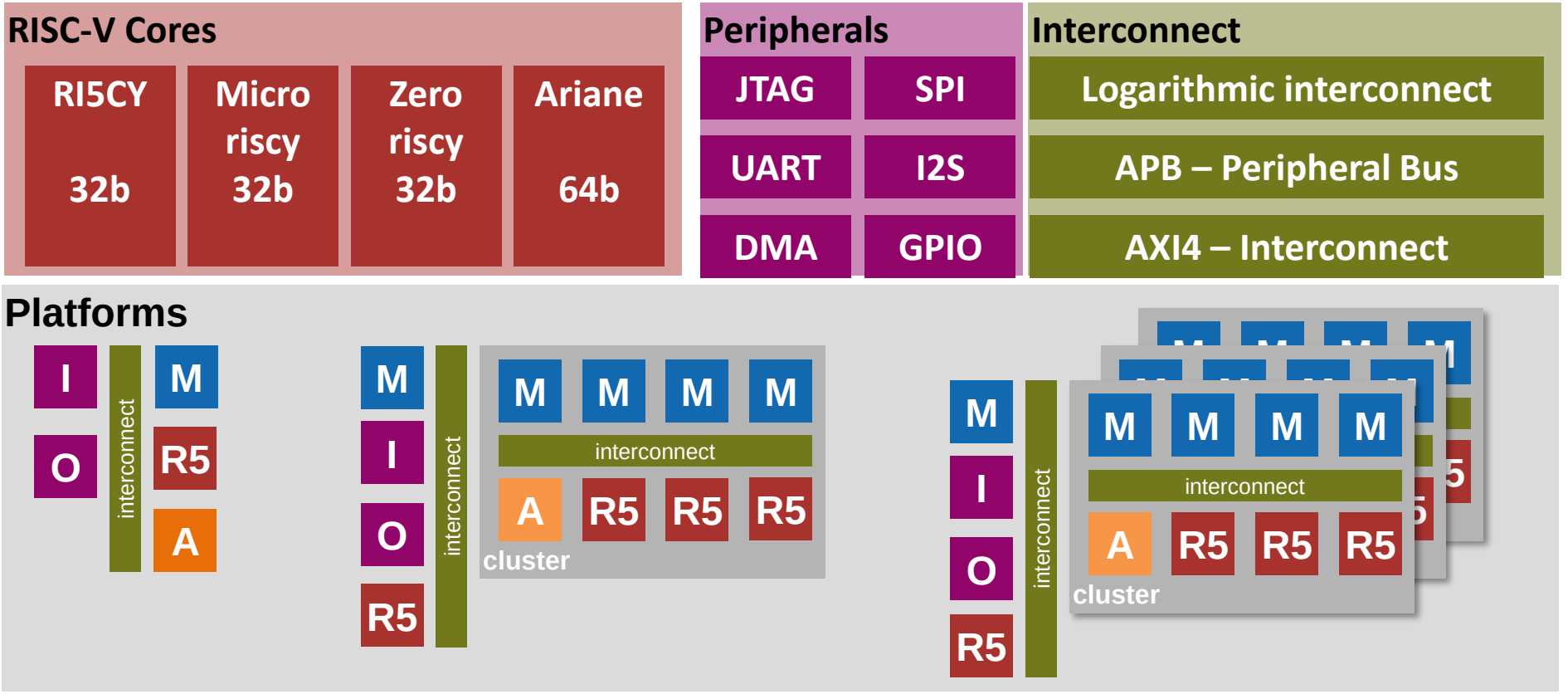




IoT → HPC

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

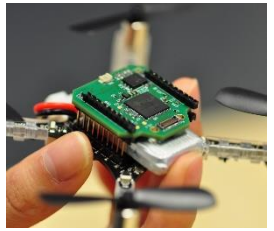
Prototyping Ecosystem & Capabilities of the lab



+ FPGA prototyping platforms

+ Silicon Prototypes (30+)

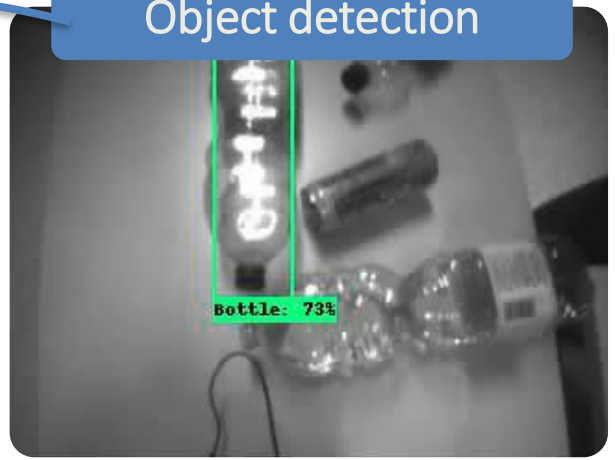
+ End to end applications



Example: Achieving true autonomy on nano-UAVs

Execute complex, heterogeneous tasks at high speed and robustness **fully on board**

Object detection



Obstacle avoidance & Navigation

Environment exploration

Multi-GOPS workload at extreme efficiency → P_{max} 100mW

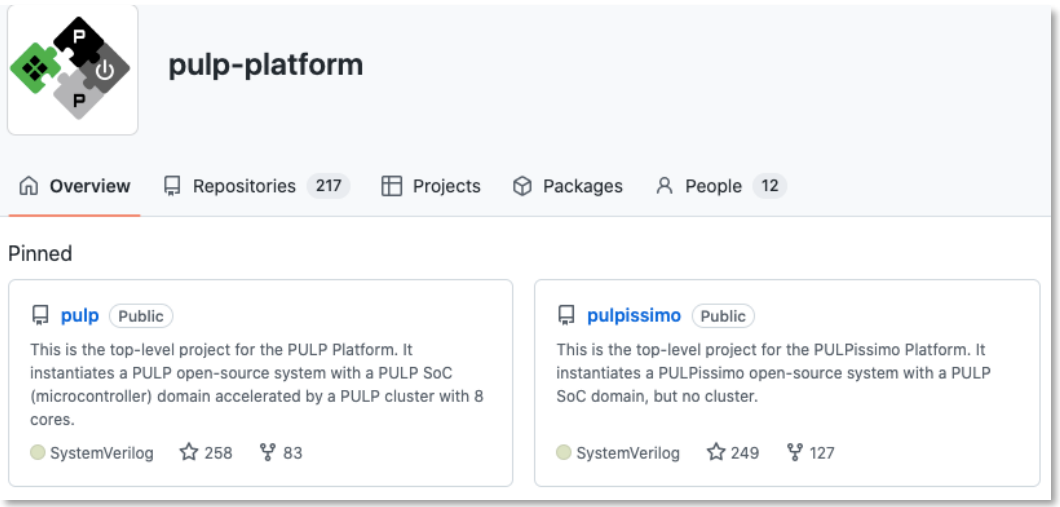
7

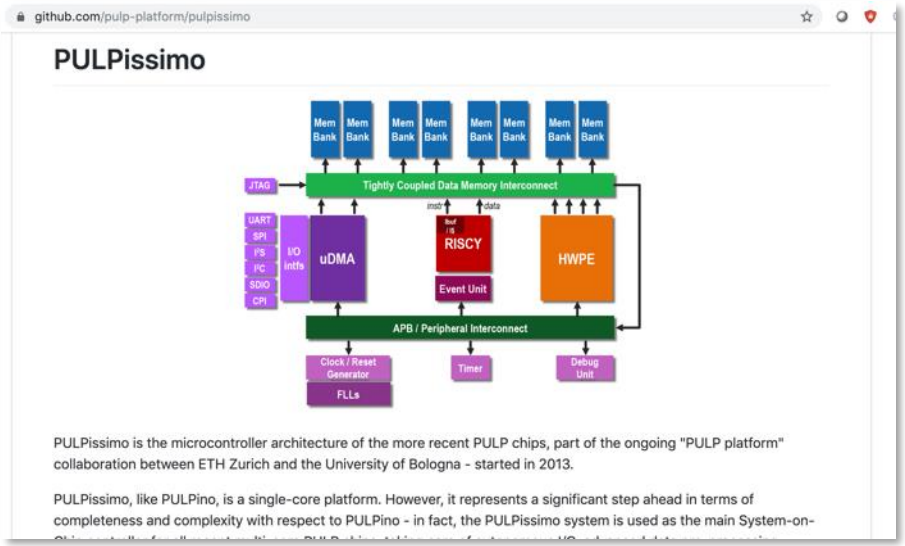
We use a permissive open source license for our IPs

- All our development is on GitHub
  - HDL source code, testbenches, software development kit, virtual platform

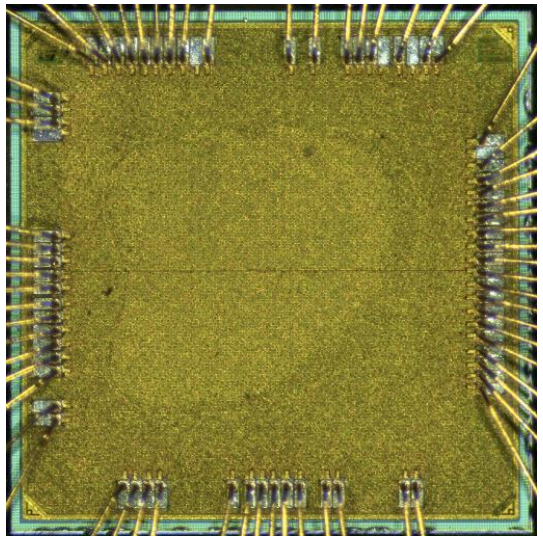
https://github.com/pulp-platform

- PULP is released under the permissive Solderpad (Apache-derived) license
  - Allows anyone to use, change, and make products without restrictions.





Industrial Collaborations

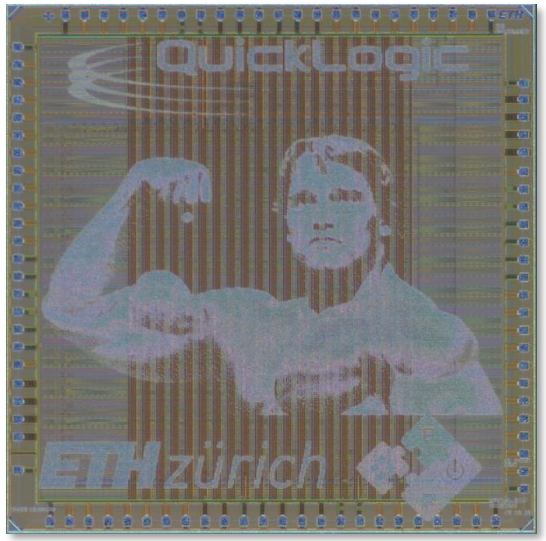


life.augmented

PULPv1,2,3 (ST28 FD)

Demonstrators of 28nm FD-SOI capabilities

Various publications ‘15 – ‘18

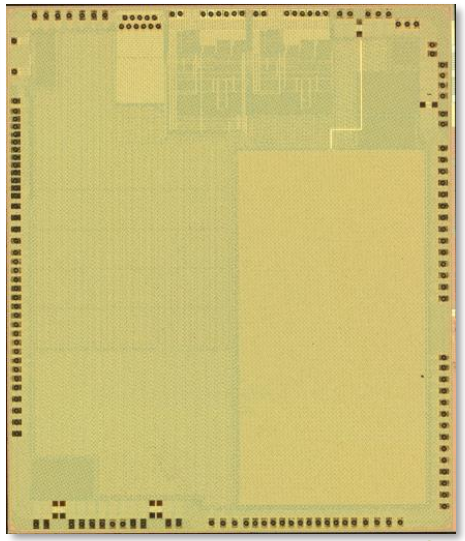


QuickLogic®

Arnold (GF22)

IoT SoC combining eFPGA with RISC-V core

Schiavone et al TVLSI ‘19

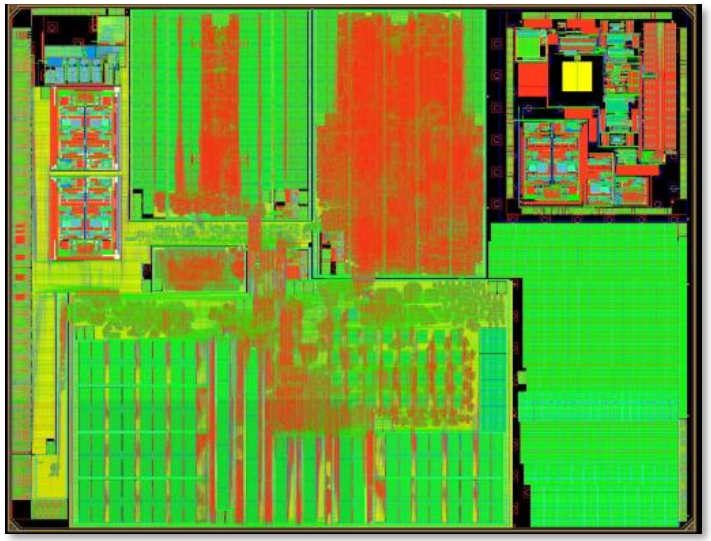


GREENWAVES TECHNOLOGIES

Vega (GF22)

IoT Processor with ML acceleration

Rossi et al ISSCC ‘21
Rossi et al JSSC ‘22



DOLPHIN INTEGRATION
The enabler of low-power Systems-on-Chip

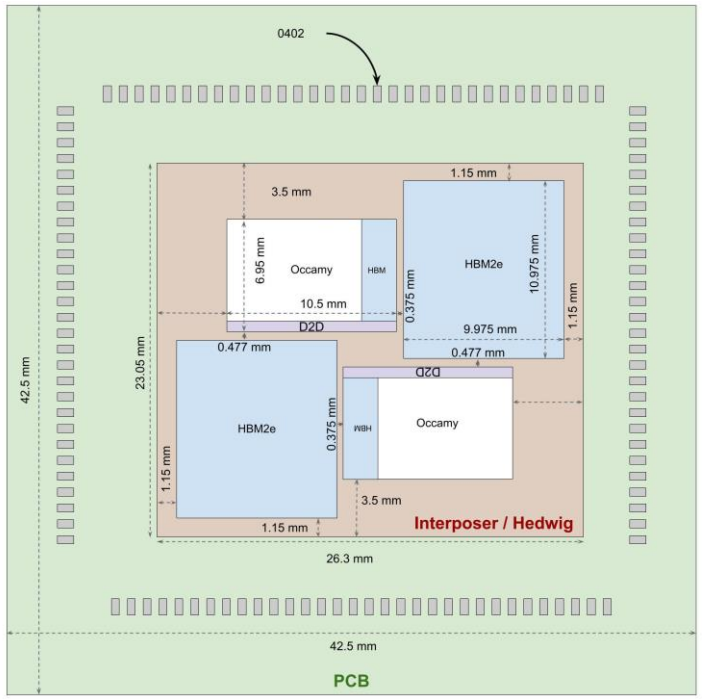
Marsellus (GF22)

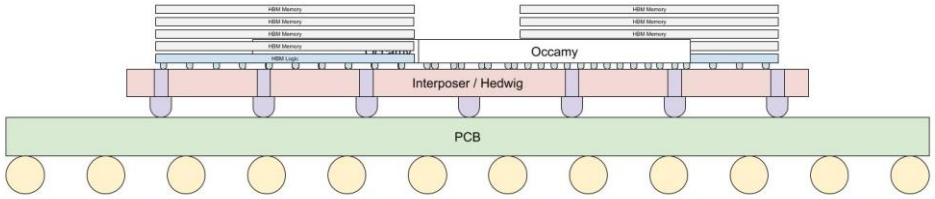
IoT Processor with low power modes and AI Accelerators

Conti et al ISSCC ‘23
(to appear)

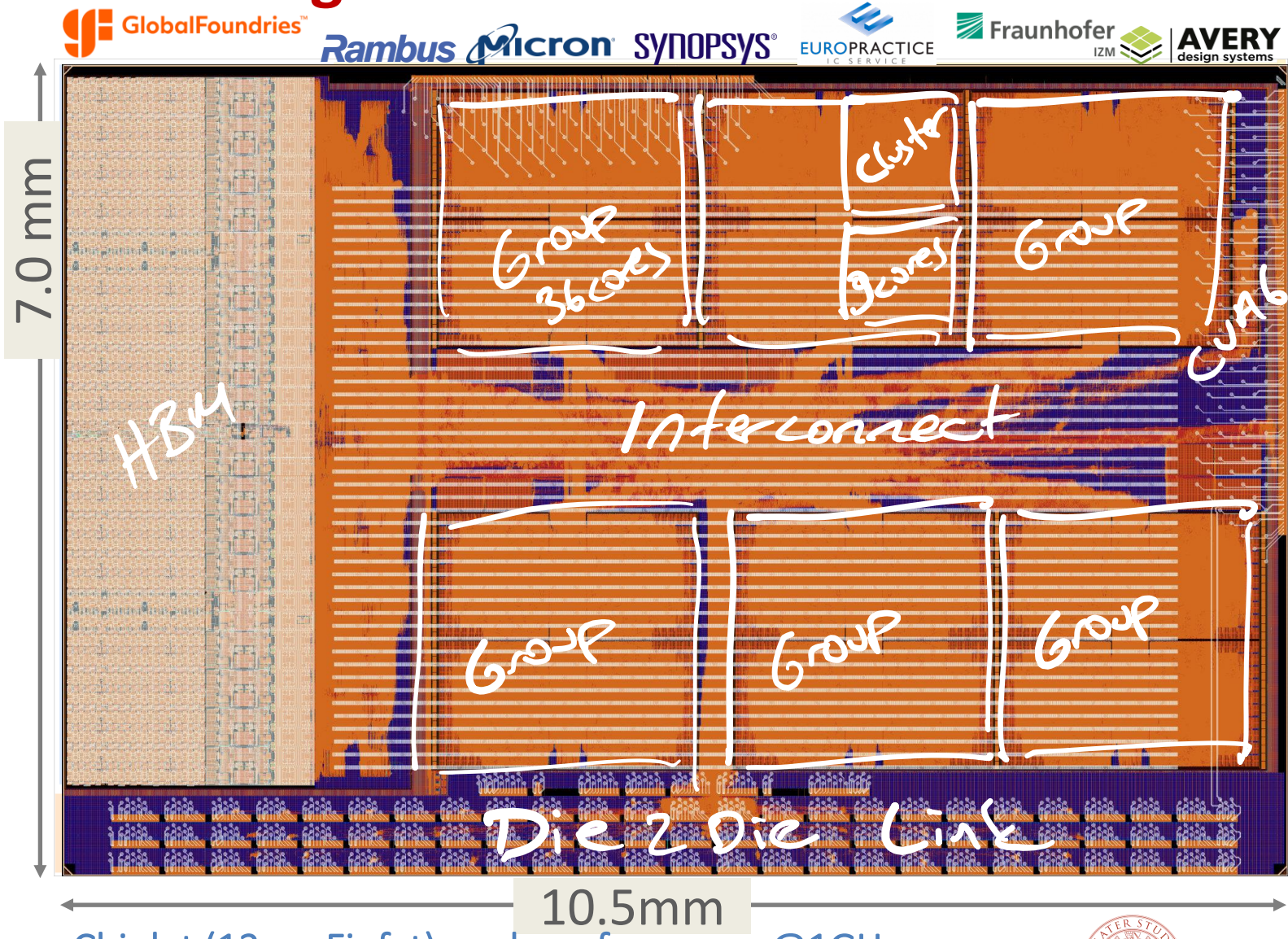
Let’s now see some on going projects…

Occamy: Accelerator Chiplet for next-gen Automotive & HPC





2.5D 2xOccamy+2HBM2 on interposer



Chiplet (12nm Finfet) peak performance @1GHz:

FP64: 384 GFLOp/s FP16: 1.536 TFLOp/s

FP32: 768 GFLOp/s FP8: 3.072 TFLOp/s

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

From Academia to Industry

RISC-V

OPENHW GROUP™ — PROVEN PROCESSOR IP

www.openhwgroup.org

- **OpenHW Group** is a not-for-profit, global organization (EU,NA,Asia) where HW and SW designers collaborate in the development of open-source cores, related IP, tools and SW such as the **Core-V** family
- **OpenHW Group** provides an infrastructure for hosting high quality open-source HW developments in line with industry best practices.

PULP — Parallel Ultra Low Power

RI5CY, ARIANE RISC-V cores, FPU, AXI4 Components.

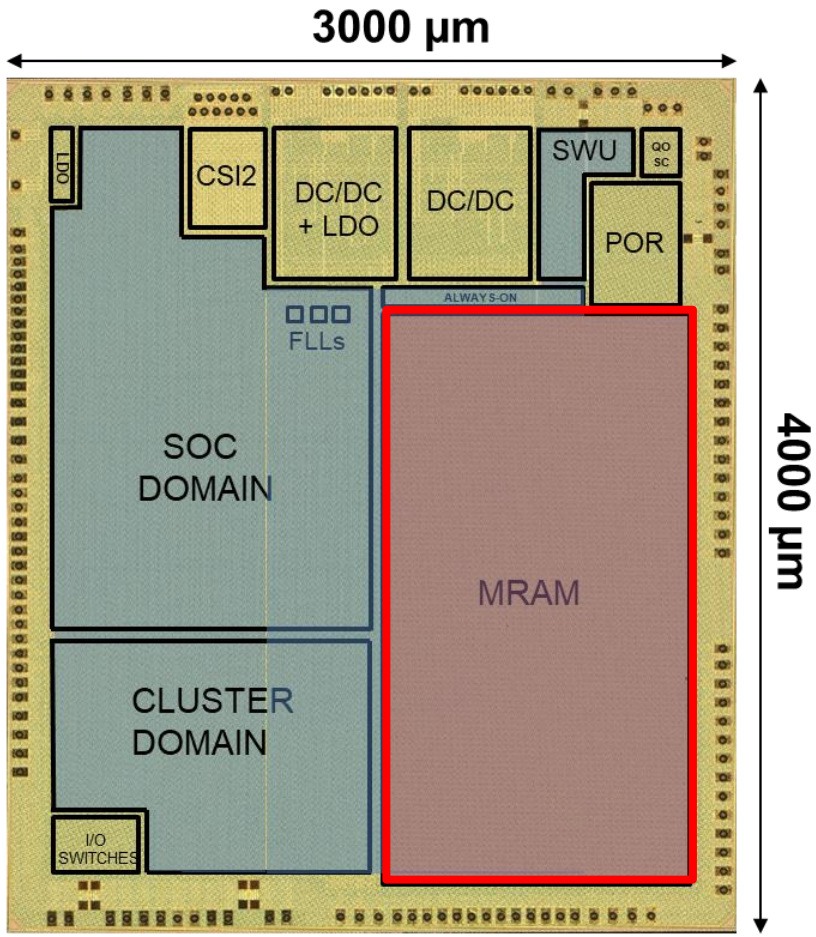
OPENHW GROUP™ — PROVEN PROCESSOR IP

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

VEGA: Extreme Edge IoT Processor

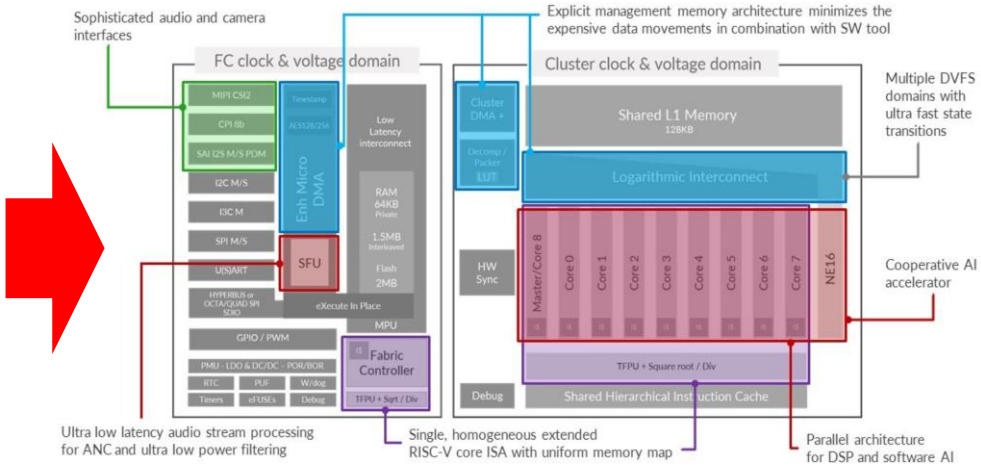
In cooperation with GREENWAVES TECHNOLOGIES

- RISC-V cluster (8cores +1) 614GOPS/W @ 7.6GOPS (bit DNNs), 79GFLOPS/W @ 1GFLOP (32bit FP appl)
- Multi-precision HWCE(4b/8b/16b) 3×3×3 MACs with normalization / activation: 32.2GOPS and 1.3TOPS/W (8bit)
- 1.7 µW cognitive unit for autonomous wake-up from retentive sleep mode
- **Fully-on chip DNN inference with 4MB MRAM (high-density NVM with good scaling)**



[D. Rossi, ISSCC21]

GAP9 Dominates the TinyML benchmarks



ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Tristan + ISOLDE (KDT)

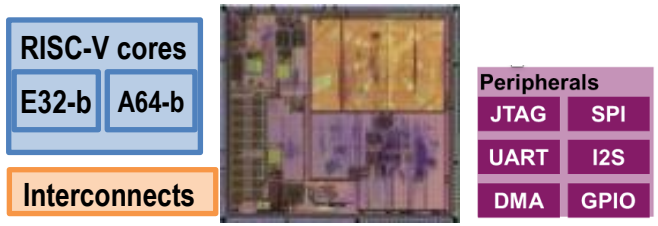
TRISTAN
RISC-V

Toegether for **RIS**C-V
Techonology and
Applicatio**N**s

RISC-V®

- More than 40 European Partners;
- 3 years project;

Project Budget: 180 M€
Group Budget: 1.4 M€



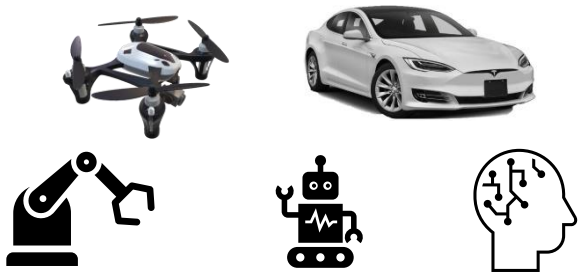
Develop open-source, industry-qualified

- **RISC-V processors and accelerators**
- Full set of interconnects, **peripheral Ips**;

OpenMP
OpenVX

Develop a full set of open-source SW tools for RISC-V:

- **Compilers, toolchains;**
- EDA tools for verification;
- **Simulators** and Emulators;
- **SW support for applications.**



Demonstrators in industrial environments (targets traditional key European Markets):

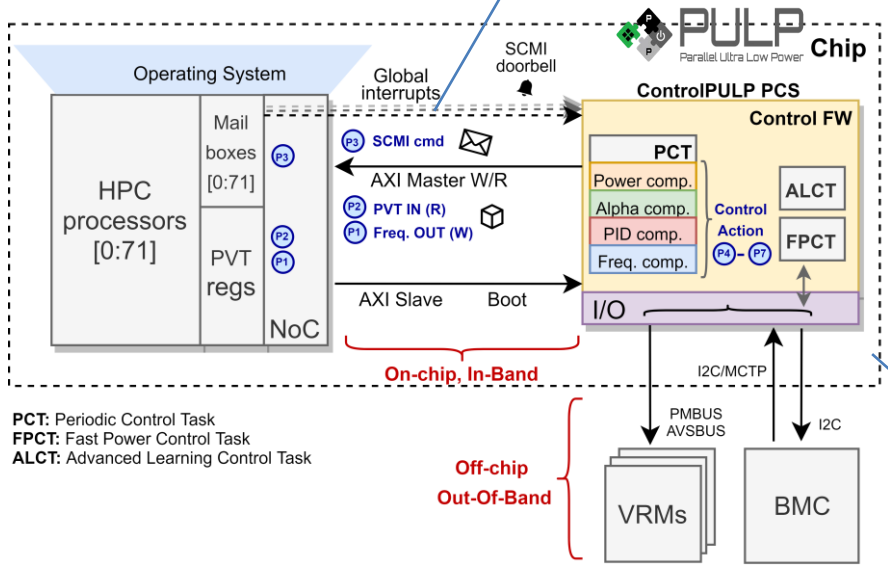
- Automotive & **Aerospace**;
- Industrial,Robotics and **Drones**;
- Mobile Communication;

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

EPI SGA1/SGA2 (EuroHPC)

European Processor Initiative

- Develop a Commercial European High Performance General Purpose Processor for HPC (ARM based)
  - **Embedded power management system based on PULP IPs**
- **High-performance RISC-V based accelerator IPs and test chips**
- Computing platform for autonomous cars
- Will also target the AI, Big Data and other markets in order to be economically sustainable



Interaction with on-chip sensors (PVT, perf.) and actuators (PLL, power gates, clock gates)

FreeRTOS

Power Capping

Thermal Capping

Interactions with off-chip voltage regulators

life.augmented

Project Budget: 160 M€
Group Budget: 4.5M €

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PNRR CN HPC, Quantum Big data

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Spoke-1 FutureHPC Budget 25M€
UNIBO's Budget: 3.3M€

LEADER

UNIBO
UNITO

PARTECIPANTI

POLIMI
POLITO
UNIPI
UNIPD
ROMA TOV
UNINA
UNICT
UNICAL
INAF
CINECA
ENEA
IIT
UNIFE

15

LAB1: Hardware and System Prototyping	LAB2: HPC SW and system software prototyping
Flagship1 - Non-functional properties	
Flagship2 - Heterogeneous acceleration	
Flagship3 - Workflow management, cloud-HPC convergence	
Flagship4 - Trustworthiness, security, privacy	
Flagship5 – Codesign, application-SW-HW targeting, benchmarking, patterns, microkernels	

Cascade funding + Innovation projects

Exteernal Insutry + Academia

- **sogei**
  1. DevSecOps
  2. Definition of HPC architectures for full lifeciclye ML and edge appls
- **autostrade per l'italia** – La passione di muovere il Paese
  1. Evaluate future HPC HW architectures and edge computing
  2. Video surveillance system using ML/AI
- **LEONARDO**
  1. Green Computing
  2. Accelerated Computing, OpenSource HW and SW
  3. Containers for Cloud continuum, Edge and IoT
  4. HPC -Cloud Convergence for Data driven Digital Twin, BigData and AI
- **Eni**
  1. Scouting future HPC HW architectures and system prototyping
  2. Parallel programming models and HPC system software
- **ENGINEERING**
  1. AI andBigData frameworks, integration in HPC ecosystem, federated, distributed learning
  2. Mini-applications and benchmarking from multiple domain;
- **INTESA SANPAOLO**
  1. Enterprise Multi-tenancy Distributed learning
  2. Enterprise AI and BigData frameworks for HPC ecosystem
- **ThalesAlenia Space**
  1. High-performance RISC-V processors for satellite
  2. HPC technologies for SAR ground processing of Earth observation data
  3. Cyber-security on-board spacecraft, including embedded AI framework and radiation tolerant AI processing in space
- **UnipolSai ASSICURAZIONI**
  1. Cloud-HPC convergence: federated and distributed learning
  2. Hybrid MLOps Workflow management & cross-application streaming tools
- **FINCANTIERI**
  1. Hybrid HPC Solutions
  2. HPC smart sensors and big data
  3. Mini applications in naval domain
  4. Solutions for onboard power monitoring and control

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Joint research lab with ST SRA-ASMS

1. Design and Optimization of RISC-V Vector Processor for domain Specific Applications

2. Design of lockstep processor for safety critical applications (open-source activity)

3. Design of tightly-coupled clusters of vector processors

- Status: all three activities have started
- Research Fellows: Gianmarco Ottavi, Riccardo Tedeschi, Gianluca Sinigaglia
- Senior support: Luca Benini, Davide Rossi, Angelo Garofalo, Matheus Cavalcante (ETHZ)
- ST team: Elio Guidetti

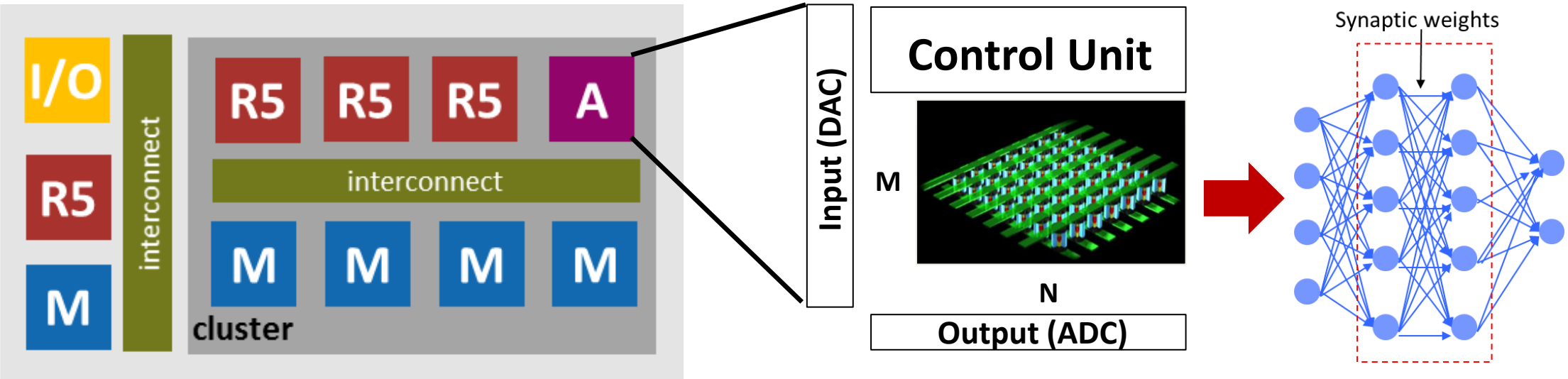
Also Ramping up cooperation with Automotive Group (M. Peri, M. Losi)
AI for Automotive Applications, RV Automotive Platform

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA


I'm Mr.
WOLF
PATRONUS
ETH
ROSETTA
Manny
FULMINE
XAVIER
2001: A SYSTEM ON CHIP
ATOMARIO
DIEGO
BAIKONUR
БАЙКОНУР
ETH zürich
PLINK

Thanks!

WiPlash & NeuroSoC (H2020 FETOPEN/RIA)



1. **Heterogeneous Clusters including Digital Cores and Analog In Memory Computing cores**
2. Demonstrate capabilities of PCM based AIMC in a 18 nm FD-SOI (ST Micro) technology
3. **Develop Software Stack**

life.augmented

IBM

Project Budget: 3 M€ + 8 M€
Group budget: 350 K€ + 150 K€

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

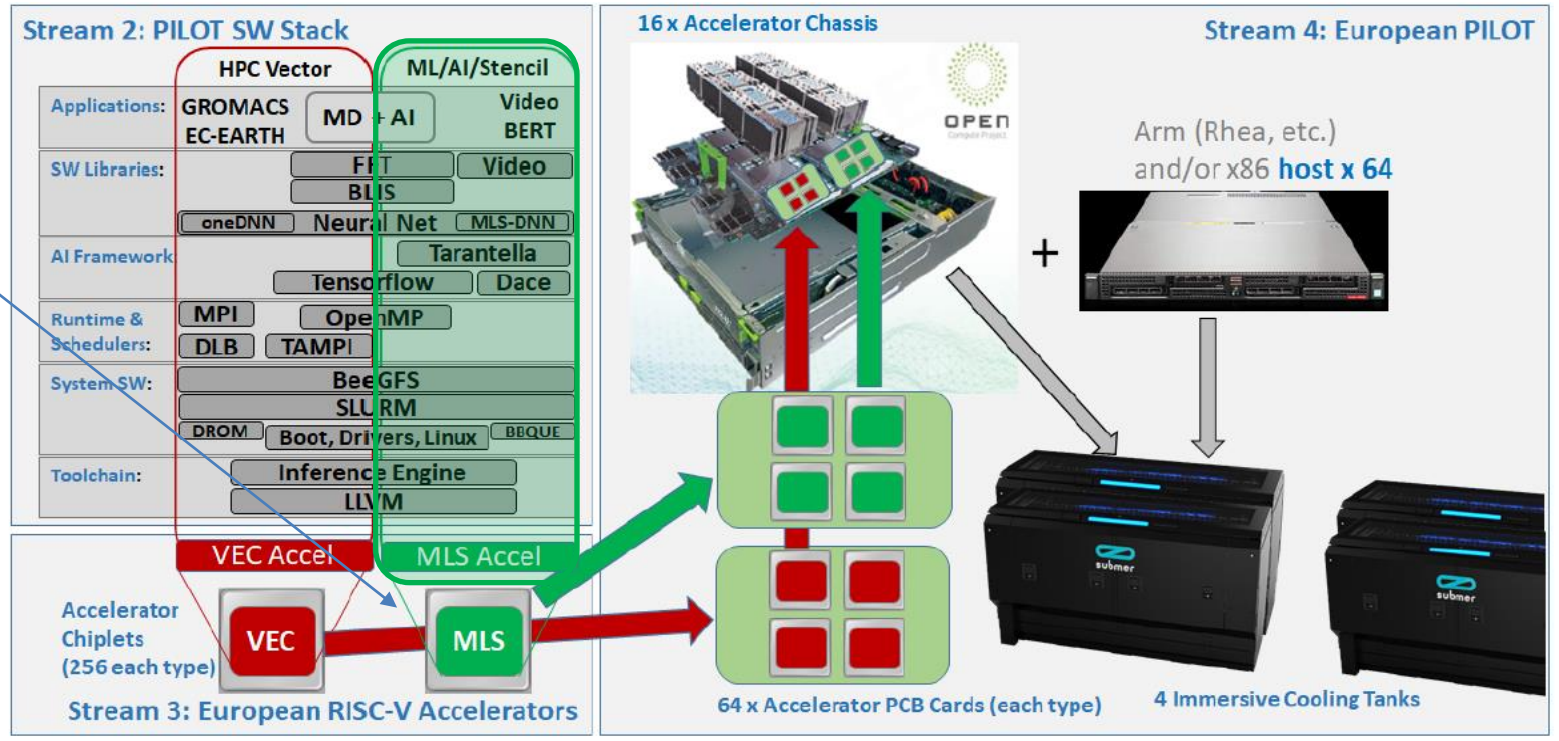
The European Pilot (EuroHPC)

THE EUPILOT

Project Budget: 80 M€
Group Budget: 800 K€

Pilot using Independent Local & Open Technologies

- Chiplet-based Accelerators for European HPC and AI
  - **RISC-V Machine Learning Chiplet**
  - RISC-V Vector Chiplet
- Full Software Stack
  - **Compilers**
  - **Libraries**
  - AI Frameworks
- Full demonstrator
  - Motherboard,
  - Rack
  - Power management



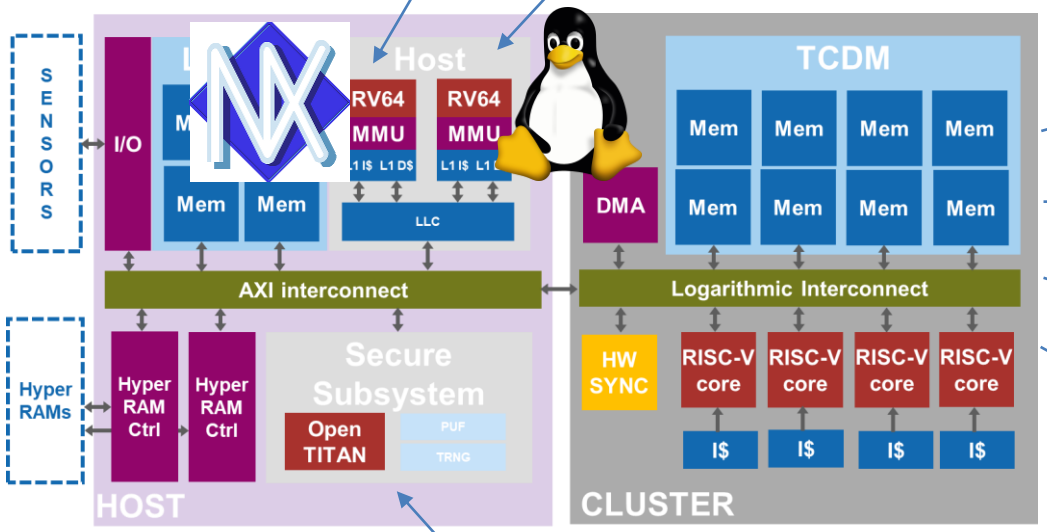
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Secure Heterogeneous SoC for UAV Navigation

64-bit RISC-V:

Flight Control Stack (NuttX)

Network Stack (Linux)



32-bit multi-core accelerator:

- On-line training
- Predictive Model Control
- Object Detection
- ...

Secure Subsystem:

Secure Boot

Secure Firmware Update

Project Budget: 5 M€

Group Budget: 1.2M€

Compatible with FMUv5 Specifications

Goal:

Drone Flying in lab

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA


ALMA MATER STUDIORUM
A.D. 1088

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

www.unibo.it