WHEN PERFORMANCE MATTERS

E4 Experience with RISC-V in HPC

Daniele Gregori Ph.D.

HPC Asia, Nagoya, Japan 25 Jan 2024

INDEX:

Company Profile

E4 IN A NUTSHELL

2002 - 2022

WHO WE ARE

E4 Computer Engineering is an **Italian** Company, designs and manufactures highly technological solutions for HPC Clusters, Cloud, Data Analytics, Artificial Intelligence and Hyper-Converged infrastructure for the Academic and Industrial markets. We have been collaborating for years with the main research centers at national and international level (Cineca, CERN, ECMWF, LEONARDO) and we are involved in national and European projects in the HPC and AI fields (EuroHPC JU EPI, EUPEX, Horizon Europe)

VISION

We explore future scenarios to find solutions for highly performing computational needs in application areas that are unimaginable today.

MISSION

We anticipate the ever-accelerating disruptive transformation of our era, providing mature solutions in sophisticated technological contexts with a dizzyingly innovative approach

APPROACH

Each E4 solution is UNIQUE, like each of our customers; TESTED in every single component; VALIDATED to verify the actual performance of each system and SERVED by technicians who provide assistance in the most extensive and complex Italian and European computing infrastructures.

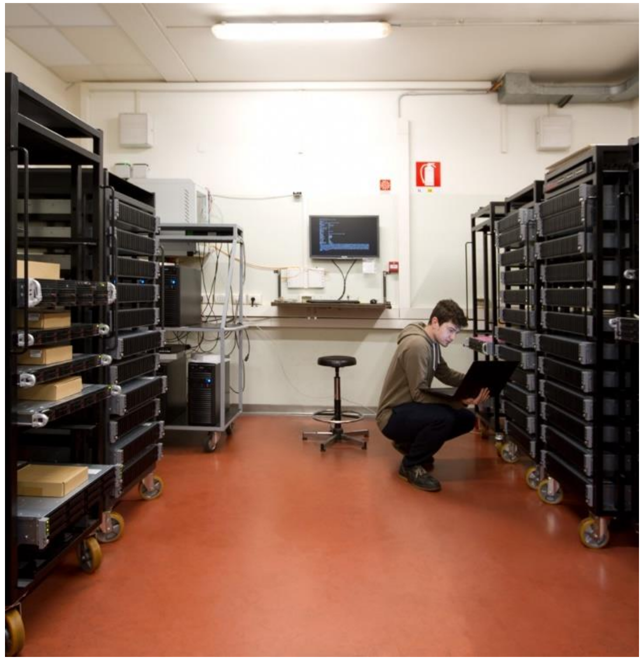
Strategic Members
https://riscv.org/

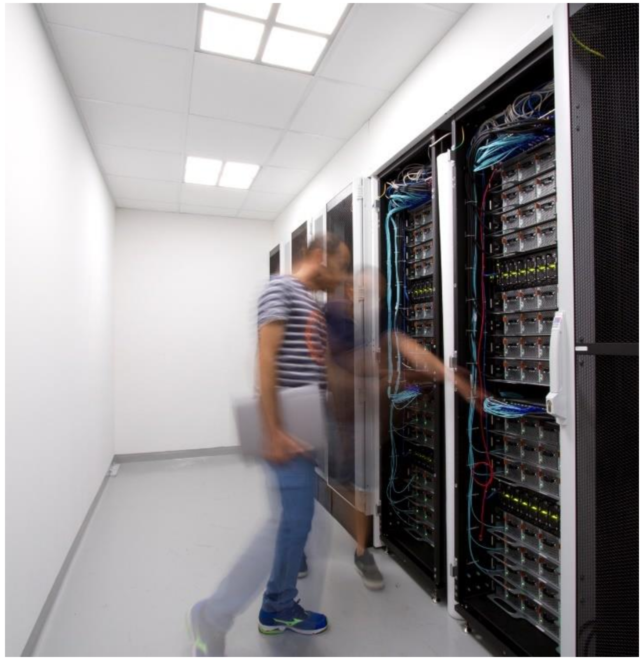
ACTUALLY… YOU ALREADY KNOW US

INFN & CERN AWARD (2012-2014)

Award of excellence in industrial collaboration for the **ATLAS** and **CMS** experiments at the Large Hadron Collider at **CERN** within the project of the discovery of the Higgs boson

E4 TECH FACTORY





- Integration Facility where our technicians build servers or storage systems
- Burn In Room to improve E4 systems reliability with at least 72 hours of test that involves all components
- R&D Lab, with 6 standard racks with heterogeneous systems, 100kW, remote access available on demand to perform benchmarking, co-design, prototyping

E4 – TAILOR MADE DESIGN AND INTEGRATION

Istituto Italiano di Tecnologia (IIT)
Genova, 2020
Expansion 2022~2023

- GPU Cluster and File systems in 16 Racks with LCP liquid cooling
- Special design with energy and temperature management tools
- Tailor-made Infrastructure integrated with home-designed Software from Uni. Bologna

LEONARDO SYSTEM

- **4th Top500**
- **HPL 240 PF + 9 PF (currently 170PF)**
- TCO Investment: **240M€** (120M€ Capex + 120M€ Opex)
- **5000 nodes** based on **BullSequana XH2000** platform technology (3500 GPU + 1500 CPU)
- Computing racks: **95% Direct Liquid Cooled**
- Data storage: >**100PB** (NVMe+HDD)
- **Warm water:** Inlet temperature of 37 degrees
- NVIDIA Mellanox **HDR 200** interconnect
  - Dragonfly+ topology

Engagement and Collaboration with EVIDEN, with 2 E4 full-time staff on-site daily

E4 – SPECIAL DEPLOYMENTAND FULL RANGE SUPPORT

ECMWF

Bologna Technopole, 2021 – ongoing

Engagement and Collaboration with EVIDEN, with 5 E4 full-time staff on-site daily

Monte Cimone:

The first HPC cluster based on RISC-V

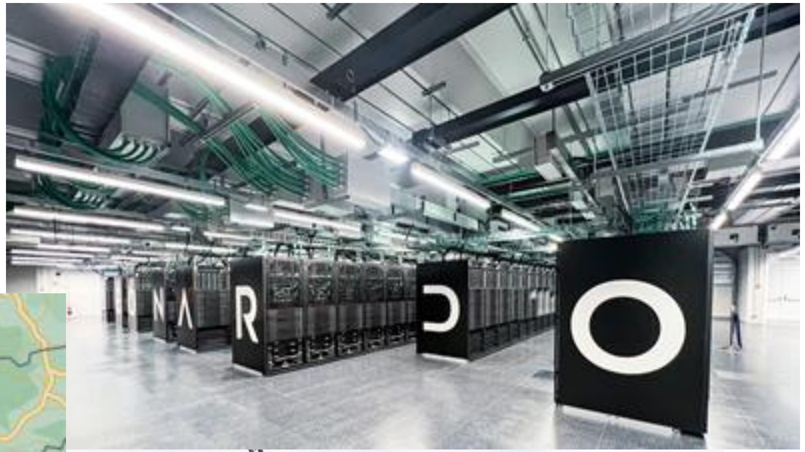
E4 – UNIBO – CINECA and the Data Valley:

A strong cooperation

Bologna New Technopole - 60MWatt datacentre

- CINECA Leonardo – The Italian Pre-exascale
  - 240 Pflops, 150PBytes, 4th Top500@Jun. 2023
- ECMWF HPCE – The new ECMWF supercomputer
  - 40+ Pflops

ECMWF CINECA



E4 is an HPC and Technology innovator company

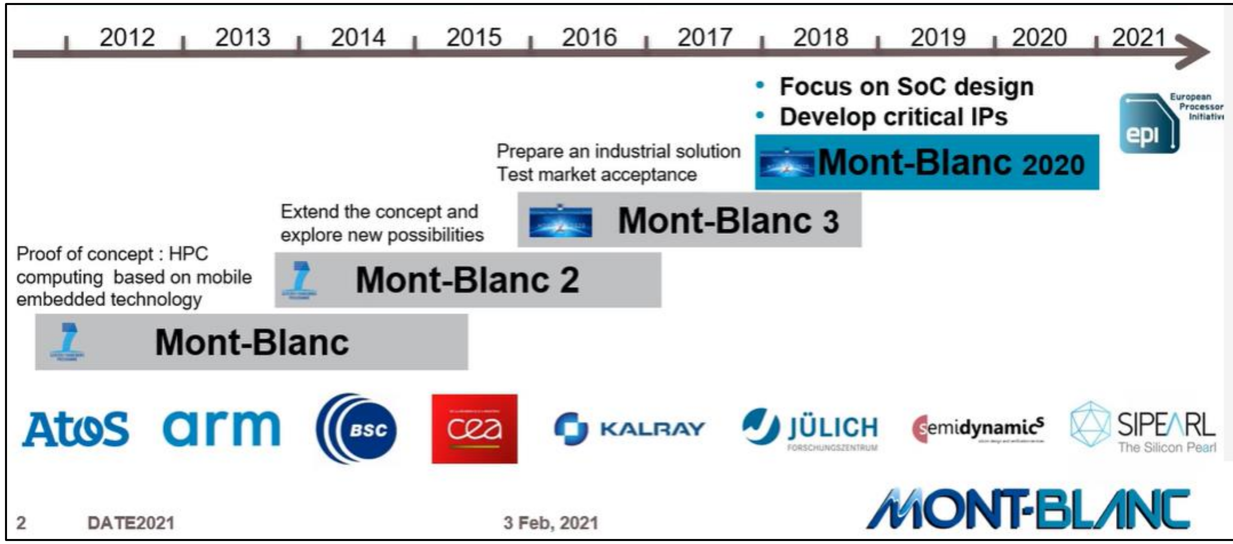
INFN

CERN Tier1 LHC
~100PBytes storage

University of Bologna
Monte Cimone
1st RISC-V Cluster

RISC-V and HPC

- New ISA & HPC:
  - 1° MONT-BLANC eu project (2012)
  - 2012 – 2020 many projects
  - 2020 Fugaku - 1° Top500 w. 415PFLOPs

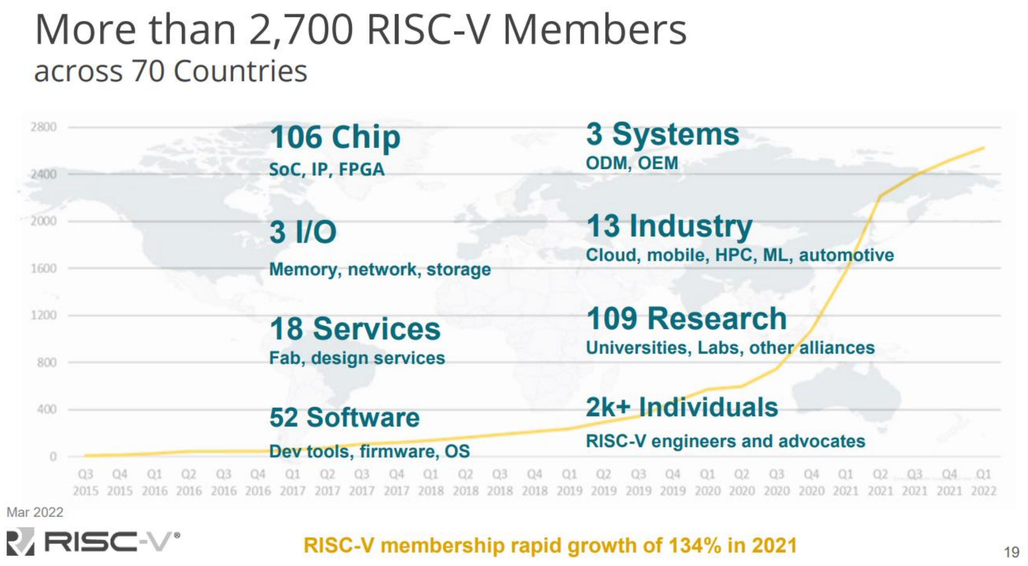


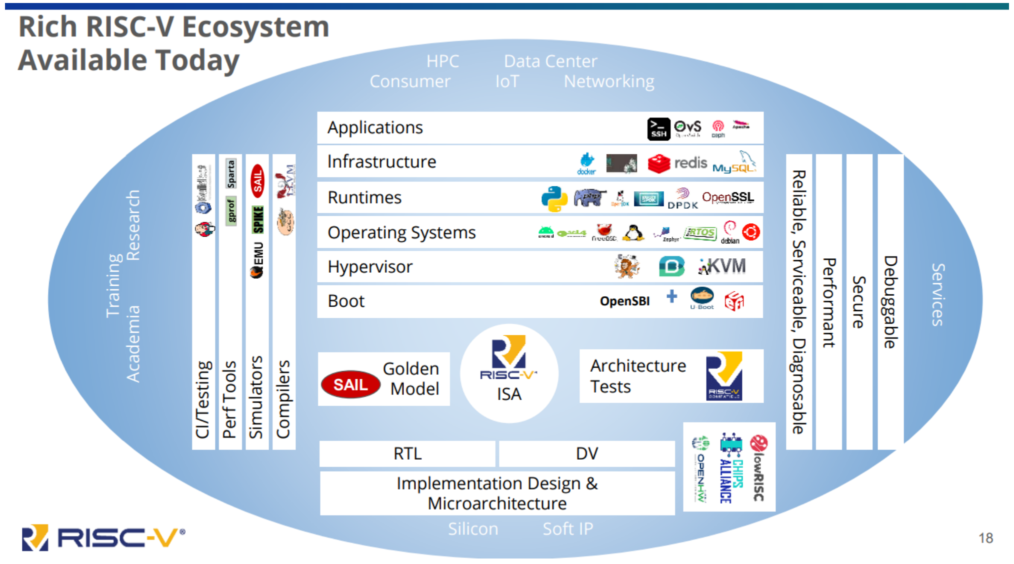
RISC-V and HPC

- New ISA & HPC:
  - 1° MONT-BLANC eu project (2012)
  - 2012 – 2020 many projects
  - 2020 Fugaku - 1° Top500 w. 415PFLOPs

- RISC-V ISA relatively new (10 years)
  - Few RV64G commercial available
  - Several announced w. Vector-extensions:
    - SiFIVE, Ventana, Esperanto, Semidynamics, Rivos, Axellera, SOPHGO…
  - Rich OpenHW ecosystem

- SW ecosystem provided by the RISC-V foundation – rich but initially focused on AI/embedded





Monte Cimone Project

The **first physical prototype** and test-bed of a **complete RISC-V (RV64) compute cluster** integrating **compute, interconnect,** *a* ***complete software stack for HPC*** *and a full-featured* ***system monitoring infrastructure****.*

1. Ported and assessed the maturity of a HPC software stack composed of:
   - SLURM job scheduler, NAS filesystem, Spack package manager
   - compilers toolchains, scientific and communication libraries,
   - a set of HPC benchmarks and applications,
   - ExaMon datacenter automation and monitoring framework.

2. Characterized the HPL and STREAM benchmarks w. the toolchain and libraries installed by SPACK.

3. Extended the ExaMon monitoring framework to monitor the Monte Cimone cluster. Power consumption characterization of Monte Cimone.

4. «In Production» since May 2021.
   1. Access to external user (>40 users).
   2. Used in University Master courses and in two PhD summer school (> 100 studends/year).
5. Now extended with SG2042 computing systems and accelerator cards

A. Bartolini *et al.*, "Monte Cimone: Paving the Road for the First Generation of RISC-V High-Performance Computers," *IEEE SOCC'22*,

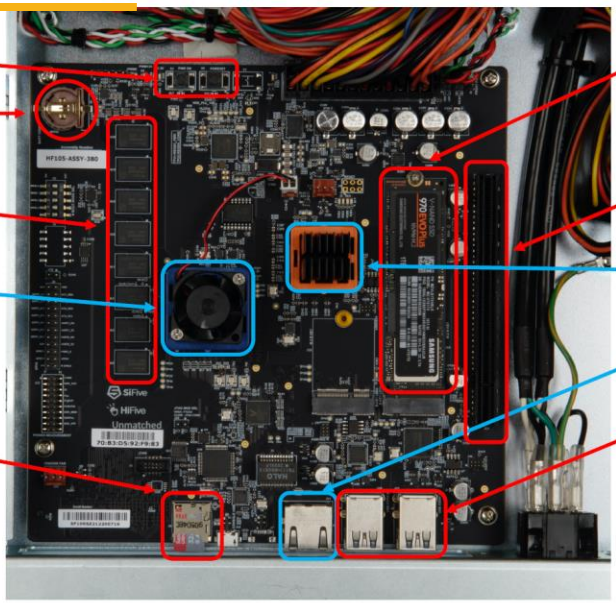
F.Ficarelli et al. «Meet Monte Cimone: exploring RISC-V high performance compute clusters,» ACM CF'22

G. Mittone et al. «Experimenting with Emerging RISC-V Systems for Decentralised Machine Learning» CF'23

Monte Cimone v1 Hardware

E4 RV007 blade prototypes

SiFive HiFive Unmatched board

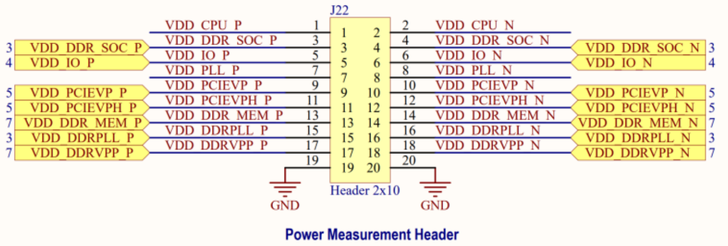


4x E4 RV007 1U Custom Server Blades:

- **2x SiFive U740 SoC** with **4x U74 RV64GCB** cores
- 16GB of **DDR4**
- 1TB node-local **NVME storage**
- **PCIe expansion** card **w/InfiniBand HCAs**
- **Ethernet + IB parallel networks**

SiFive U740 SoC w. 7 separated power rails:

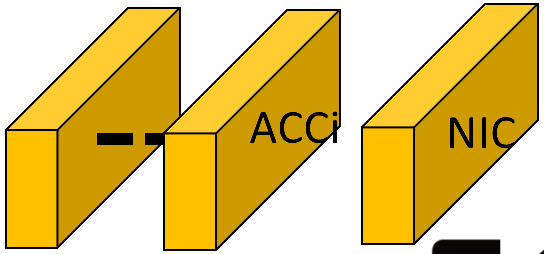
- Core complex, IOs, PLLs, DDR subsystem and PCIe one.
- Board implements distinct shunt resistors



Monte Cimone v2 Candidate HW

Now testing MILK-V Sophgo

Milk-V development server
(www.miklv.io)



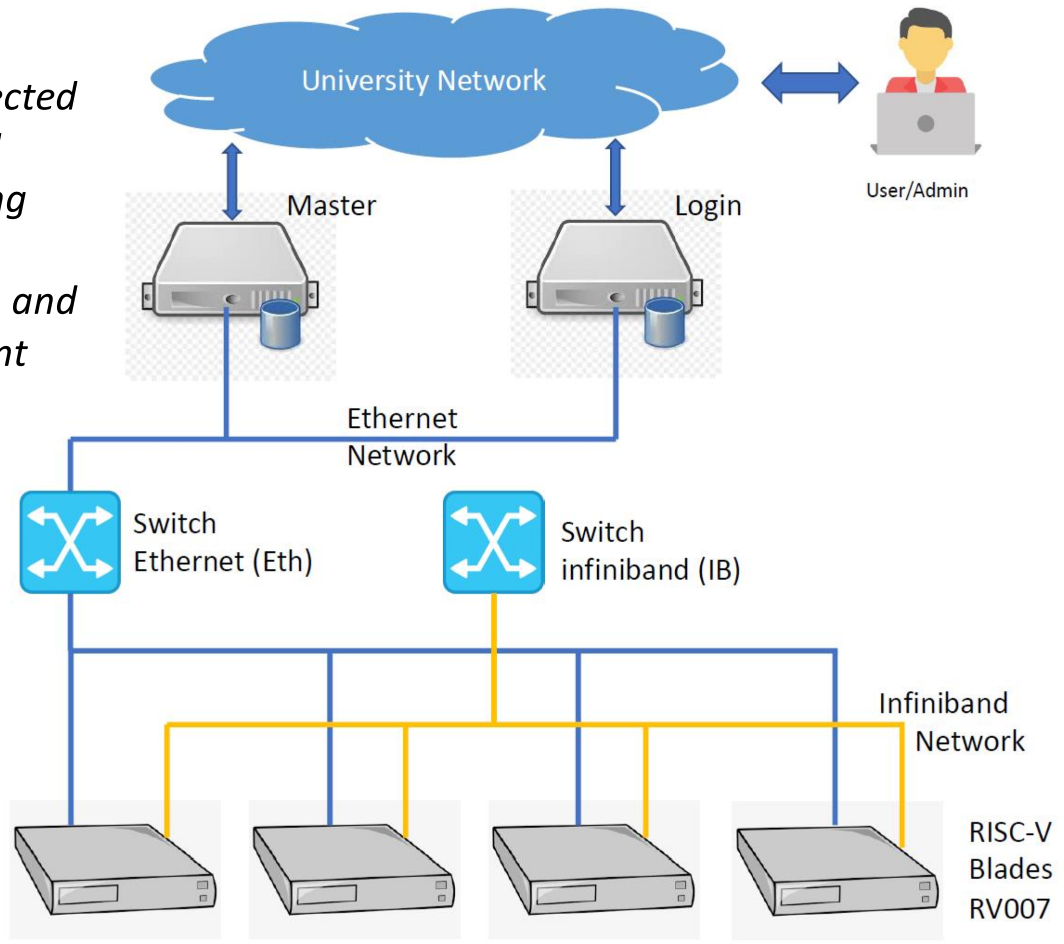
MILK-V Pioneer :
- Board based on Sophgo SG2042
- 64-cores x T-head C920 RISC-V CPU
- 2GHz main frequency
- 64MB cache
- PCIe Gen4x16
- RV vector extension 0.71
- 128GB DRAM

Monte Cimone Software Stack:

Production-level HPC software stack

- SLURM job scheduler, NFS filesystem, Nagios
- User-space deployed via **Spack** package manager
- Upstream and custom **toolchains**
- **Scientific libraries**
- Industry-standard **HPC benchmarks and applications** (e.g.: **quantumESPRESSO** suite)
- The **ExaMon datacenter automation and monitoring framework**

The cluster is connected to a login node and master node running the job scheduler, network file system and system management software.



Monte Cimone: User-facing software stack

Package	Version
gcc	10.3.0
openmpi	4.1.1
openblas	0.3.18
fftw	3.3.10
netlib-lapack	3.9.1
netlib-scalapack	2.1.0
hpl	2.3
stream	5.10
quantumESPRESSO	6.8

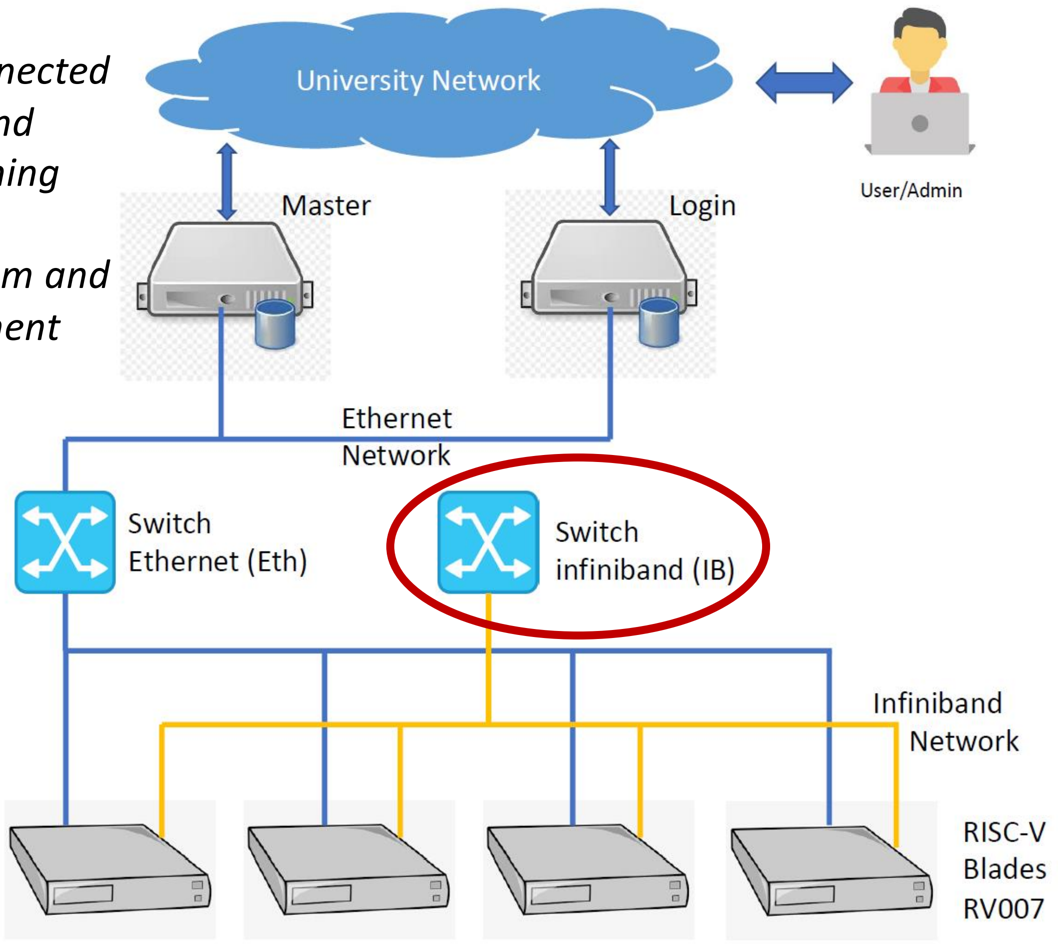
- All software stack installed w. SPACK with the already present linux-sifive-u74mc
- Ubuntu 20.04 Linux O.S. installed with riscv64 image

Monte Cimone Software Stack:

Infiniband Host Channel Adapter (HCA):
- Infiniband FDR HCA (56Gbit/s) w. RDMA
- 2x Mellanox ConnectX-4 FDR HCA
    - PCIe Gen 3 x8 lanes

Experimental results:
- Kernel recognizes the device driver
- Device Driver recognizes Mellanox OFED stack
- IB ping test → successful → Infiniband feasible
    - Between two boards
    - Between board and an HPC server.
- RDMA fails due incompatibilities of the software stack and the kernel driver.
- This is currently a feature under development.

The cluster is connected to a login node and master node running the job scheduler, network file system and system management software.

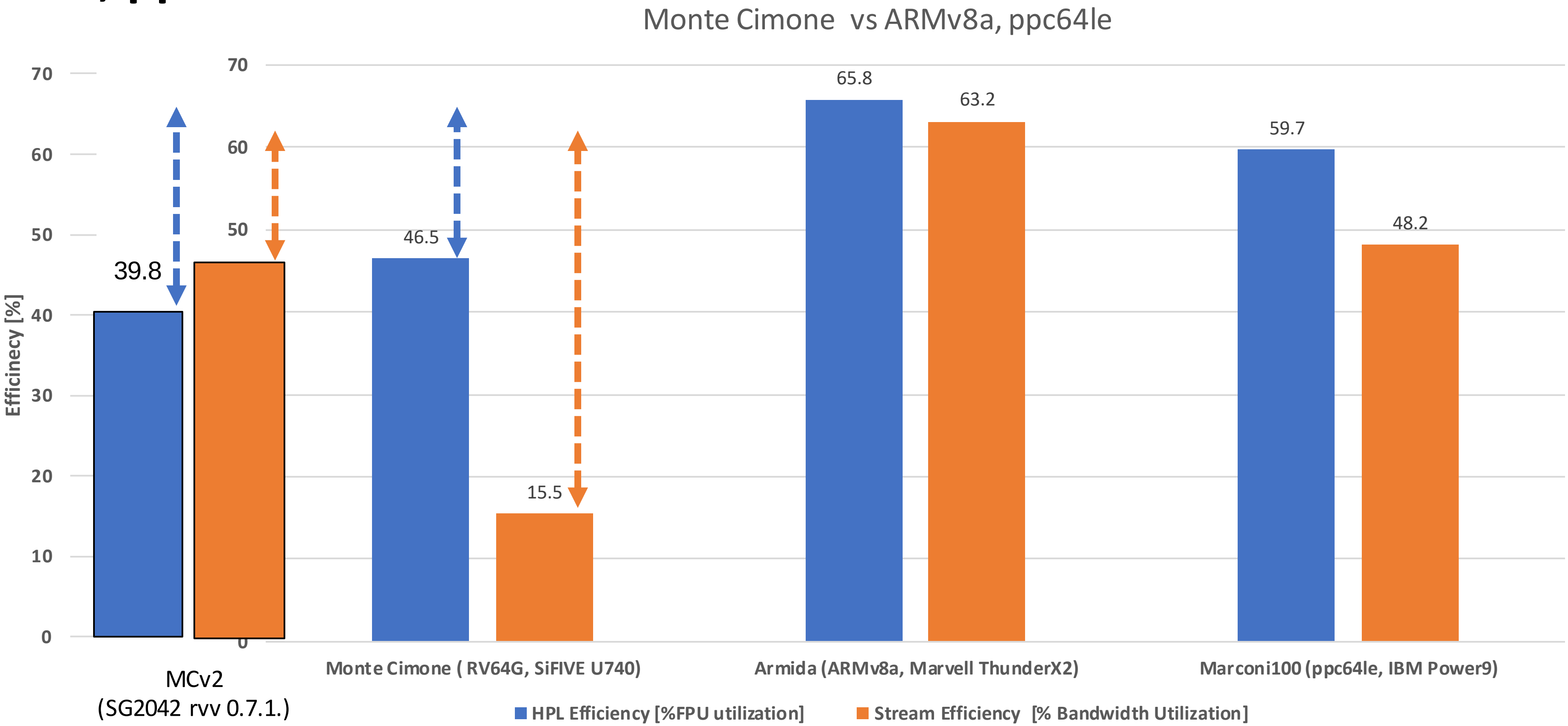


Performance Characterization

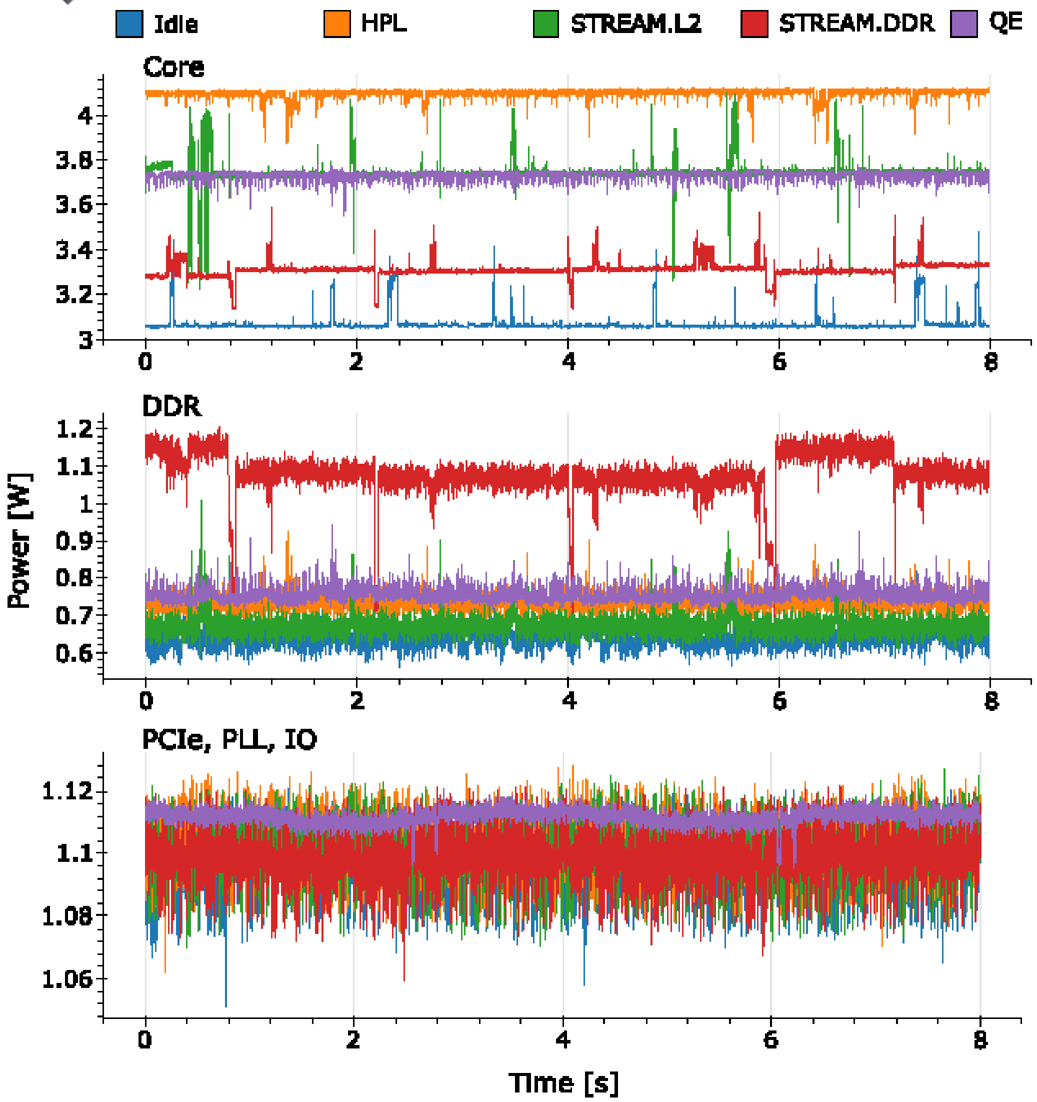
Monte Cimone vs ARMv8a, ppc64le:

HPL and Stream benchmarks on two SoA computing nodes:

- *Marconi100 (ppc64le, IBM Power9)*
- *Armida (ARMv8a, Marvell ThunderX2)*
- *Same benchmarking boundary conditions*
  - *Vanilla unoptimized libraries*
  - *software stack deployed via SPACK package manager*



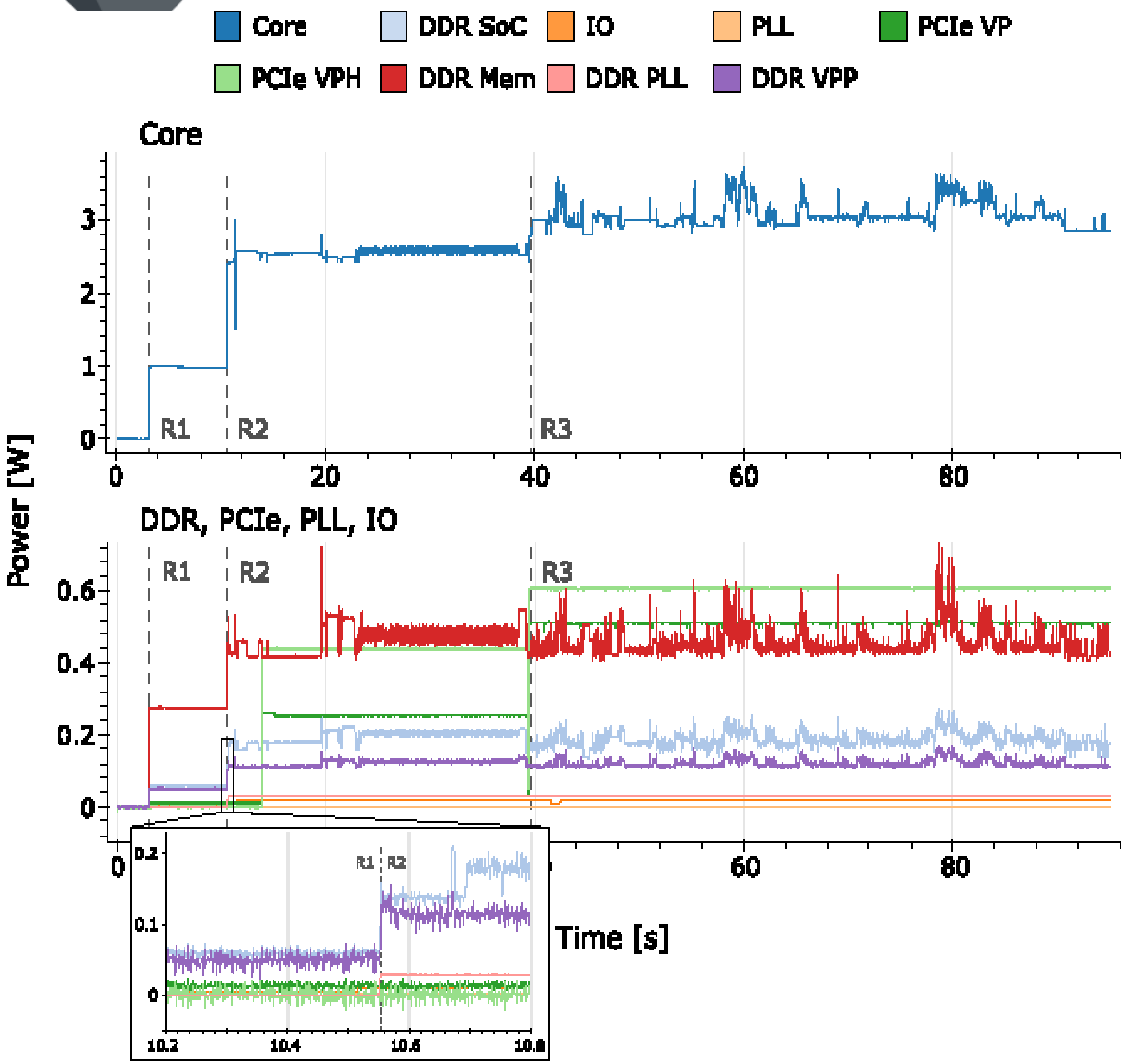
Power Characterization



Line	Idle		HPL		STREAM.L2		STREAM.DDR		QE	
	[mW]	[%]	[mW]	[%]	[mW]	[%]	[mW]	[%]	[mW]	[%]
core	3075	64	4097	69	3714	68	3287	62	3825	67
ddr_soc	139	3	177	3	170	3	232	4	176	3
io	20	0	20	0	20	0	20	0	20	0
pll	1	0	1	0	1	0	1	0	1	0
pcievp	521	11	527	9	524	10	522	10	530	9
pcievph	555	12	554	9	554	10	555	10	561	10
ddr_mem	404	8	440	7	401	7	592	11	434	8
ddr_pll	28	1	28	1	28	1	28	1	28	1
ddr_vpp	67	1	90	2	73	1	98	2	95	2
Total	4810	100	5935	100	5486	100	5336	100	5670	100

- **Idle**: 4.81W (64% of core power, 13% related to DDR and 23% of related to PCI subsystem)
- **HPL**: 5.935W (69% of core power, 14% related to DDR and 18% related to PCI subsystem)

Power Characterization



Line	Idle		HPL		STREAM.L2		STREAM.DDR		QE		Boot R1	Boot R2
	[mW]	[%]	[mW]	[%]	[mW]	[%]	[mW]	[%]	[mW]	[%]	[mW]	[mW]
core	3075	64	4097	69	3714	68	3287	62	3825	67	984	2561
ddr_soc	139	3	177	3	170	3	232	4	176	3	59	197
io	20	0	20	0	20	0	20	0	20	0	5	20
pll	1	0	1	0	1	0	1	0	1	0	0	2
pcievp	521	11	527	9	524	10	522	10	530	9	12	231
pcievph	555	12	554	9	554	10	555	10	561	10	1	395
ddr_mem	404	8	440	7	401	7	592	11	434	8	275	467
ddr_pll	28	1	28	1	28	1	28	1	28	1	0	29
ddr_vpp	67	1	90	2	73	1	98	2	95	2	49	122
Total	4810	100	5935	100	5486	100	5336	100	5670	100	1385	4024

Core complex @ boot process:

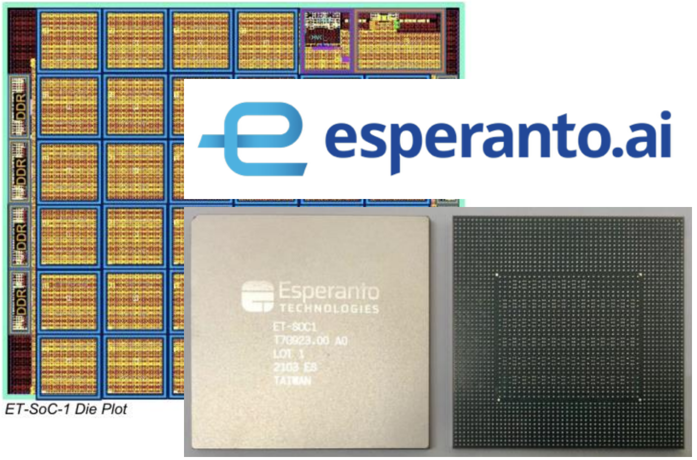
- 0.981W of leakage only power (32% of the Idle power)
- 0.514W O. S. idle power (17% of the Idle power)
- 1.577W of dynamic and clock tree power (51% of the idle power).

Looking Forward - Systems

Build a **Petascale** class RISC-V **Supercomputer**
Explore **RISC-V accelerated HPC platforms in production**.

Goal: **Currently** working with **several RV accelerator provider**. Among those:

PULP Platform energy efficient accelerators[1]
[STX, Occamy, ...]



Esperanto Technologies ET-SoC-1[2]

Axelera Metis AIPU

[1] https://pulp-platform.org
[2] Accelerating ML Recommendation With Over 1,000 RISC-V/Tensor Processors on Esperanto's ET-SoC-1 Chip, David R. Ditzel, the Esperanto team, DOI: 10.1109/MM.2022.3140674
[3] Co-Design of the Kalray Manycore Accelerator for Edge Computing, Benoît Dupont de Dinechin, HiPEAC CSW Autumn 2021

Curiosity - Why «Monte Cimone»?

Tallest moutain in the northern Apennines, w. 2.165 m

On clear days the crest is visible from the major center italian cities: Bologna, Mantova, Modena, Reggio Emilia, Firenze, Lucca …

Considering the spatial resolution of the human eye, the top of Mount Cimone is the geographical point from which the most Italian surface can be seen.

In the Monte Cimone's ski resort, the famous Italian (and Bolognese) worldwide ski champion Alberto Tomba started

Monte Cimone Team

- University of Bologna:
  - Luca Benini, Andrea Bartolini, Francesco Barchi, Emanuele Parisi, Andrea Acquaviva, Emanuele Venieri, Kashaf Ad Dooja, Giacomo Madella
- CINECA:
  - Daniele Cesarini, Federico Ficarelli
- E4 Computer Engineering:
  - Cosimo Gianfreda, Marco Cicala, Daniele Gregori, Francesco Beneventi, Mattia Paladino, Elisabetta Boella

Dissemination

E4 SUPPORTED STUDENT CLUSTER COMPETITION 2022 WITH MONTE CIMONE

https://www.nextplatform.com/2022/06/09/strong-showing-for-first-experimental-risc-v-supercomputer/

https://arxiv.org/abs/2205.03725

https://open-src-soc.org/2022-05/media/slides/RISC-V-International-Day-2022-05-05-11h05-Calista-Redmond.pdf

FIRST RISC-V HPC OPEN LAB - WORLDWIDE

As an **educational tool**:

2x courses at **Università di Bologna**:

- Computer Architectures
- Laboratory of Big Data Architectures

2x PhD schools:

- 2023 ACM Europe Summer School on "HPC Computer Architectures for AI and Dedicated Applications"
- 2023 DEI UNIBO PhD course High-performance Emerging Computing Paradigms
- 2 planned in 2024

Introduced **~120 students** to **µarch profiling**, **HPC programming**, **distributed systems** right in a **RISC-V environment**.

Ported and ran production of **widespread HPC applications** (e.g.: **quantumESPRESSO**, **OpenFOAM**).

Several **research activities currently ongoing** on Cimone.

Access open to everyone interested

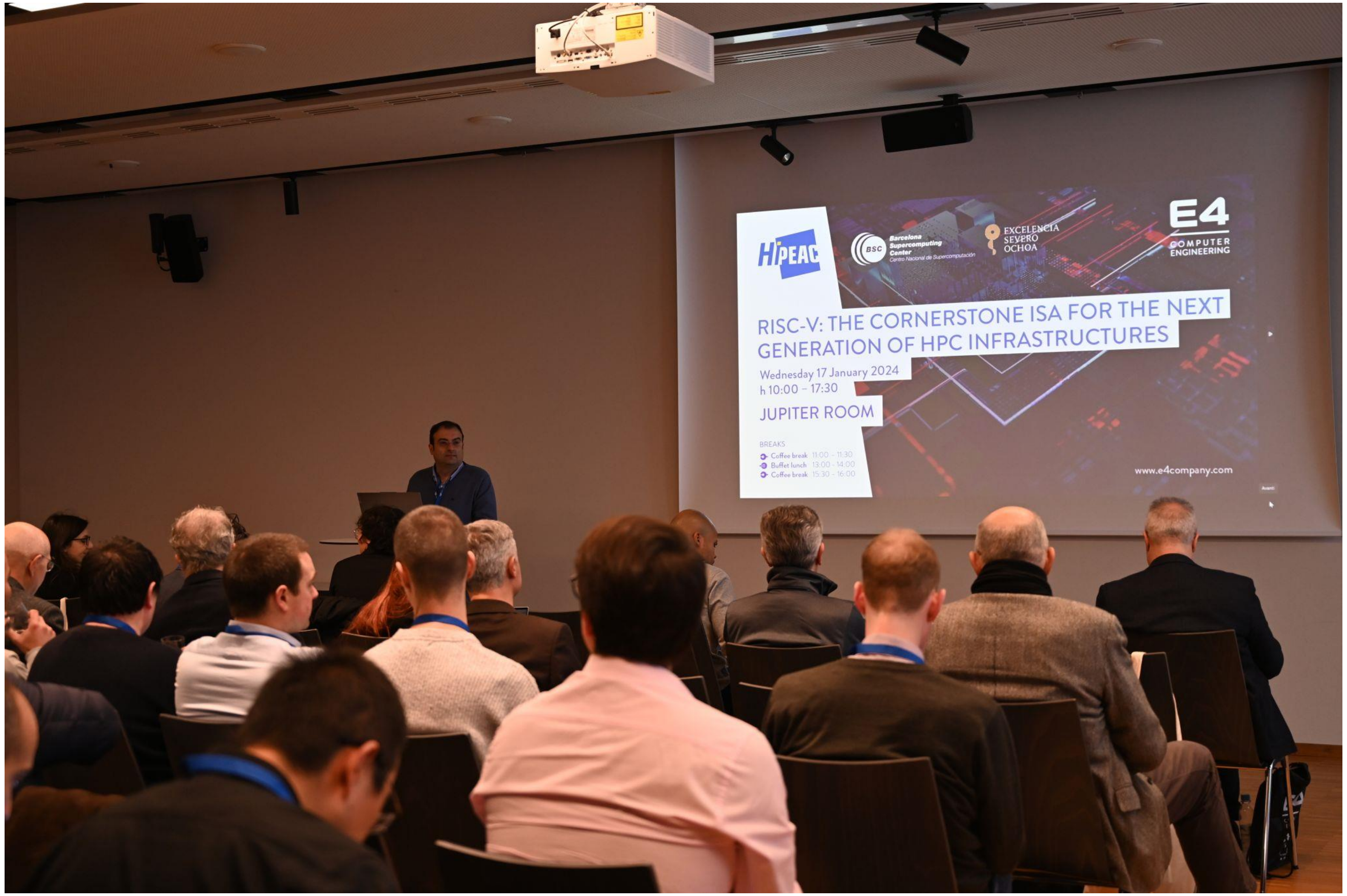
RISC-V WORKSHOP

In **2023** E4 organized the first international workshop:

«RISC-V: the cornerstone ISA for the next generation of HPC infrastructures»

at HiPEAC Conference in Toulouse France, 17/01/2023

In **2024** E4 and BSC co-organized the second edition in
Munich, Germany, 01/17/2024



EU Projects

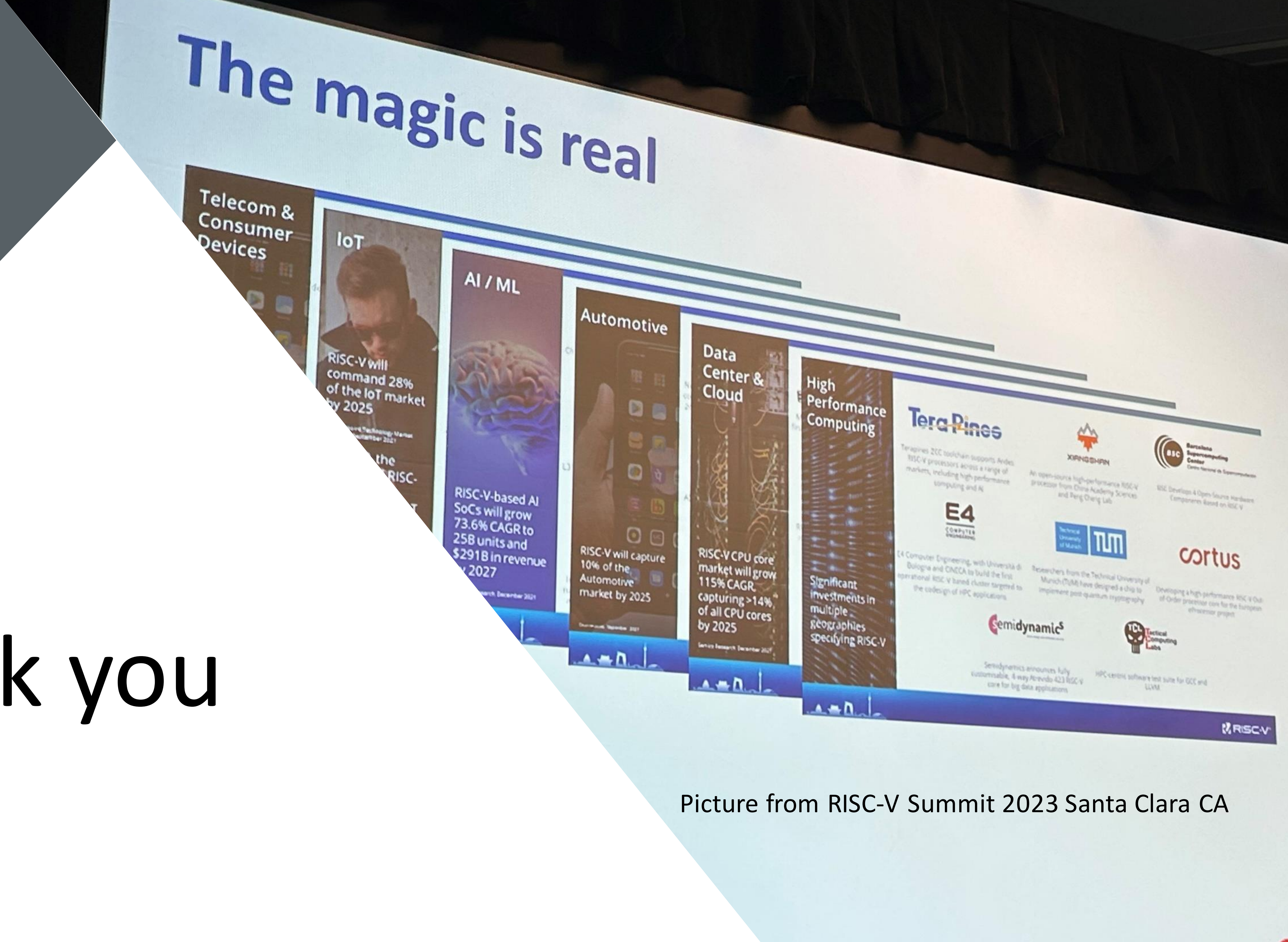
NEXT STEPS IN THE RISC-V WORLD

E4 is member of the TRISTAN & ISOLDE (2023-2025) consortia to develop a European RISC-V Framework for the Space Use Case
https://tristan-project.eu/
https://www.isolde-project.eu/

E4 is member of the DARE (2024-2030) consortium to develop a EU RISC-V Based Computing System
https://eurohpc-ju.europa.eu/system/files/2023-11/Decision%2039.2023%20Approving%20RISC-V%20Call%20Results.pdf

E4 COMPUTER ENGINEERING

Thank you



Picture from RISC-V Summit 2023 Santa Clara CA

CONTACTS

Email contacts

info@e4company.com
support@e4company.com
sales@e4company.com

E4 Computer Engineering SpA

Via Martiri della Libertà, 66 . 42019 Scandiano (RE) - Italy

Tel. +39 0522 991811