BARCELONA **ZETTASCALE** LAB

Hacia el Zettascale: Avances del Barcelona Zettascale Lab

Evento Final del proyecto BZL Madrid (España)
16 de Junio de 2026

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

BARCELONA **ZETTASCALE** LAB

Avances del Barcelona Zettascale Lab (BZL)
WP2 – Software Stack

Xavier Teruel García – BSC

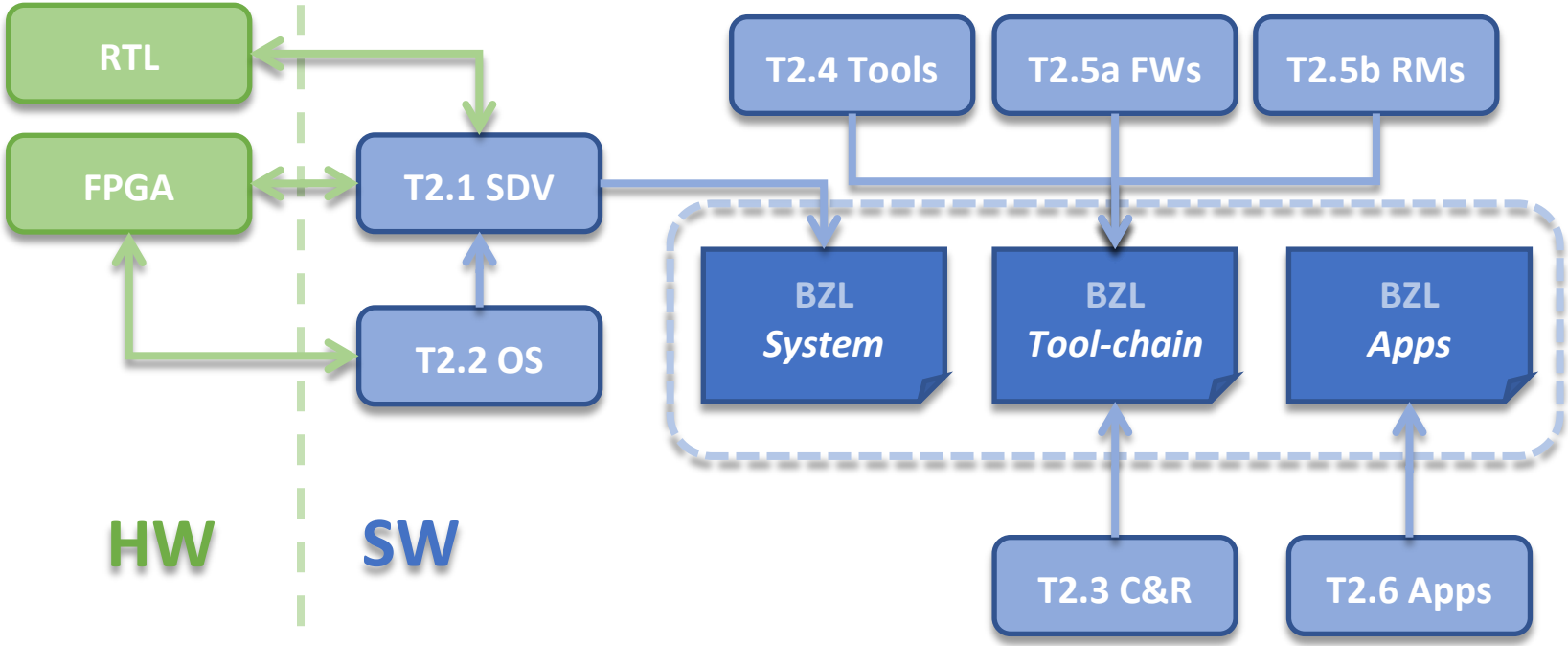
SARTECO Side Event – Madrid, Spain –
16/06/2026

Objectives

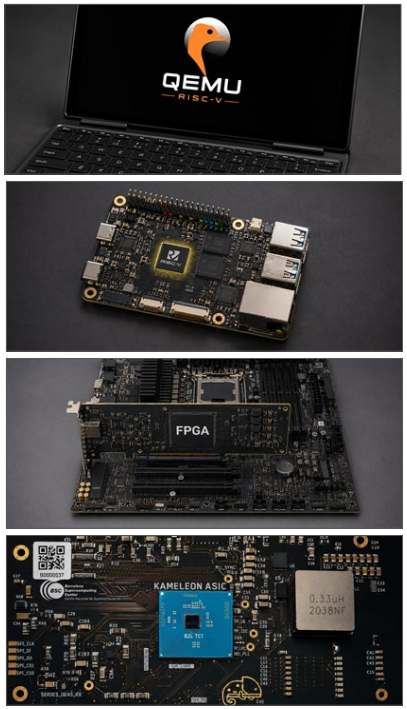
To develop a new **software ecosystem** which includes the operating system and device drivers, compilers, parallel runtimes and basic numerical libraries allowing the execution of **representative applications** in the HPC scenario and the emerging applications which combine aspects of Artificial Intelligence (AI) with the simulations based on the traditional “first principles” of HPC.

These developments will imply to adapt available technologies at BSC-CNS to the design of the proposed accelerators, honouring the criteria of **performance**, **energy efficiency** and **resource usage**.

Task structure (and workflow)



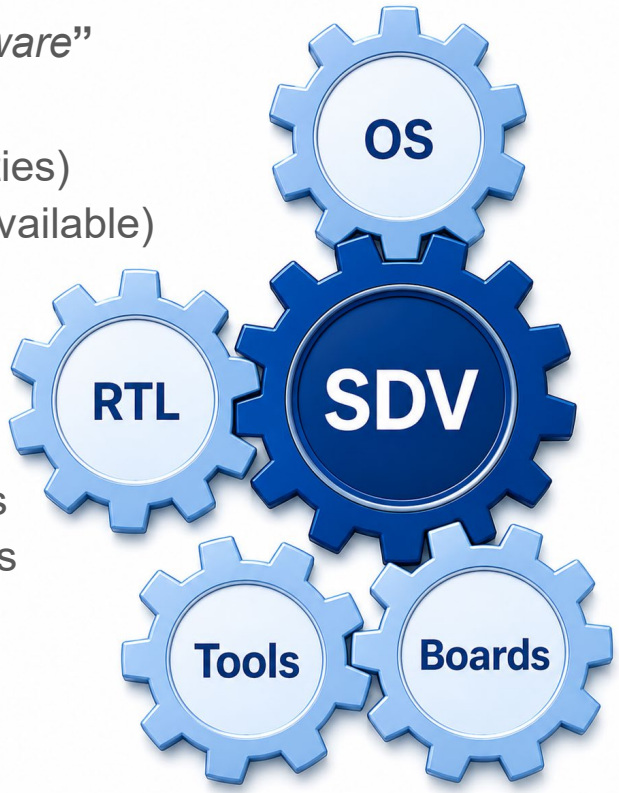
“Hardware” resources



- **QEMU:** Software emulated architecture; useful for small workloads (specially at CI/CD)
- **SDVs@Dev-Boards:** Development boards available within BSC premises (Pioneer Milk-V, Banana PI, or Premier); with tools installed in them (e.g., Vehave)
- **SDVs@FPGA:** Multiple versions of the PERTE processors, initially based on Lagarto Ox (e.g., Stand-alone, Cinco ranch, or Komodo)
- **Tape Outs (TOs):** The ASIC versions generated within the project

The importance of the SDVs

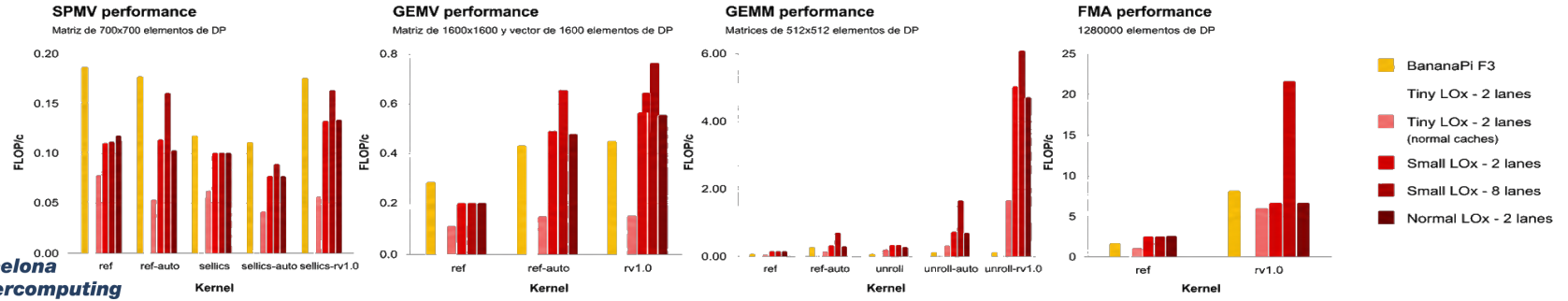
- **Spanish: (SDV) "***Vehículo para el Desarrollo de Software***"**
- **Develop software before tape-out**
  - Breaks HW/SW task dependence (unblock activities)
  - Speed-up software readiness (when HW finally available)
    - Linux execution (e.g., drivers)
    - Tool-chain features (e.g., PAPI Counters)
    - Application porting/tuning (e.g., vectorization)
- **Contiguous HW/SW validation and co-design**
  - Provides useful feedback to hardware developers
  - Reports functional issues and performance issues



Task: Software Development Vehicle

- **Infrastructure**
  - HCA cluster, maintenance of existing resources
    - Last 3 years: Pioneer Milk-V, Banana Pi and Premier
    - Previous systems: Unmatched, Allwinner D1,…
    - Local FPGAs: Pickle systems (4 hosts with FPGAs)
  - FPGA cluster, enabling systems for software users
    - Synthesis of Lagarto Ox different versions
- **Benchmarking,** Initial benchmarking in new bitstreams

Tipo	Sistema	SO	SDV	B-Mark	Usuarios
Placa Comercial	Milk-V Pioneer	*validado*	*validado*	*validado*	*hábil*
	HiFive Unmatched	*validado*	*validado*	*validado*	*hábil*
	Banana Pi BPI-F3	*validado*	*validado*	*validado*	*hábil*
	Premier	*validado*	*validado*	*validado*	*hábil*
Clúster FPGAs	Lagarto Hun	*en-curso*	–	–	–
	Sargantana (SMP)	*validado*	*en-curso*	–	–
	Lagarto Ox (stand-alone)	*validado*	*validado*	*validado*	–
	Lagarto Ox (TC2-F)	*validado*	*validado*	*validado*	*hábil*
	Lagarto Ox (Komodo-F)	*validado*	*validado*	*validado*	*hábil*
Test Chips	TC1 Sargantana	*en-curso*	–	–	–
	TC1 Lagarto Ka + VPU	*en-curso*	–	–	–
	TC2 Lagarto Ox	*en-curso*	–	–	–



Task: Operating System

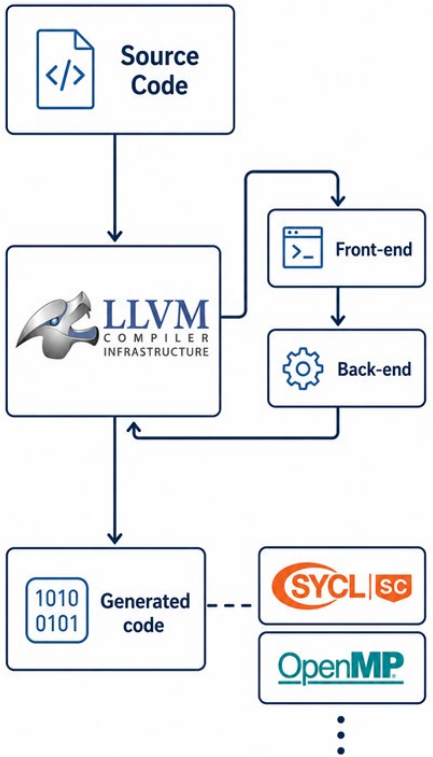
- **Linux kernel,** adapting it for the corresponding chip-sets
- **Linux booting,** and configuration
  - Feedback for HW teams, (e.g. ETH, PLIC, Cache,…)
  - Enabling images for the SDV team

- **Linux distributions**
  - NixOS: containerize executables (e.g., SPEC)
  - Buildroot: initial version to boot new bitstreams
  - Ubuntu/Fedora: higher degree of validation

- **Linux services**
  - Accelerated MPI, improving inter-process communication
  - GekkoFS, improves I/O performance

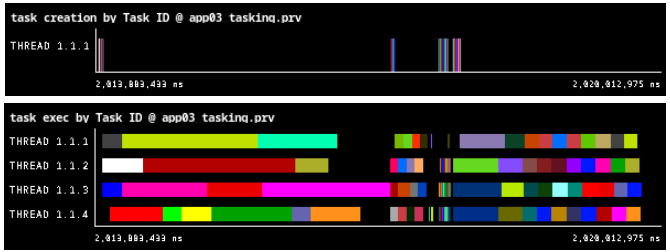
Task: Compiler and runtimes

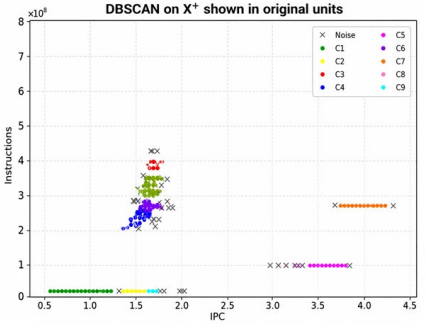
- **LLVM/Compiler**
  - Target: Clang (C/C++) and Flang (Fortran)
  - Involved in co-design activities (e.g., generated code)
  - Loop fusion, OpenMP 6.0 vs. auto-vectorization
- **SYCL/POCL**
  - Runtime improvements, enhance compatibility
  - Integration with nOS-V and TA-SYCL (PoC)
  - Leveraging vectorization
- **OpenMP@Cluster**
  - Extending OpenMP beyond the node
  - Implementing distributed taskiter (i.e., OpenMP iterations)

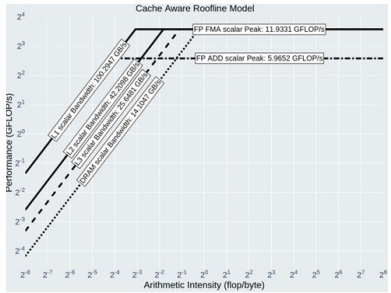


Task: Performance analysis tools

- **Extrae library,** getting Paraver traces (event-based instrumentation)
  - Capture OMPT and GOMP activity
  - PAPI-like implementation, hardware counters
- **Trace visualizers and viewers**
  - Paraver: several visual extensions
  - Clustering (GSOM): algorithm + viewer
- **Memory modelling**
  - Methodology refinement: Original RM
  - CARM, PROFET & TopDown
- **System-wide tracing**
  - API extension, trace format, OVNI

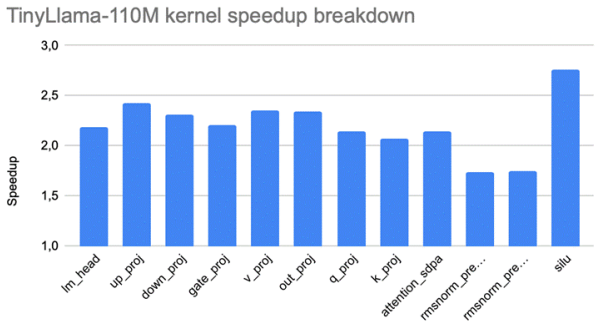


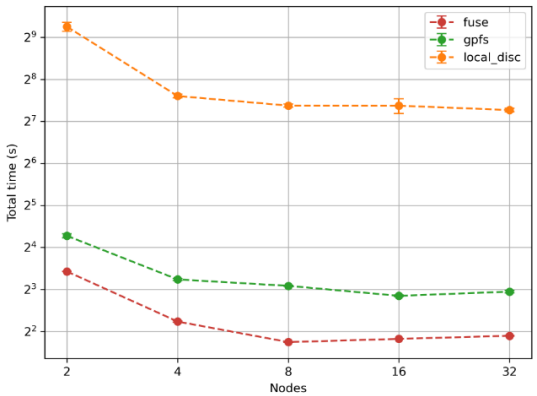




Task: Frameworks and libraries

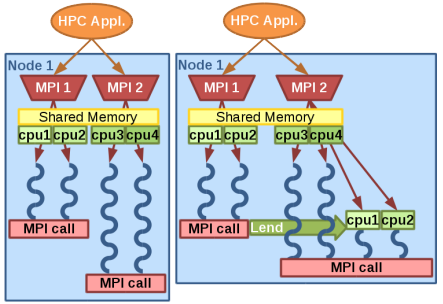
- **Numerical libraries and AI frameworks**
  - Enabling oneDNN, ATen (leveraging vector)
  - Enabling Pytorch (integrated with previous)
- **MPI Vectorization, in addition of Accelerated MPI**
  - Leveraging vector instructions in MPI services
- **Task-aware libraries,** TA-MPI, TA-SYCL
- **Workflows,** PyCOMPSs/COMPSs
  - IA, CAR, and AD decorators (scheduler impact)
  - Integration with GekkoFS





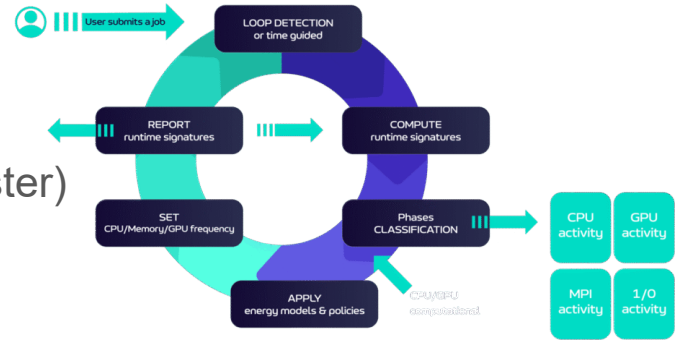
Resource managers

- **Dynamic Load Balance (DLB)**
  - Enabling DLB ecosystem at RISC-V (joint with EUPILOT)
  - MNGO: as decision maker for LeWI / TALP
- **Dynamic Management of Resources (DMR)**
  - MPI sessions; spawning improvements MPDATA adaptation (to DMRv2)
- **Energy Aware Runtime (EAR)**
  - RISC-V porting; integration with DLB
- **Dynamic Node and Core Allocation**
  - Load balance improvements (OpenMP@Cluster)



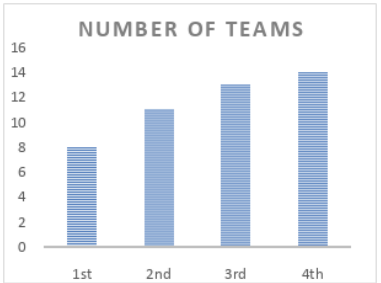
a) Unbalanced MPI application

b) Unbalanced MPI application balanced with DLB



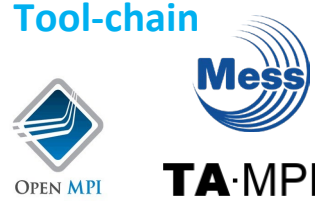
Applications and Hackathons

- **Software stack functional validation / preliminary performance**
  - Already reported in their own task as part of “porting”
  - Usually also includes an initial performance
- **Growing in complexity, analysis of performance: vectorization, multi-threading, and multi-processing**
  - Areas: Numerical libraries, Hybrid (MPI+OpenMP), Task-aware libraries, OpenMP@Cluster, Workflow applications (PyCOMPSs/COMPSs)
- **BZL Hackathons, organized since February 2026 (4 editions so far)**
  - ISA/RTL Feedback
  - System / tool-chain people
  - Department / project invites
    - CASE (Alya)
    - DARE (NEMO Dwarfs)



Considered software stack

Tool-chain

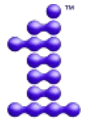


extrae
Paraver

PAPI
like

ATen

oneDNN



oneAPI

TA·MPI

TA·SYCL

flame/blis

MPICH

OpenMP

Elasticity

Dynamic
Resource
Allocation

NixOS

DMRlib

Operating System

SDV

Vehave **RISC-V**

RAVE HPL Benchmarks

SDV-CAB Benchmarks HPCG