



ChEESE

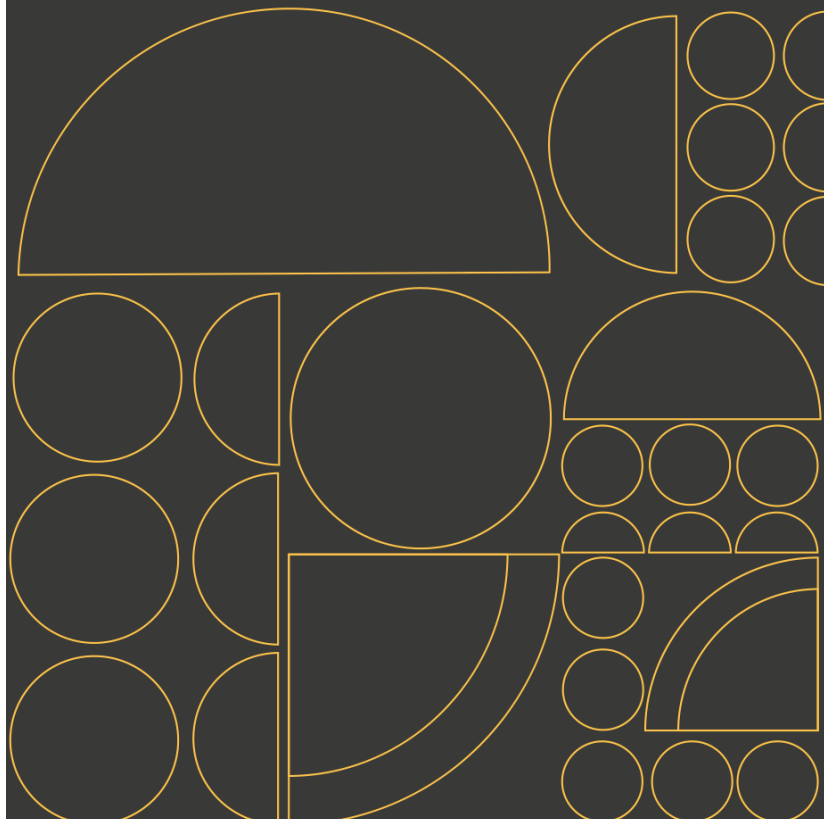
WP2 telco
January update

Piero Lanucara (representing the entire
WP2 team)

CINECA



Project funded by EuroHPC under the grant agreement No 101093038.



WP2: Tasks

Task	Name	Lead partner	Start month	End month
2.1	Code audits	HLRS	3(27)	15(36)
2.2	Increasing the performance portability of flagship codes for different accelerators	CIN	1	48
2.3	Optimisation for Exascale	CIN	6	48
2.4	Continuous Performance Efficiency Monitoring	BSC	18	48



WP2: Deliverables

D	Deliverable name	Short description	Lead	Type	Level	M
2.1	First code audit report	First code audit round results and the recommendations arising from the analysis	HLRS	R	PU	12
2.2	First report on improving code performance at exascale	Report on porting, optimization and tuning for the selected applications on the available hardware	CIN	R	PU	24
2.3	Second code audit report	Update of D2.1 and further recommendations arising from the analysis	HLRS	R	PU	30
2.4	Final report on improving code performance at exascale and Best Practice Guide	Technical report describing the achieved scalability level and related performances obtained for the selected applications on the available hardware including a final Best Practice guide	CIN	R	PU	48



D2.2 outcomes (in a nutshell)

- ❑ 117 pages in the final version
- ❑ Results in terms of:
 1. Methodology
 2. Improvements
 3. PP campaign
 4. Lesson learnt

- ❑ **With the term mini-apps we refer to mini apps, nano apps or kernels (see D2.2 for details)**

for a variety of flagship codes and mini-apps in different (GPU) flavours on suitable hardware

Flagship code	Mini apps	GPU flavors	GPUs for PPC	FOM/PP metrics
SeisSol	SeisSol-proxy	CUDA, SYCL, OpenMP target offload	A100, H100, GH200, MI250X, Intel Data Center GPU Max 1550	Max TFLOP/s/Hardware Efficiency (Tab. 4-1.1)
SPECfEM3D	MiniSPECfEM3D	CUDA, HIP	A100, MI250X	Time Loop/Application Efficiency (Tab. 4-2.1)
ExaHyPE	ExaHyPE kernels	CUDA, OpenMP target offload	A100, H100, Intel Data Center GPU Max 1550	Bandwidth/Architectural Efficiency (Tab. 4-3.1)
Tandem	Static mini app	CUDA, HIP	A100, MI250X	TFLOP/s/Architectural Efficiency (Tab.4-4.1)
xSHELLS	N.A.	N.A.	N.A.	N.A.
HySEA	Tsunami-HySEA	CUDA, SYCL	A100, H100, GH200, Intel Data Center GPU Max 1100, Intel Data Center GPU Max 1550	Runtime/ Application Efficiency (Tab. 4-6.1)
FALL3D	FALL3D suite	OpenACC	A100, H100, MI250X	Runtime/Application Efficiency (Tab. 4-7.4)
OpenPDAC	N.A	N.A	N.A.	N.A
LaMEM	MatMult, VecMAXPY (PETSc).	CUDA, Kokkos, HIP	A100, GH200, MI250X	MFLOP/s/Architectural Efficiency (Tab.4-9.1/2/3)
pTatin3D	SpMV MF	CUDA, HIP	N.A.	N.A.
ELMER/ICE	ELMER simplified benchmark	CUDA (AMGX), HIP (Rocalution), OpenMP target offload	A100, GH200, MI250X	Solver time (time-step) (Tab. 4-11.1)

D2.2 outcomes (in a nutshell)

Machine	GPU	Notes
Adastra	AMD Instinct MI250X	https://dci.dci-gitlab.cines.fr/webextranet/architecture/index.html
Claix	NVIDIA H100	https://top500.org/system/180295/
Leonardo (booster)	NVIDIA A100 SXM4	https://www.top500.org/system/180128/
LUMI-G	AMD Instinct MI250X	https://www.top500.org/system/180048/
MareNostrum 5 (ACC)	NVIDIA H100	https://www.top500.org/system/180238/
SuperMUC-NG Phase 2	Intel Data Center GPU Max 1550	https://doku.lrz.de/hardware-of-supermuc-ng-phase-2-222891050.html
Thea	NVIDIA GH200 "superchip"	N.A.
Unknown system	Intel Data Center GPU Max 1100	N.A.
Vista	NVIDIA GH200 "superchip"	https://tacc.utexas.edu/systems/vista/

D2.2 outcomes (in a nutshell)

Flagship code	Porting	PPC	CAD improvements	Other improvements
SeisSol			N.A.	
SPECFEM3D	N.A.			
ExaHyPE				N.A.
Tandem			N.A.	
xSHELLS	N.A.	N.A.	N.A.	N.A.
HySEA				N.A.
FALL3D	N.A.			
OpenPDAC	N.A.	N.A.	N.A.	
LaMEM			N.A.	N.A.
pTatin3D		N.A.		N.A.
ELMER/ICE			N.A.	

Joint effort from code owners+CINECA+Eviden+.....others....

D2.2 after D2.2 (TBD)

- ❑ There is still room for new results!
- ❑ Some flagship codes are missing/new mini-apps can appear in the arena
- ❑ New hardware
- ❑ New improvements (in the flagship codes and for mini-apps participating to a PP campaign)
- ❑ Updates suitable for next review (September 2025) and D2.4. Anyway, WP2 is not committed to do so after D2.2 (Best effort?)

WP2 after D2.2: WP2 vs WP3 (recap)

- WP2
 - full code, investigation (including porting, tuning) on CPU and GPU side. EuroHPC targeted. Code audit (but not only) driven.
 - + Performance (POP) metrics
 - + Performance Portability study on GPU for mini-apps on EuroHPC (and SDV)
- WP3
 - mini-apps investigation/co-design (including porting, tuning) on CPU and GPU side. EUPEX/EUPILOT proxies.
 - + Performance/energy efficiency metrics



WP2			
Application	Target	Task	Methodology
Full code	EuroHPC CPU & GPU	Porting Optimization for exascale Continuous performance efficiency monitoring	Code audit driven (not only) Performance POP metrics
Mini-apps	EuroHPC GPU SDV (?)	Porting Optimization	Performance portability

Porting experience
Optimization: from full code to mini-apps
GPU performance portability

Mini-apps
Porting experience
Optimization: from mini-apps to full code

WP3			
Application	Target	Task	Methodology
Mini-apps	- OpenSequana prototype: EUPEX (ARM+HBM & GPU) - RISC-V prototypes: EPI & EUPILLOT	Porting Performance study Optimization (Scalability & Energy consumption analysis)	Co-design between SE community and Hw, Sw and Technology providers

WP2 after D2.2: coming next

- 
- A decorative vertical bar on the left side of the slide, featuring a dark grey background with yellow geometric patterns including circles, semi-circles, and arcs.
- ☐ On-going tasks:
 - ✓ T2.1 (towards D2.3)
 - ✓ T2.4
 - ☐ Continue the activity started with the PPC in D2.2 and support Exascale ambition (relationship with WP4 and WP6, CINECA in particular)
 - ☐ Contribute to enforce a Code-centric vision in ChEESE-2P (from WP2 to...)

