E4 COMPUTER ENGINEERING

WHEN PERFORMANCE MATTERS

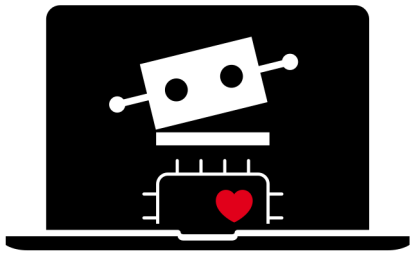
BOOSTING SCIENTIFIC APPLICATIONS PERFORMANCE ON NOVEL COMPUTING ARCHITECTURES

Elisabetta Boella, *HPC Product Specialist*

Workshop sul Calcolo nell'I.N.F.N.
May 22nd, 2024

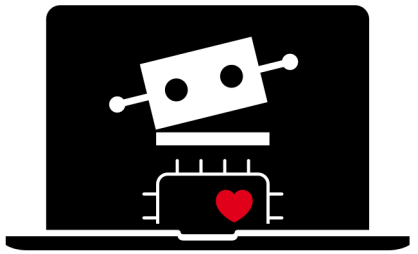
CONTENTS

- E4 computer engineering
- Synthetic tests on emerging architectures
- Performance evaluation of scientific applications on novel architectures
- Summary and perspectives



CONTENTS

- **E4 computer engineering**
- Synthetic tests on emerging architectures
- Performance evaluation of scientific applications on novel architectures
- Summary and perspectives



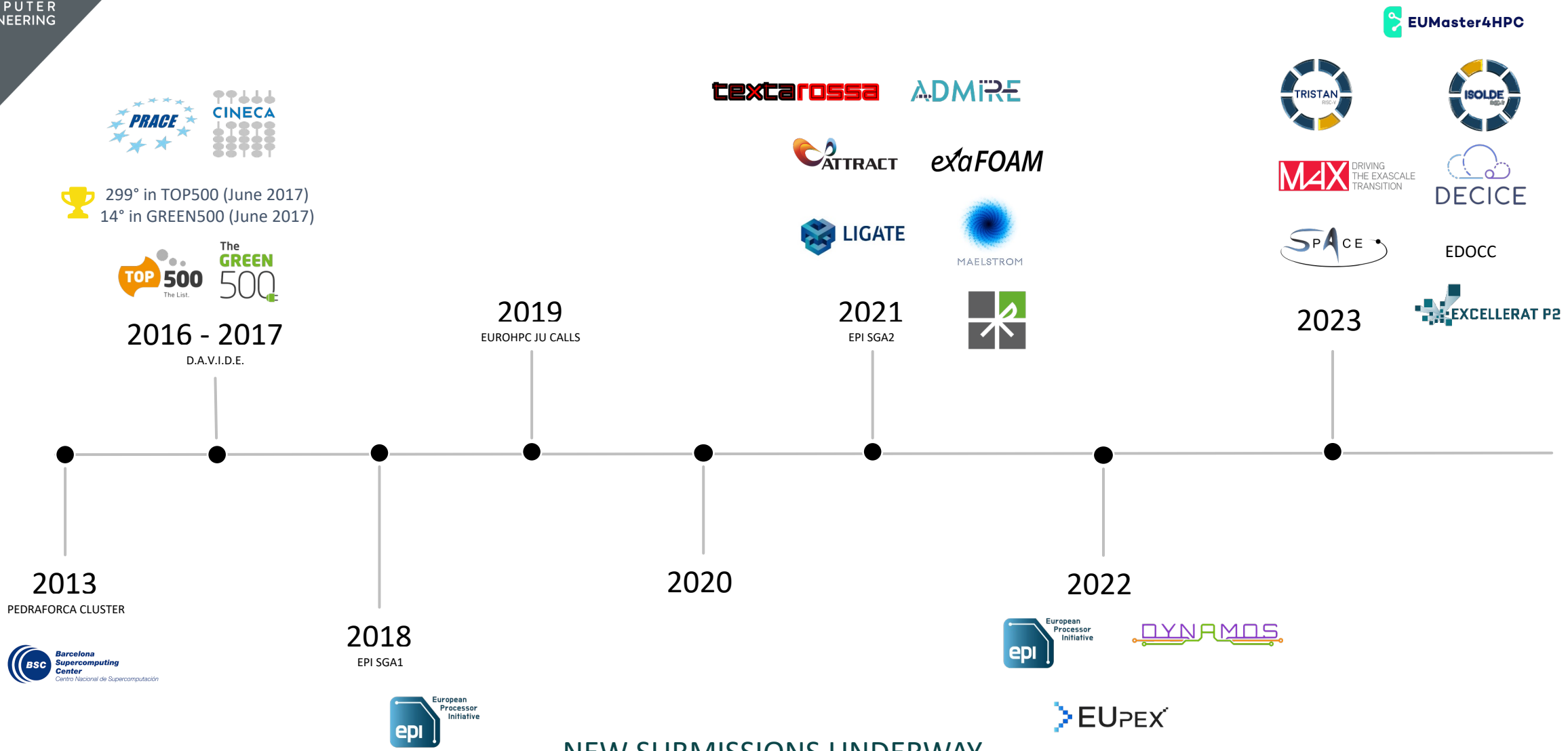
E4 COMPUTER ENGINEERING

E4 Computer Engineering 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, KDT)

Through the sister company E4 Analytics, E4 works to integrate Data Science in organizations that undertake the Digital Transformation of their business to improve products/processes and optimize resources. We operate at the intersection
between business and technology, supporting the customer in the adoption of AI solutions: with E4 Analytics, company data becomes a resource for creating value.

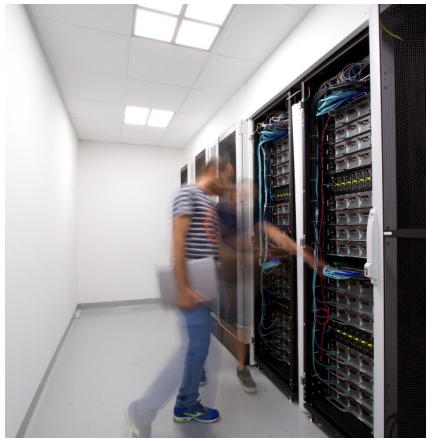
Each E4 solution is UNIQUE, like every one of our customers; TESTED in every single component; VALIDATED to verify the actual performance of each system and SERVICED by technicians who provide assistance in the most extensive and complex Italian and European computing infrastructures

E4 EU PROJECTS



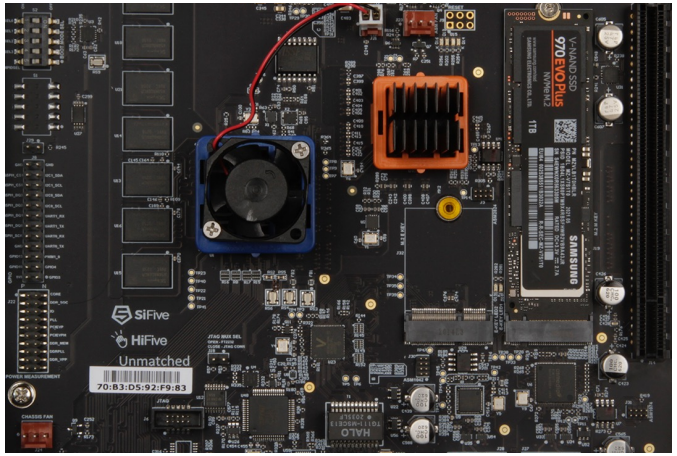
NEW SUBMISSIONS UNDERWAY

E4 TECH FACTORY



- Scout, select, design, test, integrate, configure, optimize, validation & verification and install the **full stack of HW & SW** components and infrastructures
- **Expertise and advice services towards European target customers and prospects about the optimal solution for their needs**
- Consider the **product development** as a never-ending process, accept failures but strive for success
- Apply **co-design** within a continuously changing scenario
- Agnostic approach: supports the widest possible range of hw/sw components and operating systems

SUCCESS STORIES



MONTE CIMONE
First HPC cluster
based on RISC-V

IIT FRANKLIN
HW and SW

MEDOOZA

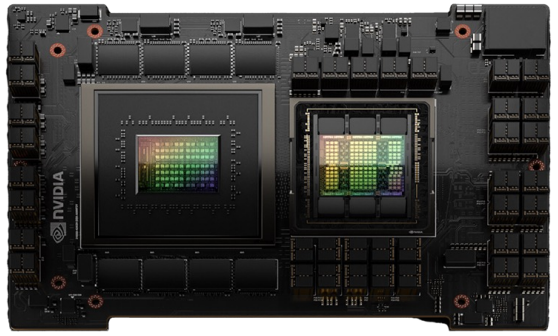
ICE4HPC

INAF ASTRI-MA
Non-supervised real time
data analysis

HETEROGENEITY IS KEY FOR MODERN HPC

AMPERE ALTRA MAX
128 cores/processor
Arm Neoverse N1
3.0 GHz
64 KB L1 cache
1 MB L2 cache
4 MB L3 cache
225 W

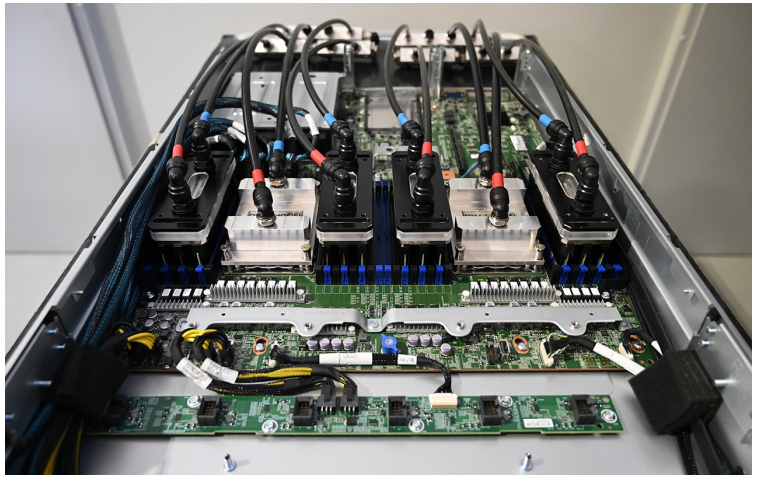
NVIDIA GRACE
144 cores/processor
Arm Neoverse V2
3.1 GHz
64 KB L1 cache
1 MB L2 cache
228 MB L3 cache
500 W



NVIDIA GRACE HOPPER
72 cores/processor
117 MB L3 cache
1 GPU H100
450 - 1000 W

SOPHON SG 2042
64 cores/processor
Risc-V
2.0 GHz
4 MB L1 cache
16 MB L2 cache
64 MB L3 cache
120 W

OUR INNOVATIVE SERVERS



MT COLLINS SERVER
2 Ampere Altra Max
256 GB RAM
Two-pahse cooling system

textarossa

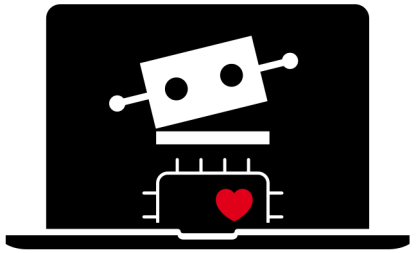
NVIDIA HGX

512 GB RAM

SOPHON SERVER
2 Sophon SG 2042
256 GB RAM

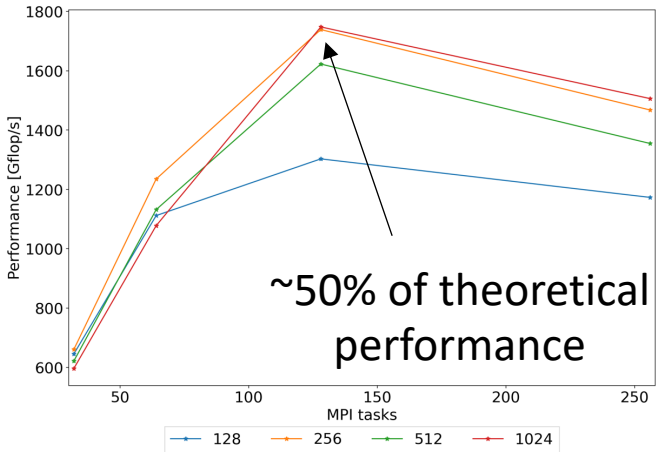
CONTENTS

- E4 computer engineering
- **Synthetic tests on emerging architectures**
- Performance evaluation of scientific applications on novel architectures
- Summary and perspectives

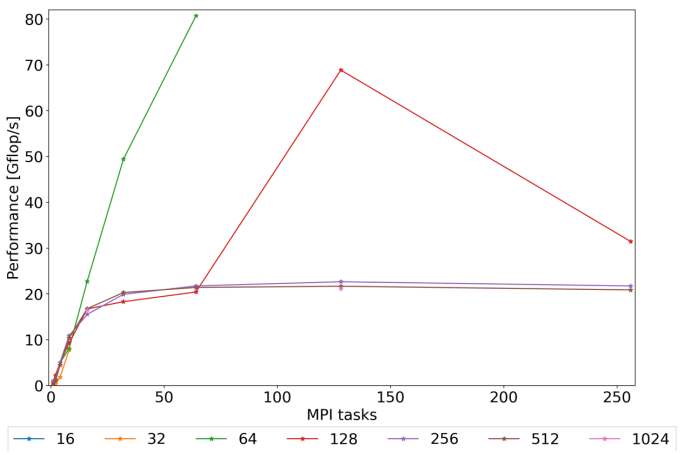


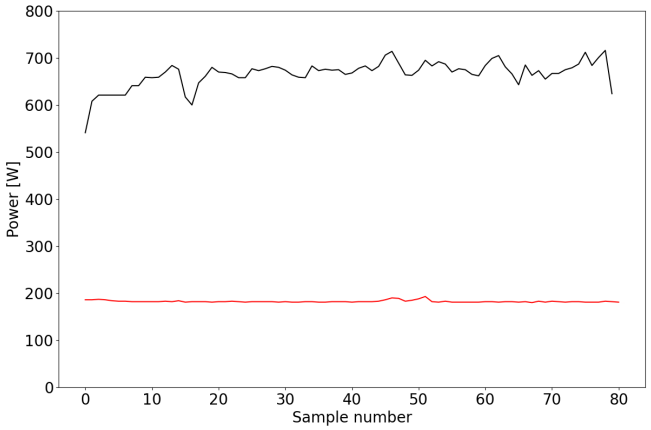
PEAK POWER OF 700 W ON MT COLLINS

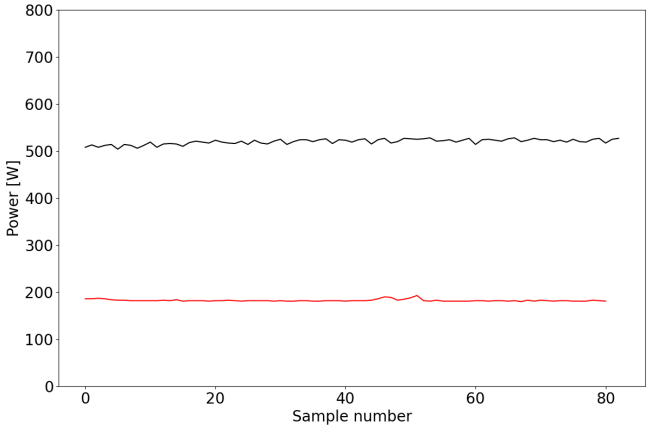
HPL



HPCG

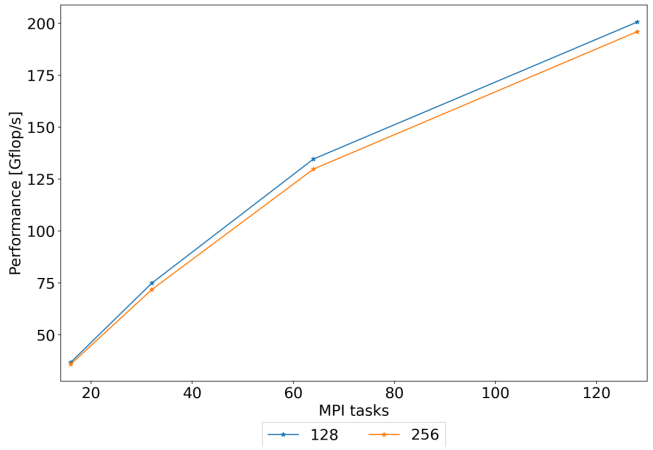


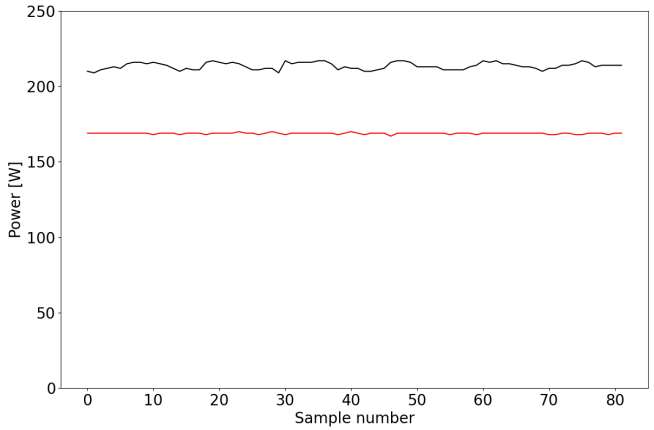




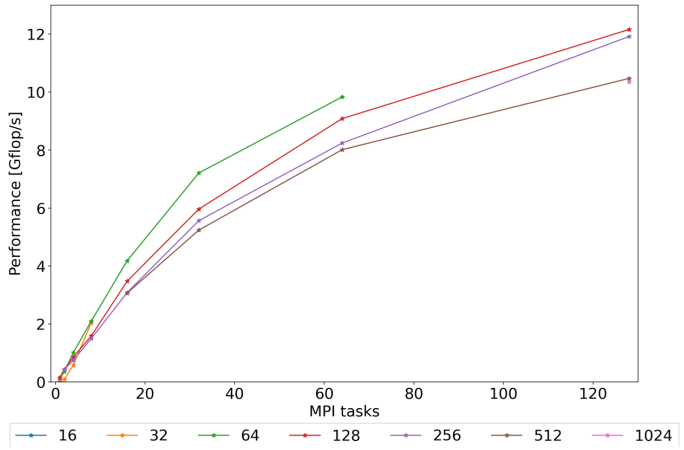
PEAK PERFORMANCE CLOSE TO THEORETICAL

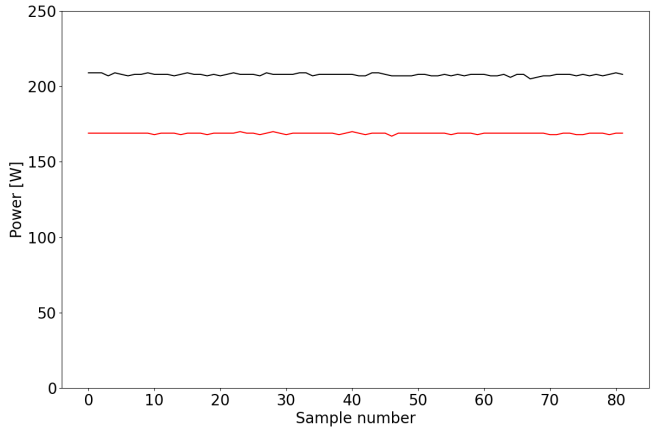
HPL





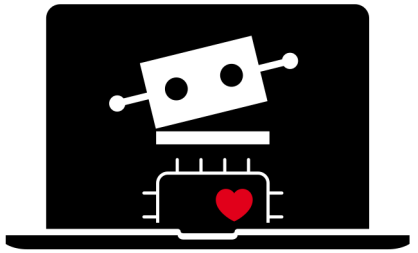
HPCG





CONTENTS

- E4 computer engineering
- Synthetic tests on emerging architectures
- **Performance evaluation of scientific applications on novel architectures**
- Summary and perspectives



OPENFOAM:
AN OPEN SOURCE CFD APPLICATION

OpenFOAM®

exaFOAM

Computational Fluid Dynamics code

Based on the cell centered finite volume method

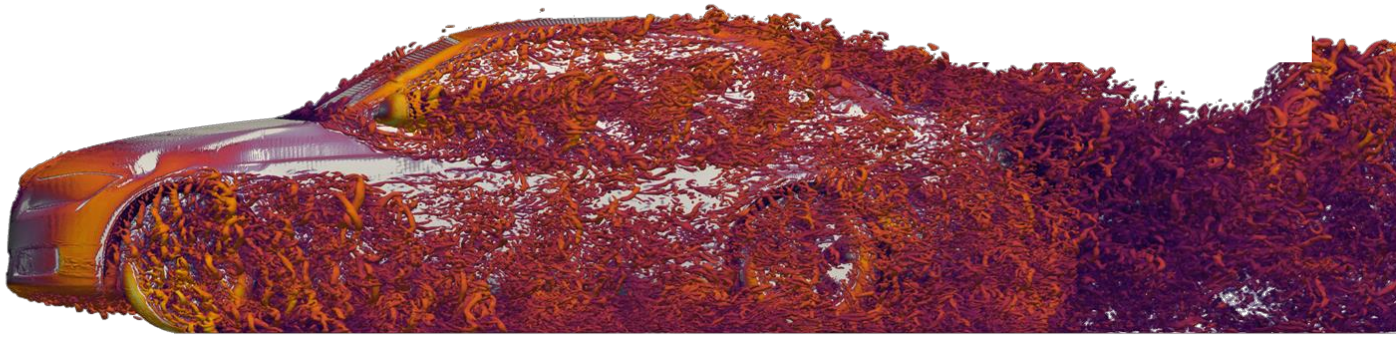
Open source

Developed by OpenCFD Ltd @ESI Group

Written in c++

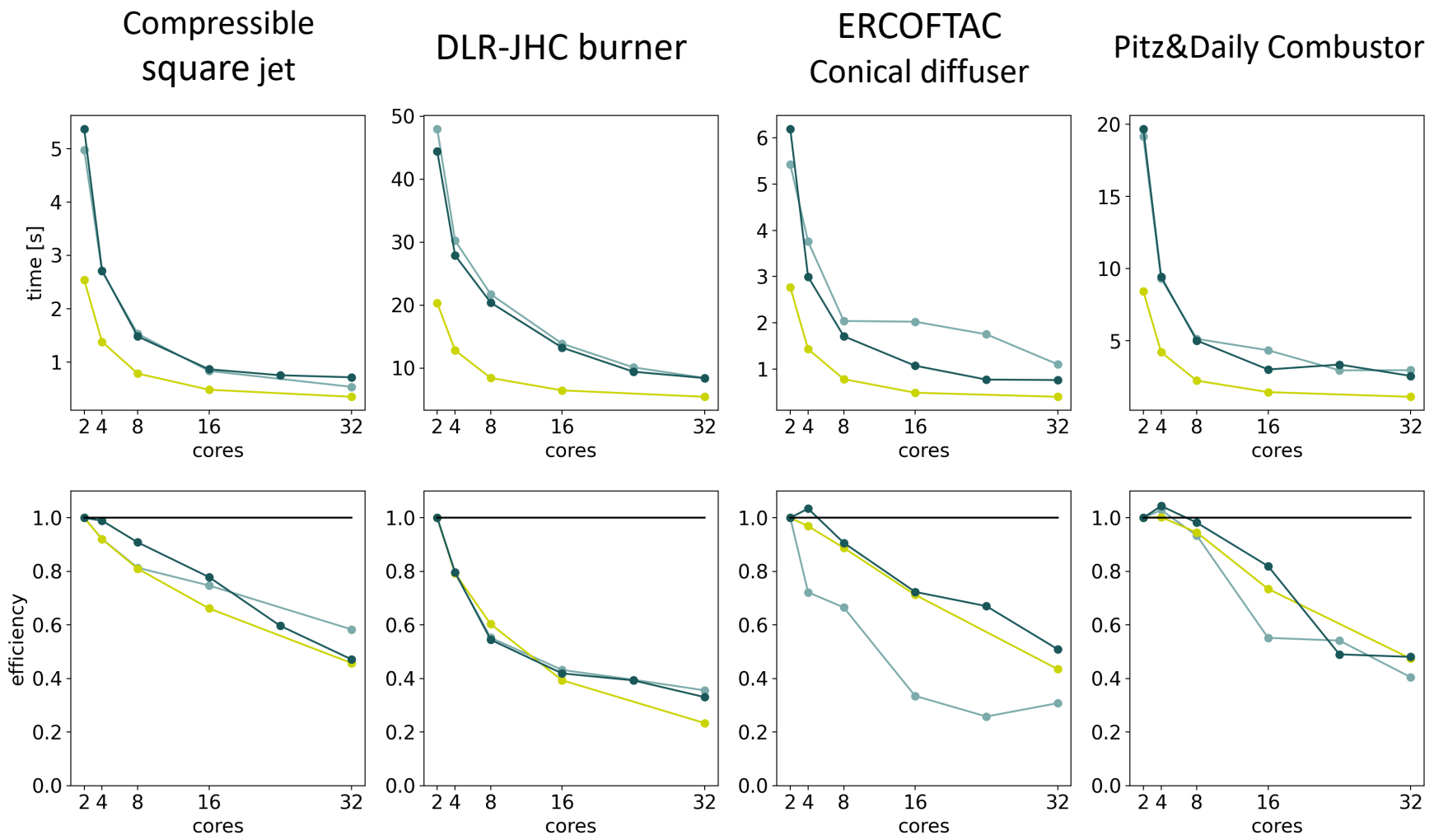
Parallelised with MPI

Some solvers now GPU-ready



upstreamCFD©

GRACE HOPPER HALVES EXECUTION TIME



--INTEL, -- AMPERE ALTRA MAX, -- GRACE HOPPER (CPU only)

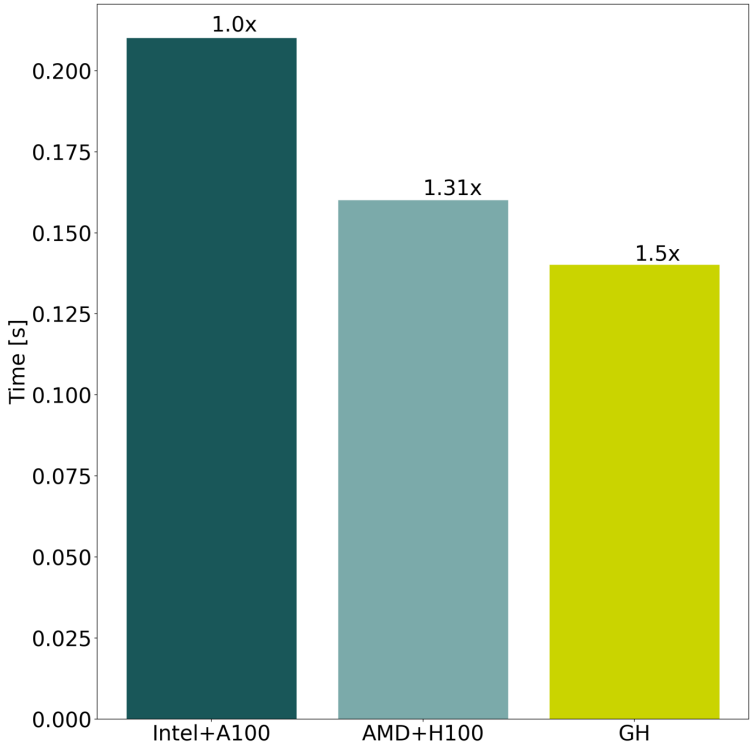
OpenFOAM v2212, gcc 8.5.0 + OpenMPI 4.1.4

~2X SPEEDUP ON GRACE HOPPER 200

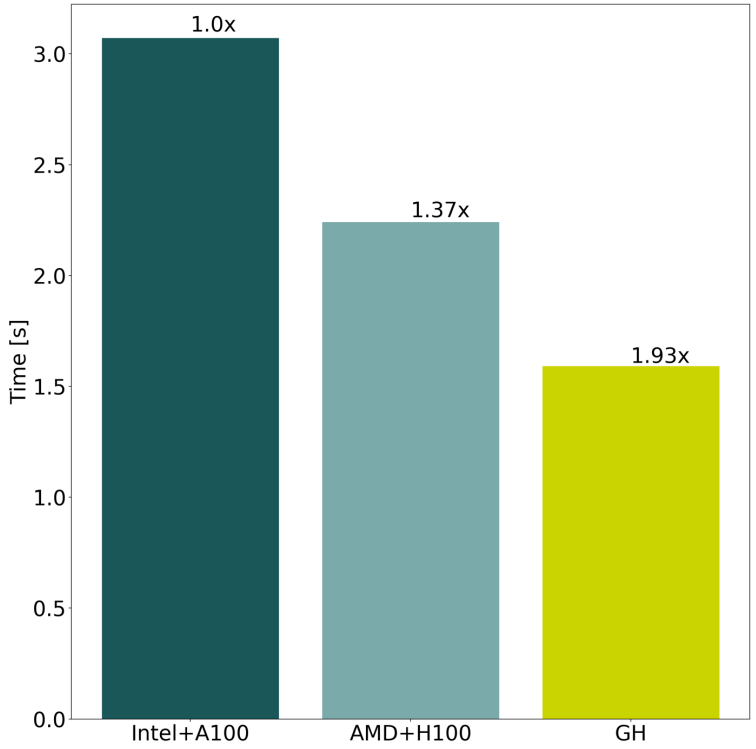
Lid driven cavity flow 3D

OpenFOAM v2306 + zeptoFOAM, NVHPC 23.11 + CUDA 12.3 + OpenMPI 4.1.4

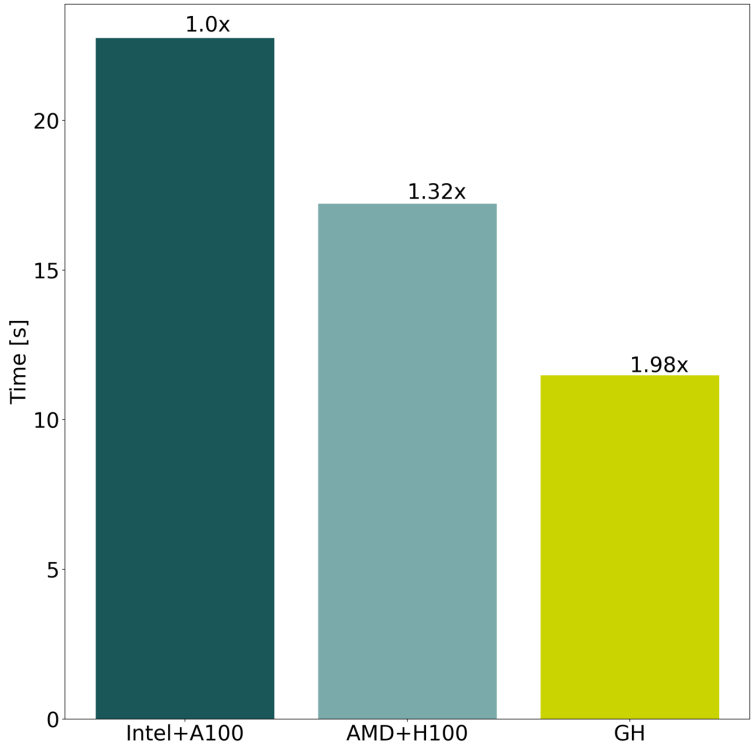
1e6



8e6 cells



27e6



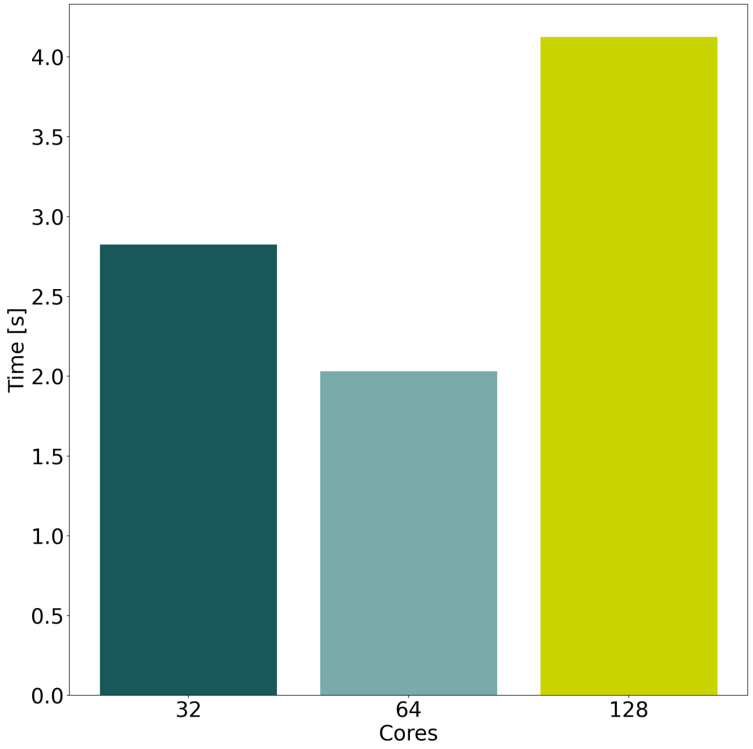
Intel Ice Lake + A100 @CINECA - AMD Genoa + H100 and GH200 @E4 - Tests performed using 1 MPI task + 1 GPU

MEMORY ISSUES HINDER PERFORMANCE ON RISC-V ATM

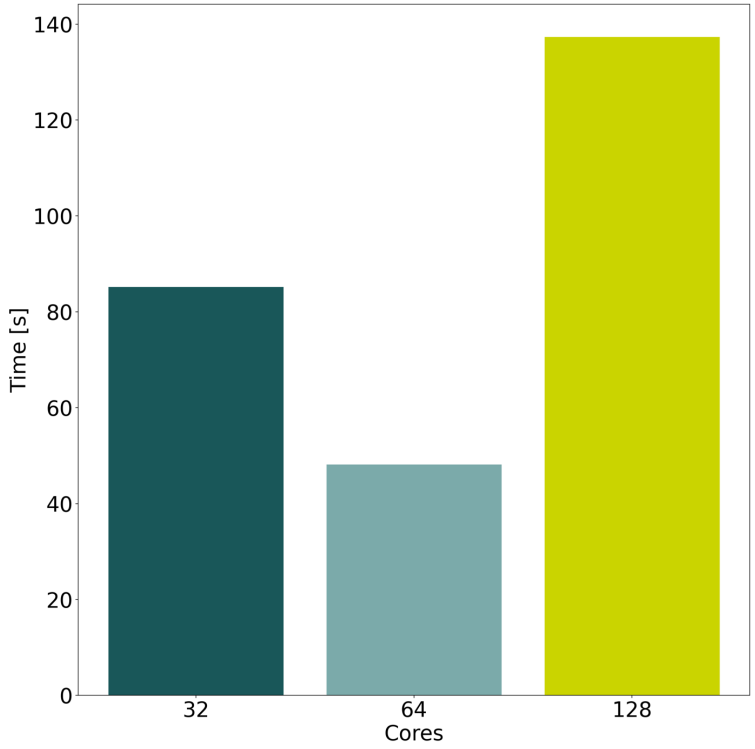
Lid driven cavity flow 3D

OpenFOAM v2312 , gcc 13.2.1 + OpenMPI 4.1.5

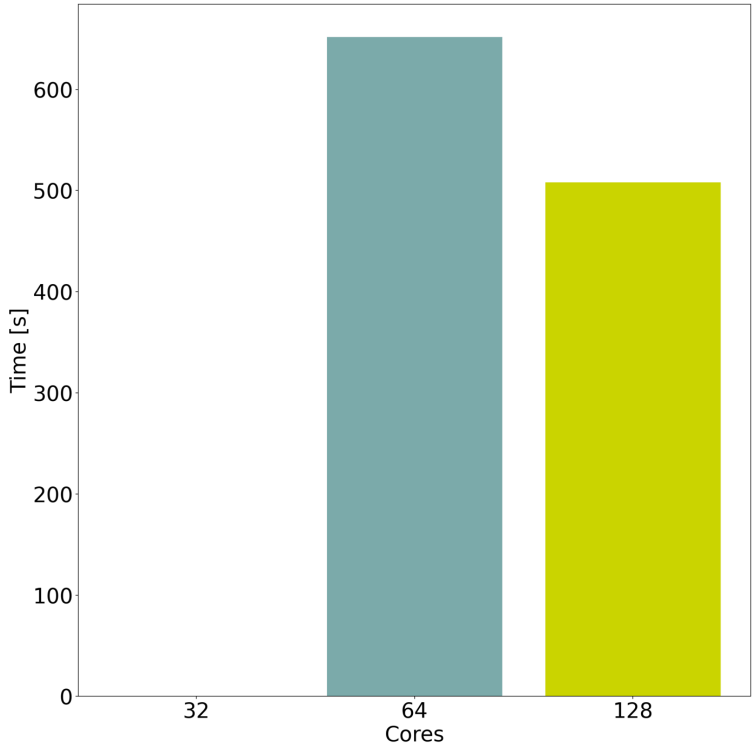
1e6 cells



8e6 cells



64e6 cells



ECSIM: A MASSIVELY PARALLEL PLASMA PHYSICS CODE

Kinetic plasma physics code

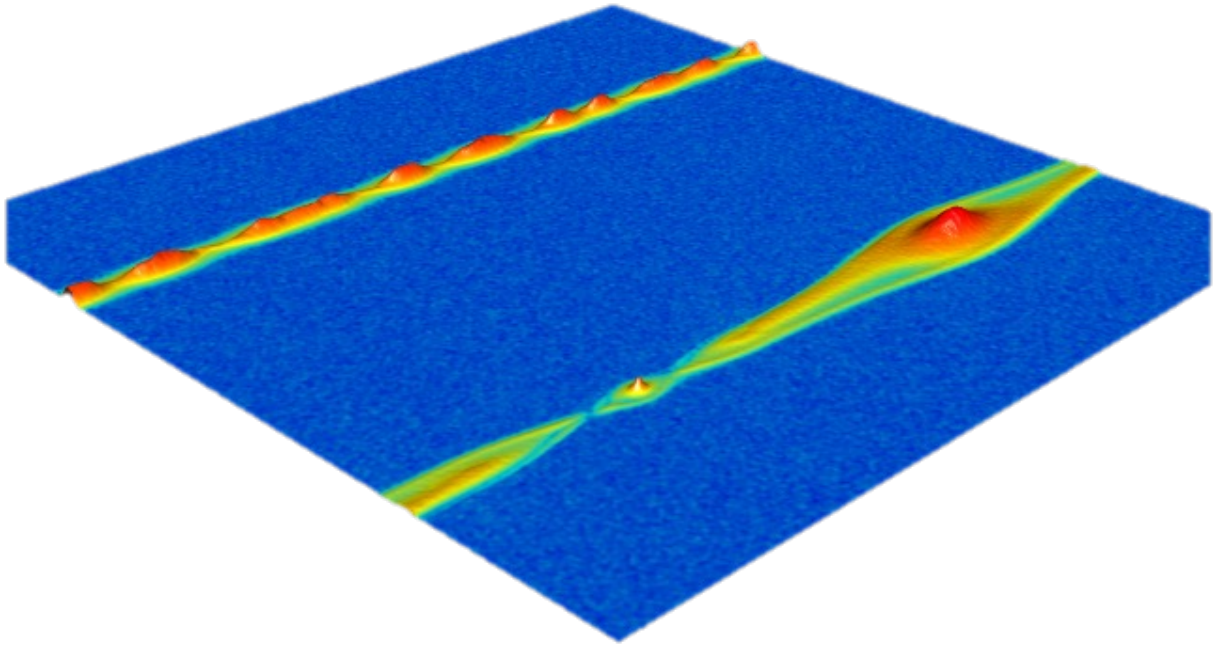
Based on the Particle-In-Cell method

Written in c/c++

Parallelised with MPI

Includes OpenACC directives

Uses PETSc to solve fields

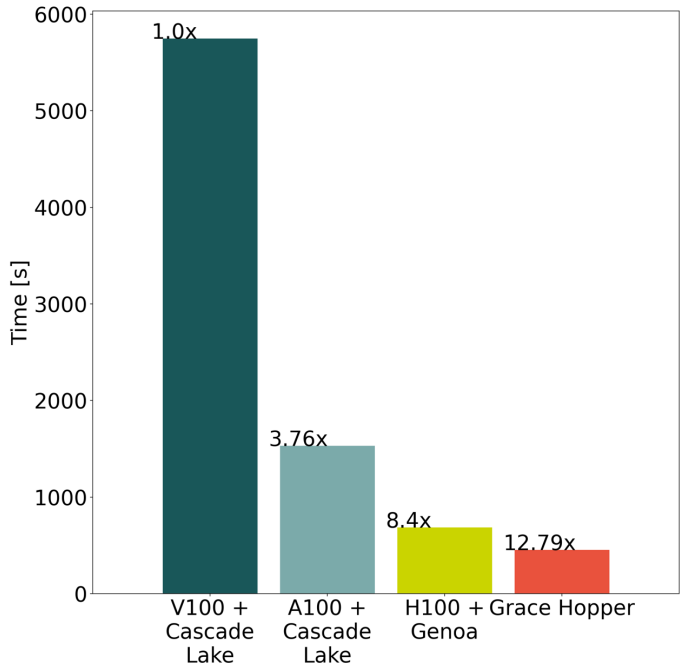


KERNELS WITH DATA MOVEMENTS BENEFITS FROM GH

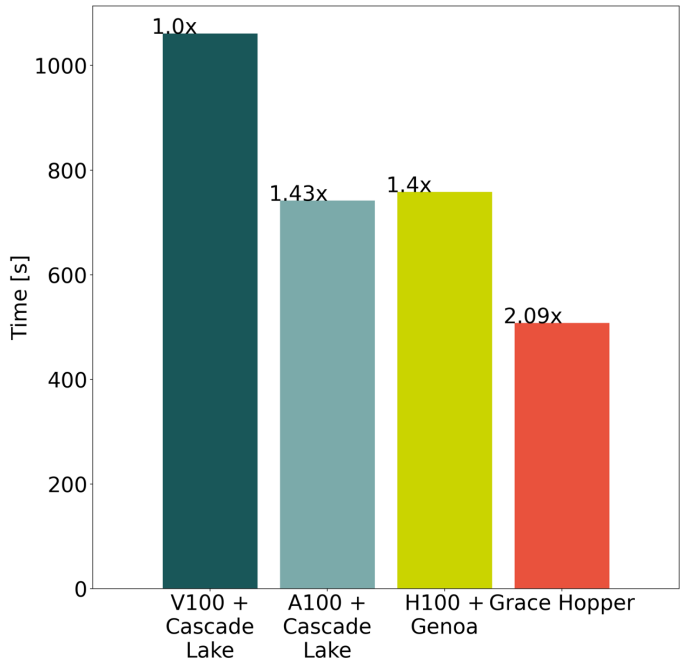
Current Filamentation Instability

NVHPC 23.5 or NVHPC 23.11, OpenMPI 3.1.5

Moment Gathering



Particle Mover

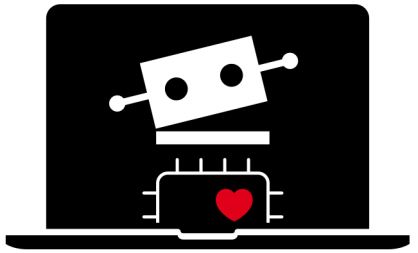


Cascade Lake + V100 @CINECA - Cascade Lake + A100, AMD Genoa + H100 and GH200 @E4

Tests performed using 1 MPI task + 1 GPU

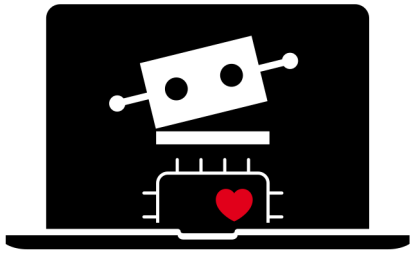
CONTENTS

- E4 computer engineering
- Synthetic tests on emerging architectures
- Performance evaluation of scientific applications on novel architectures
- **Summary and perspectives**



SUMMARY & PERSPECTIVES

- Synthetic tests and real-world applications have been tested on a variety of recent architectures
- Performance of ARM is comparable or better with respect to x86
- GPU applications show a considerable speedup on Grace Hopper
- Maturity of RISC-V for HPC applications is low, but the chip show very promising features



TESTED PLATFORMS @E4 DATA CENTRE

Architecture	CPU Model	Frequency	Cores/node	Memory/node	L3 cache
x86_64	Intel(R) Xeon(R) Gold 6226R (Cascade Lake)	2.9 GHz	32	192 GB	22 MB
x86_64	AMD EPYC 7313 16-Core Processor (Milan)	3.3 GHz	32	256 GB	**128 MB**
aarch64	Ampere Altra Max ARM Neoverse N1	3.0 GHz	256	512 GB	4 MB
aarch64	NVIDIA Grace Hopper Arm Neoverse-V2	3.1 GHz	72	480 GB	117 MB

gcc 8.5.0 + OpenMPI 4.1.4 - OpenFOAM v2212 or OpenFOAM v2306

MICROBENCHMARKS SELECTED

Microbenchmarks	Top-Level Solver	Mesh generation - Cell count - Cell type
MB1 Cavity 3D	icoFoam	blockMesh - 8M - Hexahedra
MB2 Compressible starting square jet	rhoPimpleFoam	blockMesh - 2M - Hexahedra
MB4 DLR-JHC burner	reactingFoam	blockMesh - 400k - Hexahedra
MB5 ERCOFTAC Conical diffuser	simpleFoam	blockMesh - 3M - Hexahedra
MB6 Two cylinders in line	adjointOptimisationFoam	blockMesh - 24500 - Hexahedra
MB8 Rotating Wheel	pimpleFoam	snappyHexMesh - 20M - Polyhedra
MB9 High-lift airfoil	rhoPimpleFoam	snappyHexMesh - 19796480 - Polyhedra
MB11 Pitz&Daily Combustor	XiFoam	blockMesh - 200k - Hexahedra
MB12 Model Wind Farm	pimpleFoam	blockMesh - 8M - Hexahedra
MB17 1D Aeroacoustic Wave Train	rhoPimpleFoam	1D blockMesh - 0.05M - Hexahedra
MB19 Viscoelastic polymer melt flow	viscoelasticFluidFoam	cfMesh - 1M -Polyhedra