SPACE

CENTRE OF EXCELLENCE FOR HPC
ASTROPHYSICAL APPLICATIONS

USC-VIII meeting, Oct 14-18th @ Galzignano (PD)

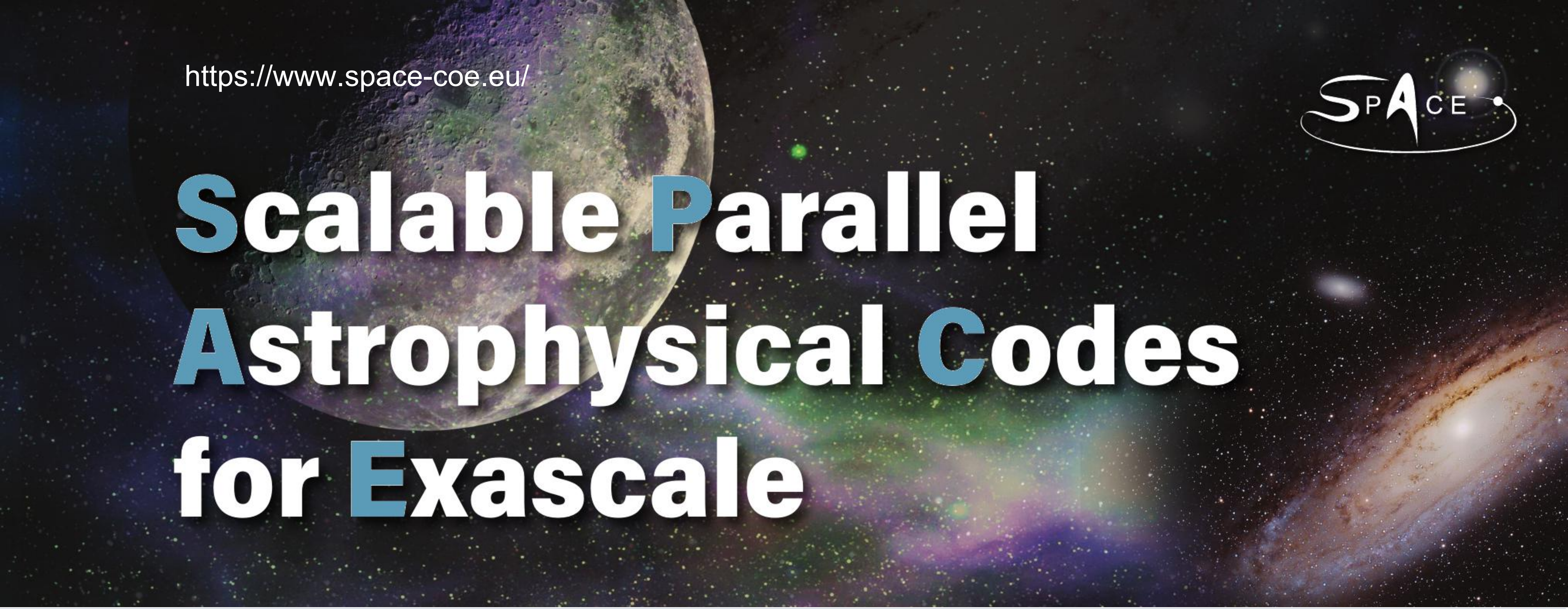
SPACE CoE

for the SPACE Team - OATs, OACt, IRA

Co-funded by
the European Union

Funded by the European Union. This work has received funding from the European High Performance Computing Joint Undertaking (JU) and Belgium, Czech Republic, France, Germany, Greece, Italy, Norway, and Spain under grant agreement No 101093441.

EuroHPC
Joint Undertaking



https://www.space-coe.eu/

SPACE

Scalable Parallel Astrophysical Codes for Exascale

1st January 2023 → **31st December 2027**
we started slightly late due to issues with national fundings

15 partners from 8 countries

The SPACE Partners

SPACE

UNIVERSITÀ DI TORINO

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

UNIVERSITY OF OSLO

LMU
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

cnrs

CRAL
CENTRE DE RECHERCHE ASTROPHYSIQUE DE LYON

KU LEUVEN

HITS
Heidelberger Institut für Theoretische Studien

GOETHE UNIVERSITÄT
FRANKFURT AM MAIN

FORTH
FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS

9
Research Institutes
from 6 countries

3
Computing Centers
from 3 countries

CINECA

IT4I

BSC
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

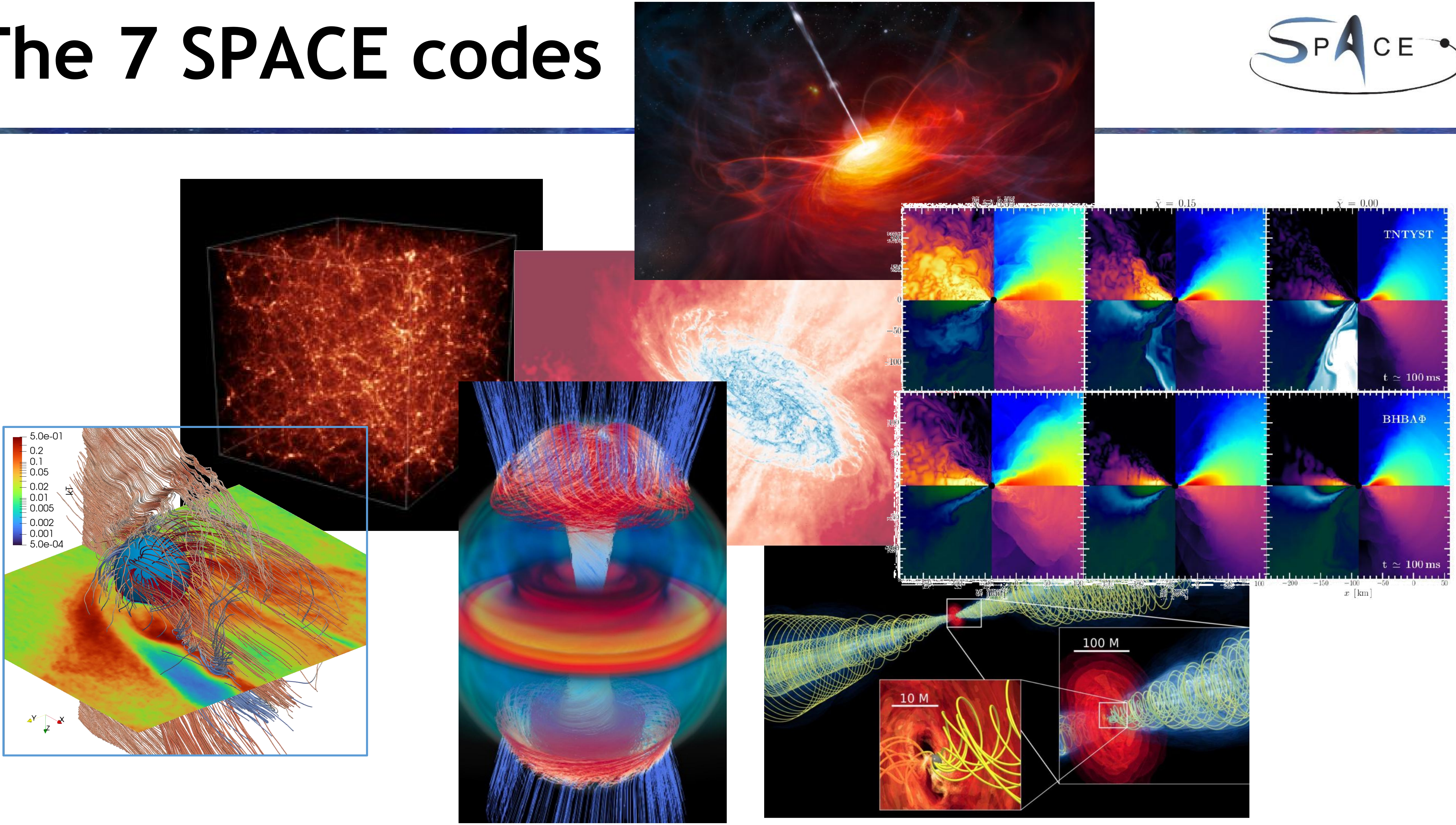
3
HPC Companies

E4
COMPUTER ENGINEERING

ENGINSOFT

Atos

The 7 SPACE codes



The 7 SPACE codes

Large-scale Cosmology and Astrophysics:

OpenGADGET 🟨 **ChaNGa** 🟨

RAMSES 🟩

Fully-Relativistic small-scale Astr & MHD:

PLUTO 🟨 🟩 **BHAC** 🟩 **Frankfurt/Illinois FIL** 🟩

Particle-in-Cell multi-scale plasma

iPic3d 🟨 🟩

🟨 **Particle**-based

🟩 **Grid**-based

The SPACE rationale

MeerKAT

LOFAR

SKA
SQUARE KILOMETRE ARRAY

Euclid
consortium

MWA
MURCHISON
WIDEFIELD
ARRAY

Precision Cosmology and forthcoming data torrent: **outstanding quality and volume of data**

exceptional challenges to their theoretical interpretation

e.g. 8 - 9 orders of magnitude in dynamic range with very different physical processes at different scales

The SPACE rationale

Precision Cosmology and forthcoming data torrent: **outstanding quality and volume of data**

exceptional challenges to their theoretical interpretation

e.g. 8 - 9 orders of magnitude in dynamic range with very different physical processes at different scales

new numerical exa-scale capable laboratories
(codes, algorithms and tools)
+
high-performance
analysis and **visualization**
for **extreme data**

innovative programming paradigms and sw solutions

The SPACE rationale

Nowadays, a limited number of numerical applications, several of which are developed and maintained in Europe, represent the state-of-the-art in A&C simulations.

However, although they are fully-productive codes used to produce cutting-edge simulations, they also require a substantial effort to evolve their computational paradigms from the petascale to the exascale era.

The main SPACE objectives

[1] **evolve 7 European codes to the exascale paradigms**

[4]
evolve **data analysis** and **visualization**

[3]
to develop **ML techniques** for post-processing and (possibly) on-line coupling

[2]
to address the **Energy Efficiency**

The main SPACE objectives

[1] evolve **7**
European codes
to the exascale
paradigms

➕

[4]
evolve **data analysis**
and **visualization**

[3]
to develop **ML techniques**
for post-processing and
(possibly) on-line coupling

[2]
to address the
Energy Efficiency

[1] evolving the codes

The codes are instrumented and prepared for periodical performance assessments

- one region for the main loop and 3-5 inner regions
- unique id for each region derived from time-step
- performance of each region was assessed using a set of metrics defined in the POP methodology
- regions have been eventually turned into kernels

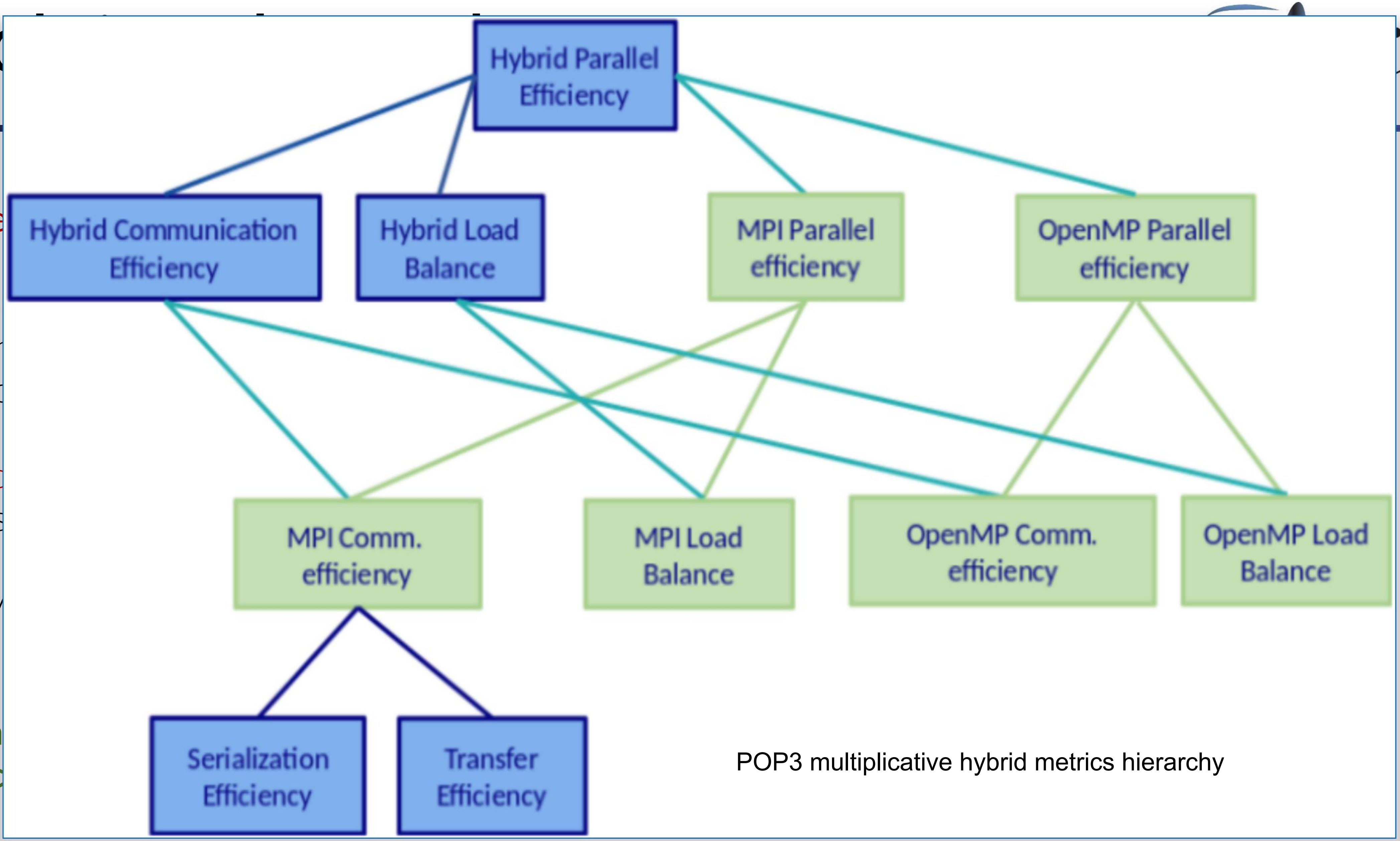
Collaboration with POP3
https://pop-coe.eu/

[1] ev

The codes are
for periodical

- one region
  inner region
- unique id f
  time-step
- performanc
  assessed us
  in the POP
- regions hav
  kernels

Collaboration
https://pop-c



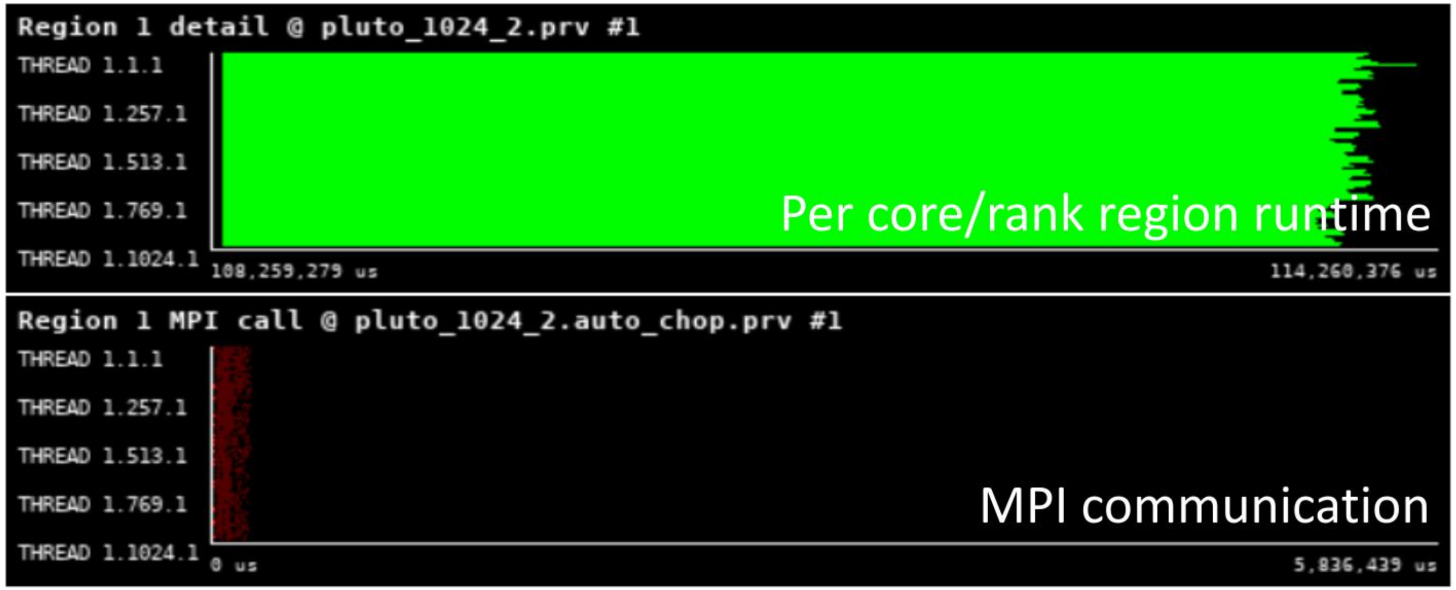
POP3 multiplicative hybrid metrics hierarchy

[1] evolving the codes

For **entire code** and for **each annotated region**, we track

- Load balance
- Communication serialization
- Transfer efficiency
- Computation scaling
- …

Using EXTRAE (BSC)



Per core/rank region runtime

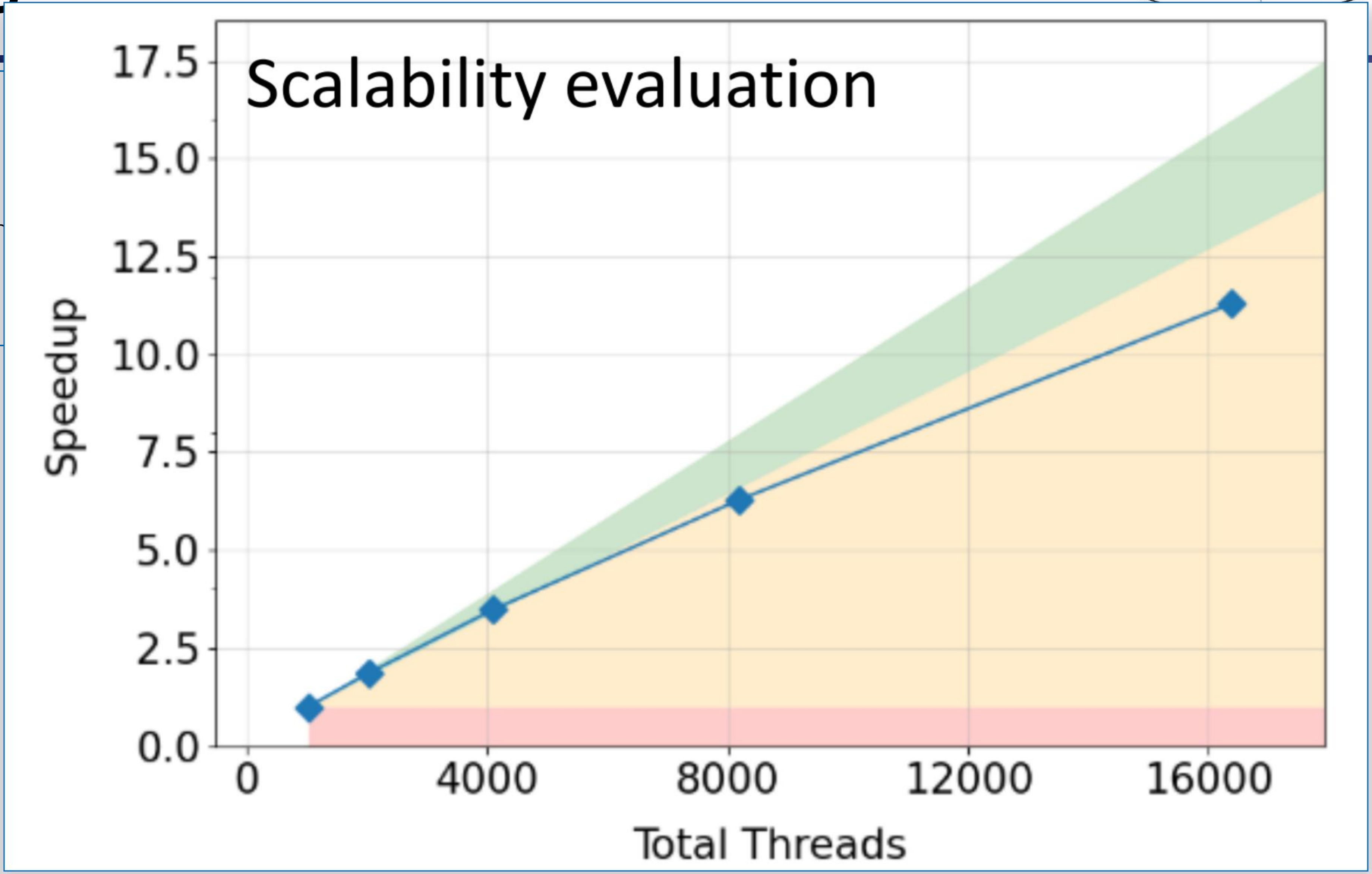
MPI communication

13

[1] evolving the codes

For **entire code** and for **each annotated region**, we track

- Load balance
- Communication serializatio
- Transfer efficiency
- Computation scaling



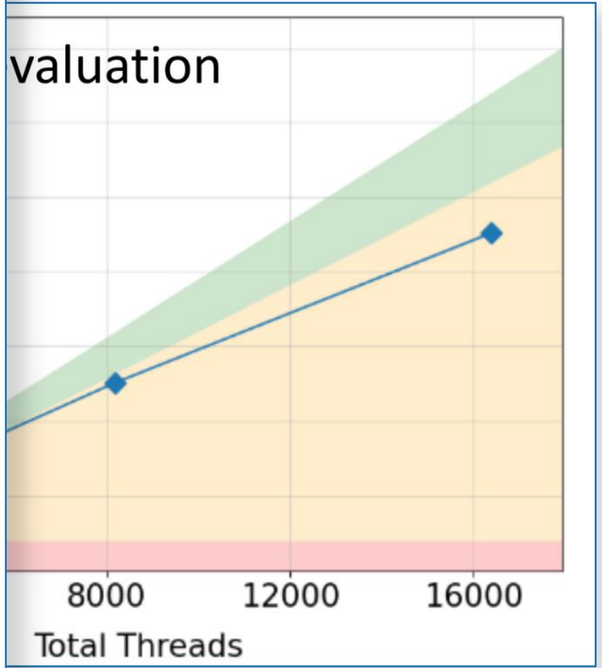
[1] evolving the codes

For **entire code** and for **each**

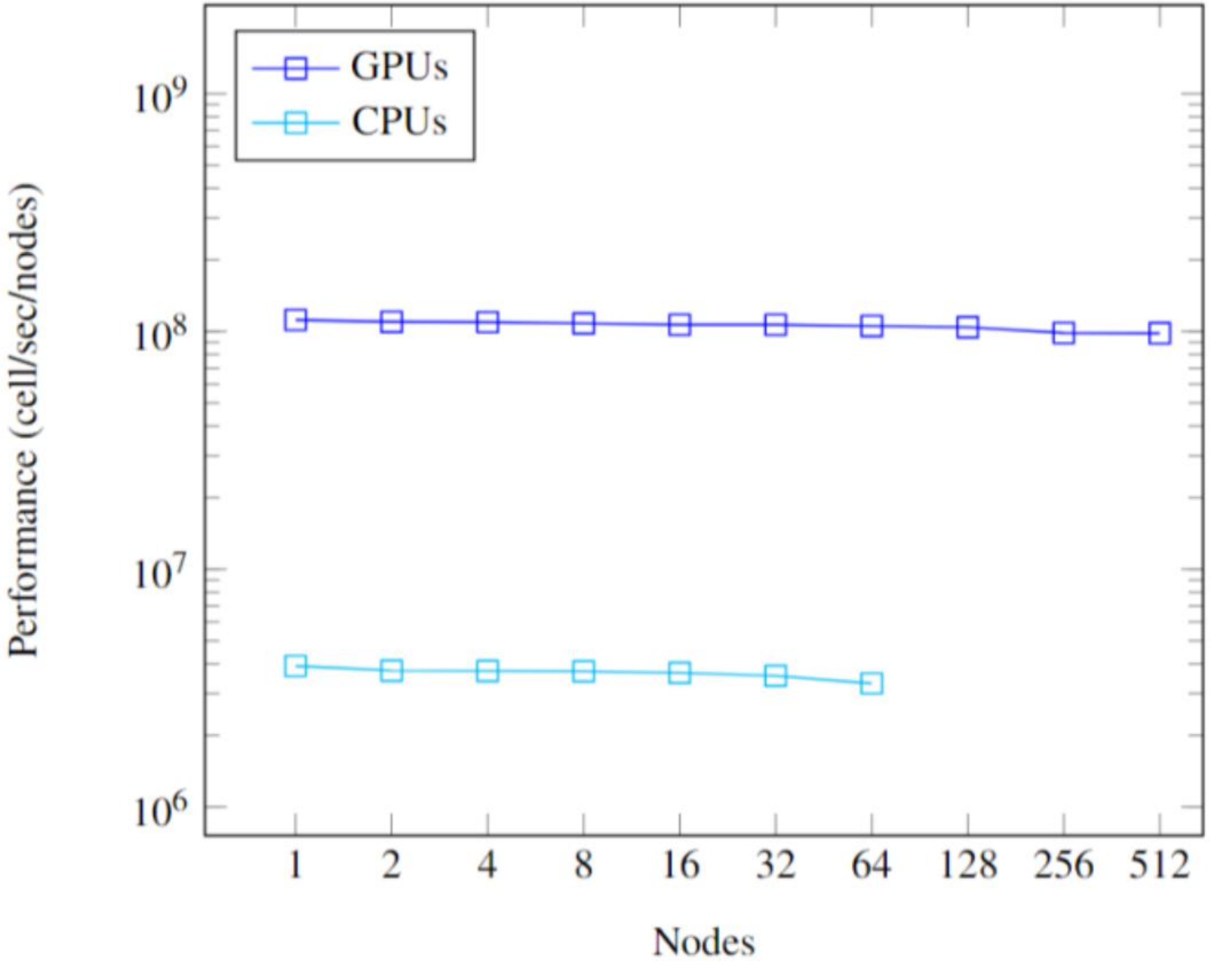
a

	1024	2048	4096	8192	16384
Global efficiency	93.15	87.34	81.06	73.70	66.41
-- Parallel efficiency	93.15	88.25	83.22	77.52	72.72
-- Load balance	94.89	91.77	86.25	81.94	78.25
-- Communication efficiency	98.17	96.16	96.48	94.60	92.94
-- Serialization efficiency	99.87	99.95	99.90	99.88	99.47
-- Transfer efficiency	98.30	96.21	96.58	94.71	93.43
-- Computation scalability	100.00	98.97	97.41	95.08	91.33
-- IPC scalability	100.00	100.94	100.83	102.13	101.37
-- Instruction scalability	100.00	98.06	96.16	92.36	88.89
-- Frequency scalability	100.00	99.98	100.47	100.80	101.36

POP metrics table per region



[1] evolving the codes : results



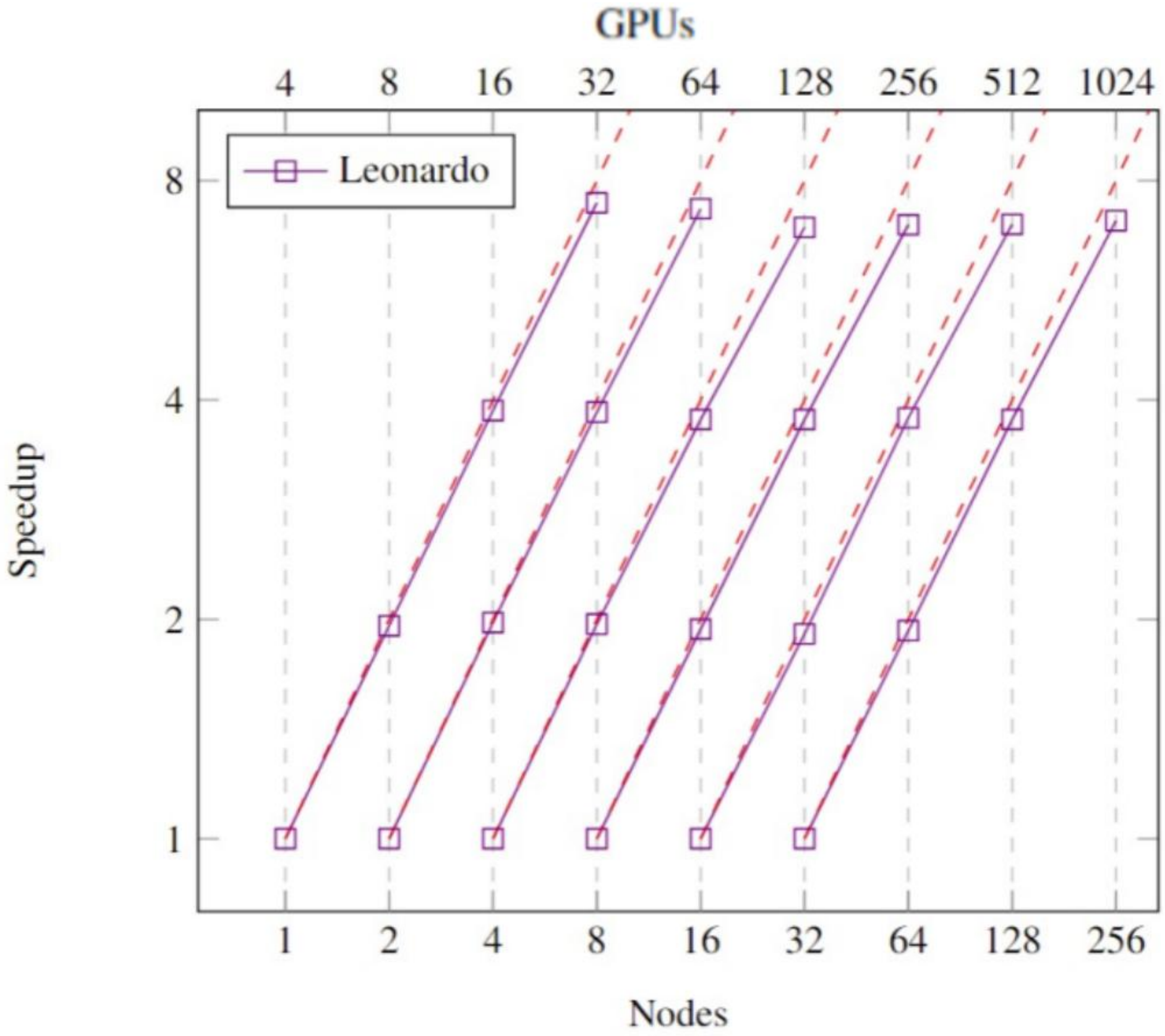
PLUTO WEAK SCALING

- very good initial state (they started 2 years before from scratch with a mini-app)

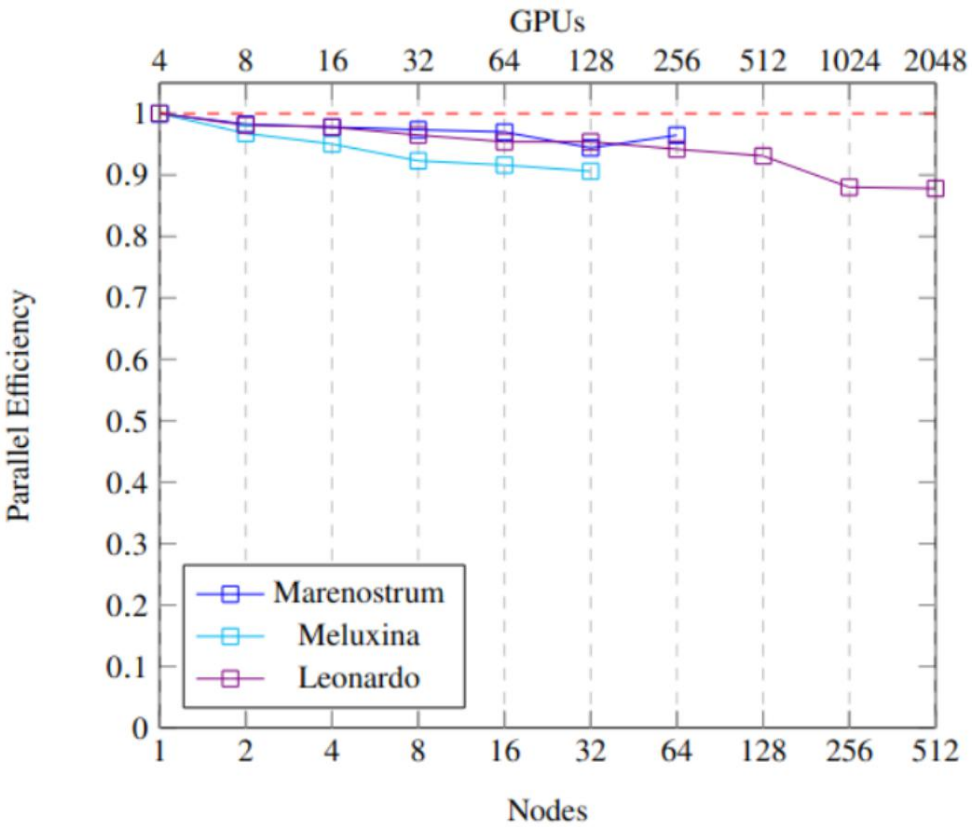
- however, the parallel efficiency dropped down to 60%
the problem was mainly in the communication pattern

- an additional 2.5x speedup with gpu with respect to the previous implementation

[1] evolving the codes : results



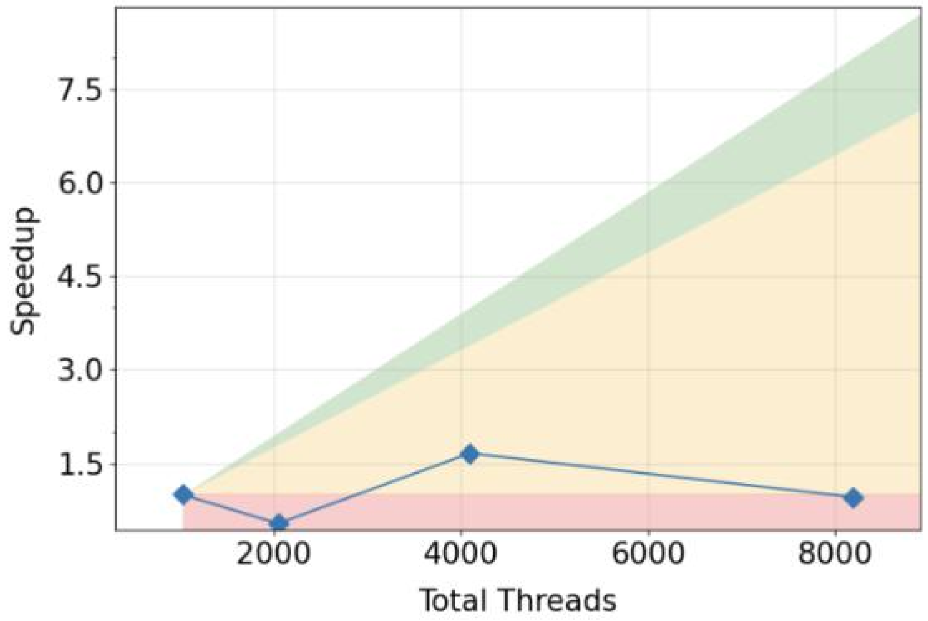
PLUTO STRONG SCALING on GPU



[1] evolving the codes : res

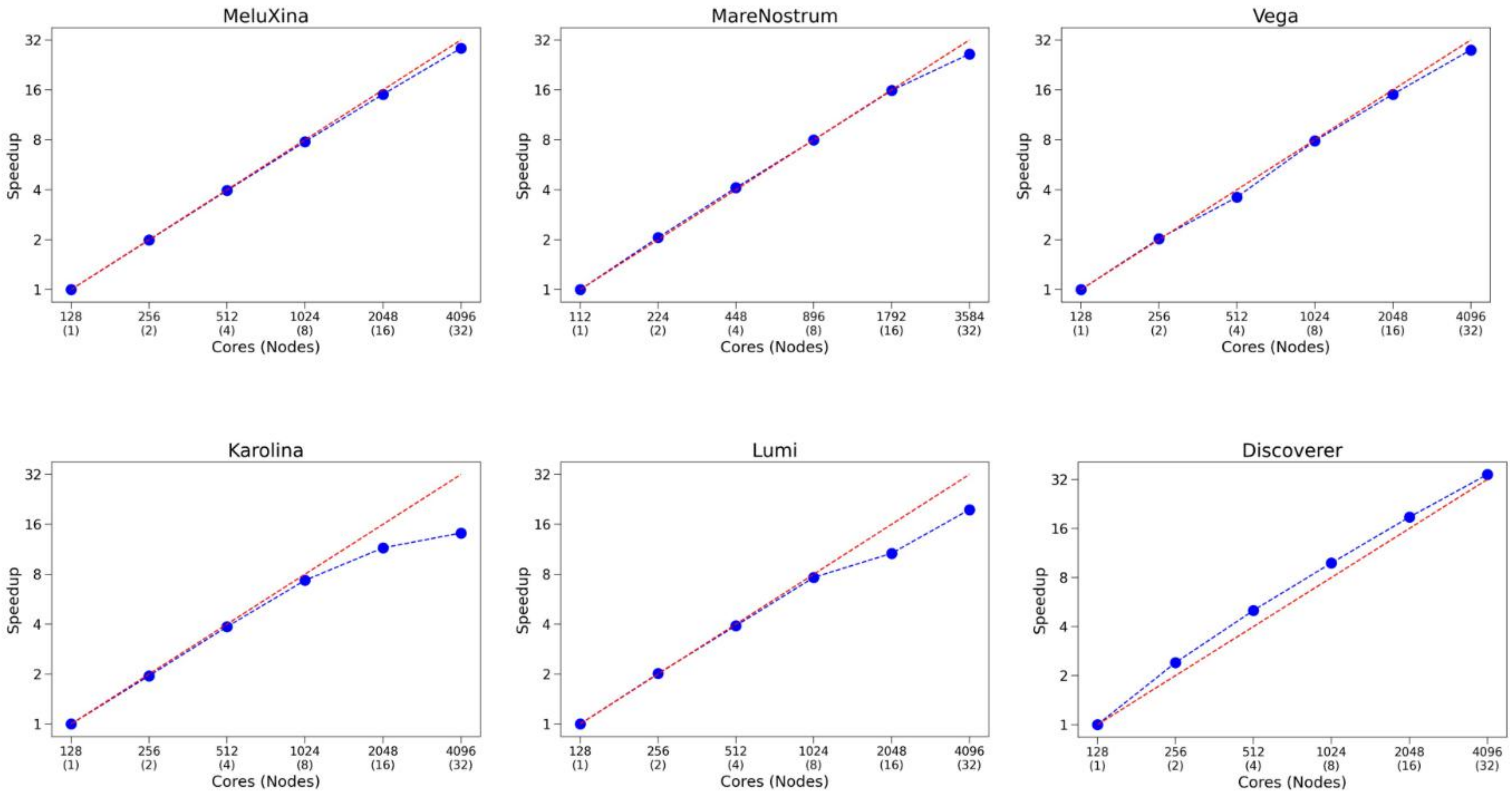
IPic3D original preformance

Strong scaling and POP efficiency metrics for Region 1 of iPic3D.

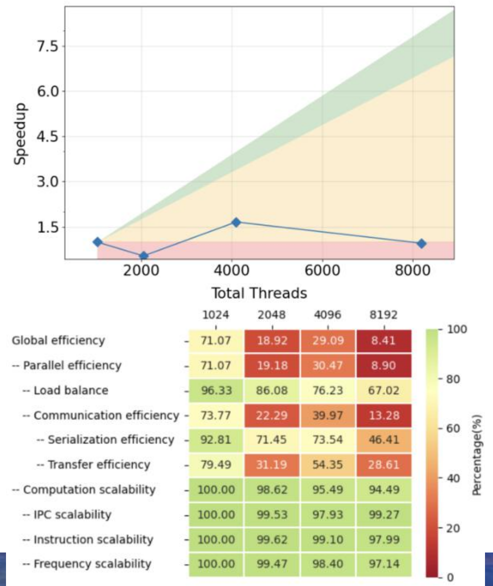


	1024	2048	4096	8192
Global efficiency	71.07	18.92	29.09	8.41
-- Parallel efficiency	71.07	19.18	30.47	8.90
-- Load balance	96.33	86.08	76.23	67.02
-- Communication efficiency	73.77	22.29	39.97	13.28
-- Serialization efficiency	92.81	71.45	73.54	46.41
-- Transfer efficiency	79.49	31.19	54.35	28.61
-- Computation scalability	100.00	98.62	95.49	94.49
-- IPC scalability	100.00	99.53	97.93	99.27
-- Instruction scalability	100.00	99.62	99.10	97.99
-- Frequency scalability	100.00	99.47	98.40	97.14

[1] evolving the codes : results



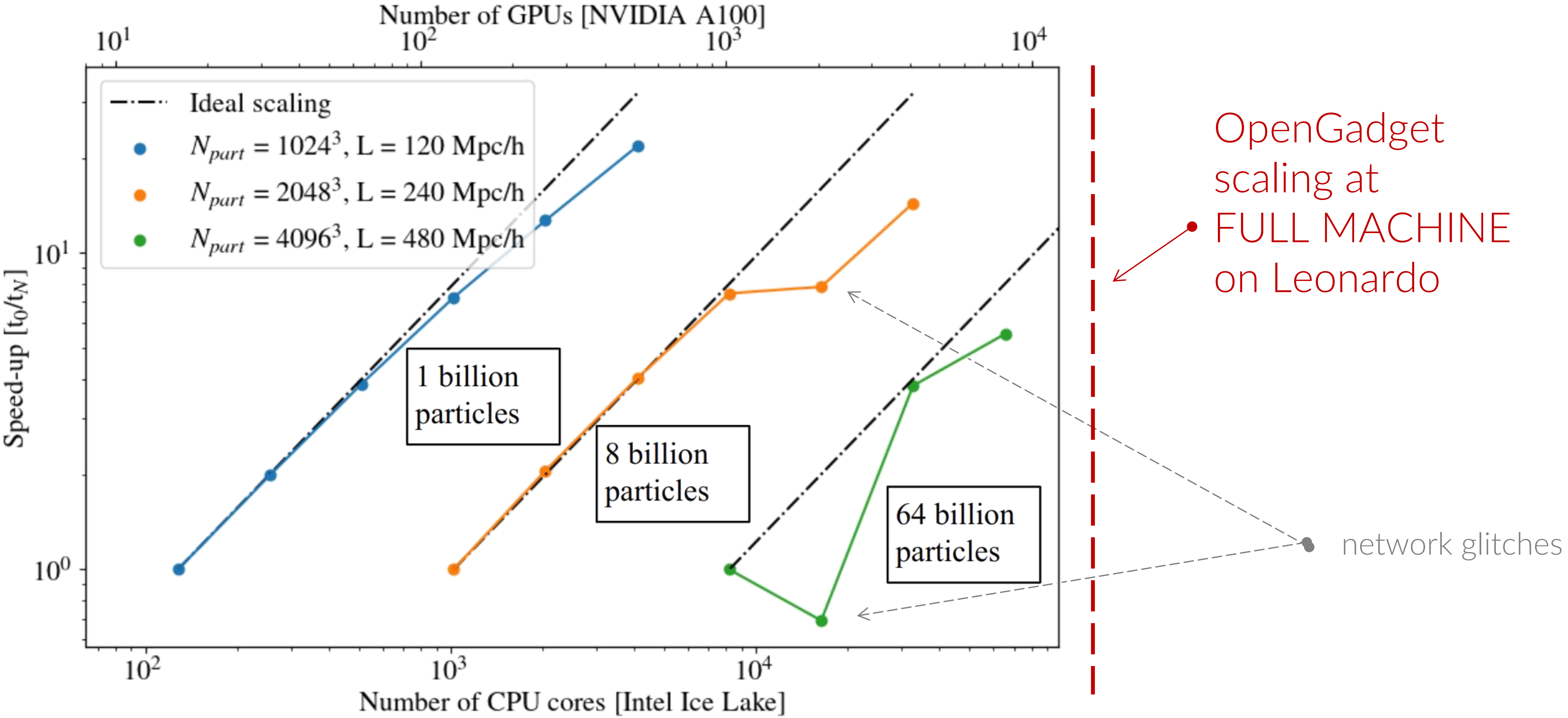
Strong scaling and POP efficiency metrics for Region 1 of iPic3D.



	1024	2048	4096	8192
Global efficiency	71.07	18.92	29.09	8.41
-- Parallel efficiency	71.07	19.18	30.47	8.90
-- Load balance	96.33	86.08	76.23	67.02
-- Communication efficiency	73.77	22.29	39.97	13.28
-- Serialization efficiency	92.81	71.45	73.54	46.41
-- Transfer efficiency	79.49	31.19	54.35	28.61
-- Computation scalability	100.00	98.62	95.49	94.49
-- IPC scalability	100.00	99.53	97.93	99.27
-- Instruction scalability	100.00	99.62	99.10	97.99
-- Frequency scalability	100.00	99.47	98.40	97.14

current scaling on JU systems

[1] evolving the codes : results



OpenGadget scaling at FULL MACHINE on Leonardo

[2] Energy efficiency : tools

H2020 READEX (2015-2018): Complex parallel application has different requirements during execution, so it gives a possibility to be dynamically tuned for energy savings without performance penalty.

READEX
Runtime Exploitation of Application Dynamism
for Energy-efficient eXascale computing

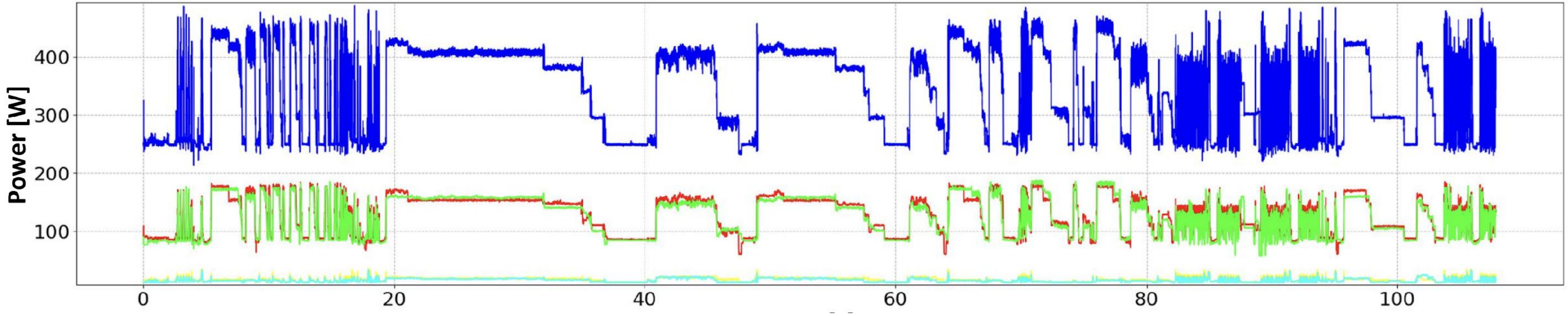
MERIC runtime system provides **dynamic application tuning**

- lightweight & easy to install & easy to use
- C/C++ API and Fortran module
- MPI, OpenMP and CUDA parallelization
- performance and power-aware
- support for a wide range of architectures and power monitoring systems

21

[2] Energy efficiency : tools

power consumption timeline on a single node (OpenGadget) and its component

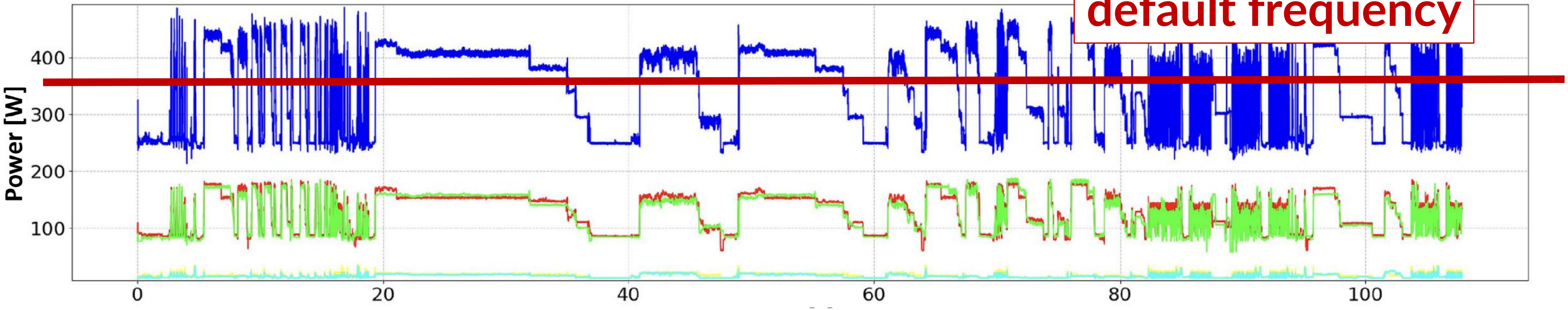


HDEEM power monitoring system on Barbora@IT4I

power samples timeline from - sampling rate is 1kHz

[2] Energy efficiency : tools

power consumption timeline on a single node (OpenGadget) and its component



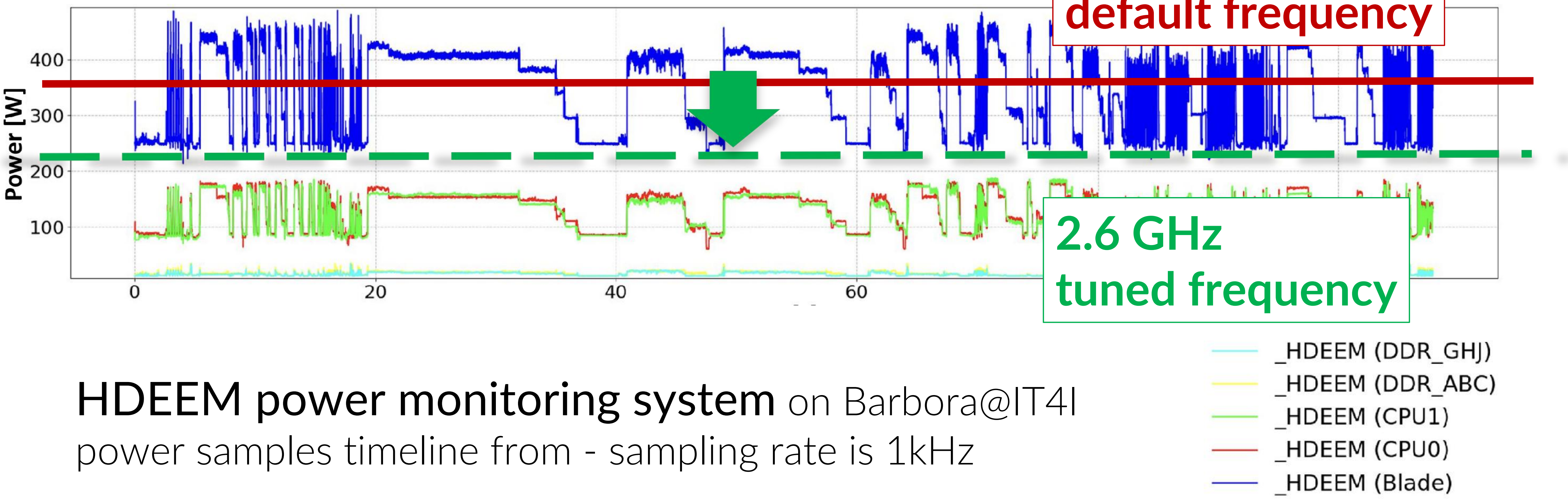
3.0 GHz
default frequency

HDEEM power monitoring system on Barbora@IT4I
power samples timeline from - sampling rate is 1kHz

23

[2] Energy efficiency : tuning

power consumption timeline on a single node (OpenGadget) and its component



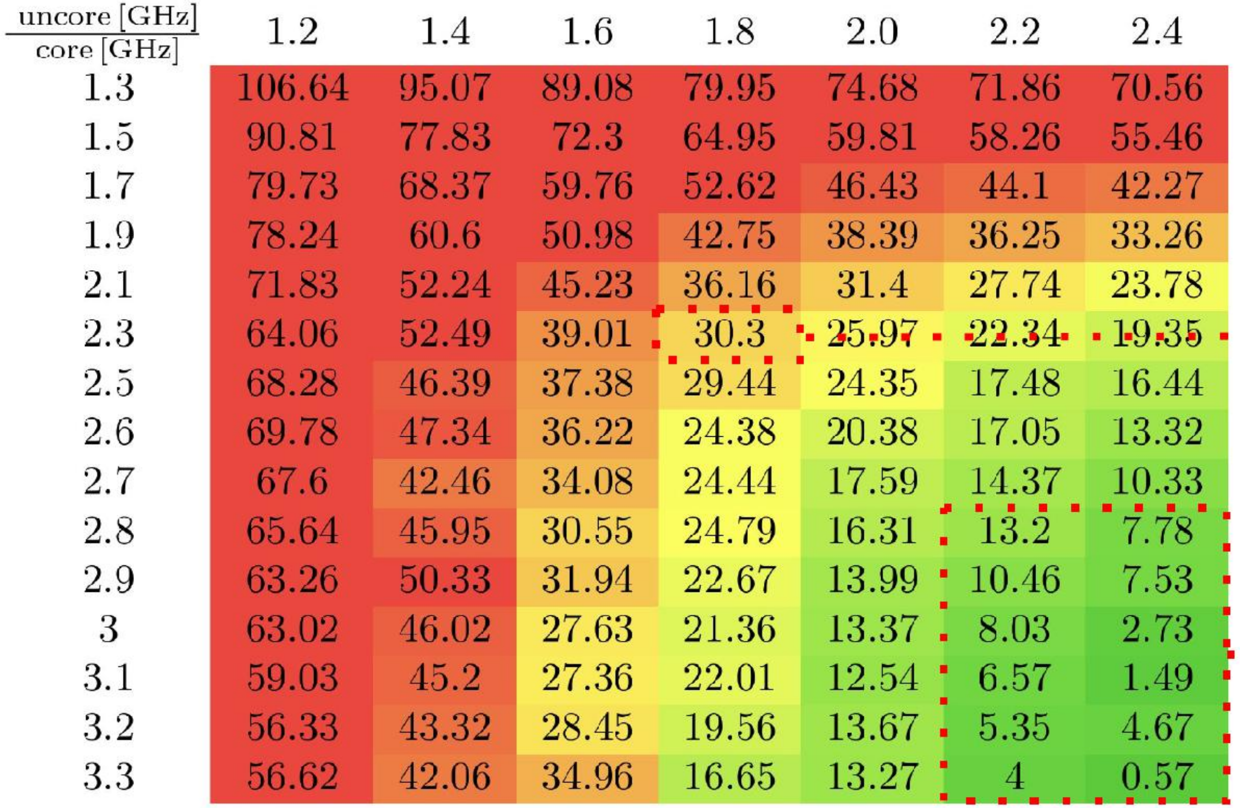
3.0 GHz
default frequency

2.6 GHz
tuned frequency

HDEEM power monitoring system on Barbora@IT4I
power samples timeline from - sampling rate is 1kHz

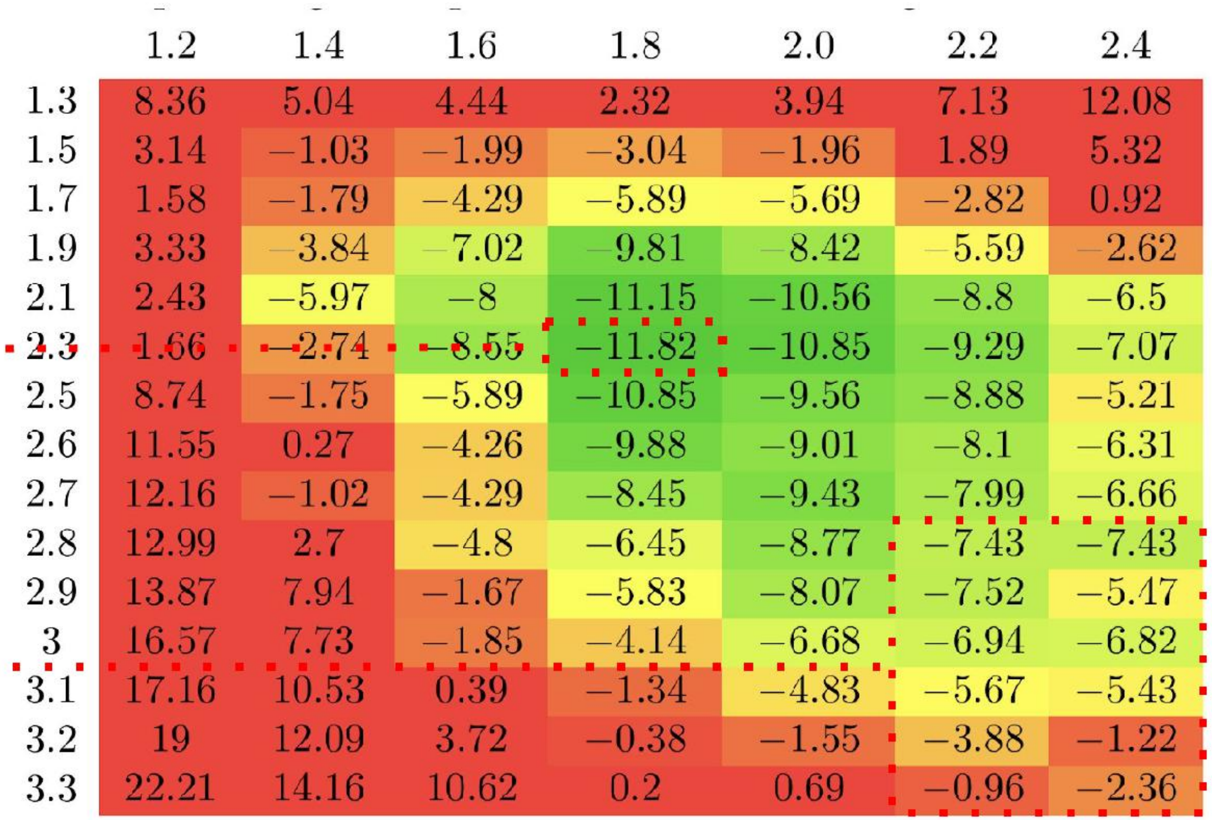
24

[2] Energy efficiency : tuning



uncore [GHz] / core [GHz]	1.2	1.4	1.6	1.8	2.0	2.2	2.4
1.3	106.64	95.07	89.08	79.95	74.68	71.86	70.56
1.5	90.81	77.83	72.3	64.95	59.81	58.26	55.46
1.7	79.73	68.37	59.76	52.62	46.43	44.1	42.27
1.9	78.24	60.6	50.98	42.75	38.39	36.25	33.26
2.1	71.83	52.24	45.23	36.16	31.4	27.74	23.78
2.3	64.06	52.49	39.01	30.3	25.97	22.34	19.35
2.5	68.28	46.39	37.38	29.44	24.35	17.48	16.44
2.6	69.78	47.34	36.22	24.38	20.38	17.05	13.32
2.7	67.6	42.46	34.08	24.44	17.59	14.37	10.33
2.8	65.64	45.95	30.55	24.79	16.31	13.2	7.78
2.9	63.26	50.33	31.94	22.67	13.99	10.46	7.53
3	63.02	46.02	27.63	21.36	13.37	8.03	2.73
3.1	59.03	45.2	27.36	22.01	12.54	6.57	1.49
3.2	56.33	43.32	28.45	19.56	13.67	5.35	4.67
3.3	56.62	42.06	34.96	16.65	13.27	4	0.57

Runtime extension [%]

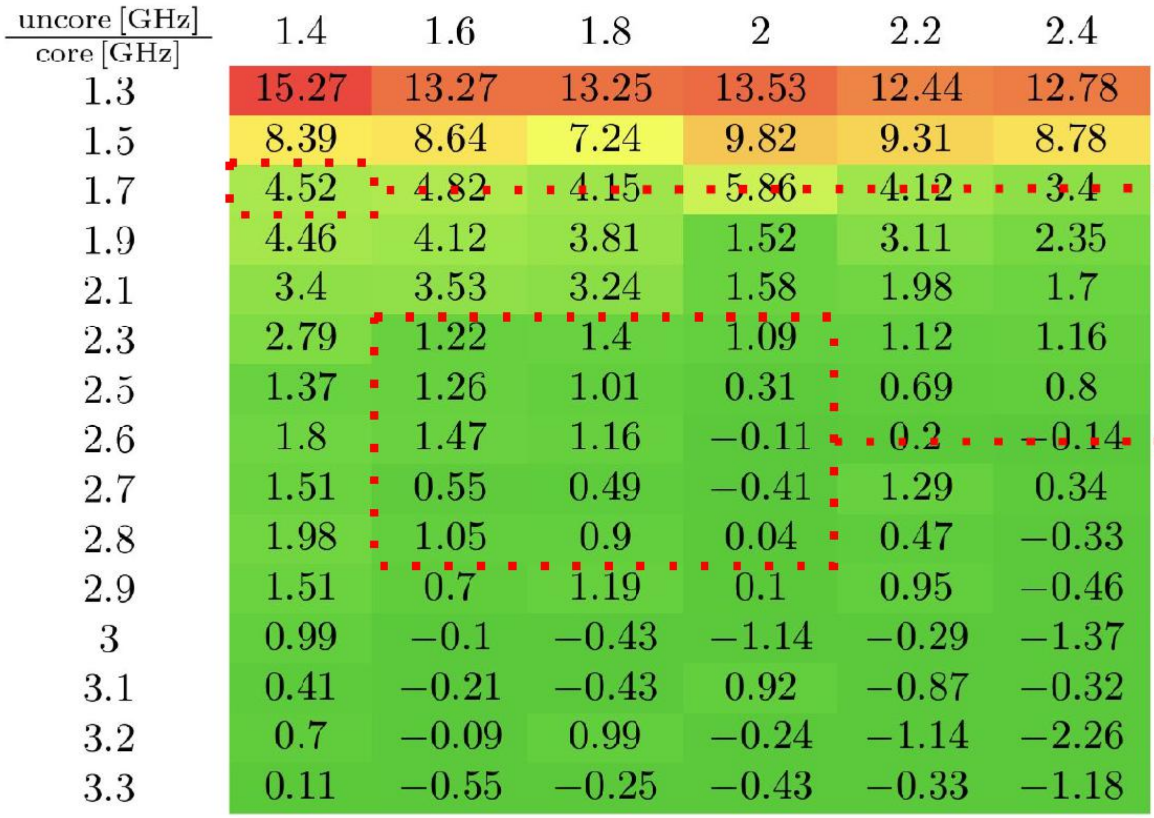


	1.2	1.4	1.6	1.8	2.0	2.2	2.4
1.3	8.36	5.04	4.44	2.32	3.94	7.13	12.08
1.5	3.14	−1.03	−1.99	−3.04	−1.96	1.89	5.32
1.7	1.58	−1.79	−4.29	−5.89	−5.69	−2.82	0.92
1.9	3.33	−3.84	−7.02	−9.81	−8.42	−5.59	−2.62
2.1	2.43	−5.97	−8	−11.15	−10.56	−8.8	−6.5
2.3	1.66	−2.74	−8.55	−11.82	−10.85	−9.29	−7.07
2.5	8.74	−1.75	−5.89	−10.85	−9.56	−8.88	−5.21
2.6	11.55	0.27	−4.26	−9.88	−9.01	−8.1	−6.31
2.7	12.16	−1.02	−4.29	−8.45	−9.43	−7.99	−6.66
2.8	12.99	2.7	−4.8	−6.45	−8.77	−7.43	−7.43
2.9	13.87	7.94	−1.67	−5.83	−8.07	−7.52	−5.47
3	16.57	7.73	−1.85	−4.14	−6.68	−6.94	−6.82
3.1	17.16	10.53	0.39	−1.34	−4.83	−5.67	−5.43
3.2	19	12.09	3.72	−0.38	−1.55	−3.88	−1.22
3.3	22.21	14.16	10.62	0.2	0.69	−0.96	−2.36

Energy savings [%]

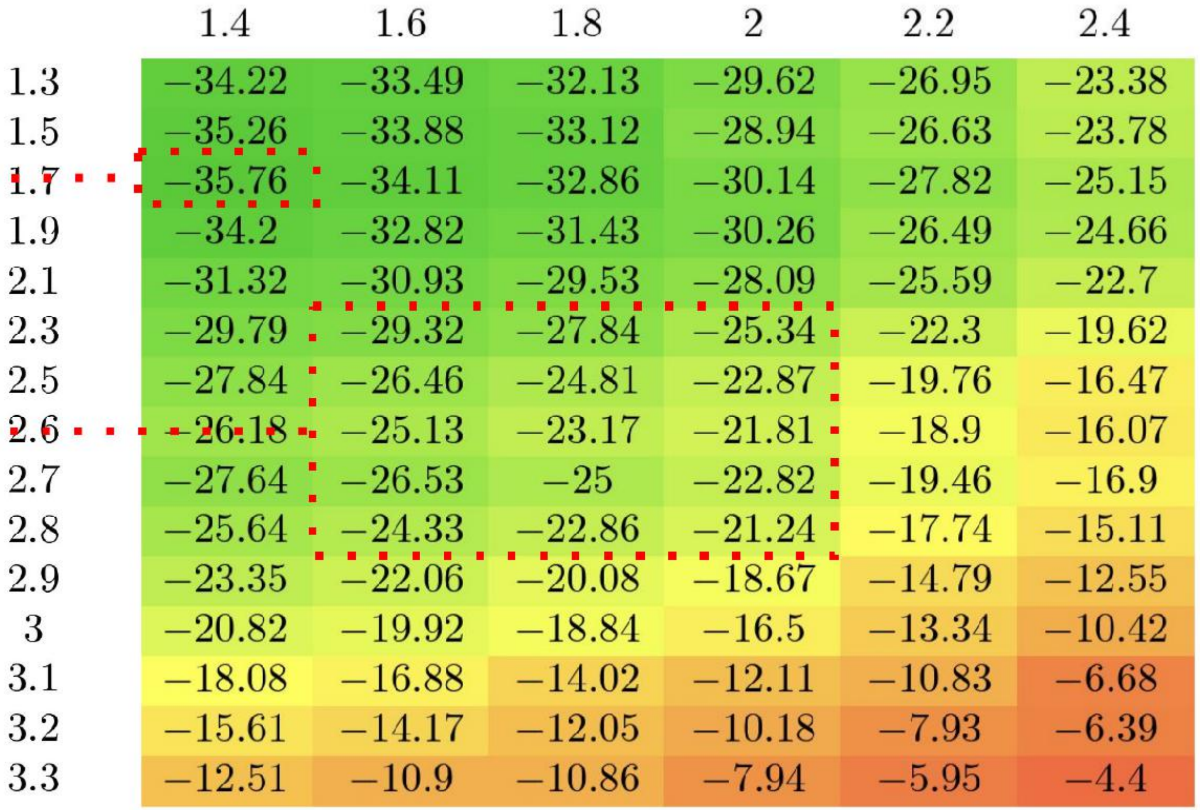
results for OpenGadget

[2] Energy efficiency : tuning



uncore [GHz] / core [GHz]	1.4	1.6	1.8	2	2.2	2.4
1.3	15.27	13.27	13.25	13.53	12.44	12.78
1.5	8.39	8.64	7.24	9.82	9.31	8.78
1.7	4.52	4.82	4.15	5.86	4.12	3.4
1.9	4.46	4.12	3.81	1.52	3.11	2.35
2.1	3.4	3.53	3.24	1.58	1.98	1.7
2.3	2.79	1.22	1.4	1.09	1.12	1.16
2.5	1.37	1.26	1.01	0.31	0.69	0.8
2.6	1.8	1.47	1.16	−0.11	0.2	−0.14
2.7	1.51	0.55	0.49	−0.41	1.29	0.34
2.8	1.98	1.05	0.9	0.04	0.47	−0.33
2.9	1.51	0.7	1.19	0.1	0.95	−0.46
3	0.99	−0.1	−0.43	−1.14	−0.29	−1.37
3.1	0.41	−0.21	−0.43	0.92	−0.87	−0.32
3.2	0.7	−0.09	0.99	−0.24	−1.14	−2.26
3.3	0.11	−0.55	−0.25	−0.43	−0.33	−1.18

Runtime extension [%]

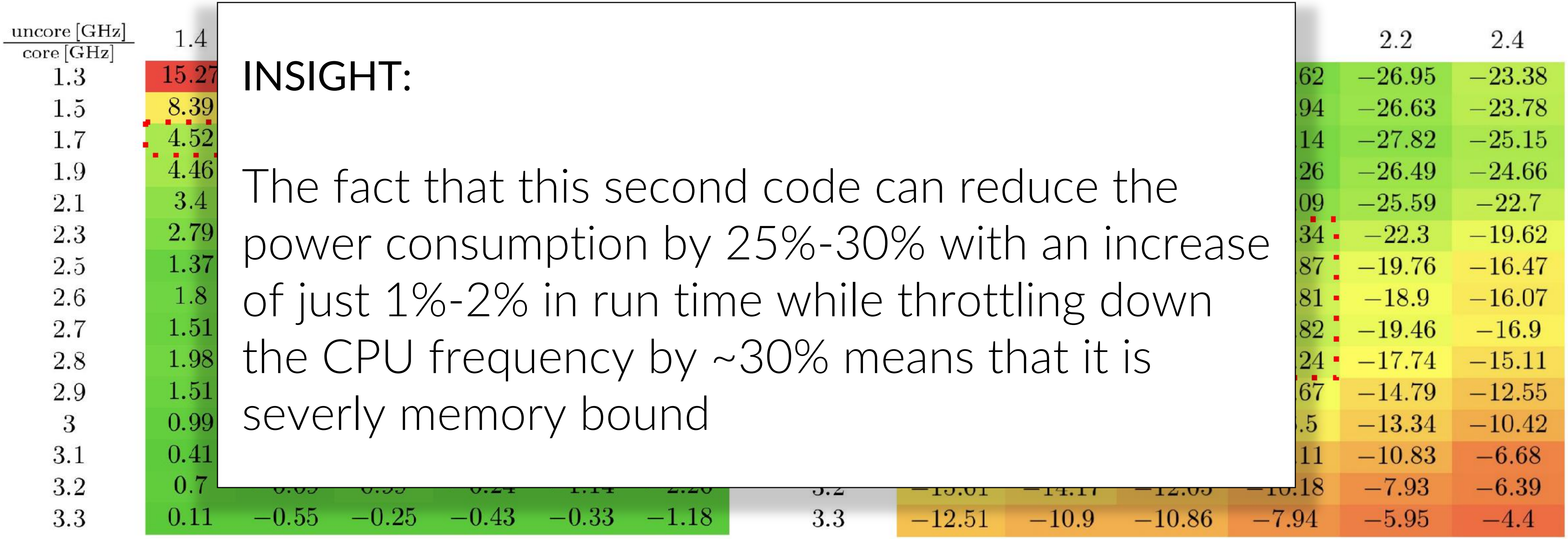


	1.4	1.6	1.8	2	2.2	2.4
1.3	−34.22	−33.49	−32.13	−29.62	−26.95	−23.38
1.5	−35.26	−33.88	−33.12	−28.94	−26.63	−23.78
1.7	−35.76	−34.11	−32.86	−30.14	−27.82	−25.15
1.9	−34.2	−32.82	−31.43	−30.26	−26.49	−24.66
2.1	−31.32	−30.93	−29.53	−28.09	−25.59	−22.7
2.3	−29.79	−29.32	−27.84	−25.34	−22.3	−19.62
2.5	−27.84	−26.46	−24.81	−22.87	−19.76	−16.47
2.6	−26.18	−25.13	−23.17	−21.81	−18.9	−16.07
2.7	−27.64	−26.53	−25	−22.82	−19.46	−16.9
2.8	−25.64	−24.33	−22.86	−21.24	−17.74	−15.11
2.9	−23.35	−22.06	−20.08	−18.67	−14.79	−12.55
3	−20.82	−19.92	−18.84	−16.5	−13.34	−10.42
3.1	−18.08	−16.88	−14.02	−12.11	−10.83	−6.68
3.2	−15.61	−14.17	−12.05	−10.18	−7.93	−6.39
3.3	−12.51	−10.9	−10.86	−7.94	−5.95	−4.4

Energy savings [%]

results for Changa

[2] Energy efficiency : tuning



uncore [GHz] / core [GHz]	1.4					
1.3	15.27					
1.5	8.39					
1.7	4.52					
1.9	4.46					
2.1	3.4					
2.3	2.79					
2.5	1.37					
2.6	1.8					
2.7	1.51					
2.8	1.98					
2.9	1.51					
3	0.99					
3.1	0.41					
3.2	0.7					
3.3	0.11	−0.55	−0.25	−0.43	−0.33	−1.18

Runtime extension [%]

					2.2	2.4
					−26.95	−23.38
					−26.63	−23.78
					−27.82	−25.15
					−26.49	−24.66
					−25.59	−22.7
					−22.3	−19.62
					−19.76	−16.47
					−18.9	−16.07
					−19.46	−16.9
					−17.74	−15.11
					−14.79	−12.55
					−13.34	−10.42
					−10.83	−6.68
					−7.93	−6.39
3.3	−12.51	−10.9	−10.86	−7.94	−5.95	−4.4

Energy savings [%]

INSIGHT:

The fact that this second code can reduce the power consumption by 25%-30% with an increase of just 1%-2% in run time while throttling down the CPU frequency by ~30% means that it is severly memory bound

results for Changa

Final notes

- HPC, especially towards “exascale”, is a winding road that requires specific knowledge and professional tools and support
  - need of a **parallel efficiency > 80% (90%)** over large intervals both in weak and strong scaling
  - need of very **specialised codes for different devices** (CPU, GPU, vector accelerators, FPGA, …)
  - **major impacts from “fine” details**: memory affinity, topology-awareness, data structures, vectorization, …

28

Final notes

- HPC, especially towards “exascale”, is a winding road that requires specific knowledge and professional tools and support
- INAF is increasingly acquiring this knowledge and is developing the unique and unvaluable convergence of scientific and computer-science realms;
  however: very specific figures needed?
- Federating our common resources, experience and knowledge is an obvious accelerator but a not-so-obvious reaction to be catalyzed (?)

Acknowledgement & Disclaimer

Funded by the European Union. This work has received funding from the European High Performance Computing Joint Undertaking (JU) and Belgium, Czech Republic, France, Germany, Greece, Italy, Norway, and Spain under grant agreement No 101093441.

Co-funded by the European Union