Six and a half years of POP CoE: What remains?

Jesús Labarta, BSC

EU H2020 Centre of Excellence (CoE)

Grant Agreement No 676553
Grant Agreement No 824080

1 October 2015 – 31 March 2018
1 December 2018 – 31 May 2022

Age before beauty (disclaimer)

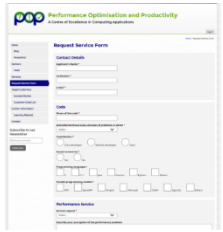
• Behavior (insight/models)	before	syntax
• Detail performance analytics	before	aggregated profiles
• Work instantiation and order	before	overhead
• Malleability	before	fitted rigid structure
• Possibilities	before	how tos
• Elegance	before	one day shine

El abuelo cebolleta ataca de nuevo

POP services

Online application

https://pop-coe.eu/request-service-form



User or developer

Who obtains traces?

expert

Client provides code, input and instructions on how to compile and run.

User or developer

POP expert provides instructions on how to install tracing tools and obtain traces.

traces

Parallel Application Performance Assessment

expert

Report

The report includes suggestions to improve performance

User or developer

Proof of concept (PoC)

expert

POP expert implements optimizations

code

Report

Code developer implements optimizations

Follow on audit

What remains?

- Mindset
  - Performance is a part of the whole development cycle
  - Beyond profiles → efficiency culture and models
  - Syntax → fundamental behavior
  - Towards DEEP analyses
    - One experiment … a lot of information
- Material
  - Training
  - Success stories
  - Co-design resources
- A feeling of
  - Belonging, learning, …
  - Appreciation by customers of external view
    - A view from an external “observer”
    - Awareness, quantification of behaviors

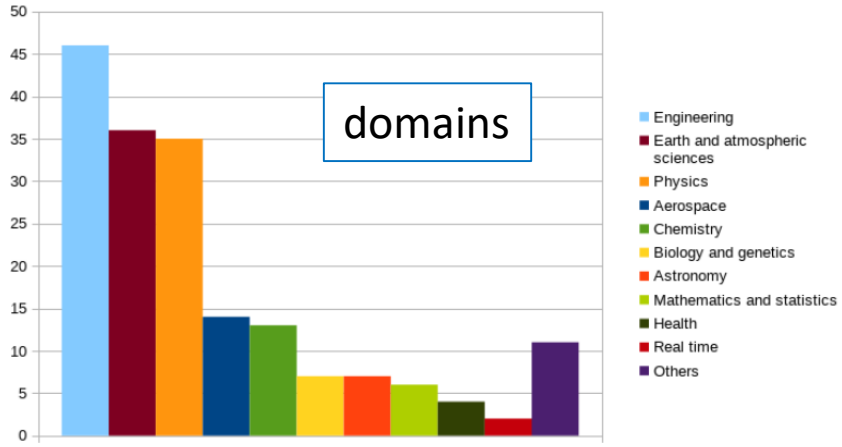
//pop-coe.eu

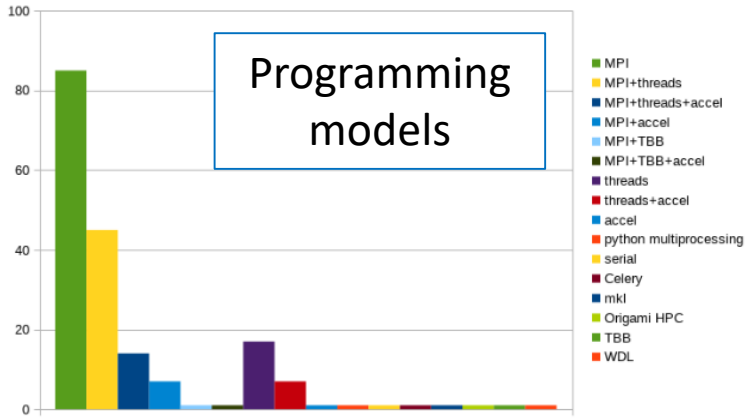
POP

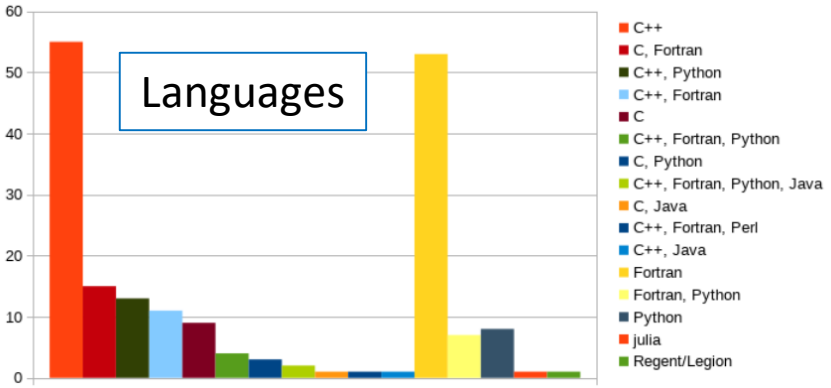
- Large number of services

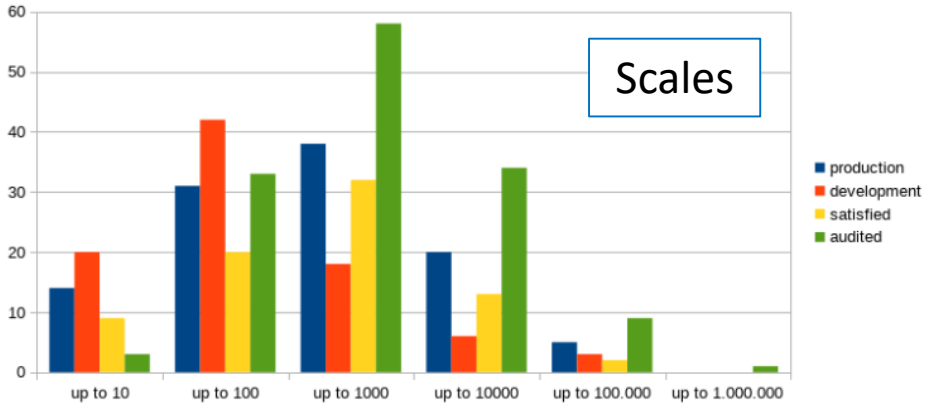
POP1 & POP2 each
- > 140 Assessments
- > 25 Proof of Concept

customer
- ~80% Research
- ~20% Industry



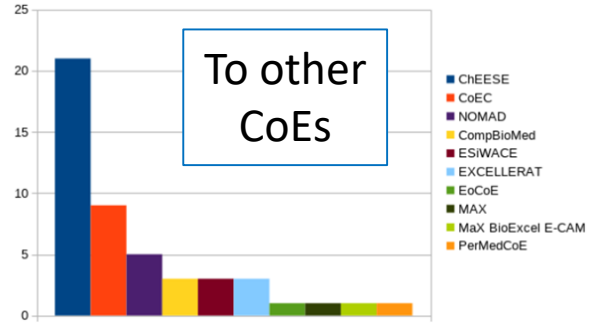


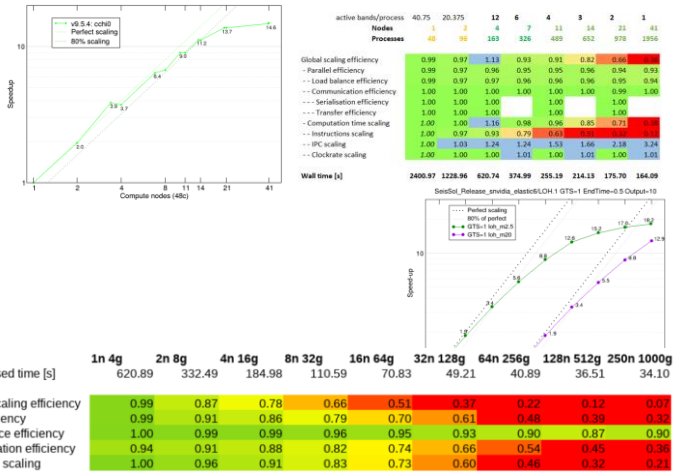




POP and the EuroHPC community

- Collaborations and other engagements with HPC CoEs, HPC³ & developer communities in Europe
  - Business development
    - HPC CoE flagship assessment campaigns
    - Ad hoc application performance assessments
  - Training
    - Bespoke training
    - Joint training
  - Dissemination
    - Minisymposium on HPC CoE exascale readiness
    - Participation in HPC³-organised events
  - Co-design insight → other projects !!

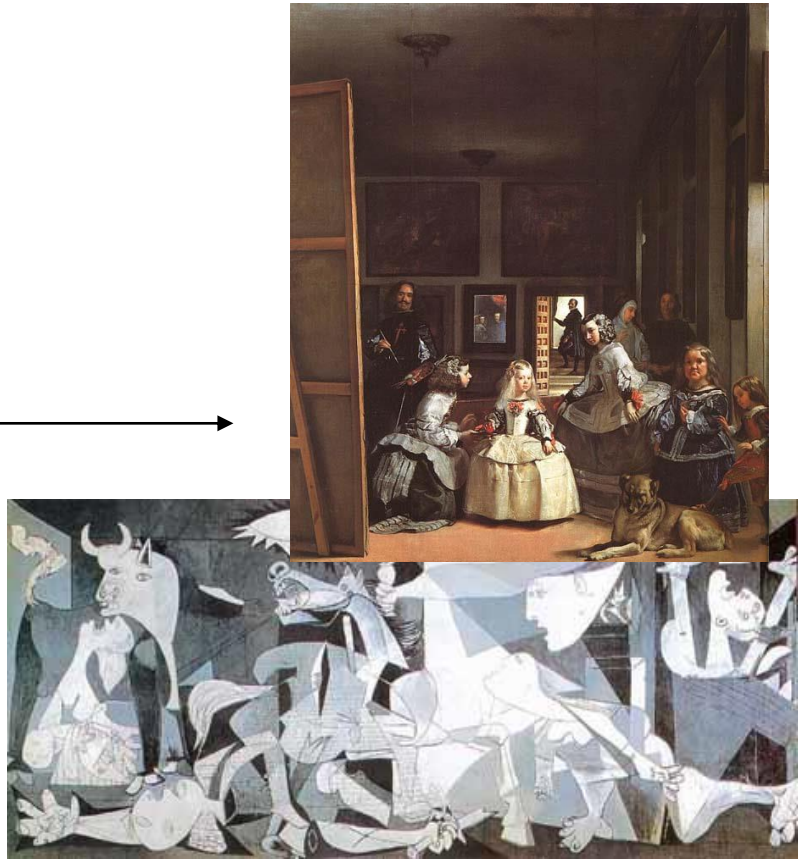




	1n 4g	2n 8g	4n 16g	8n 32g	16n 64g	32n 128g	64n 256g	128n 512g	250n 1000g
Simulate elapsed time [s]	620.89	332.49	184.98	110.59	70.83	49.21	40.89	36.51	34.10
GPU Global scaling efficiency	0.99	0.87	0.78	0.66	0.51	0.37	0.22	0.12	0.07
- Parallel efficiency	0.99	0.91	0.86	0.79	0.70	0.61	0.48	0.39	0.32
- - Load balance efficiency	1.00	0.99	0.99	0.96	0.95	0.93	0.90	0.87	0.90
- - Communication efficiency	0.94	0.91	0.88	0.82	0.74	0.66	0.54	0.45	0.36
- Computation scaling	1.00	0.96	0.91	0.83	0.73	0.60	0.46	0.32	0.21

Performance analysis

- Art …
  - Not a unified approach

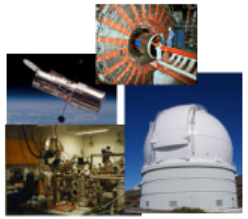


Performance analysis

- Art …
  - Not a unified approach
- … or science?
  - Precise measurement tools
  - Models
  - Method

Tools in …

- Science



- Advances linked to capability of observation and measurement

- Computer Science
  - printf()
  - timers
- Performance analysis:
  - A lot of speculation
  - We see $\int_a^b f(t)dt$
  - we talk about $f(t)$

Shoemaker's son …

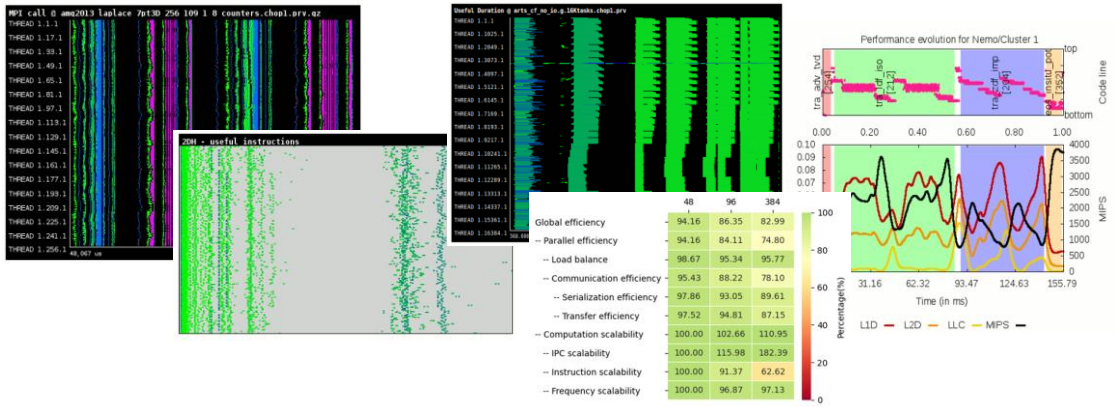
…always goes barefoot

En casa del herrero …

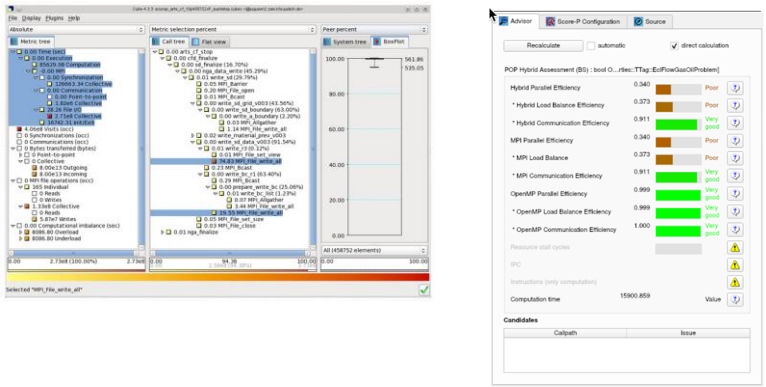
… cuchara de palo

Tools

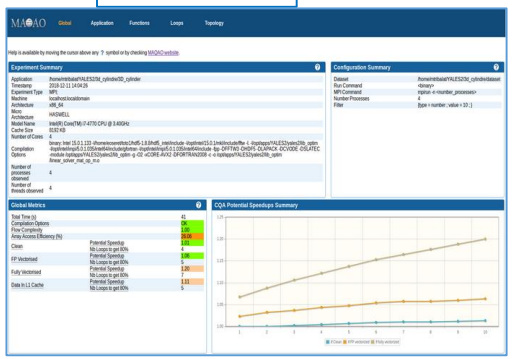
BSC-tools: Extrae, Paraver, model factors, …



JSC tools: Scalasca, cube, Score-P,…



MAQAO



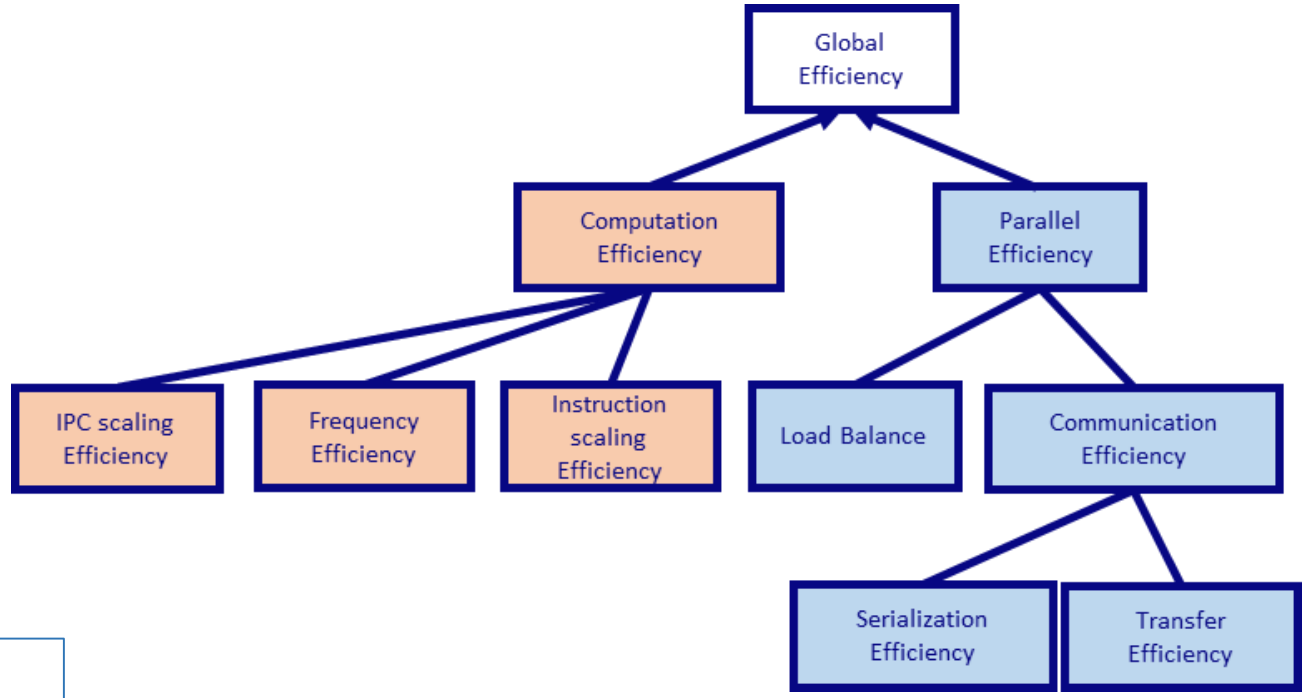
PyPOP

Correctness:
MUST (for MPI) and Archer (for Threading)

Other, vendors, …

Efficiency Model

Efficiency metrics : ~ (0,1]
Multiplicative model



Behavioral description
"Language" transversal across domains

$$CompEff = Ieff * IPCeff * Feff$$

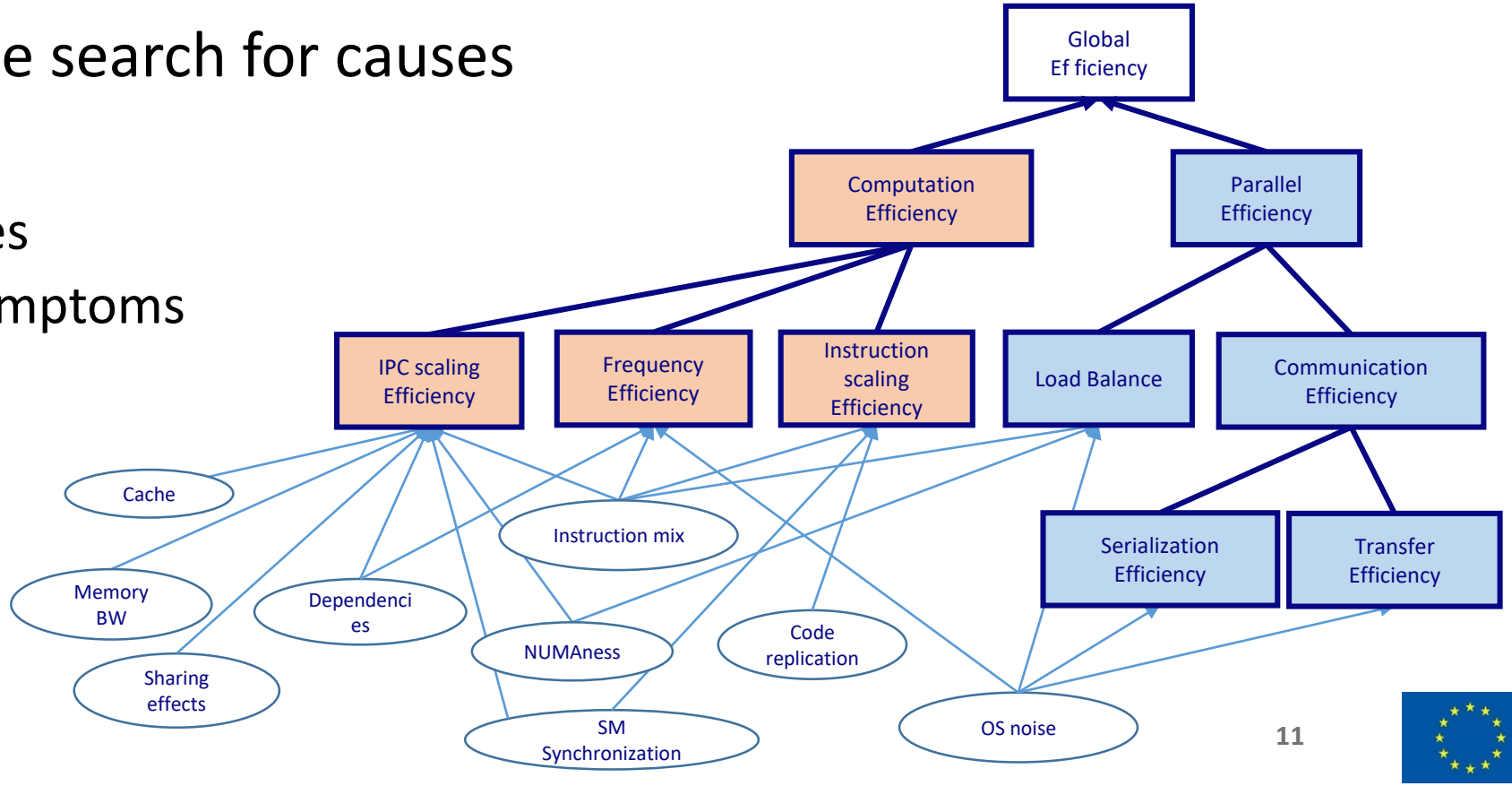
$$\eta_{\parallel} = LB * Ser * Trf$$

CommEff

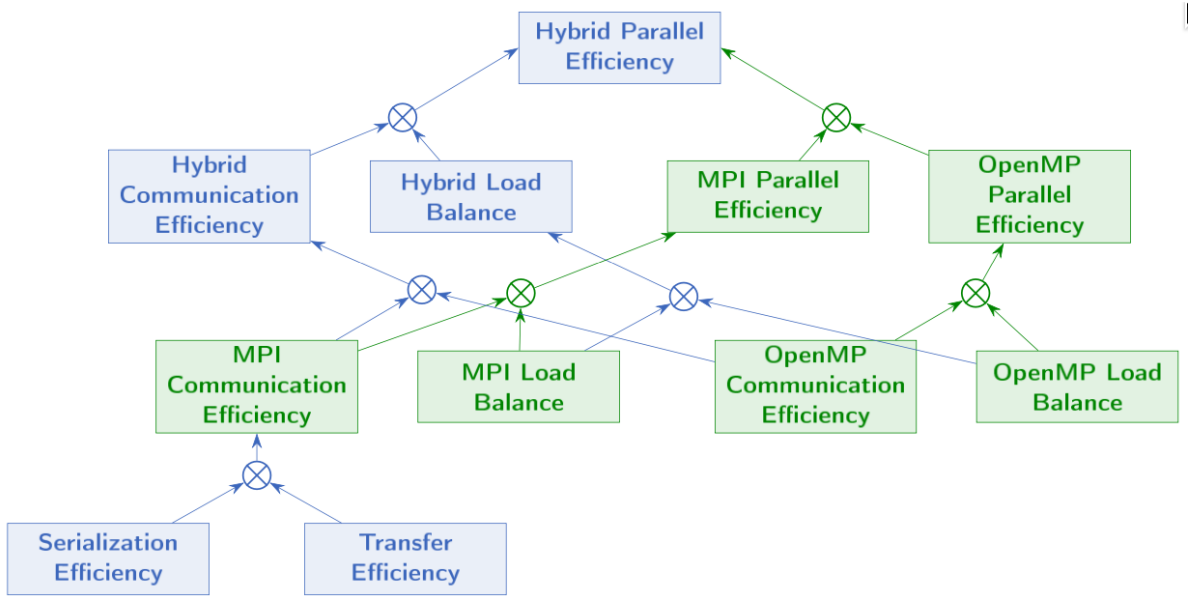
M. Casas et al, "Automatic analysis of speedup of MPI applications". ICS 2008.

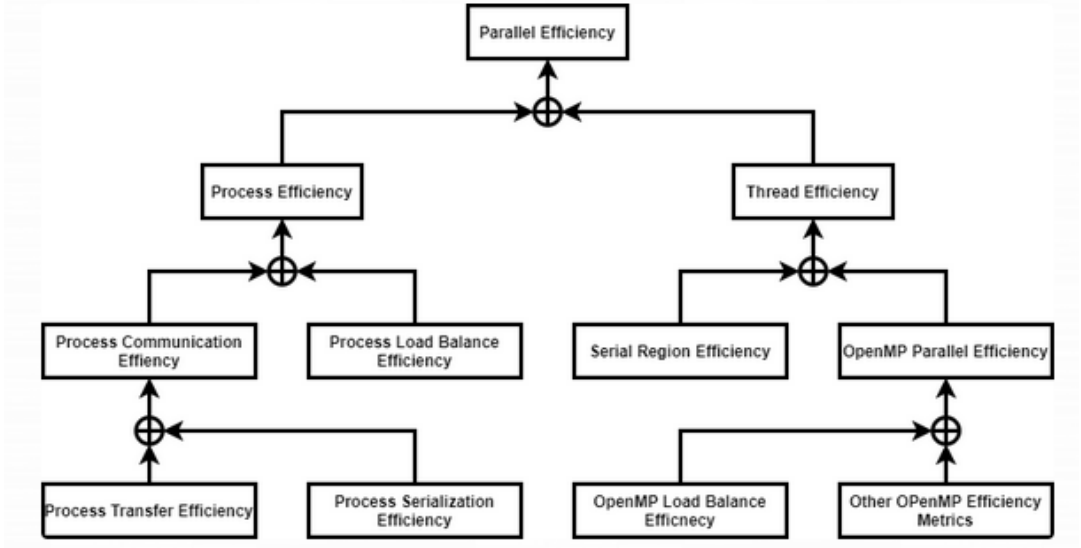
POP metrics and causes

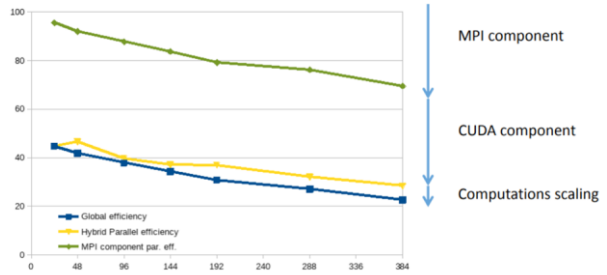
- POP Metrics: quantitative abstract description of fundamental effect
- Causes actual underlying behaviour or phenomena whose effect is captured in the factors
- Metrics to steer the search for causes
- Refactoring:
  - Address the causes
  - Counteract the symptoms



Model of hybrid programs

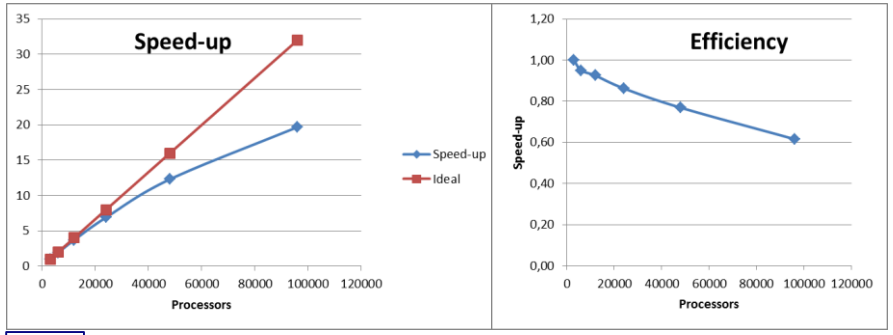






Typical practice

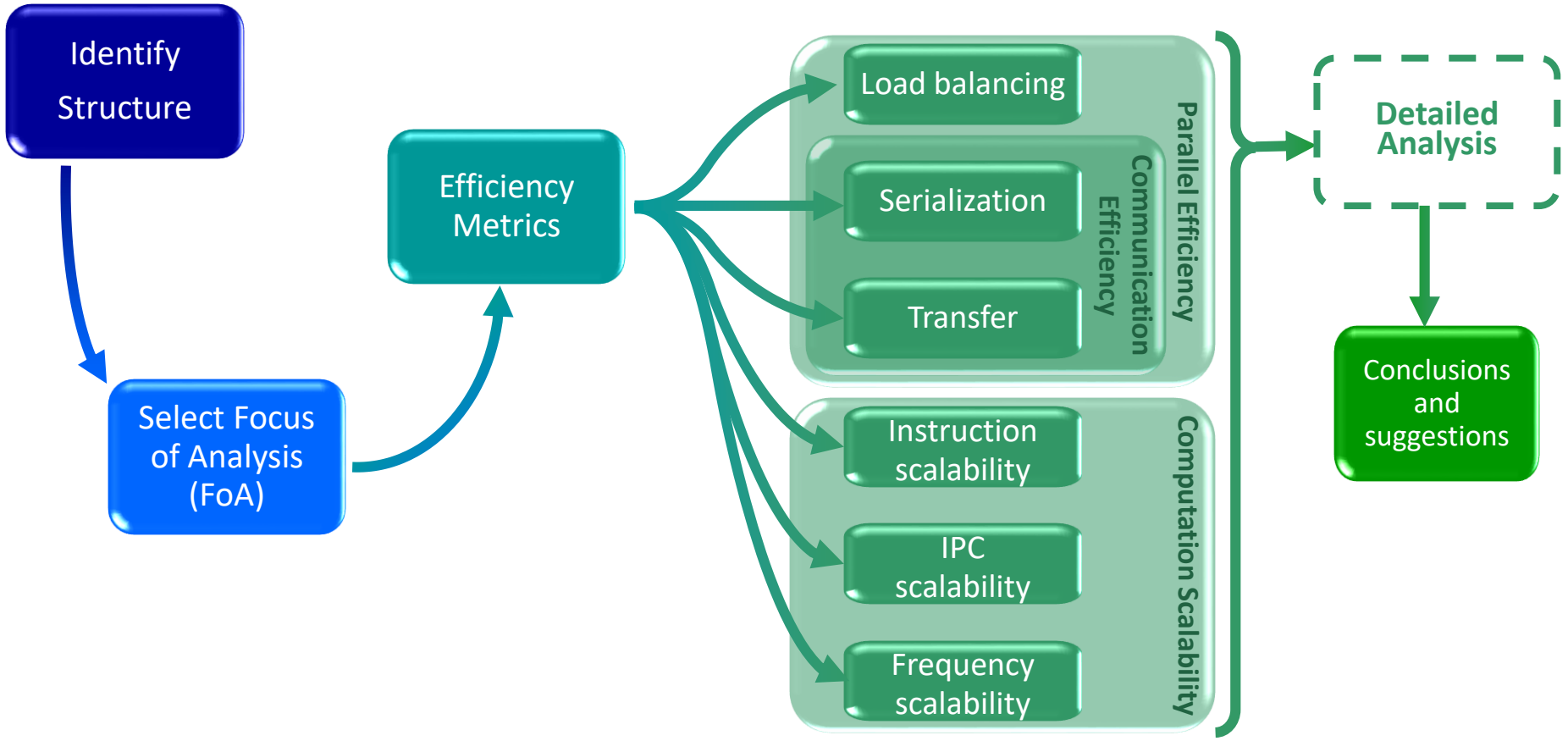
- Elapsed time
- Profiles



T(1)

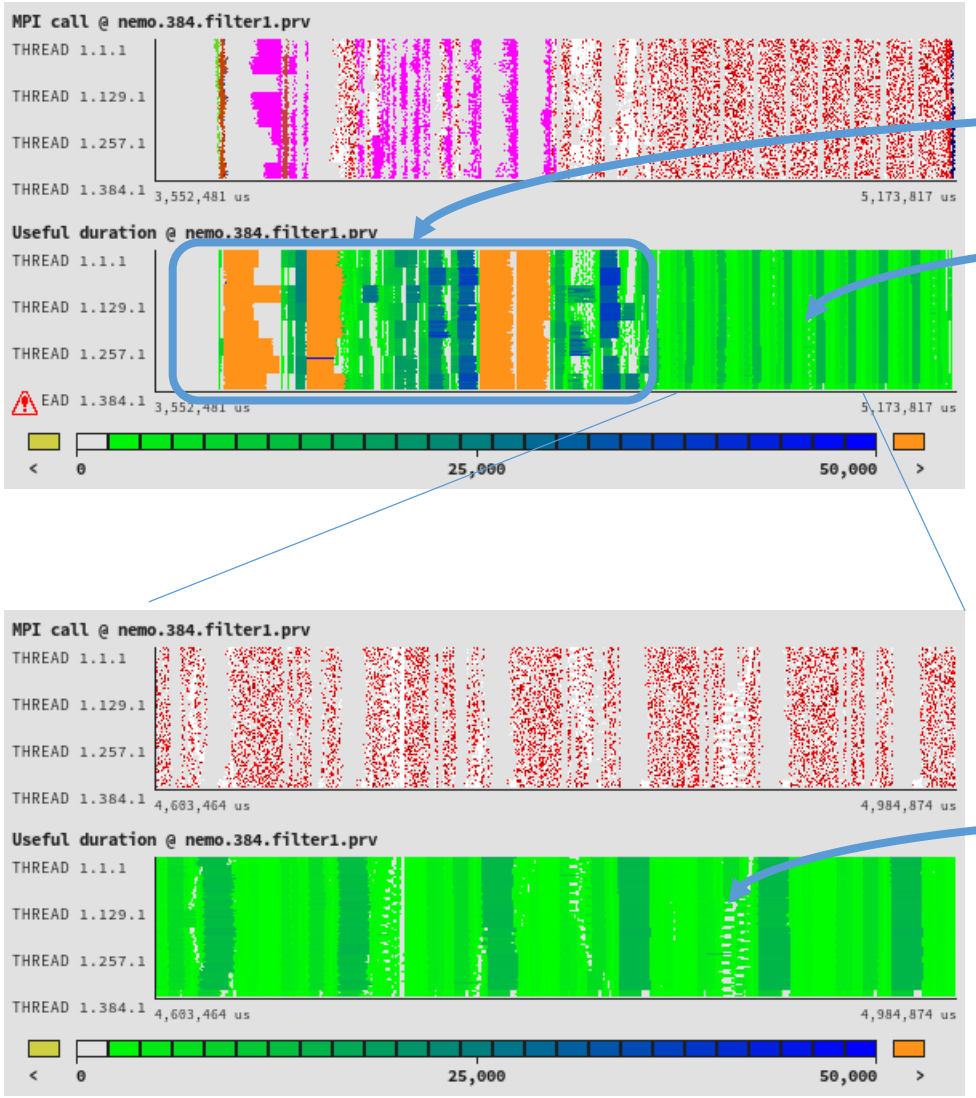
- But …
  - Lacks structural insight
  - Too coarse aggregation
  - **Time != blame**
  - **Profile != (fundamental behavior, causes)**

The performance analysis journey



Structure

384 cores

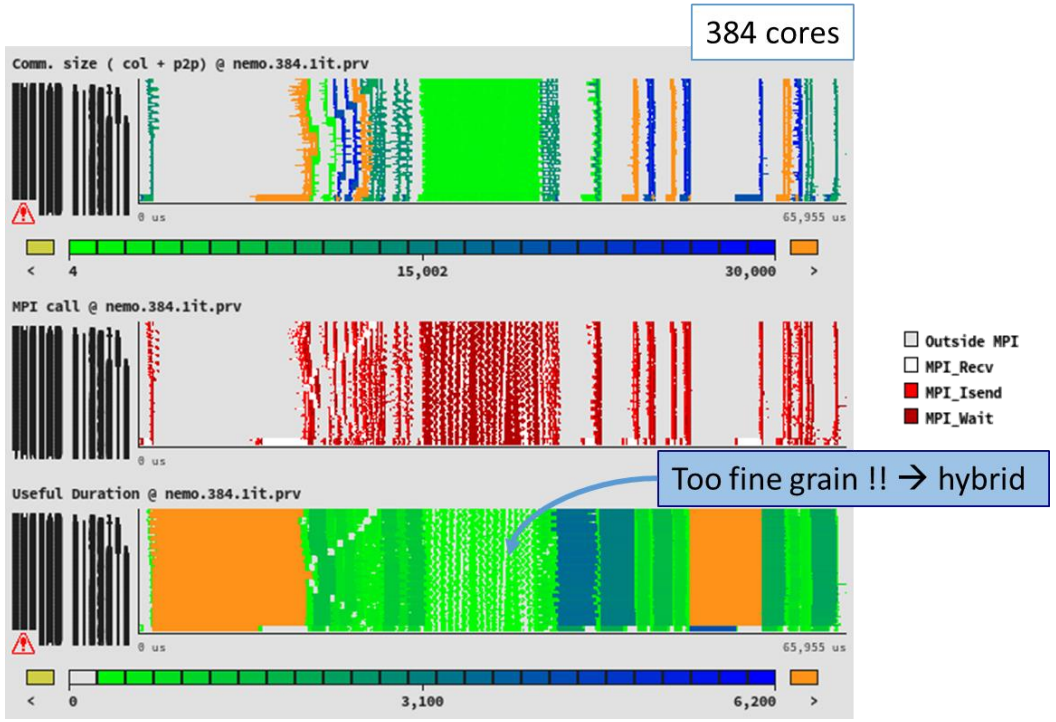


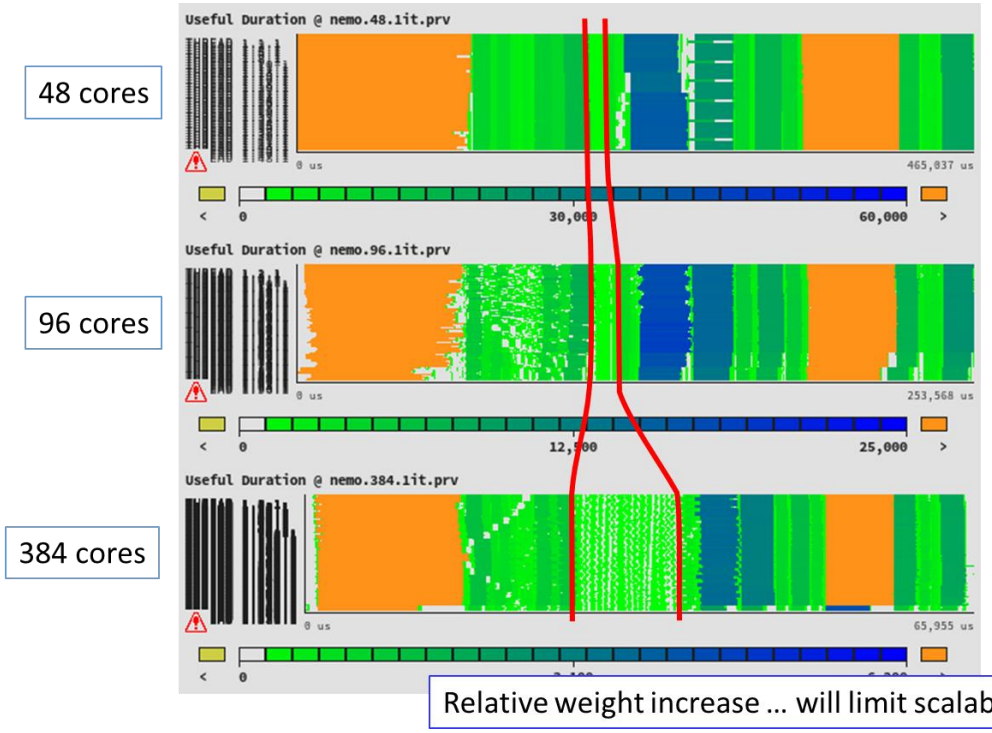
Initialization

Iterative structure

perturbation

Structure & Focus of Analysis





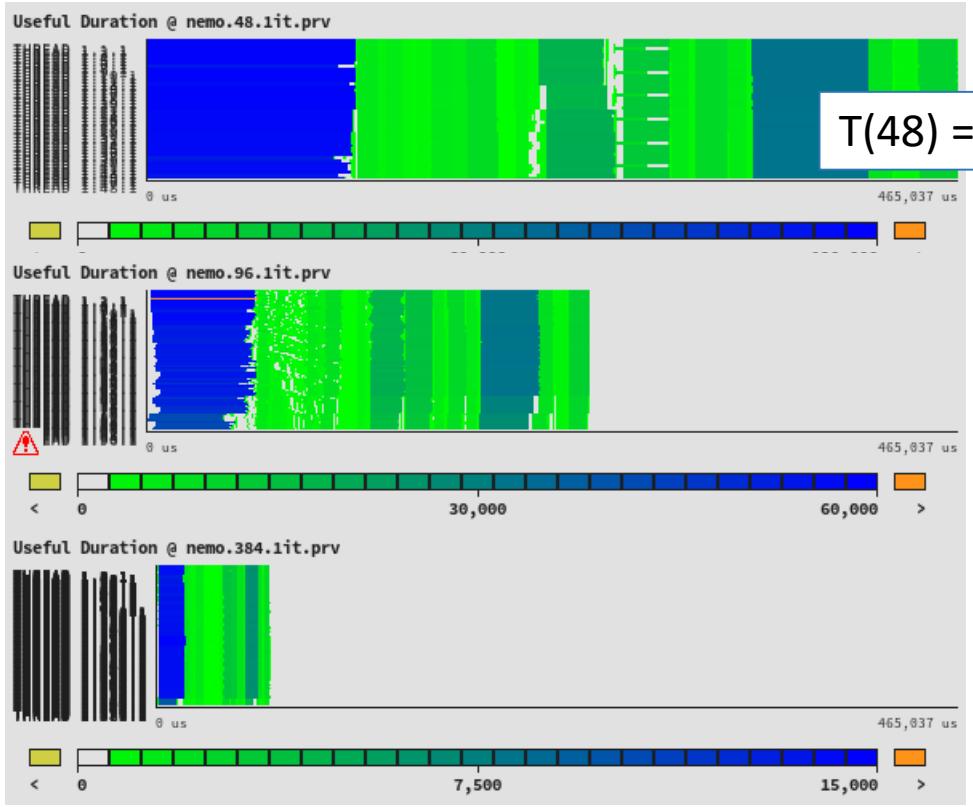
Scaling

- Time based speedup and efficiency

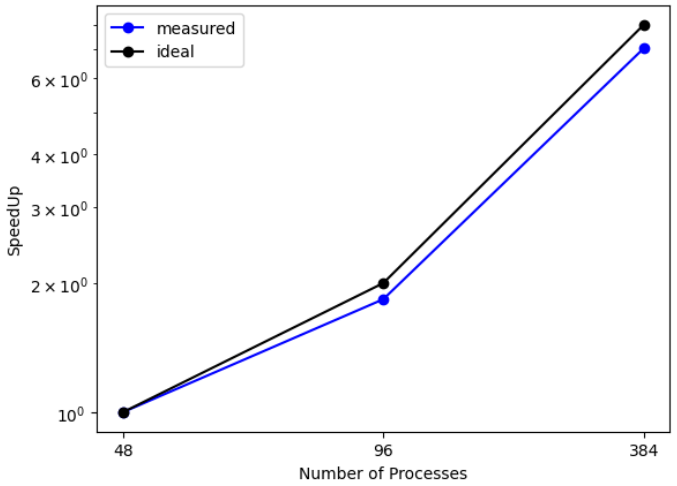
48

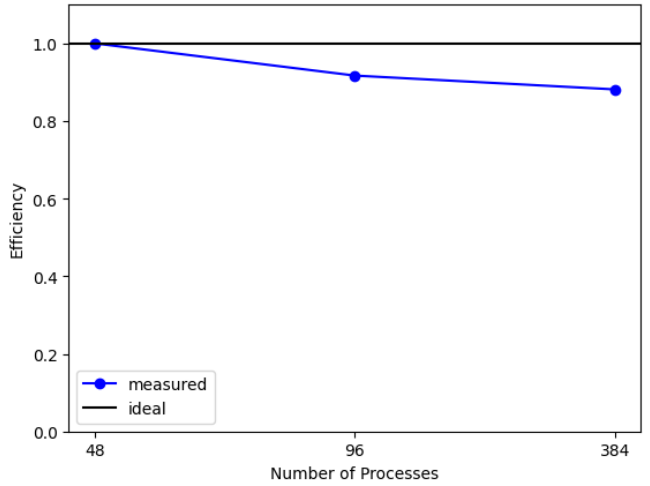
96

384



T(48) =465 ms



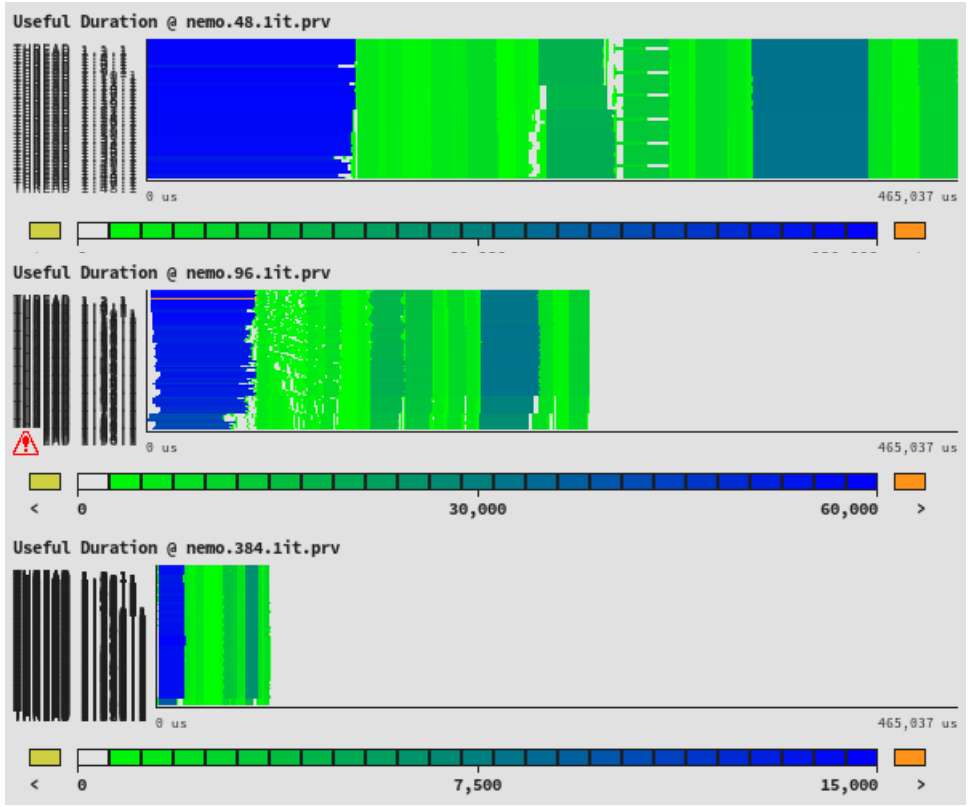


Efficiency model

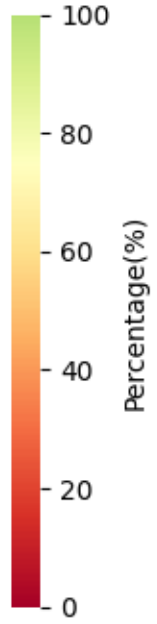
48

96

384



	48	96	384
Global efficiency	94.16	86.35	82.99
-- Parallel efficiency	94.16	84.11	74.80
-- Load balance	98.67	95.34	95.77
-- Communication efficiency	95.43	88.22	78.10
-- Serialization efficiency	97.86	93.05	89.61
-- Transfer efficiency	97.52	94.81	87.15
-- Computation scalability	100.00	102.66	110.95
-- IPC scalability	100.00	115.98	182.39
-- Instruction scalability	100.00	91.37	62.62
-- Frequency scalability	100.00	96.87	97.13

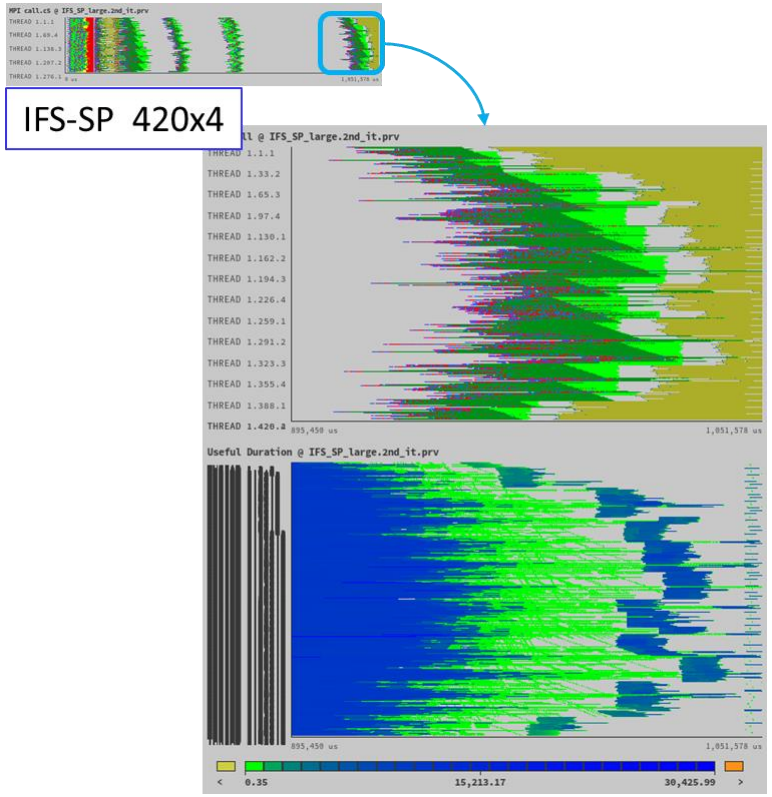


Avg Useful IPC(48) =0.67

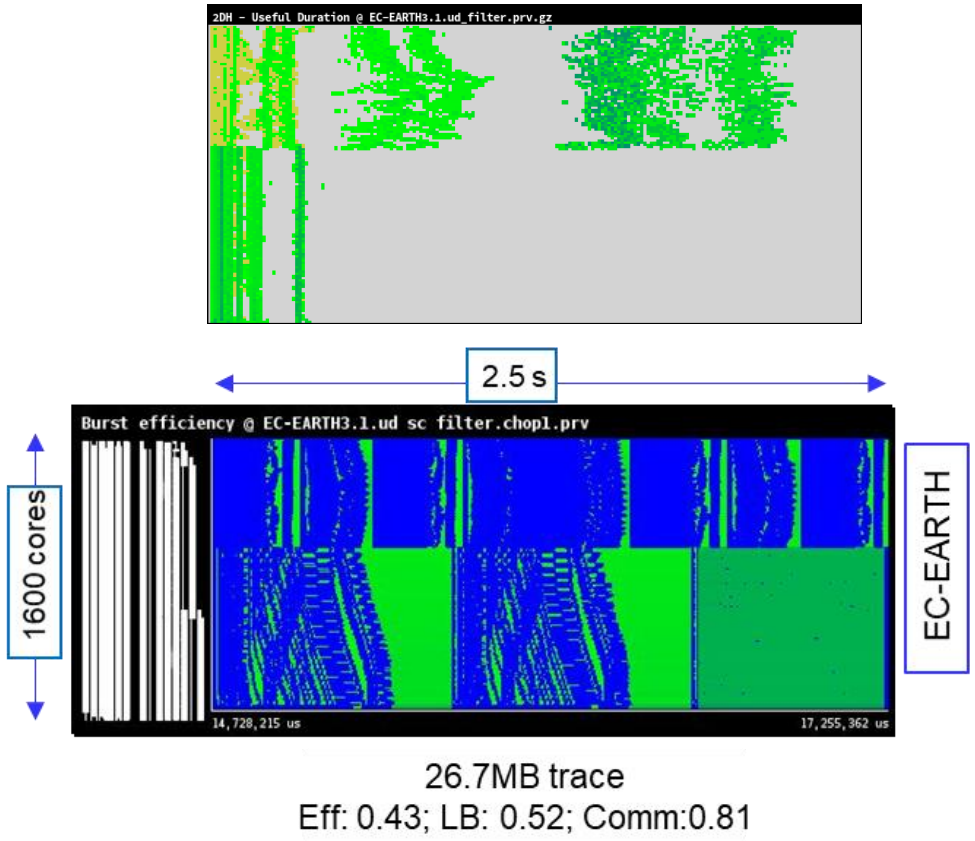
Avg Useful Frequency(48) =2.061 GHz

Load Balance

- Timelines/histograms of duration, instructions, IPC, …



Within a model …

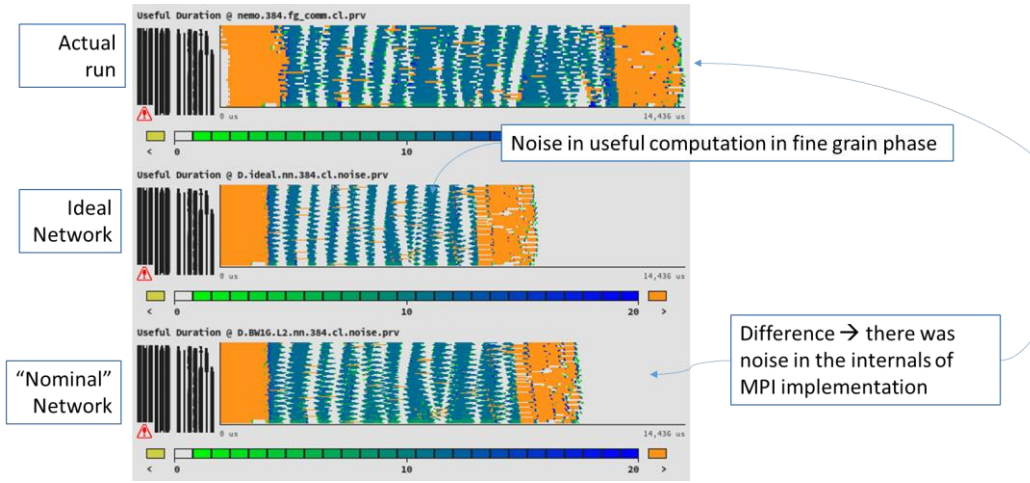


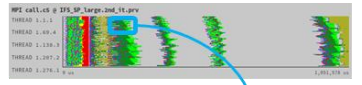
26.7MB trace
Eff: 0.43; LB: 0.52; Comm:0.81

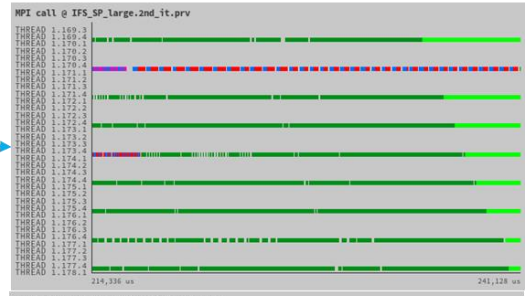
… and coupled runs

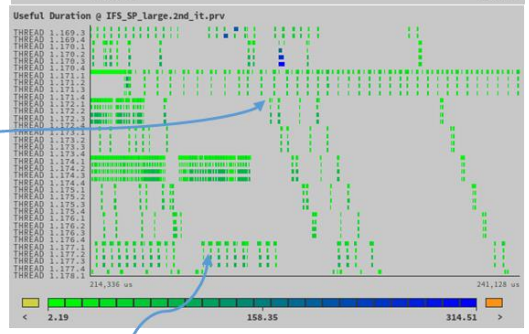
Communication analysis

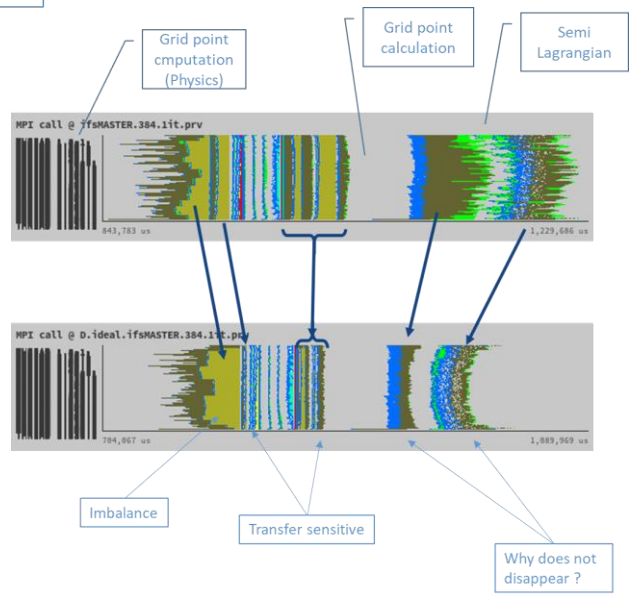
- Real vs. modeled ideal communications?



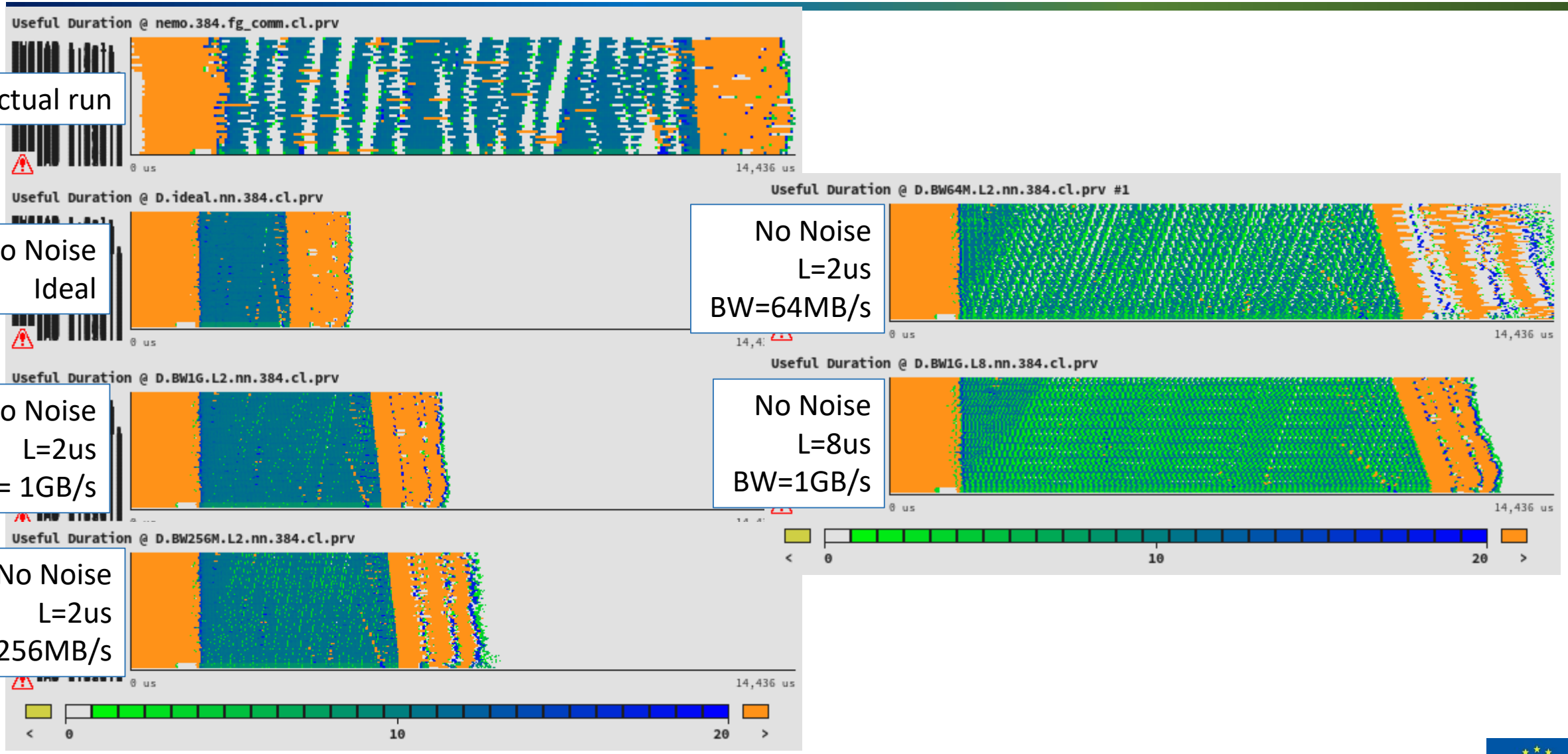








What ifs


Actual run
Useful Duration @ nemo.384.fg_comm.cl.prv
0 us
14,436 us
No Noise
Ideal
Useful Duration @ D.ideal.nn.384.cl.prv
0 us
No Noise
L=2us
BW= 1GB/s
Useful Duration @ D.BW1G.L2.nn.384.cl.prv
No Noise
L=2us
BW=256MB/s
Useful Duration @ D.BW256M.L2.nn.384.cl.prv
0 us
14,436 us
< 0 10 20 >
No Noise
L=2us
BW=64MB/s
Useful Duration @ D.BW64M.L2.nn.384.cl.prv #1
0 us
14,436 us
No Noise
L=8us
BW=1GB/s
Useful Duration @ D.BW1G.L8.nn.384.cl.prv
0 us
14,436 us
< 0 10 20 >

Scalar code scaling

- Evolution of hardware counter derived metrics with scale

48x1

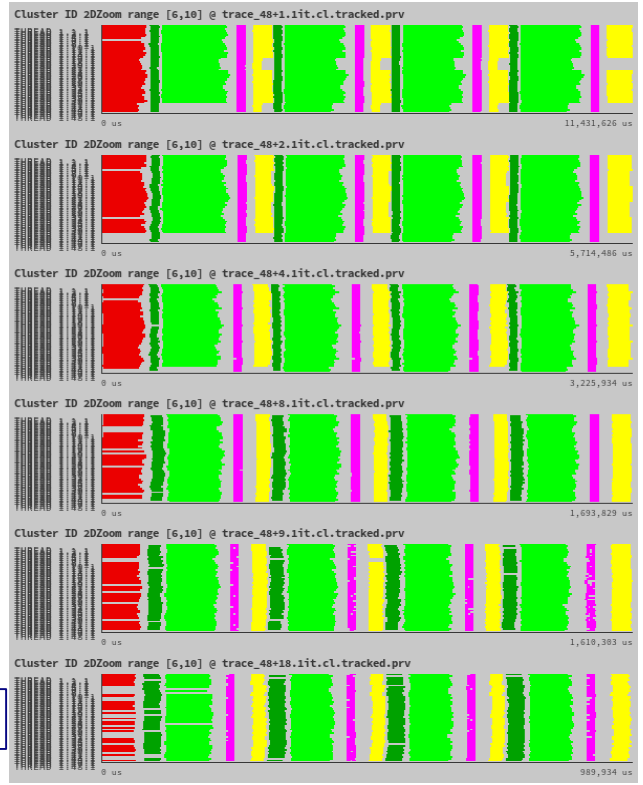
48x2

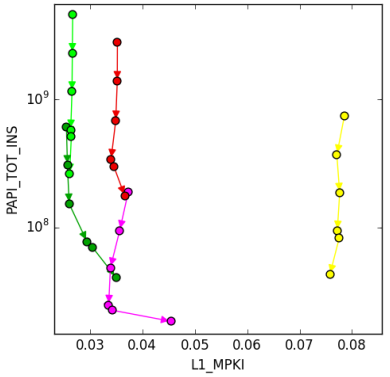
48x4

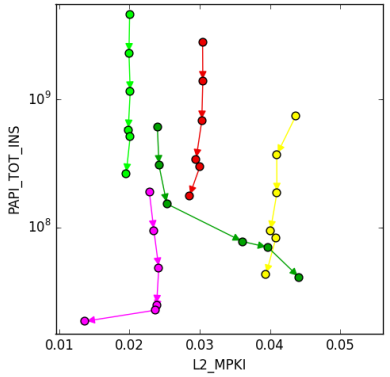
48x8

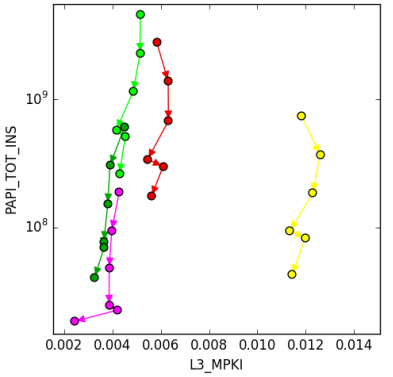
48x9

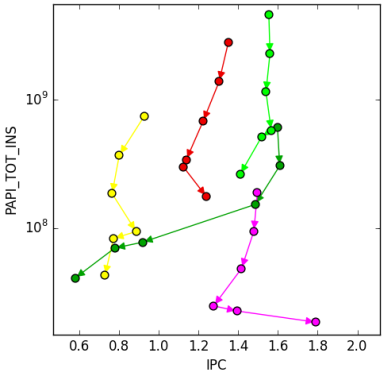
48x18

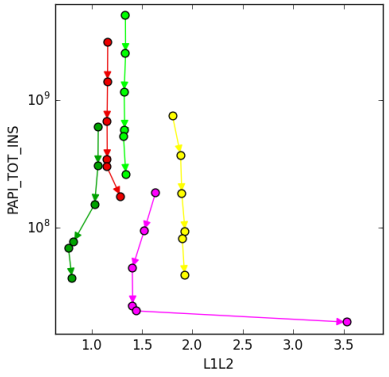


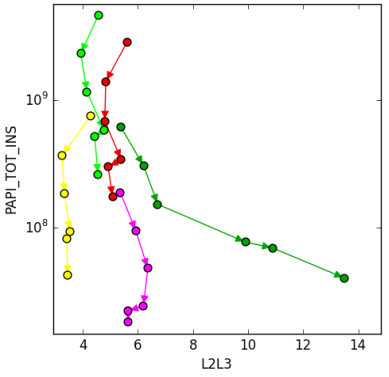








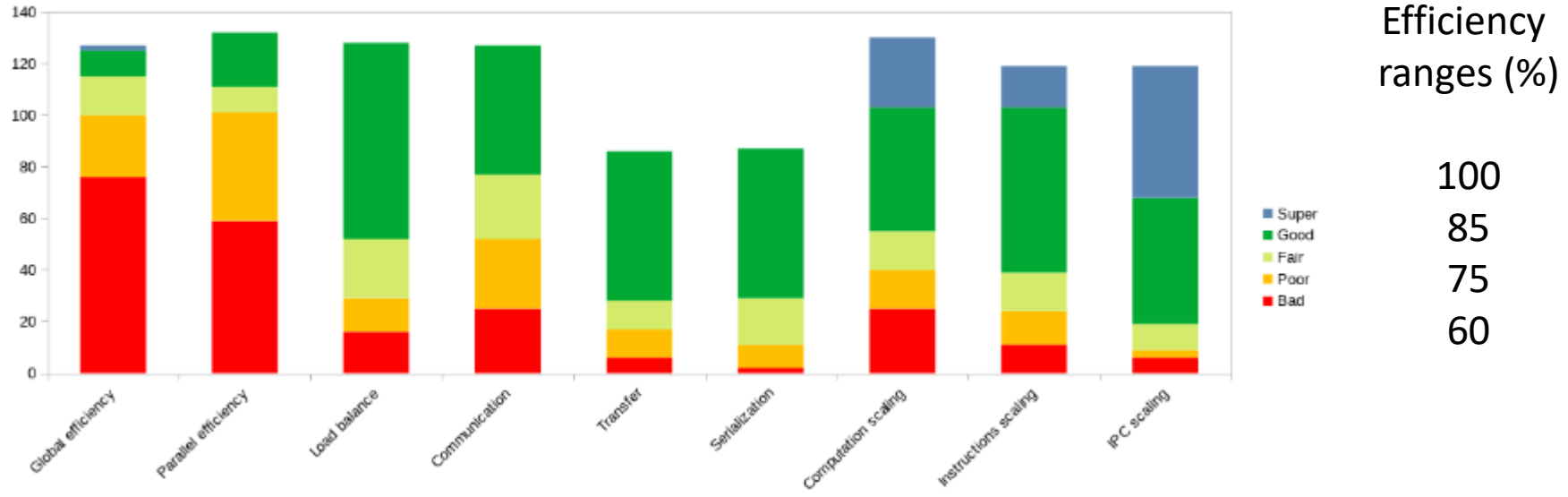




Observed efficiencies

- Room for improvement
- Coupled effects
- Each code/configuration a case on its own

@ largest scale analyzed for each assessment



Efficiency ranges (%)

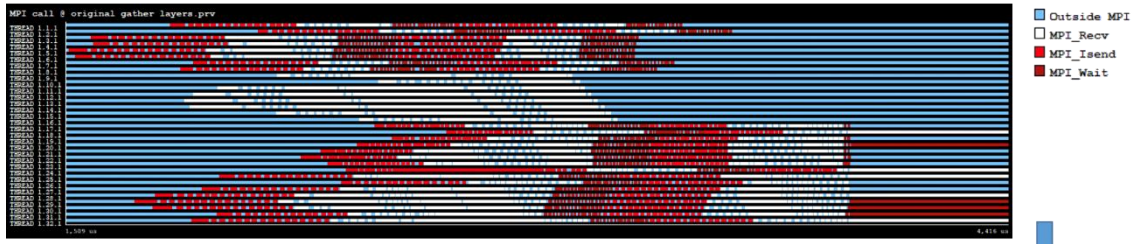
100
85
75
60

Proofs of Concept

- Cooperation with customer
  - Demonstrate on their code

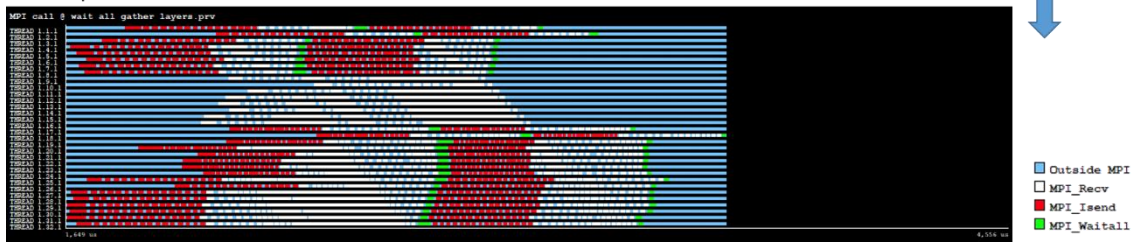
MPI calls and their order

Original:

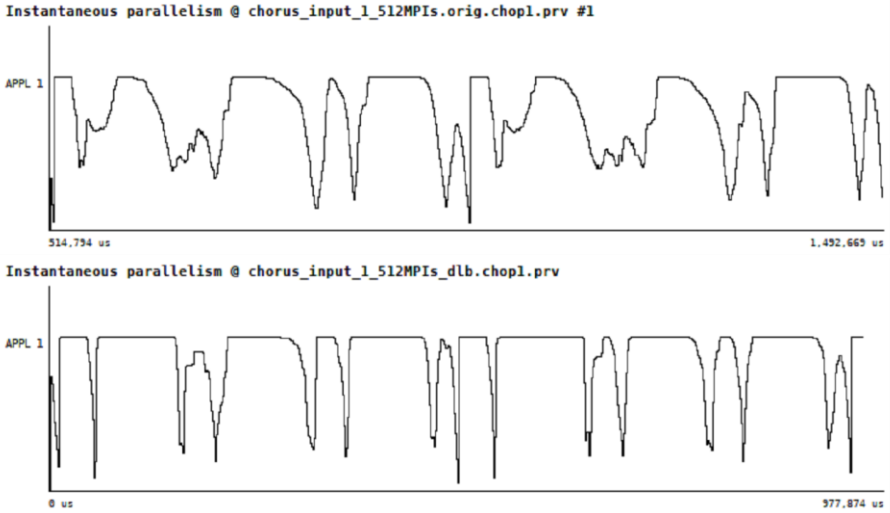


30% improvement

Proposed:



Malleability and Dynamic load balance

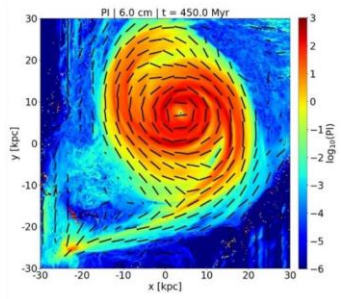


Gains

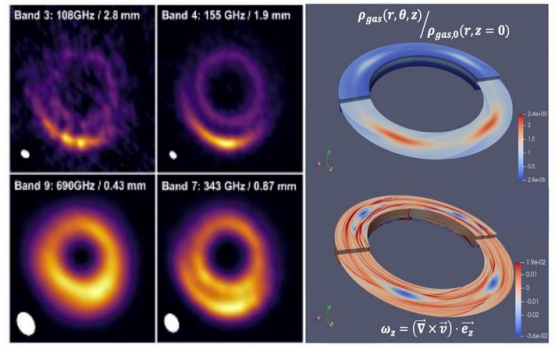
Non negligible

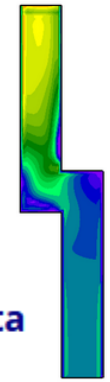
Good

POP for Astronomy - 40% Reduction in Execution Time

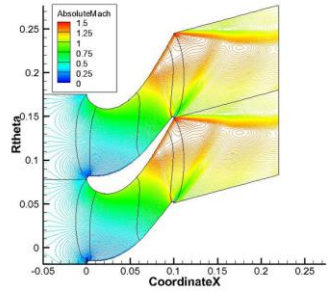


Near to 5x speedup of the RoSSBi astrophysics code



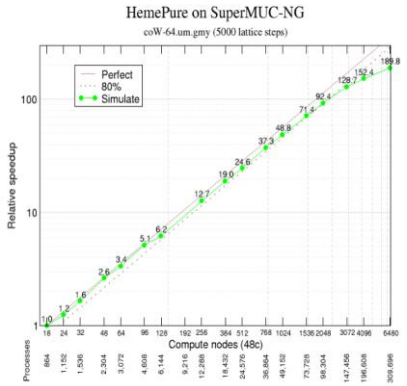


Run time halved for OpenMP code

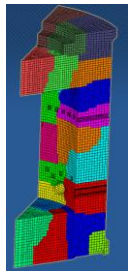


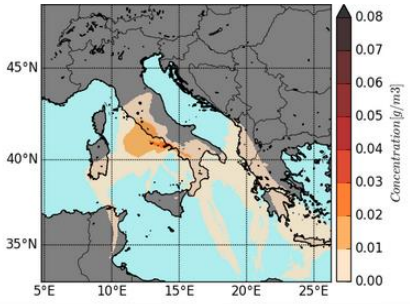
Performance Improvements by More Than 30% and a Data Race Fixed for CalculiX Code

Impressive



30x Speedup of the Matrix Factorization when Applying a Math Library



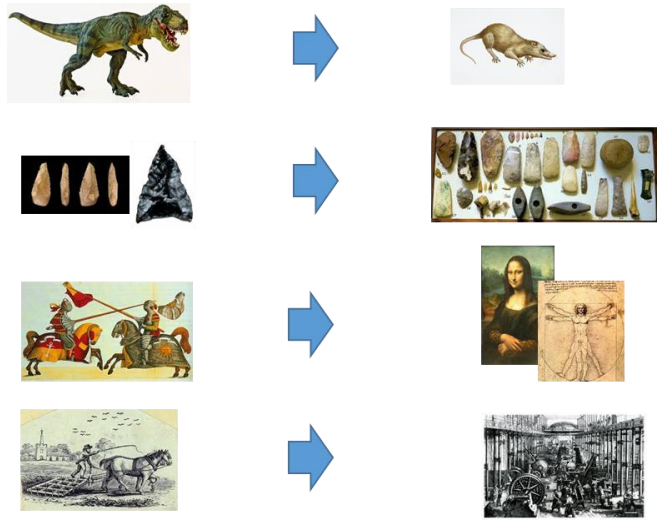


588x and 488x Execution Time Speedups

190x strong-scaling speed-up of HemeLB simulation on SuperMUC-NG

What remains to be done?

- Revolutions/mindset transitions are important …
- … but take time (years, decades, …)



- Applied for POP3
  - More frequent interaction with customer
    - Better appreciate and steer analyses
  - More interaction between partners
    - Sharing background, homogenizing methodology, mindset
  - More material
    - Training
    - Analyses
    - Co-design

A collective effort

Performance Optimisation and Productivity

A Centre of Excellence in HPC

Contact:

https://www.pop-coe.eu

mailto:pop@bsc.es

@POP_HPC

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676553 and 824080.