Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology
MeteoSwiss

CSCS
Swiss National Supercomputing Centre

Operational Numerical Weather Prediction in Switzerland and Evolution Towards New Supercomputer Architectures

Philippe Steiner (MeteoSwiss)

Michele De Lorenzi (CSCS)

Angelo Mangili (CSCS)

14th ECMWF Workshop on the Use of HPC in Meteorology, November 2010

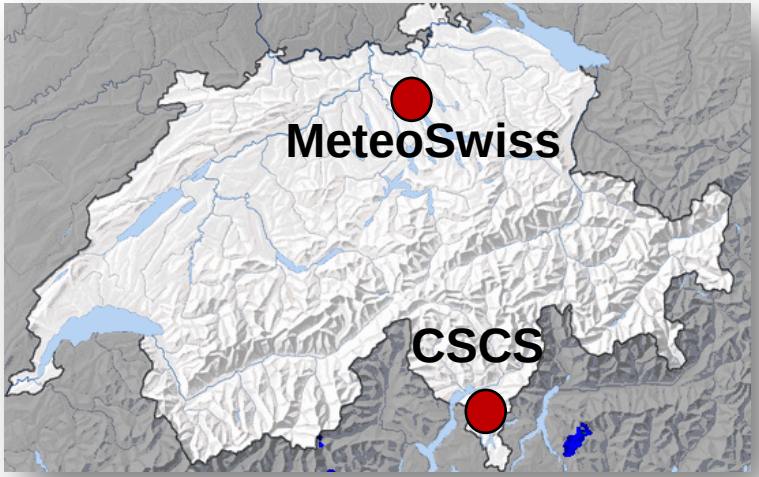
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Outline

- Actual operational suite of MeteoSwiss at CSCS and performance
- Performance on different Cray architectures
- Swiss HP2C Initiative
- HP2C Project for COSMO
  - Performance Analysis
  - Investigated Hybrid Parallelization via OpenMP/MPI
  - Parallelizing I/O
  - New Dynamical Core
- Outlook of future suites of MeteoSwiss at CSCS
- Conclusion

MeteoSwiss-CSCS Collaboration

- The *Swiss National Supercomputing Centre* (CSCS) as unit of ETH Zurich is hosting the operational model of MeteoSwiss.
- More than 10 years of collaboration which has been proven to be fruitful for both institutes.
- MeteoSwiss profits of the expertise of CSCS in HPC and can focus on the NWP tasks.
- CSCS solves the technological and engineering aspects and provides knowledge transfer to other scientific domains.



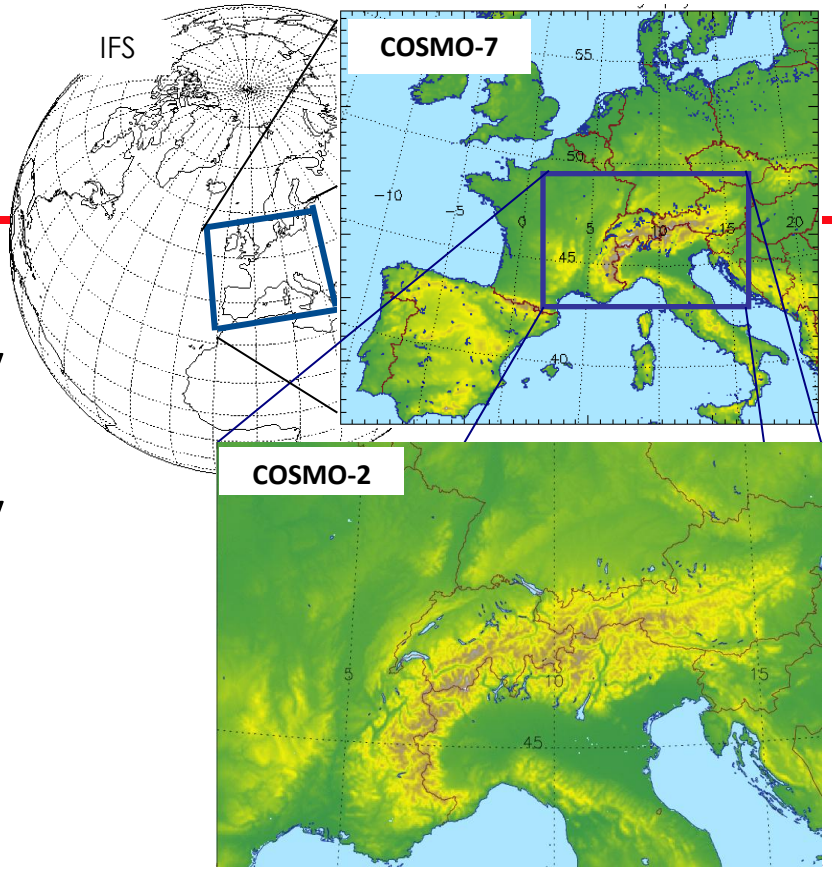
Actual Operational Suite

COSMO-7 :

6.6km, 393x338x60 GP, 3 x 72h fcst /day

COSMO-2 :

2.2km, 520x350x60 GP, 8 x 24h fcst /day



Production at CSCS

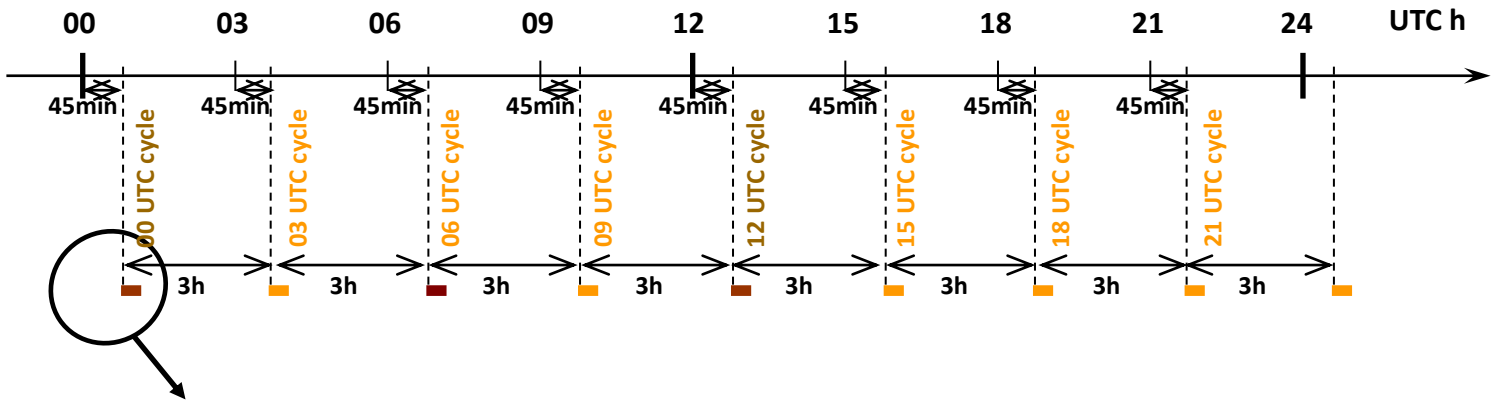
Main machine: **Cray XT4 Buin**, 1056 cores

Fall-back: **Cray XT4 Dole**, 688 cores

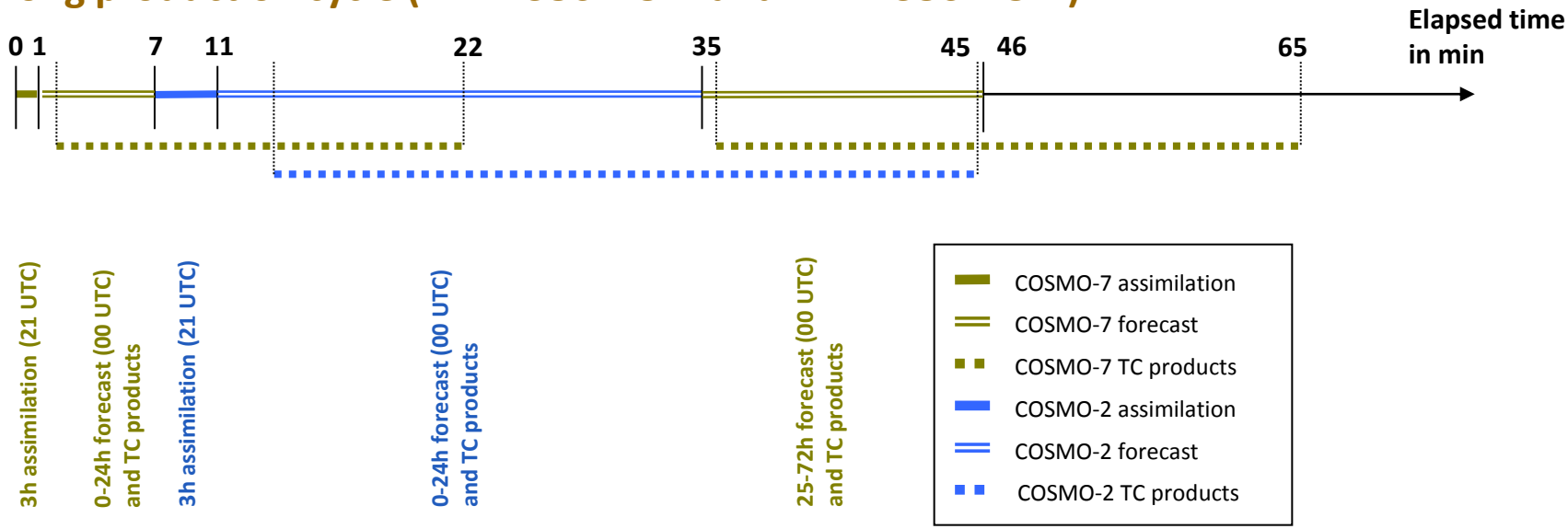
Service nodes used for front-end functions

Separate servers for DB

Production Scheme



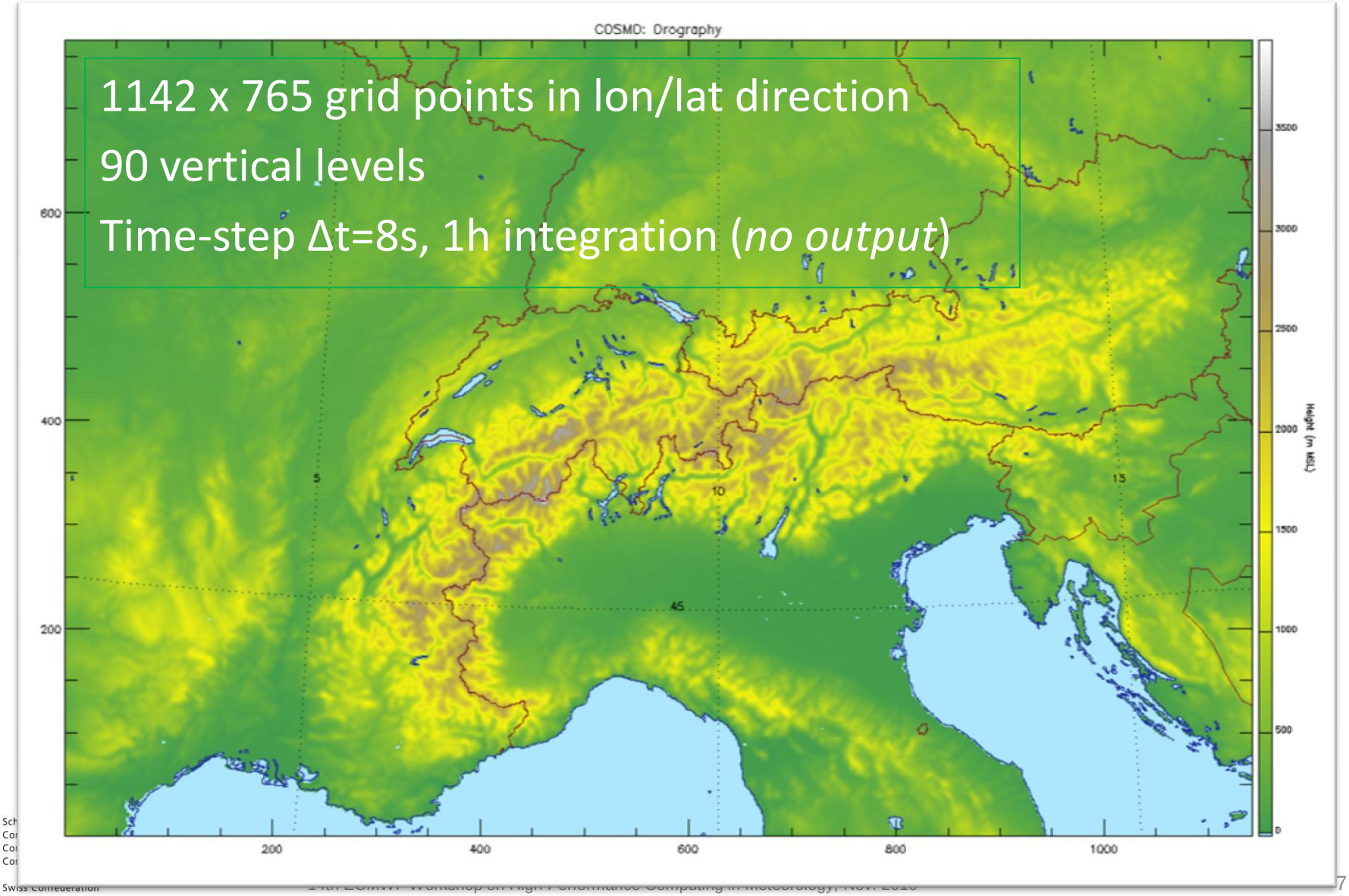
Long production cycle (+72h COSMO-7 and +24h COSMO-2):



- Actual operational suite of MeteoSwiss at CSCS and performance
- Performance on different Cray architectures
- Swiss HP2C Initiative
- HP2C Project for COSMO
  - Performance Analysis
  - Investigated Hybrid Parallelization via OpenMP/MPI
  - Parallelizing I/O
  - New Dynamical Core
- Outlook of future suites of MeteoSwiss at CSCS
- Conclusion

COSMO 1 km Performance on Cray XT4/XT5/XE6

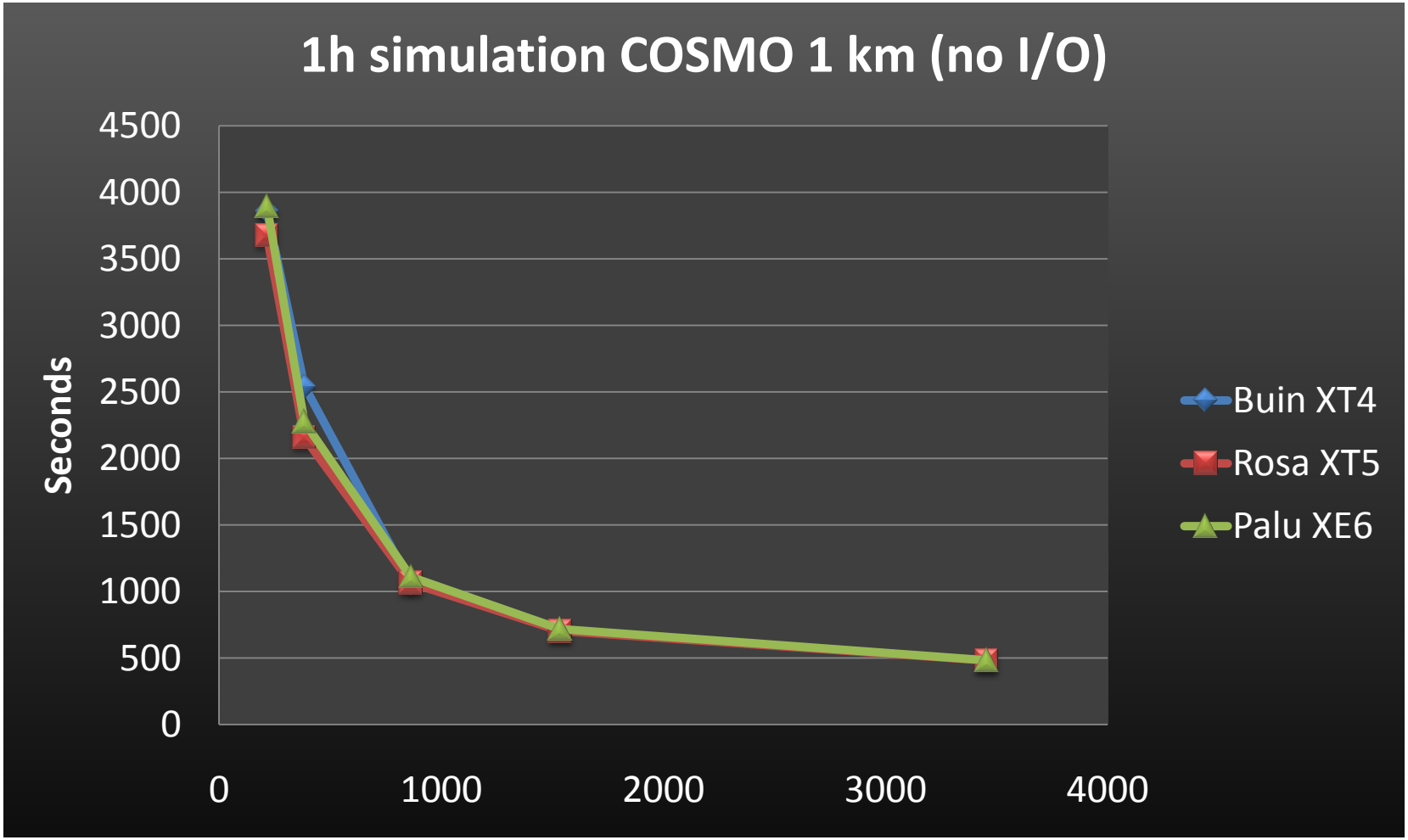
Benchmark



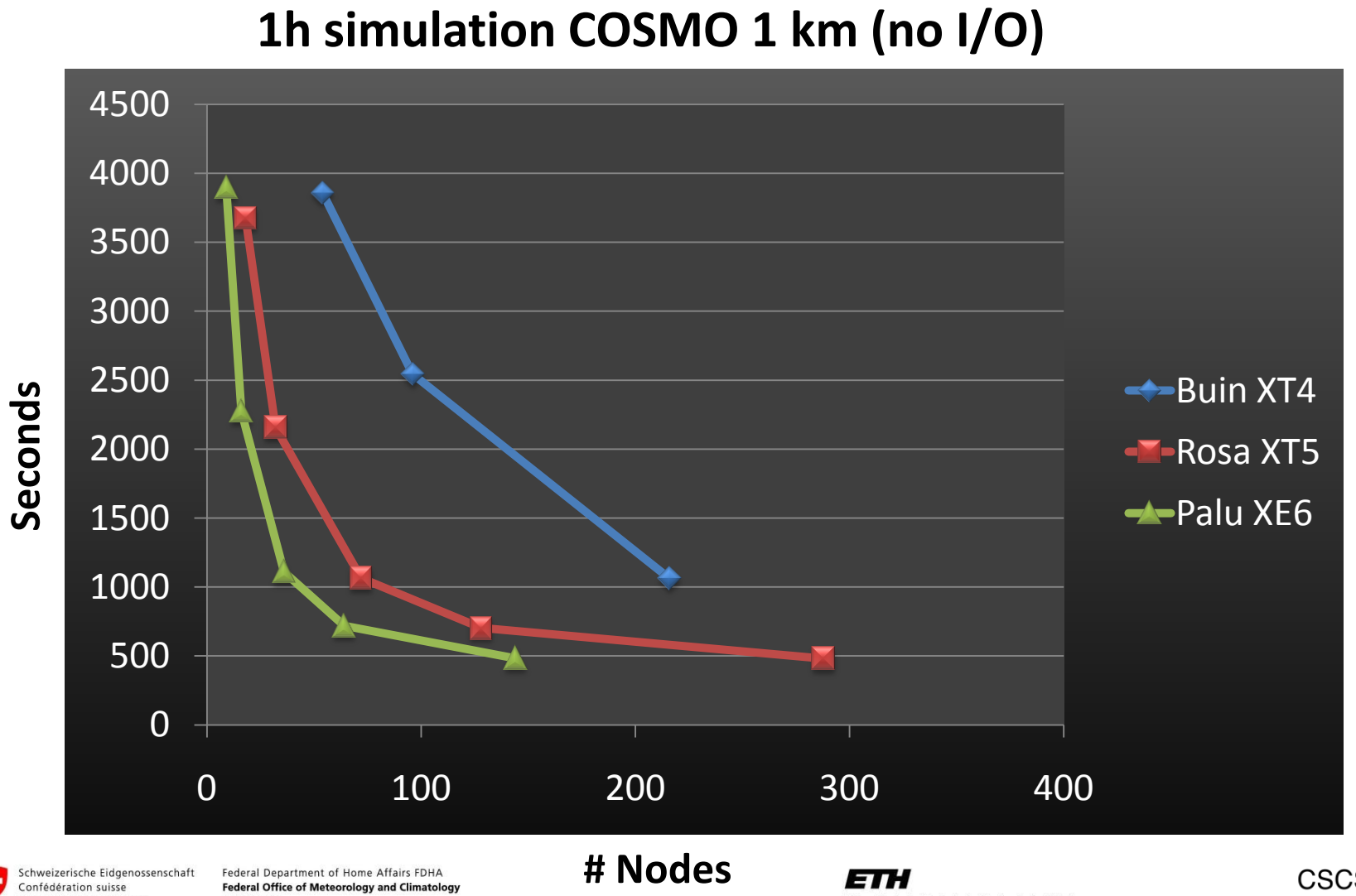
COSMO 1 km Performance on Cray XT4/XT5/XE6
Considered Systems

Some Technical Details	*Cray XT4 (Buin)*	*Cray XT5 (Rosa)*	*Cray XE6 (Palu)*
# cores (total)	1'056	22'128	4'224
# sockets	264	3'688	352
# cores/socket (name)	quad-core AMD Opt. *Barcelona*	hexa-core AMD Opt. *Istanbul*	12-core AMD Opt. *Magny-cours*
core freq. [GHz]	2.4	2.4	2.1
MEM/core [GB]	2	1.33	1.33
MEM Total [TB]	2.1 (DDR)	28.8 (DDR2-800)	5.6 (DDR3-1333)
MEM Bandw./core[GB/s]	2.6	2.6	3.6
Interconnect [GB/s] (name)	9.6 (<5 µs) *SeaStar2*	9.6 (<5 µs) *SeaStar2*	10.6 (~1.2 µs) *Gemini*
# racks	3	20	2

COSMO 1 km Performance on Cray XT4/XT5/XE6
Benchmark Results (# Cores)



COSMO 1 km Performance on Cray XT4/XT5/XE6
Benchmark Results (# Nodes)



- Actual operational suite of MeteoSwiss at CSCS and performance
- Performance on different Cray architectures
- Swiss HP2C Initiative
- HP2C Project for COSMO
  - Performance Analysis
  - Investigated Hybrid Parallelization via OpenMP/MPI
  - Parallelizing I/O
  - New Dynamical Core
- Outlook of future suites of MeteoSwiss at CSCS
- Conclusion

HP2C Initiative
Overview

Swiss Platform for High-Performance and High-Productivity Computing (HP2C , *www.hp2c.ch*)

- Prepare the Swiss HPC community for disruptive and potentially revolutionary developments in computer architectures during the coming decade.
- Supported in the frame of the Swiss National Initiative for High-Performance Computing and Networking by:
  - Swiss University Conference
  - ETH Domain (the 2 Swiss Federal Institutes of Technology)

HP2C Initiative
Framework

- Based on a **co-design** concept, which bring togheter:
  - domain scientists,
  - computer engineers,
  - hardware architects

  to design a *"system".*

  (= application + software + hardware)

HP2C Initiative

Projects Granted (2010-2012)

- BigDFT - Large scale *Density Functional Electronic Structure Calculations* in a Systematic Wavelet Basis Set
- Cardiovascular - HPC for *Cardiovascular System Simulations*
- COSMO - *Regional Climate and Weather Modeling* on the Next Generations High-Performance Computers: Towards Cloud-Resolving Simulations
- Cosmology - Computational *Cosmology* on the Petascale
- CP2K - New Frontiers in *ab initio Molecular Dynamics*
- Ear Modeling - Towards the Building of new *Hearing Devices*
- Gyrokinetic - Gyrokinetic Numerical Simulations of Turbulence in *Fusion Plasmas*
- MAQUIS - Modern Algorithms for Quantum Interacting Systems;
- Petaquake - Large-Scale Parallel Nonlinear Optimization for High Resolution 3D-Seismic Imaging
- Selectome - Selectome, looking for Darwinian Evolution in the Tree of *Life*
- Supernova - Productive 3D Models of *Stellar Explosions*

- Actual operational suite of MeteoSwiss at CSCS and performance
- Performance on different Cray architectures
- Swiss HP2C Initiative
- HP2C Project for COSMO
  - Performance Analysis
  - Investigated Hybrid Parallelization via OpenMP/MPI
  - Parallelizing I/O
  - New Dynamical Core
- Outlook of future suites of MeteoSwiss at CSCS
- Conclusion

HP2C Initiative
COSMO Project

Task 1: New high-resolution cloud-resolving climate model.

Task 2: Refactoring the COSMO model:

- *Performance analysis* of the current version.
- Investigate hybrid parallelization using *OpenMP/MPI.*
- Investigate *Parallel I/O* possibilities.

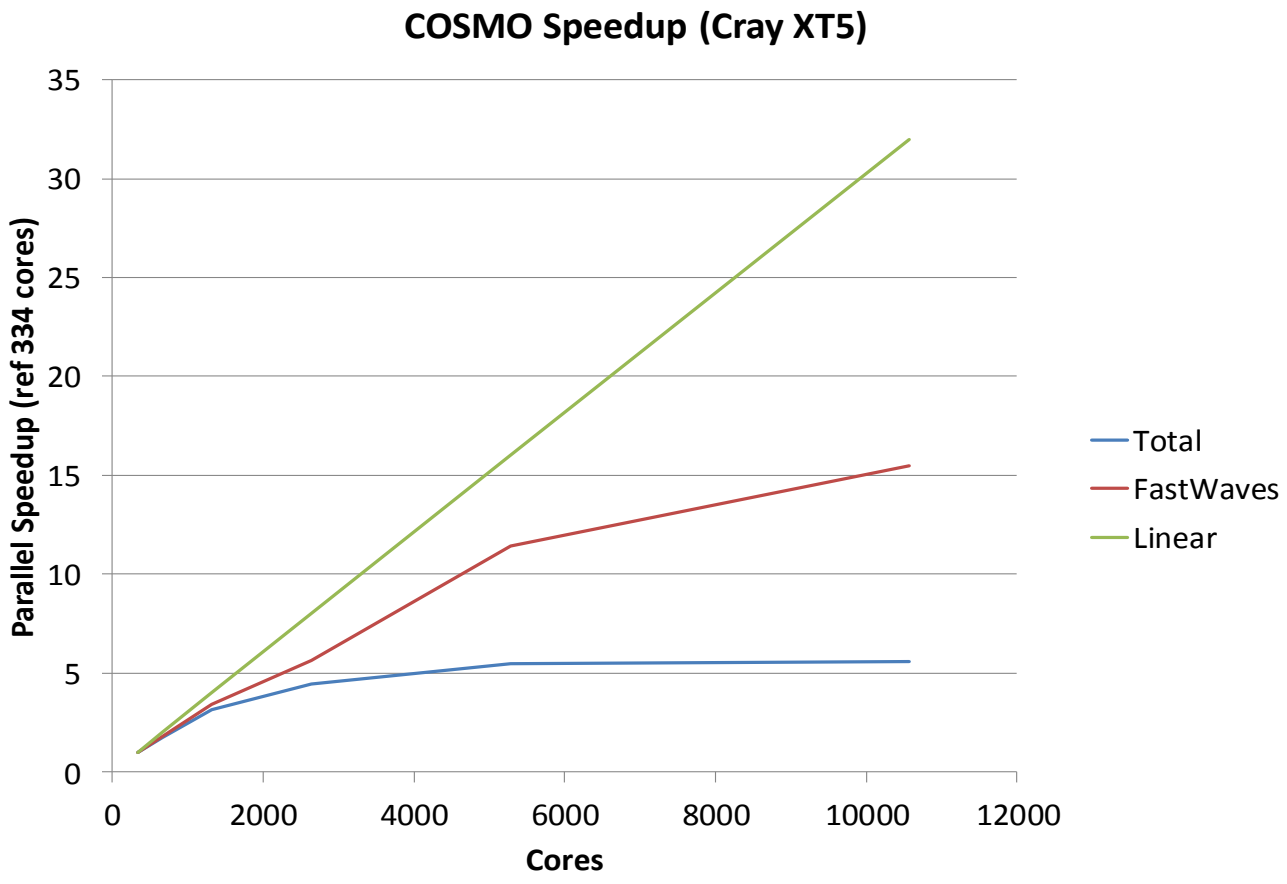
Task 3: Rewrite the COSMO dynamical core:

- Rewrite the dynamical core for current/emerging HPC architectures.
- Develop one code base for both CPU and GPU.
- Adapt physical parametrizations to new code design.

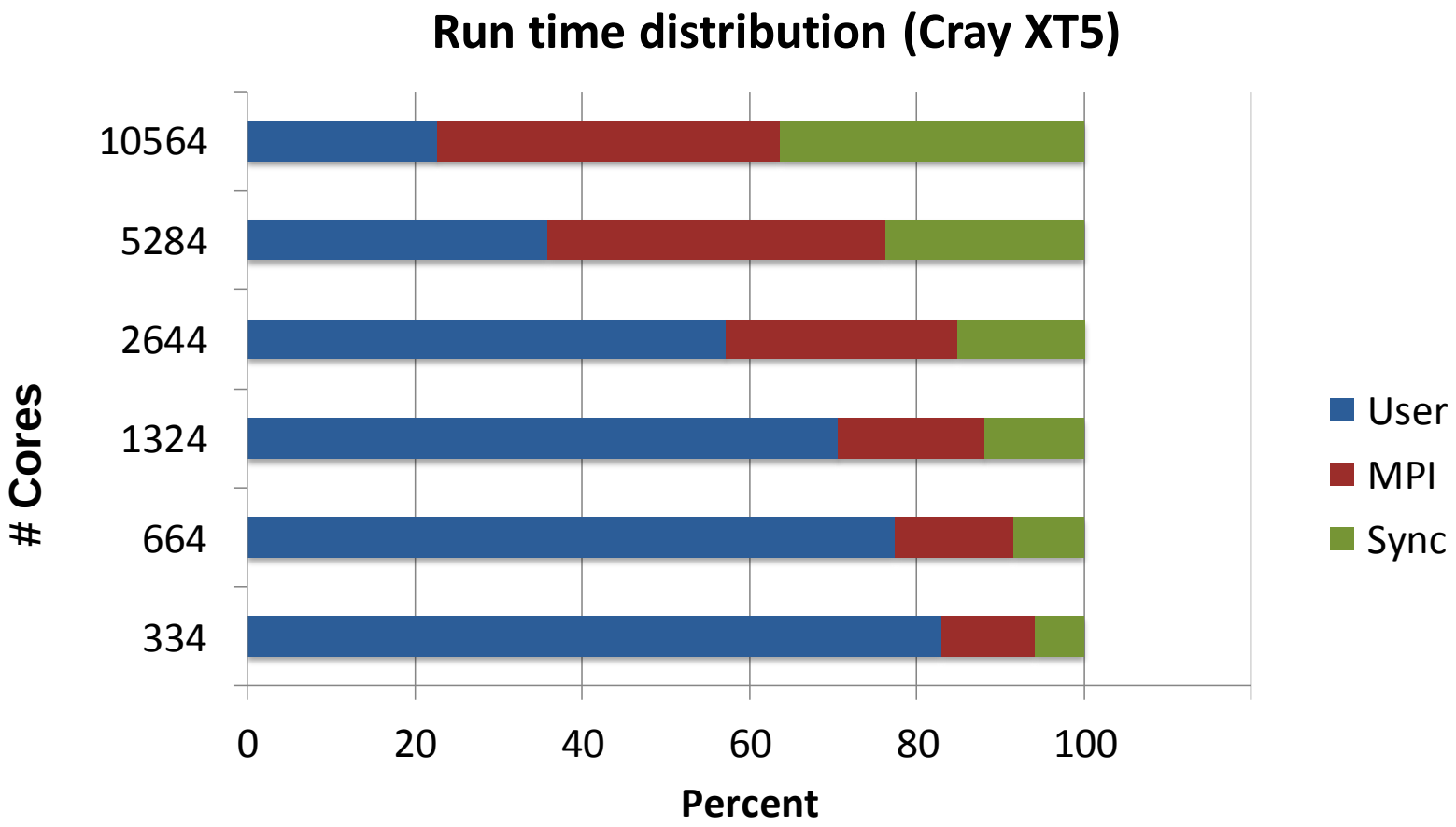
Collaboration between ETH Zurich, CSCS, MeteoSwiss, DWD, Supercomputing Systems AG and other partners.

- Actual operational suite of MeteoSwiss at CSCS and performance
- Performance on different Cray architectures
- Swiss HP2C Initiative
- HP2C Project for COSMO
  - Performance Analysis
  - Investigated Hybrid Parallelization via OpenMP/MPI
  - Parallelizing I/O
  - New Dynamical Core
- Outlook of future suites of MeteoSwiss at CSCS
- Conclusion

COSMO Scaling

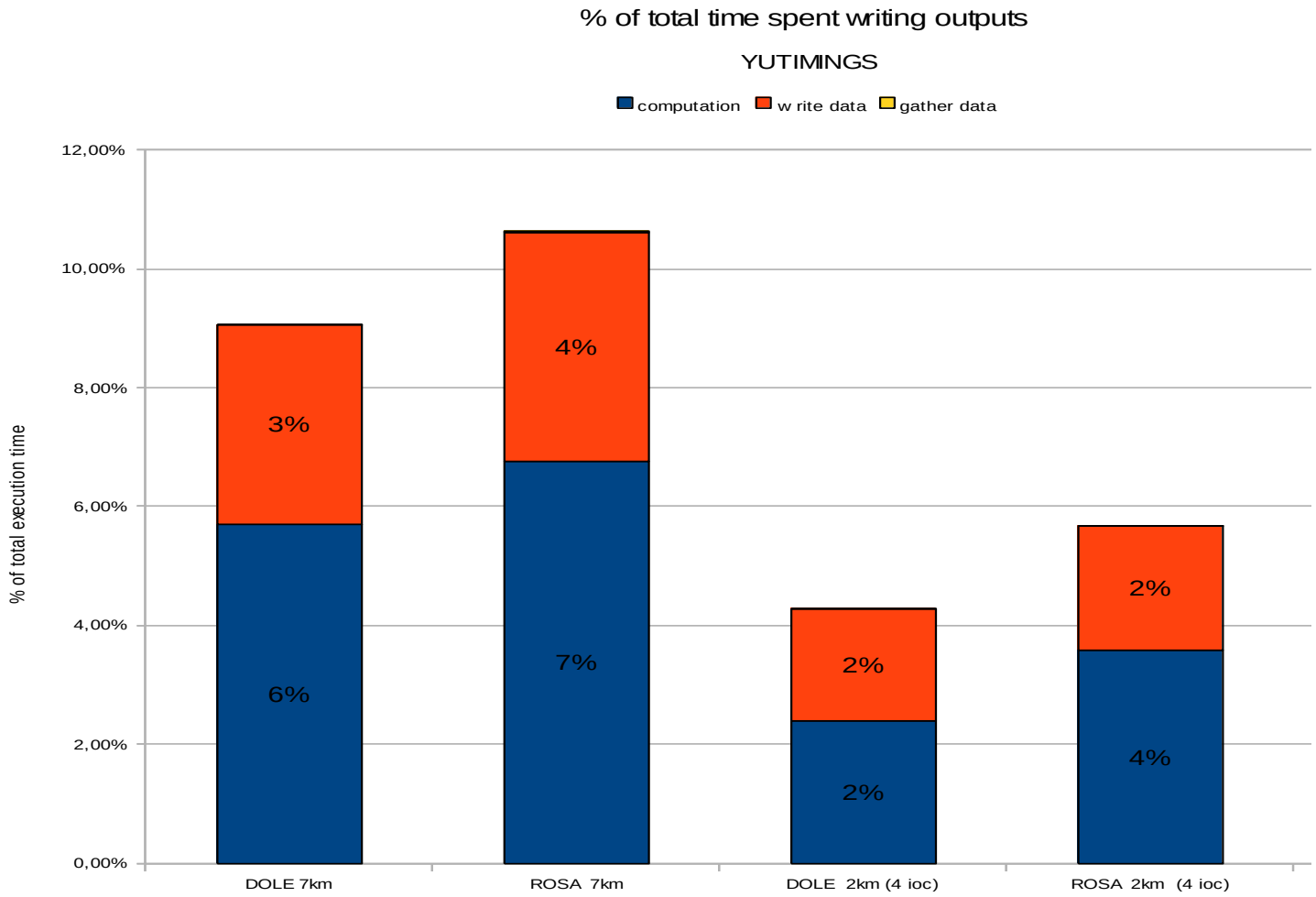


COSMO Run-time Distribution



HP2C COSMO Project – Performance Analysis

COSMO Output Overhead (7&2 Km)



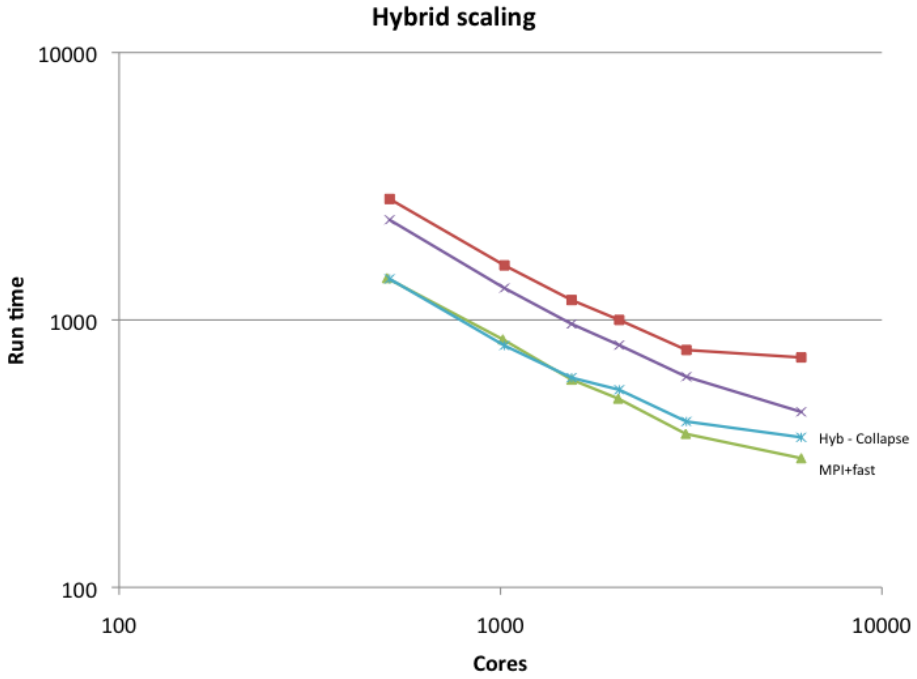
- Actual operational suite of MeteoSwiss at CSCS and performance
- Performance on different Cray architectures
- Swiss HP2C Initiative
- HP2C Project for COSMO
  - Performance Analysis
  - Investigated Hybrid Parallelization via OpenMP/MPI
  - Parallelizing I/O
  - New Dynamical Core
- Outlook of future suites of MeteoSwiss at CSCS
- Conclusion

HP2C COSMO Project – Hybrid OpenMP/MPI
Investigated Hybrid Parallelization

- Inserted OpenMP PARALLEL DO directives on outermost loop:
  - Over 600 directives inserted.
  - Also attempted to use of OpenMP 3.0 COLLAPSE directive.
  - Also enabled use of SSE instructions on all routines (previously only used on some routines).

HP2C COSMO Project – Hybrid OpenMP/MPI

Main Outcome

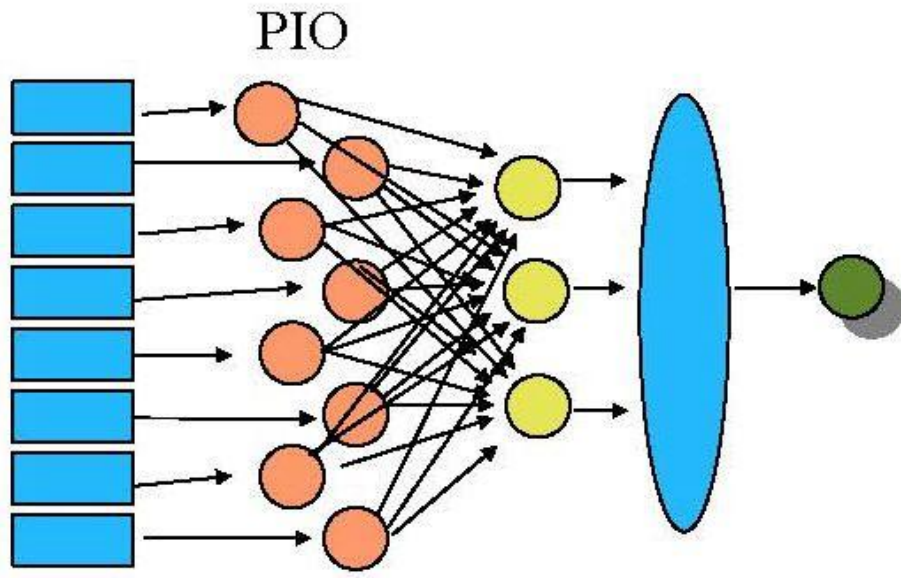


- Loop level parallelism can achieve some modest performance gains:
  - Can require many threaded loops -> OpenMP overhead.
  - Can require a lot more software engineering to maintain.

- Actual operational suite of MeteoSwiss at CSCS and performance
- Performance on different Cray architectures
- Swiss HP2C Initiative
- HP2C Project for COSMO
  - Performance Analysis
  - Investigated Hybrid Parallelization via OpenMP/MPI
  - Parallelizing I/O
  - New Dynamical Core
- Outlook of future suites of MeteoSwiss at CSCS
- Conclusion

Approach Using Parallel I/O Libraries

- Used properly, parallel I/O might alleviate I/O bottlenecks.
- NetCDF-4, pNetCDF, PIO are parallel I/O libraries which can be quite easily used for improving I/O performance.



HP2C COSMO Project – Parallelizing I/O
Investigate In-situ Visualization Techniques

The traditional post-processing model *"compute-store-analyze"* *does not scale* because I/O to disks is the slowest component.

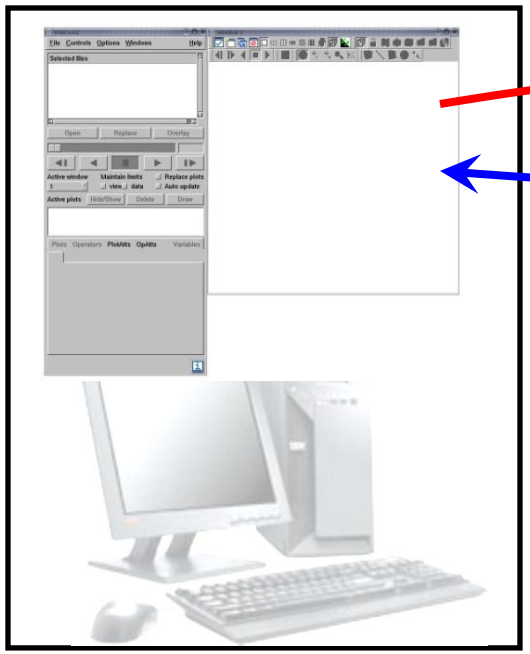
Consequences

- Datasets are often under-sampled on disks.
- Many time steps are never archived.
- ***It often needs a supercomputer to re-load and visualize supercomputer results.***

VisIt

https://wci.llnl.gov/codes/visit

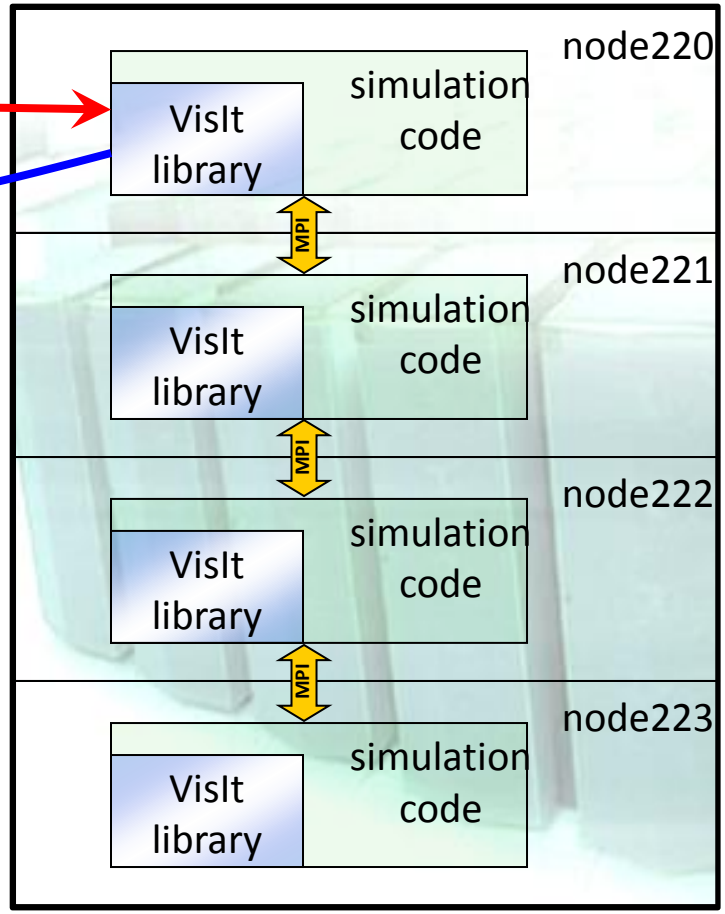
Desktop Machine



Commands

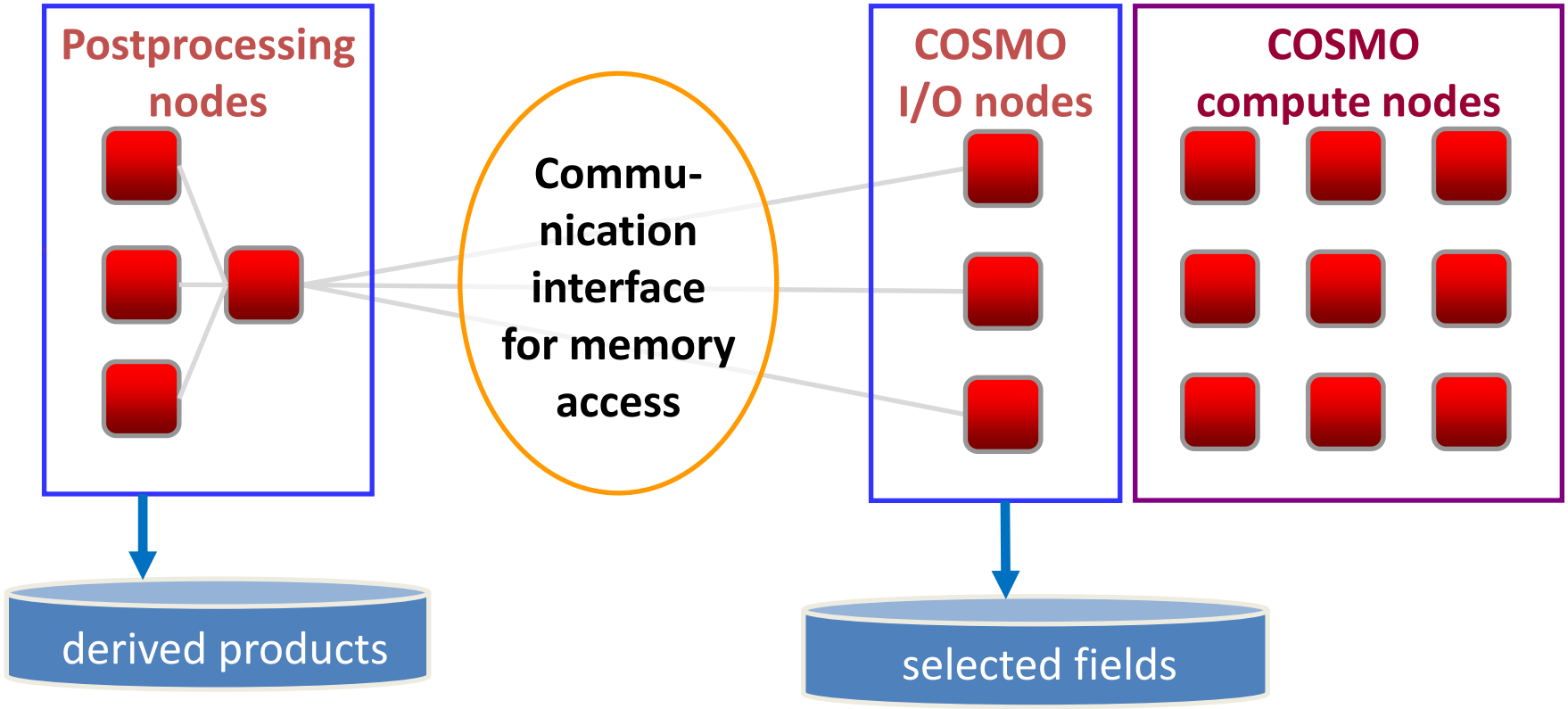
Images

HPC System



No pre-defined visualization scenario needs to be defined.

Using In-situ Techniques for Data Postprocessing

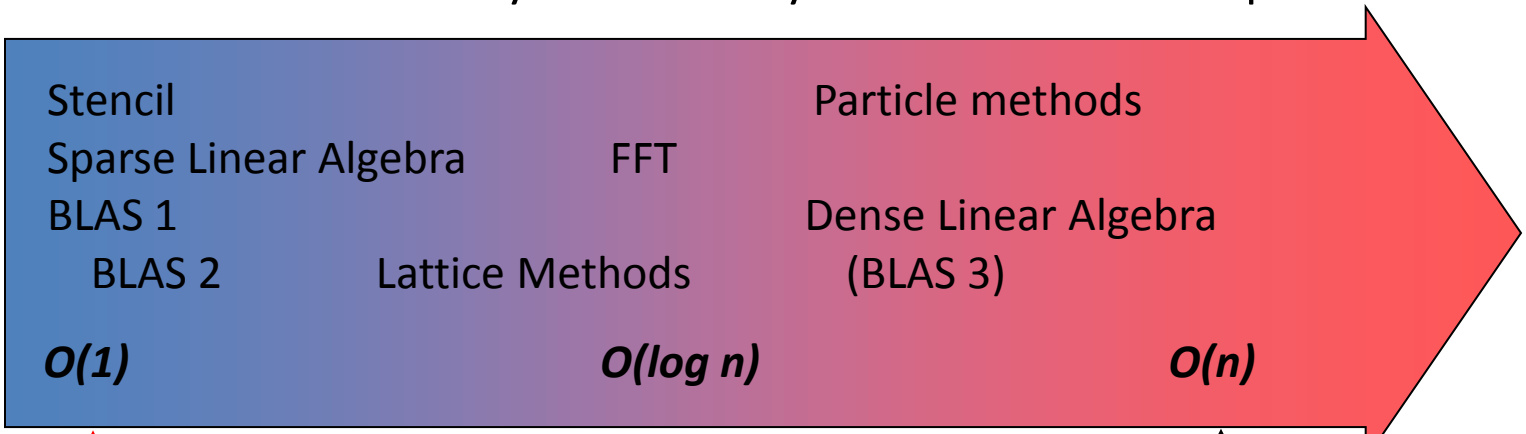


- Substantial I/O reduction is expected.
- Parallel I/O techniques can be easily applied.

- Actual operational suite of MeteoSwiss at CSCS and performance
- Performance on different Cray architectures
- Swiss HP2C Initiative
- HP2C Project for COSMO
  - Performance Analysis
  - Investigated Hybrid Parallelization via OpenMP/MPI
  - Parallelizing I/O
  - New Dynamical Core
- Outlook of future suites of MeteoSwiss at CSCS
- Conclusion

Motivation

- **Arithmetic Intensity (= FLOPs per memory access)**
  - High arithmetic intensity ⇨ processor bound ⇨ high %peak
  - Low arithmetic intensity ⇨ memory bound ⇨ low %peak



COSMO dynamical core (stencils on structured grid)

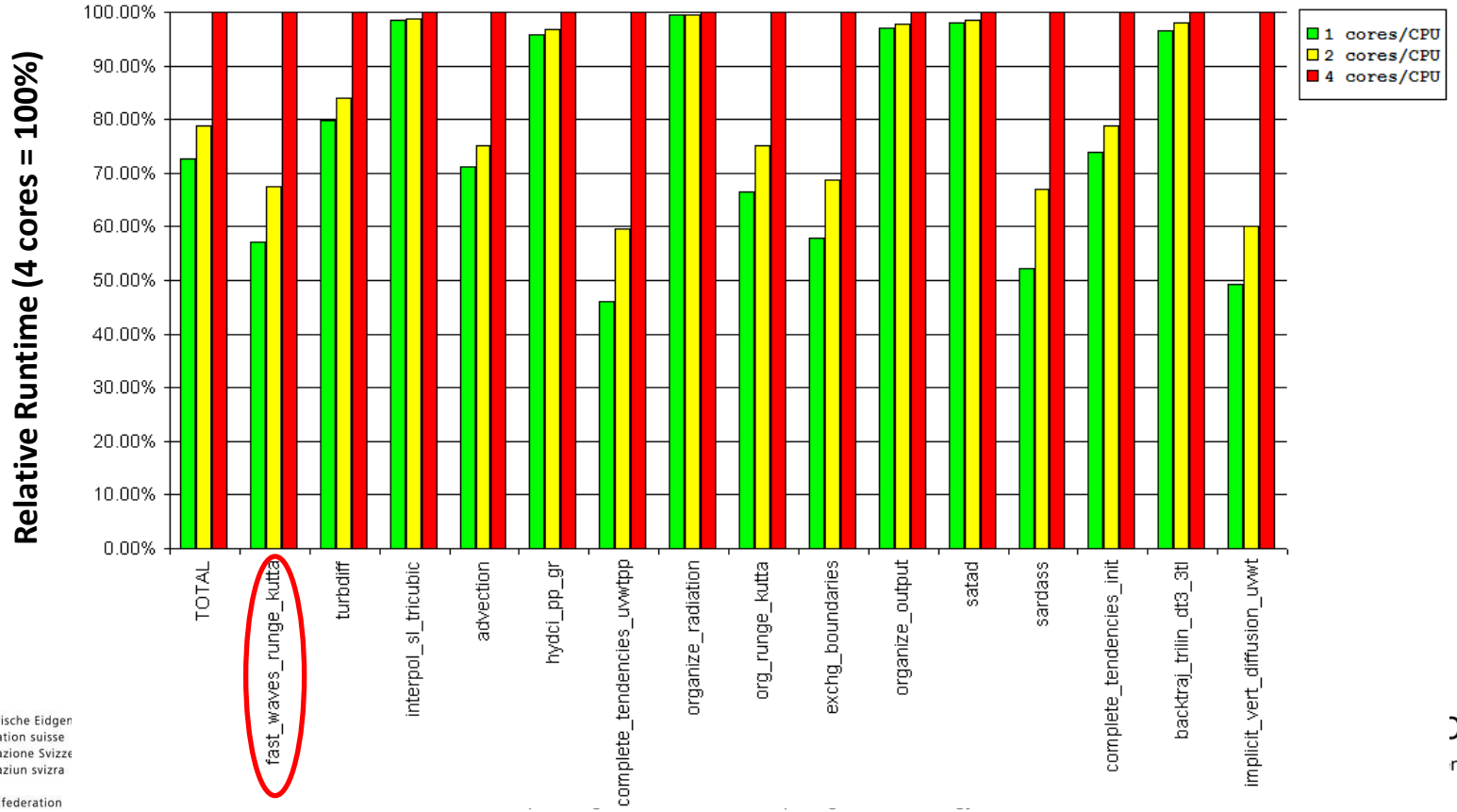
Top500 (Linpack)
Focus of HPC system design

- **COSMO-2 runs with ~3% peak on Cray XT4**

HP2C COSMO Project – New Dynamical Core

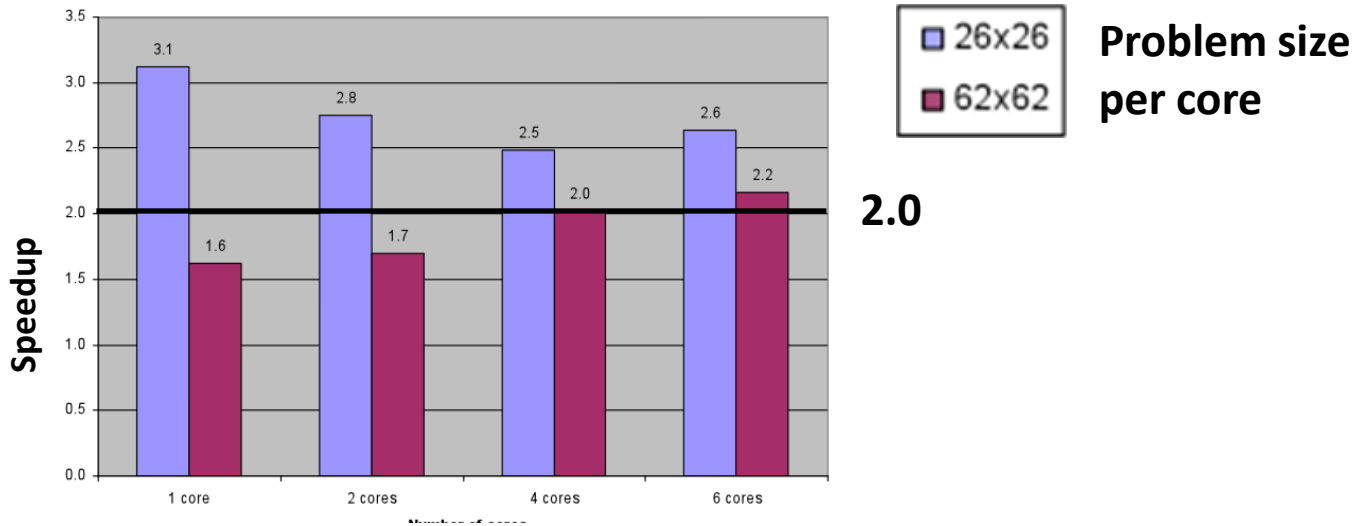
Analysis: Memory Scaling

- Benchmark on Cray XT4 (Quad-Core AMD Opteron Budapest 2.3 GHz).
- Keep total #cores constant but change #cores/CPU used.
- Cores on a CPU have to share memory bandwidth.
- Baseline (100%, indicated by red bars) is all cores used (4 per CPU).



HP2C COSMO Project – New Dynamical Core
Feasibility Study

- **Goals**
  - Rewrite a reduced version of dynamical core (containing all important algorithmic motifs) in a prototype code.
  - Reduce number of for memory accesses.
  - Do not use any optimizations that could not be accepted by the COSMO community.
- **Results**



- Actual operational suite of MeteoSwiss at CSCS and performance
- Performance on different Cray architectures
- Swiss HP2C Initiative
- HP2C Project for COSMO
  - Performance Analysis
  - Investigated Hybrid Parallelization via OpenMP/MPI
  - Parallelizing I/O
  - New Dynamical Core
- Outlook of future suites of MeteoSwiss at CSCS
- Conclusion

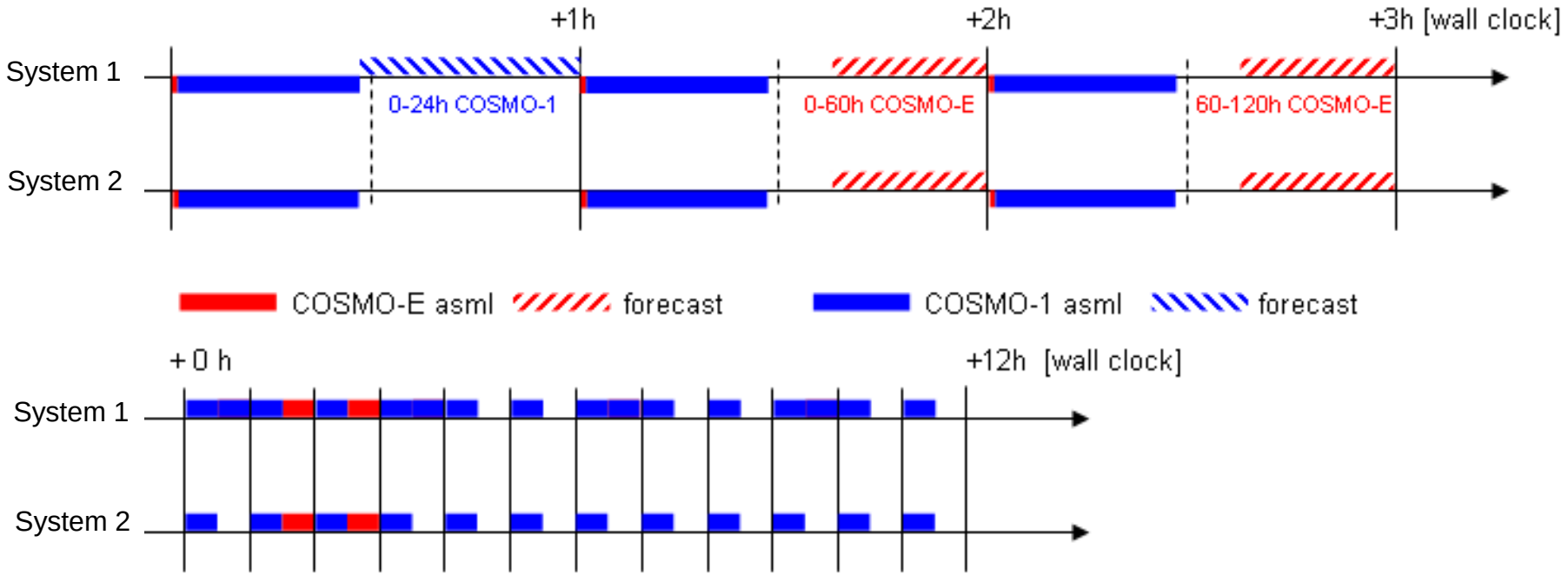
Outlook of Future Suites at MeteoSwiss

Improve Regionalization & Provide Probability Information

COSMO-1: 1 km mesh size, 8 x 24h deterministic forecast

COSMO-E: 3 km mesh size, 2 x 5d ensemble forecast

- Ensemble Data Assimilation (LETKF) with **40** members for both version.
- Possible Implementation on twin systems:
  - System 1 for operational deterministic parts, System 2 as fail-over.
  - Ensemble size reduced in case of failure of one System.



Conclusion

- We are investing in the preparation of the future COSMO to higher resolution and new HPC systems.
- A co-design project involves domain scientists, computer engineers and hardware architects.
- Approach
  - Detailed performance analysis of existing code.
  - Refactoring of the code on current hardware architectures.
  - Rewriting of the application core to run on CPU and GPU (even moving to new programming languages).
  - Addressing I/O problem with innovative approaches.

Acknowledgements

Thomas Schulthess (CSCS)

Ulrich Schättler (DWD)

Andre Walser (MeteoSwiss)

William Sawyer (CSCS)

Oliver Fuhrer (MeteoSwiss)

Tobias Gysi (SCS)

David Müller (SCS)

Matthew Cordery (CSCS)

Davide Tacchella (CSCS)

Jean-Marie Bettems (MeteoSwiss)

Neil Stringfellow (CSCS)

Petra Baumann (MeteoSwiss)

Jean Favre (CSCS)

Alam Sadaf Roohi (CSCS)

Jean-Guillaume Piccinali (CSCS)