The Abdus Salam International Centre for Theoretical Physics

High-Performance Computing

Ivan Girotto – igirotto@ictp.it

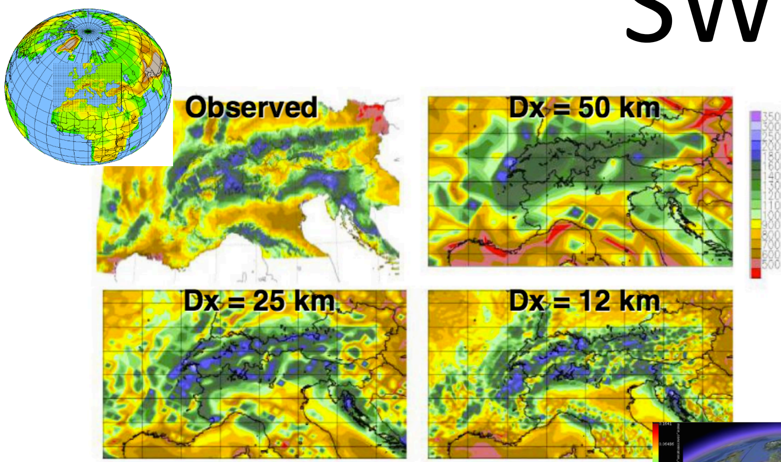
Information & Communication Technology Section (ICTS)

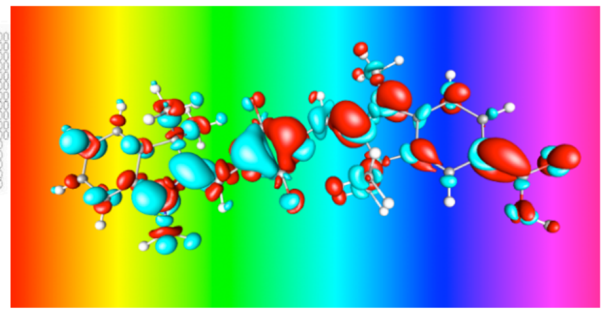
International Centre for Theoretical Physics (ICTP)

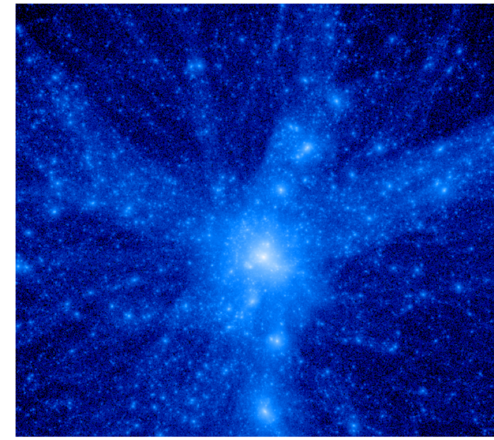
Why use Computers in Science?

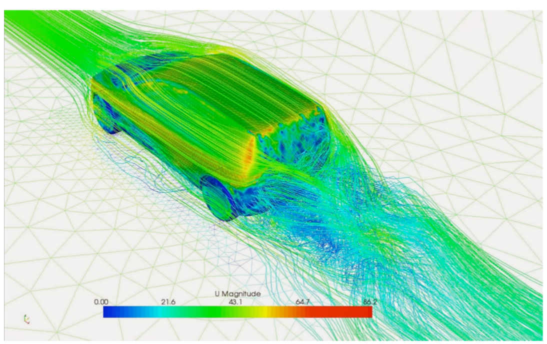
- Use complex theories without a closed solution: solve equations or problems that can only be solved numerically, i.e. by inserting numbers into expressions and analyzing the results
- Do "impossible" experiments: study (virtual) experiments, where the boundary conditions are inaccessible or not controllable
- Benchmark correctness of models and theories: the better a model/theory reproduces known experimental results, the better its predictions

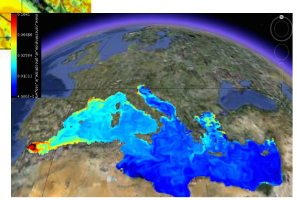
SW in Science



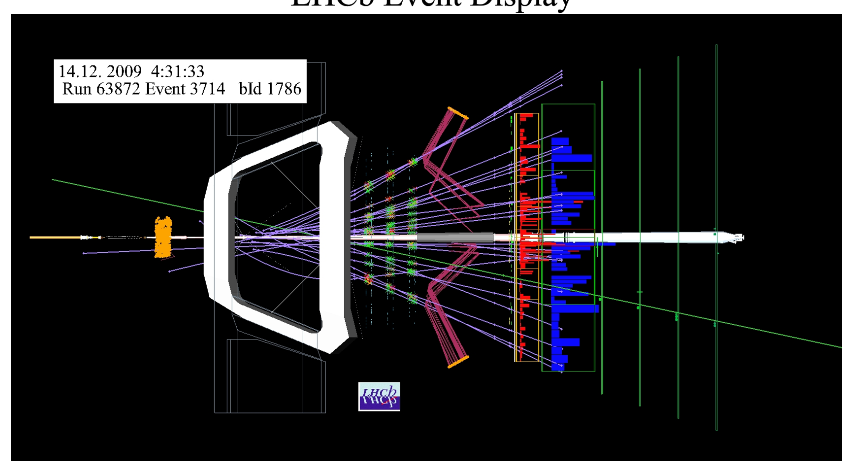








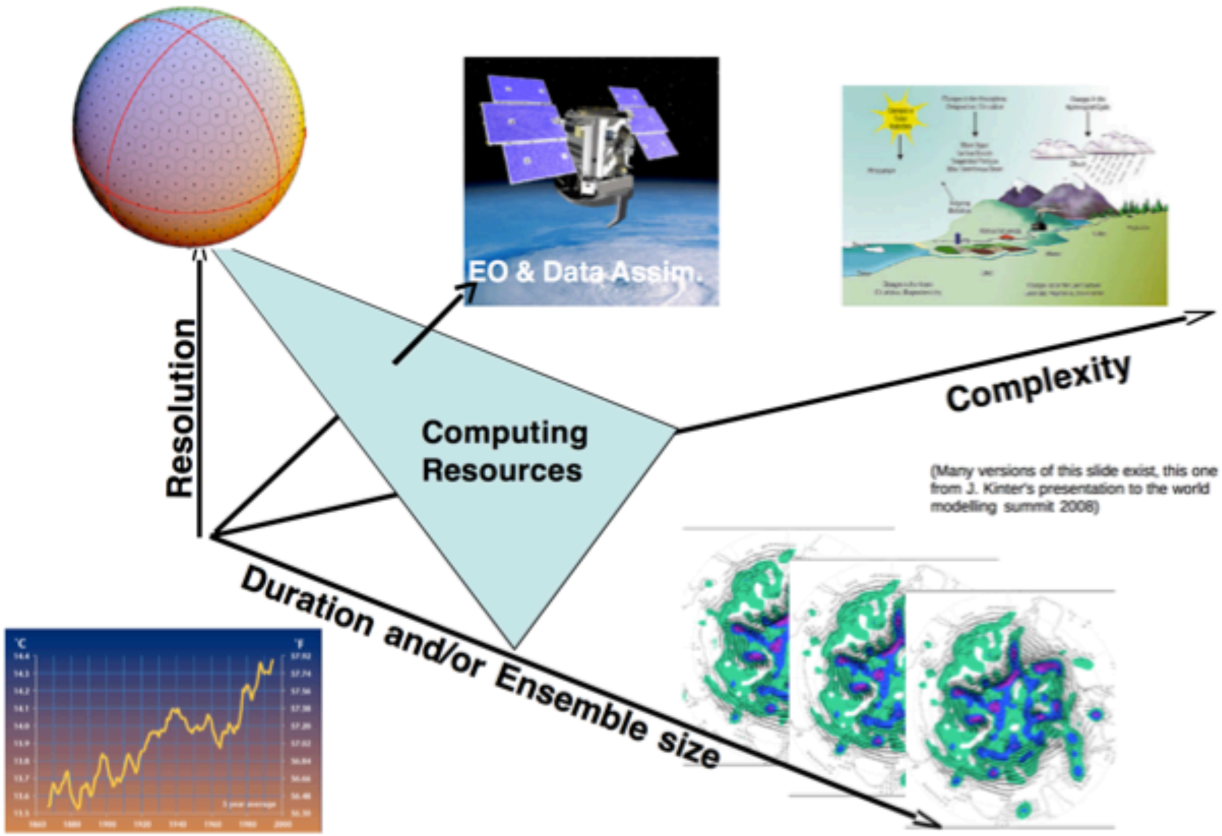
LHCb Event Display

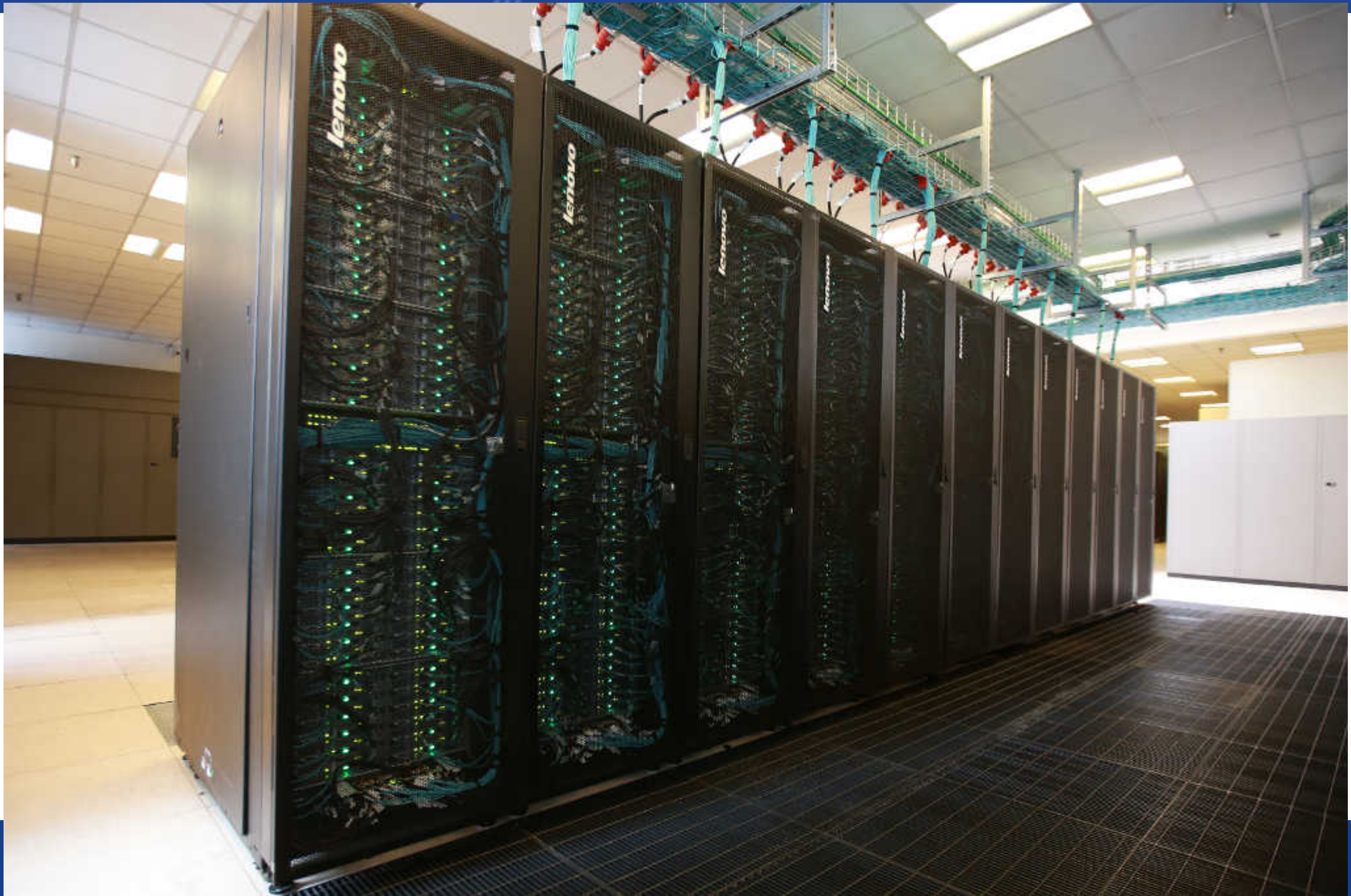


What is High-Performance Computing (HPC)?

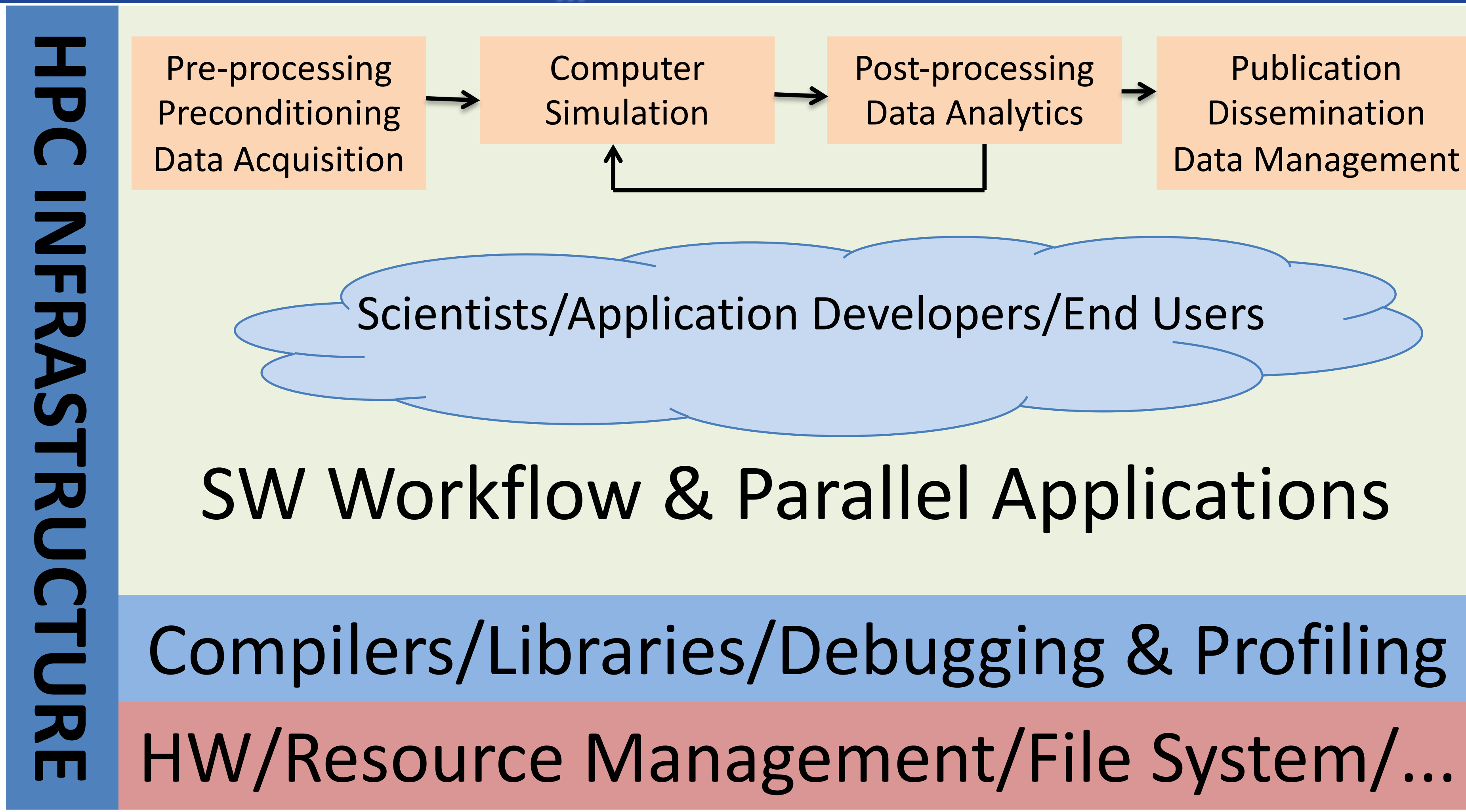
- Not a real definition, depends from the prospective:
  - HPC is when I care how fast I get an answer
  - HPC is when I foresee my problem to get bigger and bigger
- Thus HPC can happen on:
  - A workstation, desktop, laptop, smartphone!
  - A supercomputer
  - A Linux Cluster
  - A grid or a cloud
  - Cyberinfrastructure = any combination of the above
- HPC means also High-Productivity Computing

More & More Computing ...

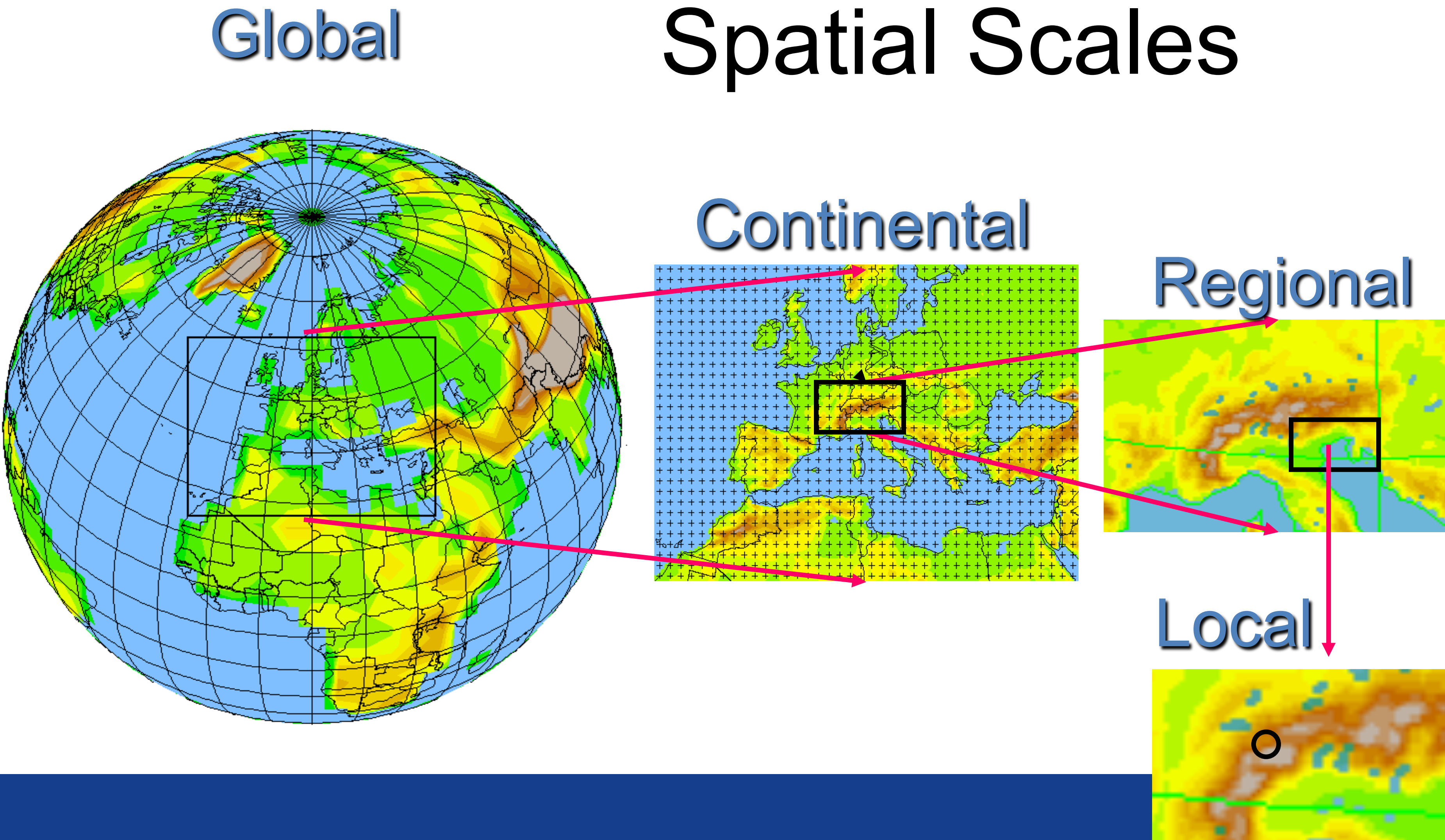



lenovo
lenovo
lenovo

The Abdus Salam International Centre for Theoretical Physics

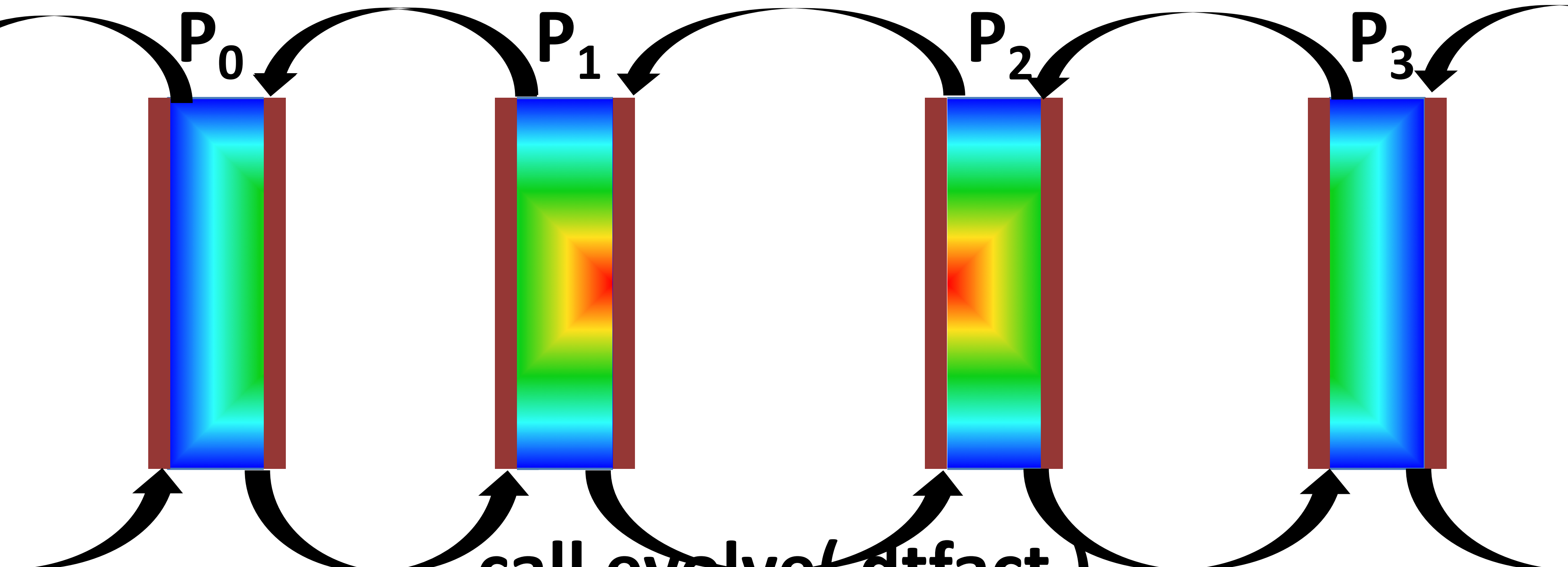


Spatial Scales



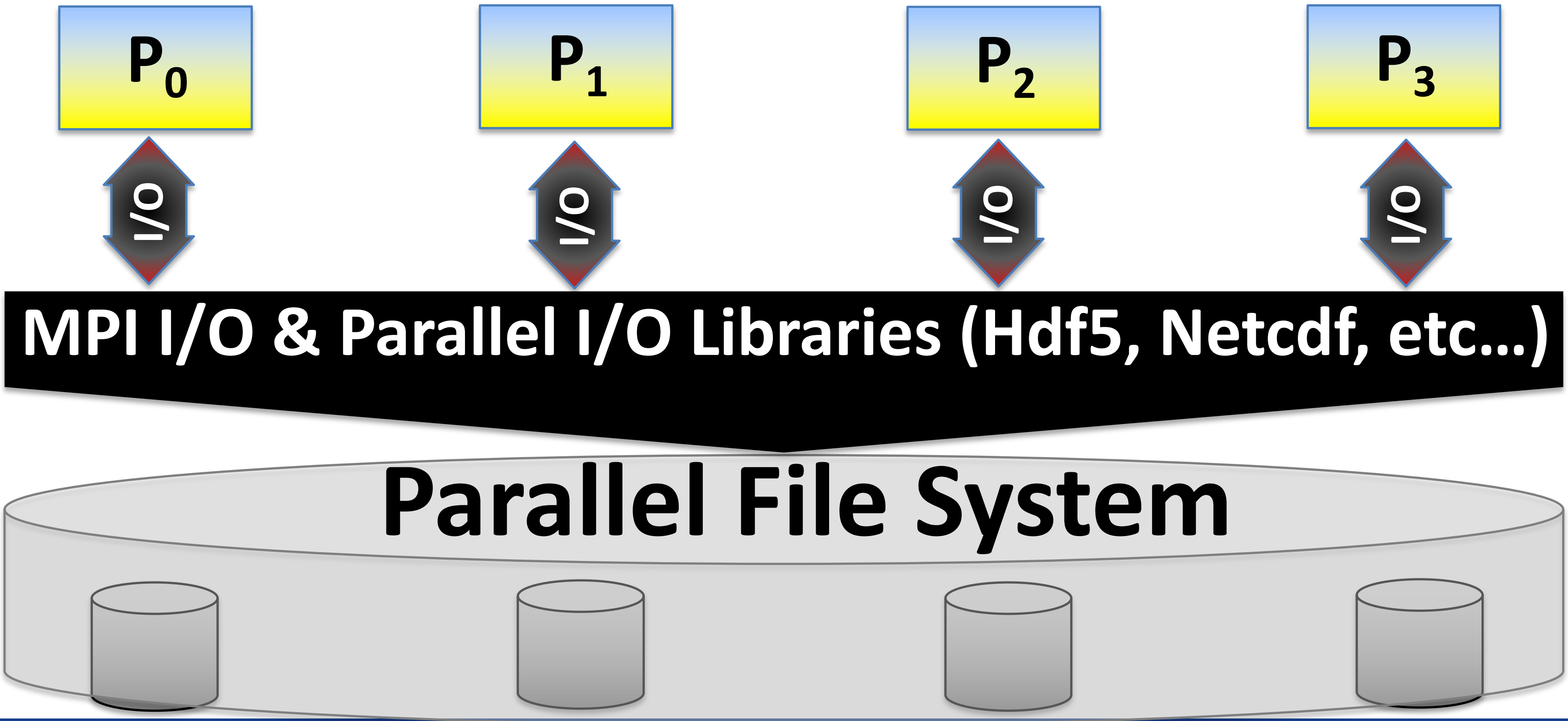
P_0

Domain Distribution



call evolve(dtfact)

Parallel I/O

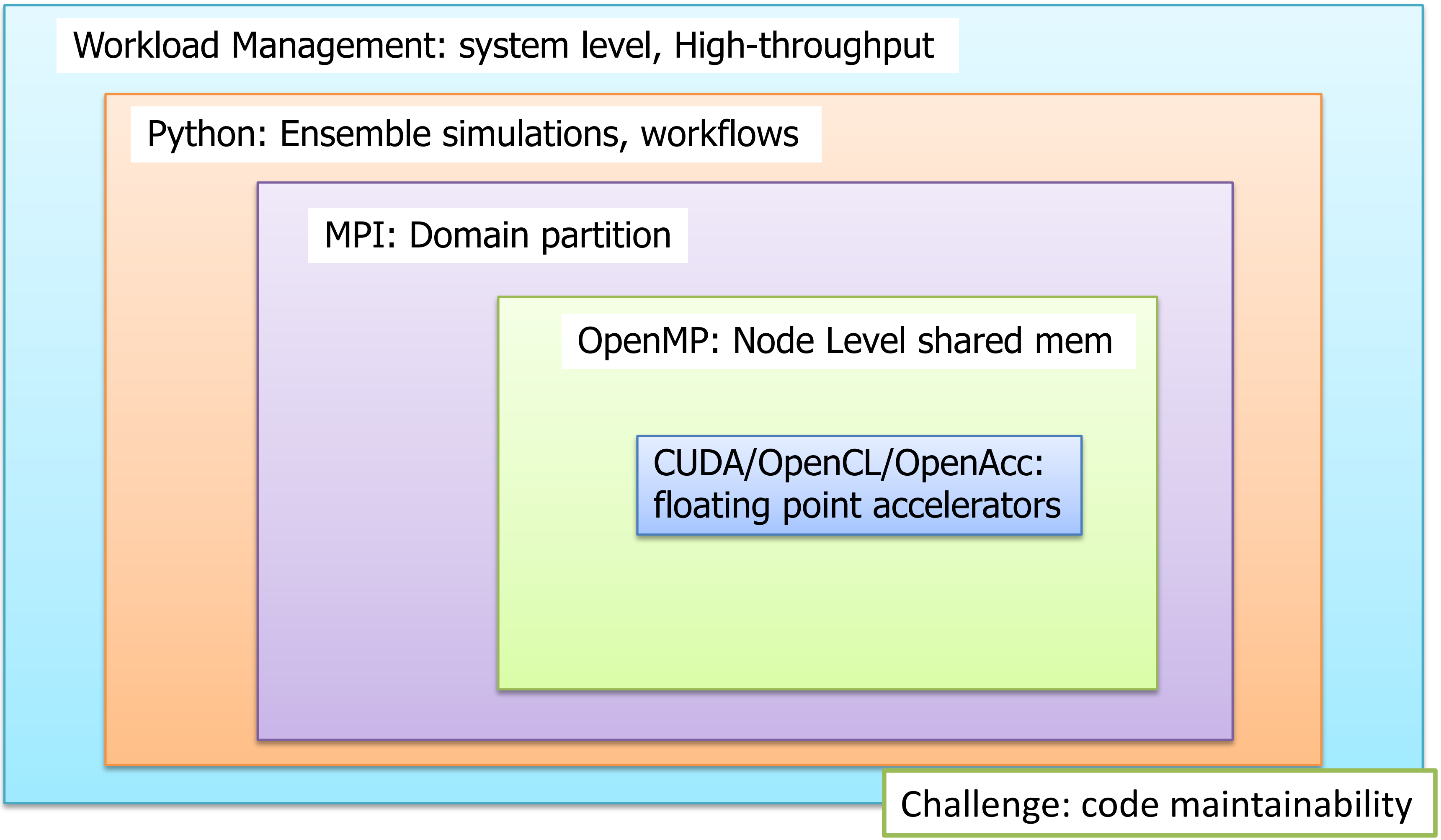

P0
P1
P2
P3
I/O
I/O
I/O
I/O
MPI I/O & Parallel I/O Libraries (Hdf5, Netcdf, etc…)
Parallel File System

A true story …

- [...] each month of simulation for Africa/South America, with 500 processors, is expected to take about 4 hours of compute time and produce 120GB of raw output files. We plan to run three 40 year present-day simulations followed by six future scenario simulations
- […] for Europe, we plan to run seven 140 year simulations. Each month of simulation, with 500 processors, is expected to take about 2.9 hours of compute time and produce 60GB of raw output files
- […] for the Alpine, we plan to run four 10 year simulations. Each month of simulation, with 500 processors, is expected to take about 5 hours of compute time and produce 120GB of raw output files

And in data ...

- The maximum peak raw I/O requested for running 20 parallel simulations is expected to be about 12TB/day. This data is going to be post-processed on-line and moved to a separate archive automatically, with a peak throughput of around 2.5TB/day on average. The total amount of raw data storage required for all simulations is 2PB over the 2 years period, which will be reduced to below 500TB after data processing.


Workload Management: system level, High-throughput
Python: Ensemble simulations, workflows
MPI: Domain partition
OpenMP: Node Level shared mem
CUDA/OpenCL/OpenAcc: floating point accelerators
Challenge: code maintainability

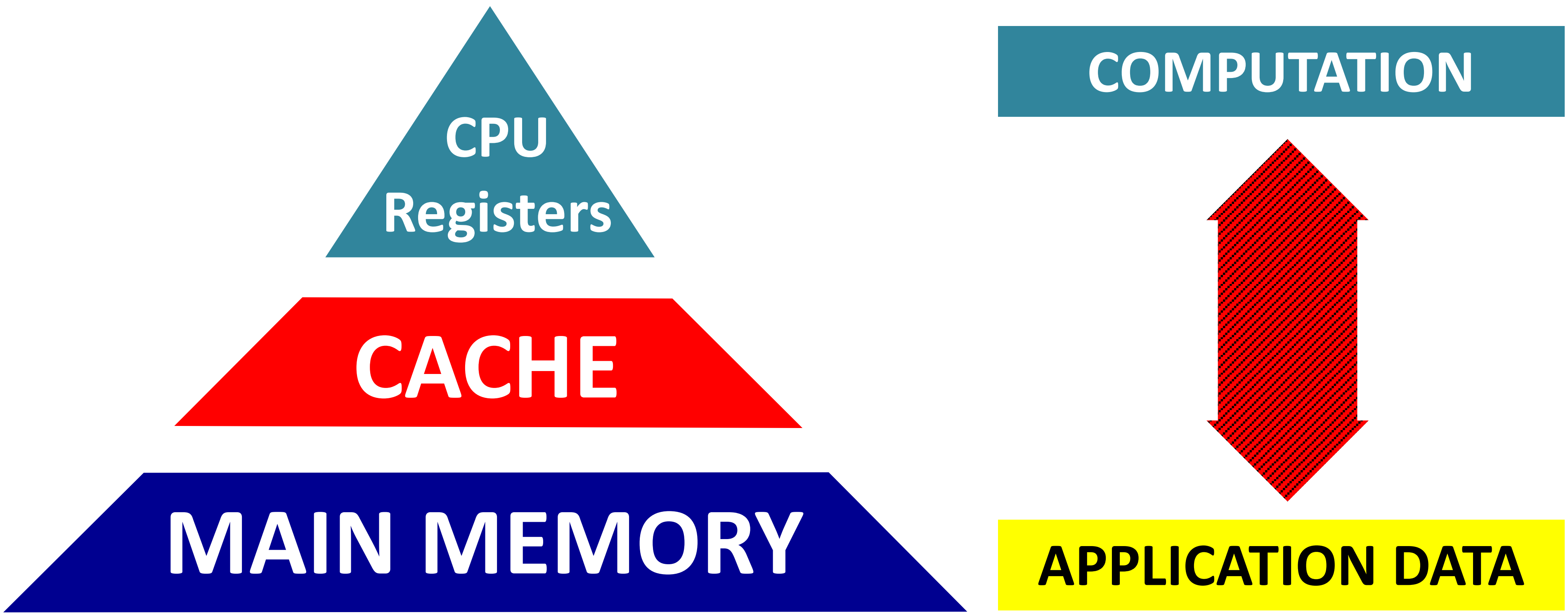
How can we make all of this happen in a single code?

Not a question of feasibility but of how we develop software:

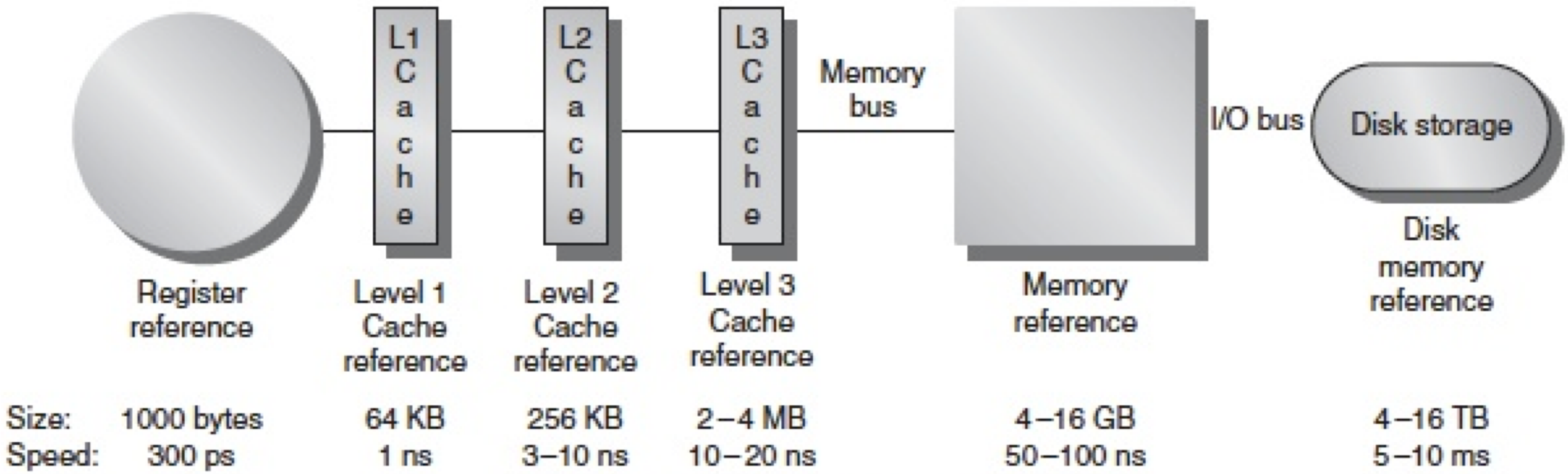
- Is every student developing their own software?
- Or are we re-using what others have done?
- Do we insist on implementing everything from scratch?
- Or do we build our software on existing libraries?

There has been a major shift on how we approach the second question in scientific computing over the past 10-15 years!

The CPU Memory Hierarchy



The CPU Memory Hierarchy



(a) Memory hierarchy for server

Conclusion

- Modern HPC applications can generate a huge amount of information
- It is an opportunity to better understand physical behaviors only if we master the production as well as the analysis of those information
- Real time monitoring and efficient data analysis are inescapable to make such scientific production sustainable at large scale

MHPC in pills: www.mhpc.it

- High-level educational program: not an Ms.C. program!
- Intensive training aimed to build knowledge in solving complex problems with an HPC approach
- Innovative, hands-on based training
- Aimed to people with strong interest in:
  - advanced programming for scientific computing
  - software optimization
  - management of computing platforms
  - data management and data analytics

Master in High Performance Computing

Thanks for your attention!!