e-infrastructure

CAPACITIES

MAPPER e-Infrastructure Use Case

10/04/2013

Derek Groen, Ilya Saverchenko

The Mapper project receives funding from the EC's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° RI-261507.

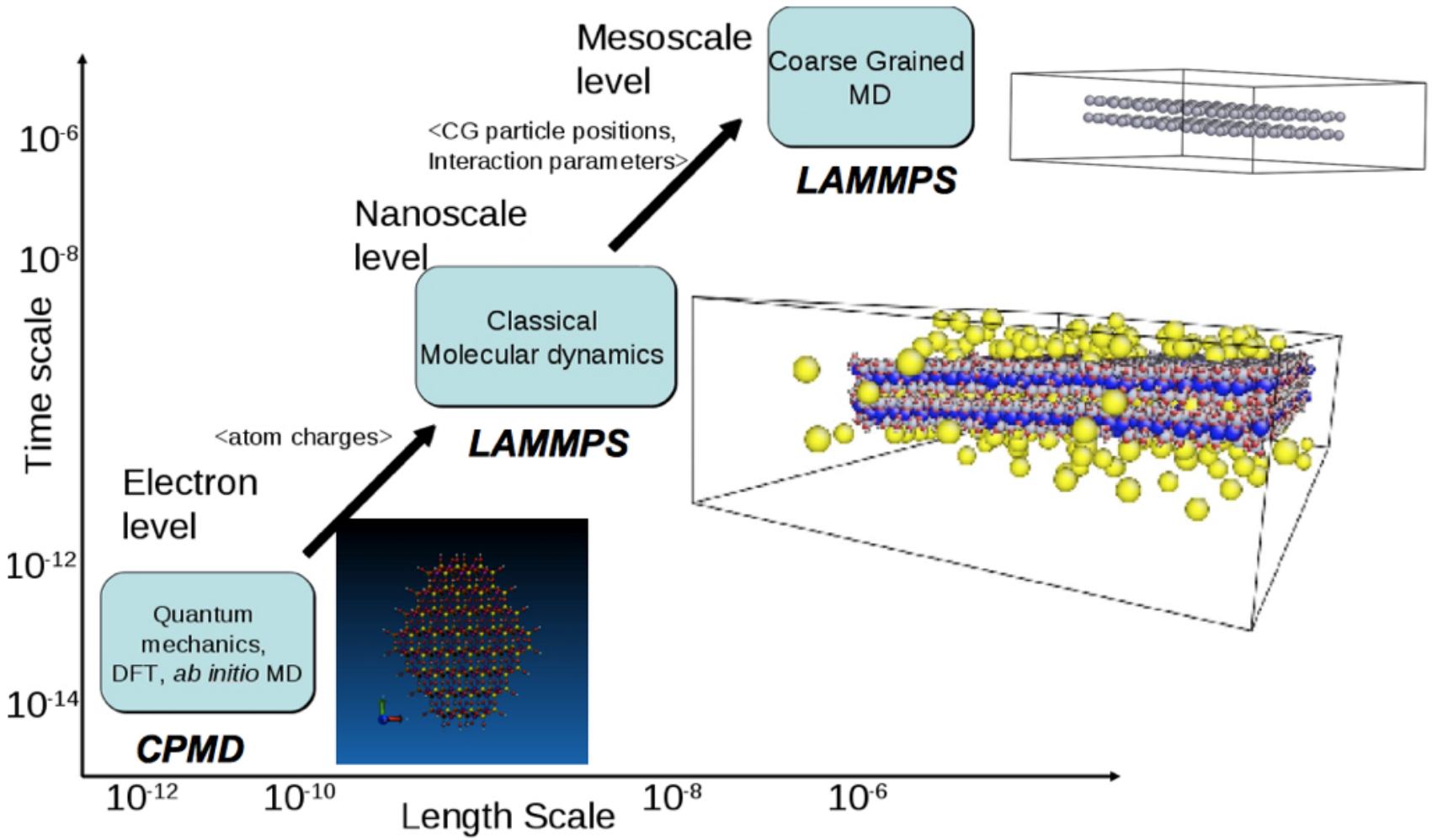
MAPPER & Multiscale Systems

- MAPPER project
  - Studying multiscale systems
  - 7 multiscale scenarios from 5 application areas
- MAPPER objectives:
  - Develop models for multiscale computing
  - Work and develop services supporting novel computational paradigms
  - Enable multiscale computing on European e-Infrastructures

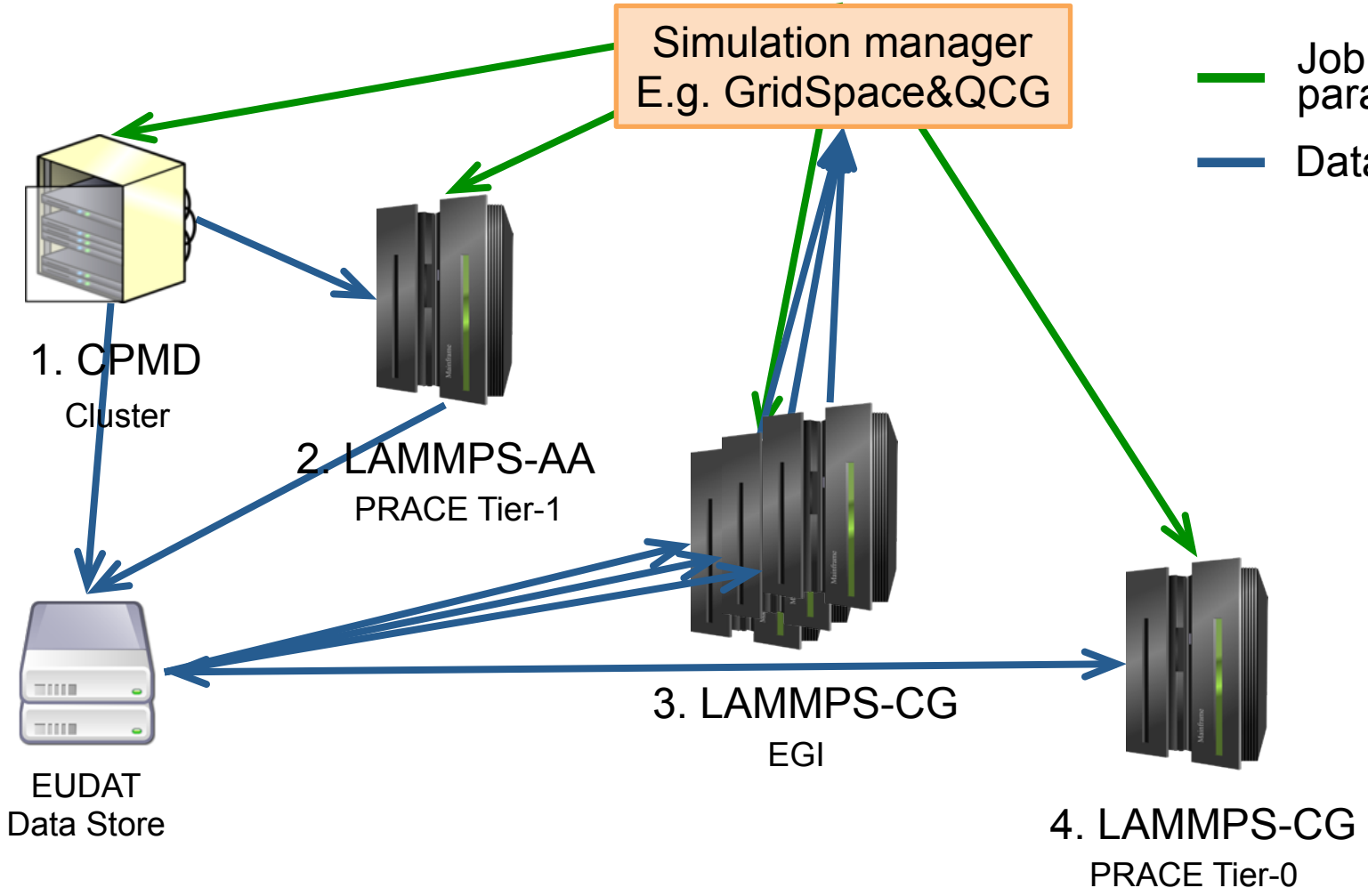
Nanomaterials use case

- We investigate mixtures of two different materials looking for an enhanced performance, e.g.:
  - Improved fire retardant properties
  - Similar performance to other composites at much lower filler volumes
- Steps:
  - Quantum mechanical simulation for calculating potentials for clay and polymer molecules
  - Atomistic MD simulation to determine accurate microscopic system properties
  - Iterative coarse-grained simulations intend to find the right potential for the coarse-grained system
  - Large-scale coarse-grained simulation to model the final system and obtain important macroscopic properties

Scale separation map



Execution scenario



Use case requirements

- Data
  - A uniform data access interface across infrastructures
  - High-performance file staging and file transfers
  - Reliable performance in cross-site network communications
  - Data services for long term storage and analysis of data
  - Publishing simulation results to support scientific publications and ensure reproducibility

Use case requirements

- Compute
  - Uniform access to compute resources
  - Resource selection according to computational requirements
- Information system
  - Availability and functionality of services
  - Properties of compute and storage services
    - Quota, capacity, performance
  - Network performance between service providers

mapper-project.eu

Step 1

- Goal: calculate energy potentials to be used in step 2
- Code: CPMD (www.cpmd.org), optionally CASTEP
- # of simulations: 1 (or a few)
- # of cores per simulation: <64
- Duration per simulation: ~24 hours
- Data produced per simulation: typically MBs, although the restart file is ~3GB
- Data transfer required: MBs before and after the simulations
- Site type: local cluster

Step 2

- # of simulations: 1
- # of cores per simulation: 1,024-8,192
- Duration per simulation: ~24h
- Data produced per simulation: ~1GB
- Data transfer required per simulation between PRACE site and the manager: ~1GB
- Access mechanisms required/supported:
  - Required: GridFTP, support for remote job submission using QCG Broker by adhering to common job submission interfaces

Step 3

- # of simulations: ~20-40 (one after another)
- # of cores per simulation: 16 – 256
- Duration per simulation: 1h - 4h
- Data produced per simulation: 75MB - 4GB
- Data transfer required per simulation between EGI site and the manager: 25 - 1GB
  - Rest are particle positions, which can be stored for future reference
- Access mechanisms required/supported:
  - Required: GridFTP, must have support for QCG job submission, preferably have QCG Computing installed
  - Advance reservation provides a performance benefit here

Step 4

- # of simulations: 1
- # of cores per simulation: 8,192-65,536 cores
- Duration per simulation: ~8h
- Data produced per simulation: 1TB+
- Data transfer required: MBs to start, 1TB+ afterwards
  - The particle positions of these simulations are to be stored for future reference and analysis
- Access mechanisms required/supported:
  - Required: GridFTP, support for QCG job submission