Exascale Oriented Solutions and Their Application to the Exploitation of Energy Sources

Rafael Mayo-García
CIEMAT Assoc Prof and Harvard-IACS Fellow

Rensselaer Polytechnic Institute
March, 28th 2017
• Will this be useful for me?

• Why HPC and energy production

• A hint on some exascale trends

• Some examples
• Energy is one of the major societal challenges that we must afford

• New computational methodologies are suitable to be applied to other fields

• Potential collaborations could be an outcome

• In the worst case, this will not take too long, I promise…
Energy is not a joke

“WHERE THE DEVIL IS ALL THAT SOLAR HEAT WE STORED UP LAST SUMMER?”
• CIEMAT (http://www.ciemat.es) is the main Spanish public organism devoted to research on energy and its related technologies
  – The Spanish DOE...

  – Really????
    - DOE budget: $32.5 Billion
    - CIEMAT budget: $140 Million
• CIEMAT was created in 1951 and was pioneering in supercomputing activities in Spain (1959)
  – It is logic to infer that it focuses on merging/comparing experimental and simulated results

• IACS is skilled in Stochastic Optimization Methods, Monte Carlo Methods for Inference and Data Analysis
  – Working on HPC & HPDA convergence

• Multidisciplinary aspects are a must

• Simulation (Computer Science) is probably the most transversal and complementary field nowadays
<table>
<thead>
<tr>
<th>Facilities</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total scientific-technical facilities</td>
<td>60</td>
</tr>
<tr>
<td>Number of laboratories</td>
<td>161</td>
</tr>
<tr>
<td>Other facilities (training, protection, medical service, ..)</td>
<td>23</td>
</tr>
</tbody>
</table>

- **Heliac Flexible TJ-II**
- **Pilot pelleting plant (300-500 kg/h)**
- **Dosimetric measurements in mannequin**
- **Safety System Analysis Laboratory**
  (Peca Facility)
- **Pilot membrane gas separation plant**
- **Solar Platform of Almería (PSA)**
- **Wind Testing Laboratory**
- **PET/CT of small animals for biomedical applications**
Approaching a Disruptive Exascale Computer Architecture
• In order to fully exploit a future exascale architecture, it is needed to study the mapping and optimization of the codes (and integrated kernels) of interest in terms of
  – New infrastructure
    ▪ New designed architectures
    ▪ GPUs, Phis, embedded processors
  – New developments in the underlying software infrastructure

• The goal is to optimize the performance but keeping a high degree of portability
  – The ratio flops/watt obtained should be analyzed as well

• Co-design
• Porting to architectures based on symmetric multicore processors with NUMA memory
  – Optimizing the performance

• Current main target architectures are Intel and AMD
  – New platforms based on ARM processors are being analyzed

• A key point will be load balancing and data placement
  – New scheduling algorithms able to improve locality
Some trends and current actions

- The management of the MPI level parallelism must be guaranteed for achieving a high scalability of the applications on HPC clusters with millions of cores
  - Creation of tools for migration of running parallel tasks inside clusters
  - Hierarchical MPI structures to manage coupled multiphysics problems
  - Parallel I/O optimization
  - Design of efficient check-pointing strategies
  - Fault tolerance strategies at system and application level
    - Creation of tools for migration...
Some trends and current actions

- Performance analysis ought to bear in mind all the parallel levels
  - Paraver, Triva, Ocelotl, TAU, etc

- Roof-line analyses inside a computational node to understand the bottlenecks of the architectures

- At the cluster level, network traffic, I/O traffic, and load balancing to pursue application scalability

- Performance prediction tools can be used to analyze the potential benefits of architecture or algorithm modifications
  - DIMEMAS, BOAST, etc.
Simulators for Exascale Computations
Main lines of research

- **Numerical schemes for Partial Differential Equations (PDE)**
  - Scalable implementations of high order schemes for wave propagation models
- **Scalable sparse linear solvers**
  - Generic (i.e. algebraic) parallel solvers for large sparse linear systems of equations
- **Adaptivity**
  - Mesh and (local) time-step adaptive algorithms in order to optimize the use of computational resources.
- **Data management**
• High order finite element methods and (standard and mimetic) finite difference schemes should be considered for both time- and frequency-domain

• A high level of parallelism is kept if suited to a mixed coarse grain/fine grain (MIMD/SIMD) parallelization targeting many-core heterogeneous systems
  – Time domain: Multiscale Hybrid-Mixed (MHM) methods combined with Discontinuous Galerkin (DG) or Stabilized Continuous Galerkin (SCG)
  – Frequency-domain: Hybridized DG formulations for reducing the number of globally coupled degrees of freedom
Both direct and hybrid direct/iterative solvers will be needed

The implementation of algebraic domain decomposition is a must

- See MaPHyS for example
- Parallel sparse direct solvers such as PaStiX2 for each sub-problem can be added too
- Krylov subspace methods can also be included on top
• Mainly applied to PDEs (numerical schemes) via algorithms

• They involve adapting the grids in space and time to minimize errors in the simulation
  – The numerical simulation of PDEs can be performed with arbitrary unstructured discretizations on serial and parallel platforms
  – Adaptive time stepping controlling strategies
• **BD / HPDA and HPC convergence is fully required**
  – Vast field to work on

• **Data Science with stochastic methods can provide excellent approximations**

• **Levering the core calculi with proactive techniques**
  – Pre-processing of data to be performed on site
  – Spatial-temporal time series predictive data from numerical simulations
  – Management of the uncertainty on numerical simulation data, integrated with a probabilistic database system

• **Scientific workflow management system focused on managing scientific data flow with provenance data support**
Some CS Results
• The ALYA multi-physics code presents speed-ups (w.r.t. CPU code)
  – 36x GPU, 16x OmpSs, 2x Phi with 2 new BOAST-based kernels

• Elastic-acoustic wave propagation kernels
  – 11x Phi, 3x GPU, both with Motif’s classes for characterization
• **Wind model applied to Olan mountain for weather forecast**
  - 40x GPU

  ![Olan (mountain), France Weather Forecast](image)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Wednesday 11</th>
<th>Thursday 116</th>
<th>Friday 117</th>
<th>Saturday 118</th>
<th>Sunday 119</th>
<th>Monday 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

• **Performance analysis of memory hierarchy usage of NVIDIA Kepler**
  - 2x using either shared or texture memory
    - Occupancy is twice better with texture memory
• Mapping threads and pages according to memory locality (Intense Pages Mapping)
  – Improvements of 39% in performance and 12% in energy eff.

• Elastic-acoustic wave propagation kernels ported to Intel’s many-core architectures
  – Haswell surpassing KNC on acoustic propagation (6.8x performance and 2.4x energy efficiency)
  – Elastic propagation (Sandy Bridge): 6.5x performance
• **Affinity clause for task directives**
  – OpenMP task directive and scheduler extended
  – New OpenMP function for affinity and HW topology queries
  – Performance and scalability improves between 2x and 4x
• Hybrid MPI/OpenMP performance by increasing overlapping time
  – Use a dedicated thread to improve asynchronous communications in MPI (not communication at the “wait” call): 9.7% in computational time
• Using low-power (ARM) architectures as file system servers with the Hou10ni app
  – 1 regular server by 2 ARMs doubles bandwidth and decreases energy consumption by 85% without performance loss
• Coordinate the access of I/O nodes to data servers for reducing contention (TWINS scheduler proposal)
  – Read performance of shared files 28% better over other alternatives and by up to 50% over not forwarding I/O requests
• Checkpointing can be performed at both
  – Application (user-defined) level (FTI)
  – System level (DMTCP)

• Slurm workload manager is being extended with third parties plugins to these APIs
  – Resilience is increased
  – New scheduling algorithms can be designed
  – Proactive and reactive action to failures and load balancing

• Streams of data efficiently and in real time with data replication (Kafka-Apache)
Some Mathematical approaches
• **Time-domain elastodynamics (Finite Difference Methods)**
  
  – Applied to a 3D elastic wave propagation in the presence of strong topography
  – FD method based on a staggered grid of Lebedev type or a fully staggered grid
  – Uses a grid deformation strategy to make a Cartesian grid conform to a topographic surface
  – Uses a mimetic approach to accurately deal with the free-surface condition
  – High order FD stencils
  – Better ratio precision/computational cost than DG & other FDs approaches
• Frequency-domain elastodynamics (Hybridizable Discontinuous Galerkin)
  – Unstructured tetrahedral mesh, high order interpolation
  – DG type method well adapted to steady-like problems
  – Hybrid unknown defined on element boundary (numerical traces)
  – Reduced linear system for the hybrid unknown
  – Local (at the element level) independent linear systems for obtaining the main field unknowns

Isotropic geophysical 3D benchmark: $V_s$, $V_p$ and $\rho$
• **Multiscale Hybrid Mixed Method**
  – “Divide and conquer” algorithms: MHM-based DG solver
  – 2D and 3D for highly heterogeneous time-domain elastodynamics
  – Coarse and Fine levels
  – The larger the problem size, the better scalability

<table>
<thead>
<tr>
<th>12,288 × 59 = 724,992 tetrahedra</th>
<th>Baseline (24 cores)</th>
<th>MPI (192 cores)</th>
<th>MPI (768 cores)</th>
<th>Erlang (192 cores)</th>
<th>Erlang (768 cores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (sec)</td>
<td>129.81*</td>
<td>57.74</td>
<td>72.88</td>
<td>97.00</td>
<td>97.00</td>
</tr>
<tr>
<td>Efficiency</td>
<td>100.00%</td>
<td>28.10%</td>
<td>5.57%</td>
<td>16.62%</td>
<td>4.16%</td>
</tr>
<tr>
<td>1,536 × 6,046 = 9,286,656 tetrahedra</td>
<td>Time (sec)</td>
<td>489.00</td>
<td>82.57*</td>
<td>51.72</td>
<td>108.00*</td>
</tr>
<tr>
<td>Efficiency</td>
<td>100.00%</td>
<td>74.03%</td>
<td>29.54%</td>
<td>56.60%</td>
<td>23.51%</td>
</tr>
<tr>
<td>12,288 × 6,046 = 74,293,248 tetrahedra</td>
<td>Time (sec)</td>
<td>4,753.00</td>
<td>641.61</td>
<td>220.20*</td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>100.00%</td>
<td>92.60%</td>
<td>67.45%</td>
<td>89.70%</td>
<td>61.89%</td>
</tr>
</tbody>
</table>
• **Hybrid Krylov iterative schemes**
  – Krylov iterative schemes at high core counts scaling issue
    ▪ Krylov methods based on orthogonalization process
    ▪ Requires at least one global synchronization at each iteration
  – New variants based on reducing synchronization and communication
    ▪ Scales at more than 10,000 cores

*iBiCGStab*
Reduces the number of inner product needed at each iteration step

2 levels hierarchical FGMRES partitioned over 4 cores
Some Energy Results
• CFD RANS models for the ABL in complex terrains including thermal coupling, wakes, and canopy modeling

• Wind farm modelling using high-quality wind farm automatic meshing (hybrid meshes)

• ALYA-LES models for the ABL and blade/vortex bladeless wind turbines
• **OLAM (Ocean Land Atmosphere Model)**

• **Weather Research and Forecasting - WRF**
Dynamical downscaling: WRF model

Improvements in the WRF parametrizations

Impacts of a Cluster of Wind Farms on Wind Resource Availability and Wind Power Production

Improving the Representation of Resolved and Unresolved Topographic Effects on Surface Wind in the WRF Model

Wind speed as simulated with and without the drag parameterization
Statistical downscaling

Relationship between wind power production and North Atlantic atmospheric circulation

Quality Control of a surface wind observations database

Analysis of the extreme winds

Global Forecasters
Chemical Transport Models
WRF-CHIMERE-GEOSCHEM.
New model developments and validation.
Applied to Air Quality Assessment in Spain
Modeling of urban atmosphere

Street-canyon CFD modeling for urban air quality. Developments of Urban Boundary Layer parameterizations for WRF model.
1. Central receiver technology
2. Parabolic dishes + Stirling engines
3. Parabolic trough technology (thermal oil)
4. Parabolic trough technology (DSG)
5. Parabolic troughs (gas) + Molten Salt TES
6. Linear Fresnel Collector
7. Solar furnaces
8. Water desalination
9. Water detoxification
10. Passive architecture
• Computation Requirements for Solar Concentrating Technologies:
  – CFD analysis of heat transfer between solar radiation, materials and fluids (Fluent & STAR-CCM+)

• Design and simulation of solar thermal power plants, both solar tower and parabolic trough technologies
  – Detailed simulation, due to size of the plants, consumes time and resources (SAM, WINDELSOL & Matlab)

• Measurement and characterization of collectors and components by photogrammetric methods
  – High resolution required (1mm), intensive memory consumption (Matlab)
• **Radiation damage in nuclear materials**
  
  — Energies to form He-H bubbles to be lately used in simulations beyond the atomic scale (Kinetic Monte Carlo, Ray Theory...)

  — Database to be stored for multiscale simulations
    • Bubble size and atoms

  — Not as demanding as *ab initio*

  — LAMMPS code
He-H in Fe with 9 vacancies and 27 and 40 He atoms
$5 \cdot 10^4$ iterations to converge in 2 h using 256 cores (Helios)

Final bubble geometry
Fe-blue (9 vacancies), 50 He-cyan, 30 H-red
DAB algorithm, based on metaheuristics, used to optimize stellarators under several criteria: NC, Mercier, ballooning,…

Exploration of the effect of rotational transform on NC transport in stellarators: Extensive estimates of NC transport properties using DKES code (30 configurations x 100 radii). No positive influence of rotational transform on NC, despite the experimental results.
Gyrokinetic (GK) simulations of microinstabilities in stellarators with the code EUTERPE

- Global GK PIC code
- FE (spline) spatial discretization
- RK4 time integration
- PETSc solver
- VMEC interface for MHD equilibria

Example: Trapped Eelectron Modes (TEMs) in TJ-II

- Simulation of TEMs in a TJ-II period with fully kinetic ions and electrons
Monte Carlo reactor & Accelerator Driven Systems calculations:

– Complex 3D designs: high level of detail, large number of materials and isotopes.

– GBs of memory required to accurately describe all materials/isotopes/reactions.

– Large data storage volumes required for nuclear data: different compilations (e.g. ENDF, JEFF, JENDL) and temperatures.

– Large data storage volume for output files (e.g. there are millions of fuel pellets per core).

– Large CPU time to accumulate enough statistics ⇒ high processor capability

– Parallel programming.
Nuclear data processing:

- Need to process the distributed nuclear data libraries to prepare them for Monte Carlo codes: MCNP, GEANT4...

- Verification of nuclear data libraries is done for more than 400 isotopes, ~5 available major reactions and all the energy range (from meV to MeV).

- Large amount of benchmarks required for verification ⇒ large requirements of computer power.

Experiment vs. simulation (MCNP)

- Neutron flux axial distribution in the reactor VENUS-F.
- Three nuclear data sets investigated (ENDF, JEFF and JENDL).
- CPU time for each set is about 3,000 core-hours in CIEMAT’s Euler HPC cluster.
- Taken from FREYA project deliverable No. 3.2, 7th EU Framework Programme.
Development and validation of a reduced mechanism for H₂-air combustion: 21 reactions $\rightarrow$ 3 reactions

Validation in a laminar triple flame configuration.

Validation in a turbulent H₂-air flame stabilized by autoignition (RANS + transported PDF + reduced chemistry)

Large Eddy Simulation of a turbulent supersonic H₂-air flame stabilized by autoignition
Portable combustion reformers for hydrogen production from biofuels (HPC4E project H2020)

Spinning flame solutions in a solid energetic material (Le → 0)

Propagation speeds of symmetric (unstable) and non-symmetric (stable) lean H2 flames (Le < 1)
Thermochemical database generation

Flames Evolution (Sandia-D LES)
• Development and optimization of classical extrapolation schemes in 3D
  • Inversion kernels in imaging production software → Only 60 iterations required
• One more (and last) example of experiment & simulation

• Development of thin crystalline silicon by sputtering deposition of amorphous silicon (around 10-micron thick) with later laser crystallization

• Simulation required for a much better understanding of the interaction between the laser radiation and the material that undergoes solid- or liquid-phase epitaxy as a consequence of the process
  – Synthesis of dielectric surfaces for optimizing the absorption of light
  – Optical/Electromagnetic simulations for improving the reflectivity of the cell
  – Atomic FE simulation for the Si crystallization
• Energy and HPC/HTC must keep on collaborating pursuing better, more efficient and cleaner energy sources

• Simulations must be compared with experimental data
  — From simulation to reality, but opposite direction too

• Potential collaborations could be established from this moment on with...
Meteorology and Wind Energy: J. Navarro (CIEMAT), A. Folch (BSC-HPC4E)
Solar Concentration Power: J. Fernández-Reche (CIEMAT)
Photovoltaic: J. Carabe and J.J. Gandía (CIEMAT), S. Lenteri (INRIA-HPC4E)
Materials: P. García-Müller (CEMAT)
Fusion: F. Castejón (CIEMAT)
Fission (Nuclear innovation): D. Cano-Ott (CIEMAT)
Oil & Gas: S. Fernández (Repsol-HPC4E), J. Panneta (ITA-HPC4E)
Atmospheric pollution: F. Martín (CIEMAT)
Computer Science: J.M. Cela (BSC-HPC4E), P.O.A. Navaux (UFRGS-HPC4E),
A.L.G. Coutinho (COPPE-HPC4E), F. Valentin (LNCC-HPC4E), P. Protopapas and
R. Dave (IACS), M. Rodríguez-Pascual and J.A. Moríñigo (CIEMAT)

The research leading to the HPC4E results has received funding from the European Union's Horizon 2020 Programme (2014-2020) and from Brazilian Ministry of Science, Technology and Innovation through Rede Nacional de Pesquisa (RNP) under the HPC4E Project (www.hpc4e.eu), grant agreement n° 689772.
THANK YOU!!!

rafael.mayo@ciemat.es
CIEMAT
Avda. Complutense, 40 – 28040 Madrid
http://rdgroups.ciemat.es/web/sci-track/

mayogarcia@seas.harvard.edu
Harvard-IACS
52 Oxford St – Cambridge, MA 02138