Fostering Collaboration in Energy Research and Technological Developments applying new exascale HPC techniques

José María Cela
Computer Applications in Science & Engineering
Barcelona Supercomputing Center-Centro Nacional de
Supercomputación (BSC-CNS)
Barcelona, Spain

Alvaro L.G.A. Coutinho
High Performance Computing Center
COPPE / Federal University of Rio de Janeiro
(COPPE/UFRJ)
Rio de Janeiro, Brazil

Philippe O.A. Navaux
Parallel and Distributed Processing Group
Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil

Rafael Mayo-García
Centro de Investigaciones Energéticas
Medioambientales y Tecnológicas (CIEMAT)
Madrid, Spain

(On behalf of the project)

Abstract—**During the last years, High Performance Computing (HPC) resources have undergone a dramatic transformation, with an explosion on the available parallelism and the use of special purpose processors. There are international initiatives focusing on redesigning hardware and software in order to achieve the Exaflop capability. With this aim, the HPC4E project is applying the new exascale HPC techniques to energy industry simulations, customizing them if necessary, and going beyond the state-of-the-art in the required HPC exascale simulations for different energy sources that are the present and the future of energy: wind energy production and design, efficient combustion systems for biomass-derived fuels (biogas), and exploration geophysics for hydrocarbon reservoirs. HPC4E joins efforts of several institutions settled in Brazil and Europe.**

Keywords-Exascale; HPC; Wind energy; Combustion; Hydrocarbon

I. INTRODUCTION

New energy sources, if untapped, might become crucial in the mid-term. Intensive numerical simulations and prototyping are needed to assess their real value and improve their throughput. The impact of exascale HPC and data intensive algorithms in the energy industry is well established in the U.S. Department of Energy document "Synergistic Challenges in Data-Intensive Science and Exascale Computing" [1], for example.

The High Performance Computing for Energy (HPC4E) project aims to provide these new exascale HPC and data intensive algorithms to three energy sources: wind energy, biomass, and oil. To do so, several European and Brazilian institutions are closely working, fostering in this way a collaboration that can be extended to other countries in Latin America as the computational solutions provided will be useful to many scientific and industrial fields as well as to other software and middleware developers.

Moreover, the current collaboration embraces both public and private companies who are leaders in their respective sector. The industrial sector is formed by REPSOL, Iberdrola Renovables Energía, TOTAL, and PETROBAS, who present a solid activity and presence in many Latin American countries. Regarding the public sector, Brazilian (COPPE/UFRJ, LNCC, ITA, UFRGS, UFPE) and European (BSC-CNS, INRIA, UNLAC, CIEMAT) institutions belonging to the HPC and the energy fields are represented.

All of them aim to collaborate with other Latin American entities who will be interested in the developments carried out as part of HPC4E in both the computer and the energy sciences.

II. CONCEPT AND APPROACH

As previously stated, the main objective of HPC4E is to develop beyond-the-state-of-the-art high performance simulation tools that can help the energy industry to respond future energy demands and also to carbon-related environmental issues using the state-of-the-art HPC systems. This Brazilian-European collaboration also aims at improving the usage of energy using HPC tools by acting at many levels of the energy chain for different energy sources:

- Exploitation: In wind energy (respond to demand peaks, output prediction)
- Efficiency: In biomass-derived fuels (develop more efficient and renewable fuels, reduce green-house gas emissions, reduce hydrocarbon dependency and fuel cost)
- Exploration: In wind energy (resource assessment) and in hydrocarbons (improve available reserves, explore with less financial and environmental risk).

Another main objective is to improve the cooperation between energy industries from EU and Brazil, guaranteeing that the Technology Readiness Levels (TRL) of the

particular project technologies will be very high. This will be done via a strong collaboration between the project partners.

Regarding scientific technical objectives, this will be addressed in the sections below.

III. A DISRUPTIVE EXASCALE COMPUTER ARCHITECTURE

The project will set up a disruptive exascale computer architecture to study the mapping and optimization of the codes proposed for each energy domain on novel architectures for exascale, as well as developments in the underlying software infrastructure. In order to properly test this infrastructure, porting, tuning, and testing efforts of the different simulations codes will be previously carried out.

Specifically, four lines of action will be pursued.

The selected computing kernels of the codes coming from the energy sector will be optimized for architectures based on accelerators. The goal will be to optimize the performance but keeping a high degree of portability. The ratio flops/watt obtained in each platform will be analyzed. The main target architecture platforms are those based on Xeon Phi and NVIDIA GPUs, but other platforms based on embedded processors will be also analyzed. To guarantee the maximum portability of the codes we will use programing models and tools like openCL, ompSs and BOAST.

The selected kernels will be also ported to architectures based on symmetric multicore processors with NUMA memory. The goal will be to optimize the performance. The main target architectures will be Intel, AMD and SGI, but also new platforms based on ARM processors will be analyzed. Thus, the key point will be the load balancing and data placement, takin into account new scheduling algorithms able to improve locality too.

The management of the MPI level parallelism in the codes coming from the energy sector will be guaranteed for achieving a high scalability of the applications in HPC clusters with millions of cores: The main topics to be analyzed will be: creation of tools for migration of running parallel tasks inside clusters; hierarchical MPI structures to manage coupled multiphysic problems; parallel I/O optimization; design of efficient check-pointing strategies; and, fault tolerance strategies at MPI level.

Last, performance analysis will be focused on the performance analysis of the different applications and kernels. The proper environments and tools (Paraver, Triva, Ocelotl, TAU, etc.) will be deployed to analyze all the parallel levels in the applications. Inside a computational node roof-line analyses will be done to understand the bottlenecks of the architectures. At the cluster level, network traffic, I/O traffic and load balancing will be analyzed to guarantee the application scalability. Also performance prediction tools will be used to analyze the potential benefits of architecture or algorithm modifications. Different proposal of exascale architectures will be studied for the selected applications as well.

IV. SIMULATORS FOR EXASCALE COMPUTATIONS

Innovative computational algorithms well suited to the numerical simulation of complex phenomena on exascale architectures will be introduced. These refer to numerical schemes for Partial Differential Equations (PDE), sparse linear solvers, adaptivity, and data management.

Particularly the first topic is concerned with scalable implementations of high order schemes for wave propagation models. The second topic will develop and demonstrate the benefits of generic (i.e. algebraic) parallel solvers for large sparse linear systems of equations. The third topic addresses mesh and (local) time-step adaptive algorithms in order to optimize the use of computational resources. The fourth topic focuses on leveraging techniques to support simulation data management as required by the pre- and pos-processing steps involved in highly complex simulations.

Regarding scalable high order numerical schemes, two families of innovative high order finite element methods and a family of (standard and mimetic) finite difference schemes will be considered for both time-domain and frequency domain. These numerical schemes exhibit a high level of parallelism. In particular, they are well suited to a mixed coarse grain/fine grain (MIMD/SIMD) parallelization targeting many-core (Xeon Phi/GPU) systems. First, we will implement and demonstrate the benefits of a recently designed class of high order multiscale methods. The common core approximation framework is the Multiscale Hybrid-Mixed (MHM) methods combined with Discontinuous Galerkin (DG) or Stabilized Continuous Galerkin (SCG). For frequency-domain problems, we will perform the same analysis with the so called hybridized DG formulations that drastically reduce the number of globally coupled degrees of freedom. Both types of solvers (i.e. time-domain and frequency-domain) are linked to the simulation/inversion framework for subsurface imaging proposed in the geophysics domain.

With respect to scalable sparse linear solvers, the goal is to provide state of the art parallel solvers for spare linear systems of equations or numerical schemes adopted in the simulation software associated to the applications coming from the energy field. Both direct and hybrid direct/iterative solvers will be considered. Regarding the former solver type, the PaStiX software [2] will be adopted. It is based on a supernodal approach and has been implemented on top of various runtime systems enabling an efficient use of heterogeneous manycore platforms. The hybrid iterative/direct strategy will be made available through the MaPHyS software [3] that implements algebraic domain decomposition ideas and relies in parallel on parallel sparse direct solvers such as PaStiX for each subproblem. On top of those two solvers, Krylov subspace methods are implemented either for the iterative refinement steps of PaStiX or to solve the reduced Schur complement system in MaPHyS. Finally, the high performance dense linear algebra kernels on which those solvers are relying will also be part of the project but no specific action is foreseen on it. This software stack will be made available through a coherent and flexible API where the matrices can be provided by the application in various format such as centralized or distributed, assembled or unassembled.

Optimal numerical schemes for PDEs involve adapting the grids in space and time to minimize errors in the simulation. The activity on this topic will then be to explore

libraries to support adaptivity such as the libMesh library [4]. libMesh provides a framework for the numerical simulation of partial differential equations using arbitrary unstructured discretizations on serial and parallel platforms. Adaptive time stepping controlling strategies will be also studied. The objective here is to demonstrate the applicability of such strategies to large-scale parallel computations of the simulation of polydisperse mixtures typically found in the geological processes.

Big Data management and analysis of numerical simulations will be explored by the use of three systems: SimDB, UpsilonDB and Chiron. The first is being designed to manage spatialtemporal time series predictive data from numerical simulations, represented as a multidimensional array. The second system, UpsilonDB, is currently an early stage prototype aiming at managing the uncertainty on numerical simulation data, integrated with a probabilistic database system, MayBMS. UpsilonDB supports simulation post-processing analysis. Chiron is a scientific workflow management system focused on managing scientific dataflow with provenance data support. Chiron strong support in data analytics at runtime, allowing for dynamic configuration fine-tuning, including uncertainty quantification data steering.

V. ATMOSPHERE FOR ENERGY

The fundamental knowledge barriers to further progress in wind energy are defined as scientists' understanding of atmospheric flows, unsteady aerodynamics and stall, turbine dynamics and stability, and turbine wake flows and related array effects. The use of computational fluid dynamics (CFD) large-eddy simulation (LES) models to analyze atmospheric flow in a wind farm capturing turbine wakes and array effects requires exascale HPC systems.

In this way, microscale atmospheric models are based on CFD solvers adapted to simulate the Atmospheric Boundary Layer (ABL) in order to approach two fundamental wind energy problems: analysis (mainly focused on wind resource assessment and wind farm design) and forecast (mainly focused on short-term prediction for wind farm dispatch to the electricity network). Both problems will study key aspects concerning microscale modelling simulations, as standalone CFD models or in connection with mesoscale models, by developing dynamical and statistical downscaling strategies. All models, methods and techniques developed for analysis will be tuned to produce short-term online forecasts of the wind farms output.

The objective is then to have the CFD models ready to exascale systems in order to overcome the present limitations and increase the accuracy on the evaluation of technical and economic feasibility of wind farms.

Regarding dynamical downscaling in order to assess wind resource, CFD models must account for the coupled effects of complex terrain, Coriolis forces, thermal stability, presence of forests, and wind turbines. Modifications need to be made to the RANS/LES CFD models including turbulent closures for ABL. The objective is to characterize the accuracy of the different ABL-CFD models, the numerical stability (robustness) and the convergence behavior of their HPC implementation.

The statistical downscaling approach will be complementary to the dynamical downscaling. Statistical downscaling models will be developed using local observations and large scale circulation and wind fields in the wind farm region.

A compilation of the available data and an evaluation of the quality issues that might affect the succeeding analyses will be accomplished, for this purpose is necessary to identify an appropriate metrics that account for the deviations in the wind power production predictions. Transfer functions between wind and wind power should be determined and serve as reference to translate the wind into wind power estimates for the rest of analyses.

On the other side, In order to efficiently plug wind farm power production to a distribution electricity network it is mandatory a forecast of the power production that allows the network operator to manage the electricity resources. In this sense, wind power short-term prediction within hourly to daily time scales is of fundamental importance. The models, methods and techniques developed for wind farm modelling will be tuned to produce short-term online forecasts of the wind farms output. Efficient use of HPC resources is critical to have these forecasts online. We will develop a forecast based on dynamical and statistical downscaling strategies.

A. Improving CFD microscale models

In order to achieve such an improvement, the following actions will be made:

- For RANS/LES models, study the turbulence closure models for ABL simulations as an alternative to the existing parametrizations.
- Implementation of a canopy model [5]
- Validate the HPC implementation using experimental data from the New European Wind Atlas project (ERA-Net)
- Wind farm modelling: Currently, farm models simulate downwind effects of rotors by extracting axial momentum at the turbines. In this simplistic approach, the rotor characteristics are incorporated trough velocity-independent drag coefficients obtained from tunnel experiments. This tasks aims at characterizing numerically the downwind effects of rotors by solving complex turbulent rotating flows.

B. Dynamical Downscaling strategies

Boundary conditions for solving wind flow on microscale domains are typically assumed steady and homogeneous over the computational inflow. These limitations can be overcome by dynamically coupling microscale CFD models with mesoscale simulations furnishing initial and time-dependent boundary conditions at the computational boundaries. The following will be made:

- Blending between mesoscale (WRF) and CFD computational meshes to have consistent terrain information (topography and roughness) at the computational margins. The topography will be

interpolated in such a way that it will be coincident with the CFD resolution in the inner zone and coincident with the mesoscale (WRF) over the boundaries;

- Initial condition and time-dependent CFD boundary conditions consistent with the mesoscale outputs that will drive the CFD model through boundary conditions;
- Study the use of nudging strategies based on introducing a force term over the momentum equation in the CFD model close to the boundary in order to enhance consistency between models near the boundaries;
- Validate the methodologies for different site conditions, onshore and offshore, considering benchmark validation cases from other projects such as IEA-Task 31 [6] Wakebench and FP7-NEWA [7].

VI. BIOMASS FOR ENERGY

Another important challenge is to develop a validated, predictive, multi-scale, combustion modeling capability to optimize the design and operation of evolving fuels. The next exascale HPC systems will be able to run combustion simulations in parameter regimes relevant to industrial applications using alternative fuels, which is required to design efficient furnaces, engines, clean burning vehicles and power plants

Thus, in order to obtain a thorough understanding of the effects of fuel variability on energy utilization of biomass-derived gaseous fuels, a coupled approach which covers three distinct areas of development will be employed: generation of chemical kinetic mechanisms for biomass-derived fuels, integration of the schemes into a CFD code, and creation of efficient algorithms for data exchange that can run efficiently in HPC platforms.

The activities include analyzing the physical characteristics of bio-syngas flames, assessment of performance in practical systems and providing an optimized industrial guideline for biomass derived gaseous fuel compositions and performance. The application scenario corresponds to industrial devices of the energy sector: stationary gas turbines, furnaces and portable combustion devices.

Generation of detailed chemical schemes that reproduce accurately the oxidation of biomass-derived gaseous fuels will be developed and assessed via detailed chemical kinetics mechanisms that can predict the oxidation process and species formation with certain level of accuracy. Several well-established mechanisms (GRI 3.0, San Diego, Leeds, etc.) will be examined and compared for different fuel compositions and the accuracy at predicting reference species and radicals will be provided for different operating conditions. These mechanisms are the starting point for reduction or tabulation techniques.

Also, development of skeleton and reduced chemical schemes for biomass combustion for engine operating conditions will be carried out. The use of detailed chemical

mechanisms imposes an important limitation for practical applications of turbulent combustion. These mechanisms include both slow and fast chemical reactions involving a large number of species leading to a highly costly numerical problem. Besides, the effects of turbulence and flow strain also contribute to a complex interaction between chemistry and fluid mechanics that has to be accurately reproduced by CFD codes. In order to reduce the stiffness of the chemical problem, skeleton and reduced mechanisms will be developed so it can be integrated into a multiphysics code. The reduction technique will be based on flame-generated manifolds (FGM) and quasy-steady state (QSS) approximation for operating conditions of interest.

For studying the combustion dynamics of laboratory flames and comparing it with available data an integration of the chemical schemes developed in the previous points and the corresponding validation using benchmarking cases will be performed. Several cases using experimental data of laboratory flames will be investigated and the effects of fuel variability on the flame dynamics will be investigated.

Regarding industrial applications of biomass fuels in practical systems, numerical simulation of an industrial engine burning biomass will be carried out. Different operating conditions and fuel compositions will be examined. The activities will be focused on providing details of the system performance to develop industrial guidelines for the use of biomass derived gaseous fuels.

Last, applications of biomass-derived gaseous fuel combustion in portable reformers for hydrogen production will be explored as well. Limitations to the miniaturization of hydrogen production reformers are linked to the large surfaceto-volume ratio, which enhances heat losses through the walls. Additional measures, such as heat recirculation, catalytic combustion, reactant preheating etc., are needed to sustain their proper operation. The activities of this Task will consist in numerical studies of the stability of combustion in small size reformers, with the objective of improving the understanding of their operation and determining the stable and more efficient regimes as a function of the fuel composition. The reduced chemical mechanisms developed in the aforementioned paragraphs as well as the use of HPC shall prove essential to this large parametric numerical investigation.

VII. GEOPHYSICS FOR ENERGY

The third energy sector is related to oil as an energy source. Huge computational requirements arise from full wave-form modelling and inversion of seismic and electromagnetic data. By taking into account the complete physics of waves in the subsurface, imaging tools are able to reveal information about the Earth's interior with unprecedented quality. Nevertheless, actual wave physics has a high cost in terms of computational intensity, which can only be matched by using the exascale HPC systems.

In this sense, the capacity for imaging accurately the Earth's subsurface, on land and below the sea floor is a challenging problem that has significant economic implications in terms of resource management, identification of new energy reservoirs and storage sites as well as their

monitoring through time. As recoverable deposits of petroleum become harder to find the costs of drilling and extraction increase accordingly. Thus the oil and gas industry needs more detailed imaging of underground geological structures in order to find the best representation of the subsurface in terms of which model sticks better to the data recorded during acquisition surveys. This involves research based on advanced methods combining mathematics, geophysics and scientific computing. Such multidisciplinary collaboration is essential to the design of numerical simulation codes capable of delivering the clearest possible picture of the subsurface.

The data types involved in geophysical imaging are mostly seismic (acoustic or elastic) and electromagnetic. Modern imaging techniques (RTM, FWI …) rely on intensive usage of full 3D physical modeling engines. Hence, in order to attain results in a reasonable time, these engines must use, as efficiently as possible, the fastest hardware architectures in a massively parallel way. On top of that, the larger and more complex the scenarios become, techniques which attain results with low computational complexity or few degrees of freedom become preferable.

The main goal is, then, attaining the sharpest possible images of the subsurface with the best possible quantitative content (i.e. parameter estimation, uncertainty analysis) in the shortest possible time. In addition, in the exascale era, power efficiency is becoming ever more a crucial factor in establishing the usability of HPC in industrial applications. The main developments will involve geophysical inversion of elastic and electromagnetic waves using high-order structured and unstructured computational grid types. Test-driven code development will allow us to put a special focus on having detailed comparisons and benchmarks between all possible approaches in the most realistic scenarios. Architecture-oriented programming optimizations are expected to play a crucial role in establishing cost/accuracy/complexity relationships which will help delineate the future directions of geophysical imaging in the exascale era.

Specific sub-objectives will be: development and optimization of high-order finite-element schemes for 3D elastodynamics; development and optimization of classical extrapolation schemes in 3D; uncertainty estimation of petrophysical quantities; Synthetic benchmarking of exascale geophysical problems; and, Industry validation.

VIII. GENERAL AND LATIN AMERICAN IMPACT

HPC4E lifetime is of two years and has recently started in December 2015. As it can be easily inferred, its impact goes beyond the three selected energy sources as the computing solutions proposed will be able to be applied to other scientific and engineering domains.

Some general impacts derived from the efficient use of exascale HPC and simulation technology can be considered:

- Vast improvement in simulation efficiency in terms of Watts needed per execution and reduced time-to-solution. This will be applied to critical aspects of the energy value chain, with

rapid deployment in the partner' current production systems.
- Establishing transnational "numerical laboratories", which are cheaper, safer and faster than real-life experiments.

HPC4E will have some general impacts derived from collaboration of EU and Brazil teams that can be easily extrapolated to other Latin American countries:

- A framework to share know-how and solutions to common HPC problems with a long-term establishment or consolidation of collaborations between EU and Latin American academic researchers in the field of HPC in the broad sense, on multi-disciplinary projects involving applied mathematics, computer science and Big Data topics.
- Reinforce the ties between EU and Latin America in critical aspects for society such as energy, which will help making the way towards their energetic independence in the following decades.
- An impulse to EU and Latin American economy by increasing the competitiveness of their key energetic players in a globalized world

Under these expectations, several liaisons in Latin America could be established in order to foster collaboration between HPC4E and those institutions and initiatives that could be interested in the project's outcomes.

Regarding computer science, it is noteworthy the joint action that can be carried out with the Servicio de Computación Avanzada para América Latina y el Caribe (Advanced Computing Services for Latin America and the Caribbean) [9]. In this sense, HPC4E plans to periodically inform SCALAC about the different advances that will be achieved so it could be easily adopted by the Latin American HPC community; such an information exchange will be seamlessly developed as several HPC4E partners are involved in SCALAC.

It is also important to remark that the industrial partners of HPC4E have a strong business activity in Latin America, where most of their downstream research is experimentally tested. Being so, these continental regions will directly benefit from the project advances in terms of wind farms, biomass plants and oil exploration.

At the same time, new contacts could be set up with the United Nations Industrial Development Organization Observatory for Renewable Energy in Latin America and the Caribbean [10], who is currently operating in 13 countries of the region. All the advances achieved in the energy sectors identified by HPC4E could be applied by the Centers of Excellence on Renewable Energy in Latin America with the support of this Observatory.

Last but not least, several dissemination activities in scientific forums will be carried out by the project, so specific collaborations with Latin American institutions in any of the topics addressed by the project will be easily defined. In this context, at least two HPC4E meetings will be organized in Brazil.

REFERENCES

[1]   Synergistic Challenges in Data-Intensive Science and Exascale Computing, DOE ASCAC Data Subcommittee Report, March 2013

[2]   The PaStiX software, available at http://pastix.gforge.inria.fr)

[3]   The MaPHyS software, available at http://maphys.gforge.inria.fr

[4]   The libMesh library, available at http://libmesh.github.io/

[5]   A. Sogachev, M. Kelly, and M.Y. Leclerc, "Consistent Two-Equation Closure Modelling for Atmospheric Research: Buoyancy and Vegetation Implementations". Boundary-Layer Meteorology, vol 145 (2), 2012, pp..307–327.

[6]   The IEA-Task 31 Wakebench, http://www.ieawind.org/task_31.html

[7]   The NEWA project, http://euwindatlas.eu/

[8]   G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[9]   SCALAC, more information at its twiiter profile @SCALAC_Computo

[10] The Observatory for Renewable Energy in Latin America and the Caribbean, http://www.renenergyobservatory.org/en.html