D3.2 Second intermediate report on applications working groups

CONTRACT NO	EESI2 312478
INSTRUMENT	CSA (Support and Collaborative Action)
THEMATIC	INFRASTRUCTURE

Due date of deliverable: 31/05/2014

Actual submission date: 31/03/2015

Publication date: 31/03/2015

Start date of project: 1 September 2013 Duration: 30 months

Name of lead contractor for this deliverable: PRACE-GENCI, Stéphane REQUENA

Authors: Stéphane Requena (GENCI), Norbert Kroll (DLR), Ange Caruso (EDF), Giovanni Aloisio (Univ. Salento), Jean Claude André (Jca Consultance), Alan Sacha Brun (CEA), M. Orozco (BSC), R. Goni (BSC) and C. Laughton (Univ. Nottigham).

Name of reviewers for this deliverable:

Abstract: This is the second intermediate report of EESI2 WP3 Applications Working Groups.

Revision: 1.0

Project co-funded by the European Commission within the Seventh Framework Programme (FP7/2007-2013)		
Dissemination Level to be filled out		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	X

Table of Contents

Glossary

Abbreviation / acronym	Description
API	Application Programming Interface
CAD	Computer Assisted Design
CERN	Centre Européen de Recherche Nucléaire
CFD	Computational Fluids Dynamics
CI	Combustion Instabilities
CMIP5, CMIP6	Coupled Model Intercomparison Project Phase 5 or Phase 6
CoE	Center of Excellence, result of the H2020-EINFRA-2015-1 call from the European Commission
CTR	Center for Turbulence Research (Stanford)
CSE	Computational Science and Engineering
DEEP	Dynamical Exascale Entry Platform
DNS	Direct Numerical Simulation
DoF	Degree of Freedom
DSL	Domain Specific Langage
EC	European Commission
ENES	European Network for Earth System Modeling
EESI	European Exascale Software Initiative (Europe)
EPOS	European Plate Observation System
ERC	European Research Council grant
ESA	European Space Agency
ESM	Earth System Models
EXDCI	European Xtreme Data and Computing Initiative, a H2020 proposal submitted in 2014 as a follow up of EESI2
FAU	Friedrich-Alexander Universität Erlangen
FDA	(US) Federal Drugs Administration
FET	Future and Emerging Technologies, an EU pathfinder programme in information technologies
GPU	Graphical Processing Units
HPC	High Performance Computing
IDC	International Data Corporation
IT	Information Technology

ITER	International Thermonuclear Experimental Reactor
LB	Lattice Boltzmann
MOOC	Massive Open Online Courses
NAFEMS	National Agency for Finite Element Methods and Standards, an independent, not-for-profit membership association dedicated to FEA and CFD
NASA	(US) National Aeronautics and Space Administration
ISV	Independent Software Vendors
LES	Large Eddy Simulation
OSS	Open Source Software
PETSc	The Portable Extensive Scientific toolkit, a numerical library
PRACE	Partnership for Advanced Computing in Europe, the European HPC research infrastructure
SPH	Smooth Particle Hydrodynamics
UQ	Uncertainty Quantification
WCES	EESI-2 WP3 Weather, Climatology and Solid Earth Sciences working group
WG	Working Group (in EESI)
WP3	EESI2 workpackage on Applications
WP4	EESI2 workpackage on Enabling technologies
WP5	EESI2 workpackage on Cross cutting issues

1. Executive Summary

This document is the second EESI2 intermediate report on the application working groups, corresponding to WP3 with specific topic on industrial applications (WG3.1), Weather, Climatology and Solid Earth Sciences (WG3.2), Fundamental Sciences (WG3.3), Life Sciences and Health (WG3.4) and Disruptive Applications (WG3.5).

This deliverable present an update for the year 2014 of the D3.1 document for each WG and highlights the following issues:

- Industrial applications are requiring more and more HPC for dealing with complex multiscale and multiphysics methods in Oil & Gas, automotive, aeronautics, … New methods using Fast Multipole Method, H-Matrices, SPH or LBM are now coming out the field and used in production by industries leading also to a strong (re)equipment or (re)use of HPC facilities by these domains. This is particularly true for Oil & Gas and automotive with internal acquisition of massive HPC facilities or remote use of research infrastructures with collaborations (DoE in US or PRACE in Europe).
- In climate, astrophysics or fusion, the development of next generation codes (like the new dynamical cores for the climate, the latest dark matter codes like AREPO or the scale out of the fusion codes through the G8 Exascale initiative) is pushing scalability to hundreds thousand of cores, but now exhibiting burden to in situ data management of massive amount of results generated by such simulations.
- In life sciences a lot of initiatives started toward the simulation of the Brain in US, Japan or Europe (with the HBP Flagship)
- The communities have been deeply engaged toward the rise of Center of Excellence (CoE) as proposed by the European Commission call end of 2013, even if the budget finally allocated to the 8 to 10 proposals selected (40M€ in total) is largely below what EESI1 suggested, this initiative is a good sign toward European structured colorations between scientific research teams for developing Exascale class components.
- Experts from WP3 also worked into the EXDCI (European eXtreme Data and Computing Initiative) proposal with aims to continue the EESI history by bridging together all the actors of the new European HPC ecosystem, PRACE for the HPC research infrastructure, ETP4HPC for the HPC technology and upcoming CoEs.

Finally this deliverable also reviews the process of elaboration of five of the project recommendations where WP3 has been involved directly:

- Towards flexible and efficient Exascale software couplers ;
- Identification of turbulent flow features into massively parallel Exascale simulations ;
- Enhanced unified framework for model verification and validation and uncertainty quantification ;
- Parallel-in-Time: a major step forward in Parallel simulations;
- Declarative processing frameworks for big data analytics, extreme data fusion e.g. identification of turbulent flow features from massively parallel Exaflops and Exabytes simulations.

2. WP3 overall activity

This is the second deliverable of WP3 "Applications working group" which objective is to investigate on scientific and industrial application drivers for Peta and Exascale computing. WP3 is organized into five representative working groups and a last one dedicated to WP3 management. Following the activity started during the previous EESI-1 project this work package aims to pave the path by:

- Investigating on key application breakthroughs, quantifying their societal, environmental and economical impacts and performing a gap analysis between current situation and Exascale targets;
- Evaluating the R&D activity performed by scientific and industrial communities, especially in applications redesigning and development of multiscale/multiphysics frameworks;
- Fostering the structuration of scientific communities at the European level;
- Integrating within WP tasks, cartography update through continuous use of network.

WP3 is organized around the following five applications scientific and industrial working groups:

- WG 3.1: Industrial and engineering applications
- WG 3.2: Weather, Climatology and Solid Earth Sciences
- WG 3.3: Fundamental Sciences
- WG 3.4: Life Sciences and Health
- WG 3.5: Disruptive Technologies

And WG 3.6 dedicated to the coordination of WP3.

The work performed by all the different working groups relies on meetings, conference calls involving a total of 43 international experts coming from academia as well as industry with 7 companies represented. It was also important to ensure a wide geographical representation of the experts across 8 European countries as well as Russia and USA.

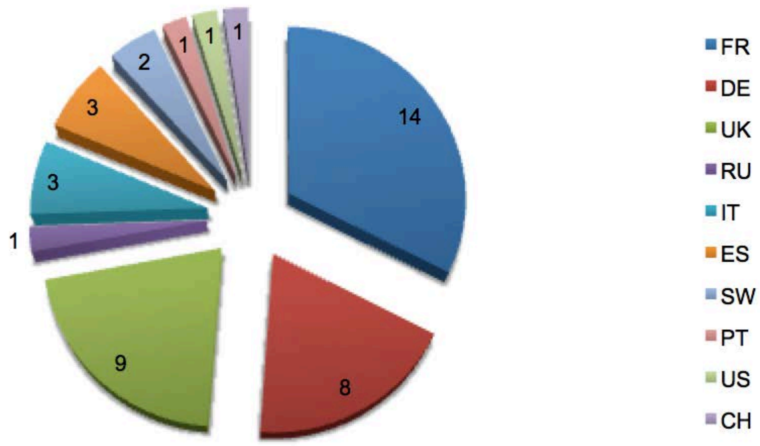


Figure 1 - Repartition per country of the 43 WP3 experts

3. Industrial and engineering applications

HPC and advanced numerical simulation are key issues for industrial and engineering applications by allowing the reduction of the time to design, increase safety and reduce cost of development of the products.

3.1 Advances, results and breakthroughs in the international context

In the oil & gas domain, investments in HPC like any investment in the oil & gas exploration are often tightly linked with the evolution of the price of oil. In less than one year the price of the crude oil barrel has been divided by 2 with an immediate consequence of freezing or cutting investments in companies.

However, oil & gas companies are pushed to invest in new technologies including HPC for being able to explore and to exploit new ultra deep offshore or non-conventional oil fields by using more and more precise seismic algorithms and more accurate reservoir modelling methods.

By consequence its interesting to notice that in the period a lot of oil & gas companies, despite the context, have been able to increase significantly their internal HPC facilities, making Oil & Gas the second largest profitable market for HPC (just after Finance) with according to DC an increase of the CAGR of 9.2% between 2012 and 2017.

As some recent examples:

- TOTAL the French Oil & Gas company just announced in March 2015 an upgrade of their SGI Pangea ICE X system to a full capability of 6.7 PFlops (for a power consumption of 4.5 MW), making this system one of the biggest HPC configuration used for Oil & Gas exploration. The storage capacity will be expanded to up to 27 PB of disk with a global bandwidth of 300 GB/s.
- Just few weeks before, PGS (Petroleum Geo-Services) a Norwegian seismic contractor announced the acquisition of a CRAY XC40 of 5 PFlops peak performance based on a homogeneous set of Intel x86 Haswell processors. This machine will be used for high-resolution seismic acquisition processing of huge datasets coming from deep water exploration in Gulf of Mexico.
- In Australia, DownUnder GeoSolutions a seismic service company purchased to SGI 6 Rackable clusters accelerated with a total of 3,800 Intel Xeon PHI manycore coprocessors for an aggregated peak performance of close to 6 PFlops.
- BP built in 2014 a new datacenter in Houston with a power capacity of 15MW (scalable to 35MW) for hosting as a first step a 2.2 PFlops HPC system for seismic and reservoir exploration.
- Exxon, Shell and Chevron also deployed in 2014 some petascale systems for seismic processing. Numbers are quite hard to obtain but for Chevron the configuration is close to 5 PFlops hosted in a new datacenter capable of hosting more than 20 MW.
- ENI installed in early 2014 a 4.6 PFlops IBM hyrid cluster composed by 1,500 dual CPU nodes (24,000 cores total) and 1,300 K20x graphics accelerators, provided by nVIDIA.
- National companies like Petrobras (Brasil), Woodside Energy (Australia), Petrochina (China) have expressed whished to expand to petascale systems their current HPC facilities. Regarding China, companies like PetroChina are already heavy users of the national research systems which are one of the largest systems installed wordwide.

As said previously this market is very important for the HPC industry which very often use to co design specific solution for addressing needs of oil & gas companies like seismic imagining where most of the algorithms are single precision based. As an illustration of that, during the last GTC 2015 conference in San José, nVIDIA introduced a new accelerator board called Titan X, stated as the word fastest

GPU, providing more that 7 TF of SP peak performance, 12 GB of memory and 336.5GB/s of memory bandwidth (and only 0.2 TF of double precision performance). This device is clearly targeted to gaming but also finance and oil & gas where single precision and memory footprint on the GPU is mandatory for performing seismic processing of large datasets.

For the others industrial segments (aeronautics, aerospace, energy, automotive) the current level of equipment stayed constant with in some case companies like EDF (energy), Airbus (aeronautics) or Mercedes-Benz and Porsche (automotive) owning or accessing remotely to petascale systems.

In the combustion domain Argonne National Laboratory (USA) used the first day of the SAE 2014 World Congress in Detroit on April 8 to launch a new program designed to help automakers design more efficient engines via optimized combustion. Called VERIFI[1] (Virtual Engine Research Institute and Fuels Initiative) is claimed to be the first and only source in the world for high fidelity, 3-D, end-to-end, combustion engine simulation and visualization using uncertainty quantification. VERIFI offers access to Mira a 10 PFlops IBM BlueGene/Q system as well as scientific instruments available at Argonne like brightest X-ray beams, Photon Source and Electron Microscopy Center, powertrain devices, …

More globally in the US, there is alongside Department of Energy (DoE), a strong initiative to develop with the support of the DoE laboratories and facilities the next generation of CFD codes for advanced combustion engines. A first workshop organised in August 2014[2] aimed to assess the needs of the automotive community and the research community toward the development of next gen LES CFD codes for combustion, including Uncertainty Quantification.

During this workshop a representative of General Motors (GM) showed a detailed roadmap related to the development of LES combustion codes for the design of next gen of advanced downsized engines allowing a drastic reduction of consumption and pollution. This roadmap relies to an access to LCF (Leadership Class Facilities) of DoE, e.g systems like Mira with a capability of 10 PFlops.

This concept relies to the following pathway to advanced combustion strategies:

- The realization of high resolution experiments using optical diagnostics, …
- The development of high fidelity validated computational models
- In depth understanding of complex physics (engine design)

[1] http://verifi.anl.gov

[2] https://blogs.anl.gov/verifi/files/2014/10/GurpreetSingh-VERIFI2014.pdf

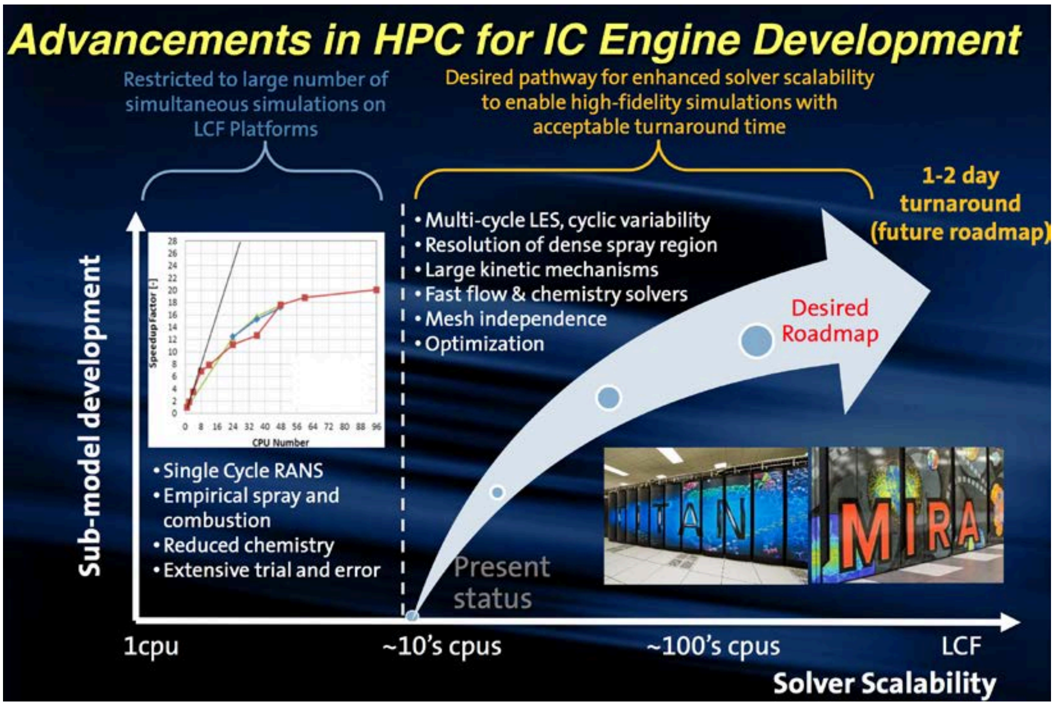


Figure 2 - Roadmap of development of models and use of HPC facilities at GM

In the automotive domain the industry is now facing for 2020 5 more main challenges: safety, environnement, driving pleasure/life on board, affordability and integration of mobility systems with 5 leading technologies for reaching these challenges : enerfy efficiency, connected cars and autonomous driving.

During a recent meeting of SIA (Société des Ingénieurs en Automobile) in France on March 2015[3], Gaspar Gascon-Abellan General Manager of R&D at Renault-Nissan, followed by Gilles Le Borgne R&D Director at PSA Peugeot Citroen gave an overview of such challenges for the whole automotive industry.

For the environment and the safety parts, the regulations are extremely high in terms of reduction of CO2 (-60% in 2025), NOx or causalities (EuroNCAP) and this will be achieved by reducing the weight of the cars (-20% expected with the use of multi-materials e.g composite, magnesium and optimised steels), optimise the aerodynamics (-30% expected) and working on powertrain improvement (downsizing of the engines, generalisation of the automatic transmission, hybrid and electric cars, …).

[3] http://www.sia.fr/actualite-b-congres-simulation323.htm

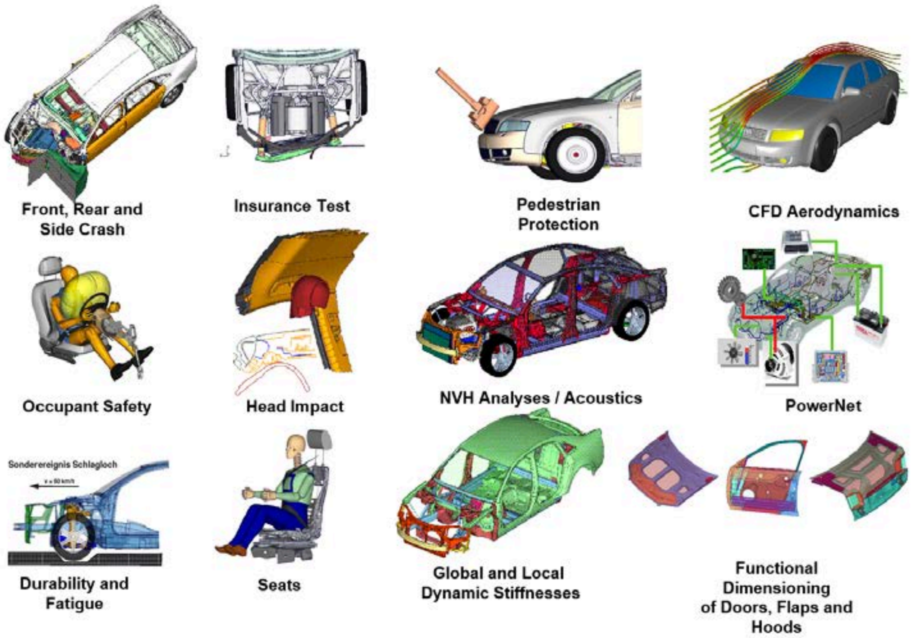


Figure 3 - Examples of use of CAE in the automotive industry (courtesy of AUDI AG)

All of these technologies are already big users of numerical simulation and HPC as illustrated by the figure above.

Addressing the challenges will require for staying competitive to increase largely the use of numerical simulation by using multi physics and multi scale simulations (especially when using composite materials) leading to increased use of HPC. This has been already shown during the last 2 years by the growing interest of companies like GM to the INCITE DoE program, of Renault, PSA and Continental for PRACE and Porsche or Daimler in Germany with HLRS.

Beyond standard simulations companies are also interested in using HPC for quantifying uncertainties for example in crash modelling where companies can estimate the influence of varying material parameters (fiber directions, sheets, scars, etc.) or varying geometry.

The life on board part is making the automotive entering into a new area with massive introduction of electronics on board with new human interface, voice/face recognition, new displays, and the development of virtual personal assistance. Here strong research in new algorithms and cognitive ergonomic will require the use of embedded HPC systems.

The challenge of affordability is very important for car companies, that means that they will need to address all these challenges by keeping an eye on the final price. This mean to perform strong optimisation studies in order to increase the percentage of re utilisation of parts, leveraging the development of multiple cars, between multiples brands which will cooperate (compete and collaborate).

Finally, the latest and one of the most HPC requiring challenges in the future will be connected cars and autonomous driving. This has been popularised by the Google car but now ALL the vendors and their Tier1 (like Dephi, Valeo, Continental, Faurecia, …) are developing these technologies.

The connected car will bridge all the on-board IT infrastructure of the car to the outside in order to provide traffic simulation, interaction with other cars, interaction with any other service for the safety or the leisure of the driver or the passengers.

The autonomous driving will come gradually with hand-on, hands-off, eyes-off and at the end even mind-off. The complexity between the first step: hand-on to the last one : mind-off is increasing by a factor of more than 1000 with obviously extremely low failure rate per hour expected, close to 10^{-9}. This is something which is already present in the aeronautics but automotive will require even more control and reliability since a pilot almost always have time to take back the commands while it could be impossible for a driver.

These 2 challenges are illustrating a strong convergence between HPC and embedded systems with severe issues on real time image/signal processing, data management, security, reliability and insurance regulations !

3.2 Developments of the working group of the past year

From PRACE allocations or use of national resources, some major achievements in scalability of industrial applications during the period were performed like:

- A team from Renault with the support of ESI Group (one of the leading ISV in the field of structure mechanics) and Ecole des Mines de St Etienne, performed thanks to a big allocation of 42 million core hours on CURIE (at GENCI@TGCC) the biggest crash modeling optimization study. They used models up to 20 million finite elements (5 times what they can currently run internally) with more than 200 different parameters studied (more than 10 times what is also usually done). Results of this multi objective optimization proven to the company the impact of advanced using numerical simulation and HPC at this scale with gains in weight while keeping or increasing safety. This project received in November 2014 the HPCWire Readers Awards for the best use of HPC in the automotive industry and gave to Renault a competitive advantage in the perspective of the EuroNCAP6 regulations to appear in 2016.
  In the same domain, a team composed by experts from PSA[4], Altair and Ecole Polytechnique performed initial works for scaling out (again on CURIE) their structure mechanics code RADIOSS to up to 16 384 cores and assessing an optimization methodology prior to submit a research project to PRACE. This result has been obtained using specific MPI and I/O optimizations performed by BULL.
- In the field of combustion, applied to explosions, the CFD team of CERFACS has been granted in 2014 with a DOE' INCITE allocation of 86 million core hours to study accidental gas explosions in buildings and develop models for turbulent pre-mixed flames. The target configurations for the proposed research on gas explosions are the original small-scale configuration of 0.25 m long, and two larger configurations of 1.5 m and 6 m long. The CFD team, with the support of TOTAL, has access to a unique experimental database (an explosion chamber at scales 1, 6 and 24 tested for multiple gases) and can validate his LES tools for a range of scales previously not accessible. To perform these computations, CERFACS relied on the AVBP code, one of the most advanced combustion solvers developed for parallel machines like the ALCF's Mira on up to 65k cores.

4 http://www.altairhyperworks.com/ResLibDownload.aspx?file_id=3794&from=ResourceLibrary.aspx%3fkeywords%3dRADIOSS%2bCase%2bStudy%2bwith%2bPSA%2bPeugeot%2bCitro%25c3%25abn%2band%2bBull

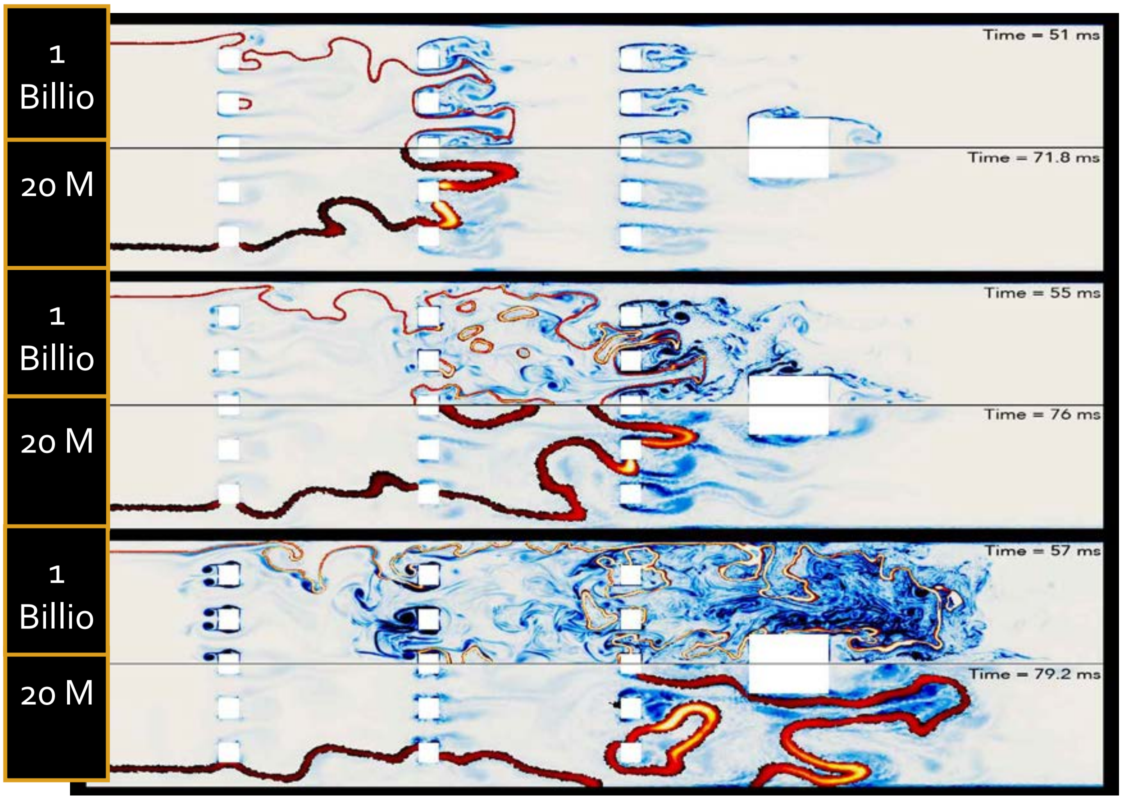


Figure 4 - low fidelity and high fidelity LES

- The same CFD team at CERFACS performed end of 2014 a coupled 3 stage aeronautic compressor LES simulation. Two instances of AVBP have been coupled using OpenPALM for a first ever simulation of such device provided by SAFRAN, one instance for the static and one for the moving parts of the compressor.

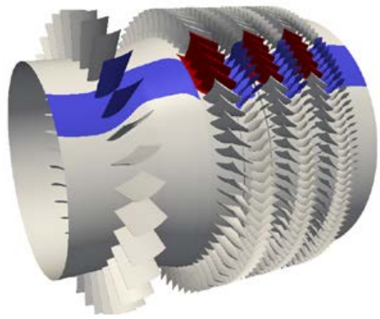


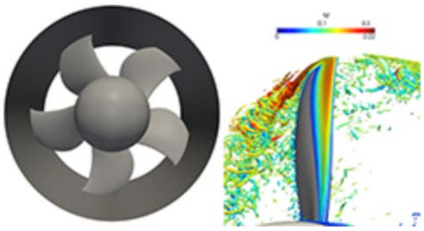
Figure 5 - Geometry of the SAFRAN CREATE compressor

Simulations on up to 7680 cores for each instance of AVBP have been conducted on OCCIGEN, a recent 2.1 PFlops cluster based on the latest Intel Hawell processors at CINES. Evolution of the turbulent flows on two configurations of 90 and 200 million meshes have been studied.

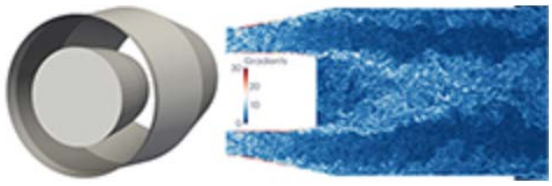
Figure 6 - Coupled AVBP simulations on 90 million meshes (left) and 200 million meshes (right)

- In the early installation phase of their new 3.8 PFlops CRAY XC40 and prior to making this machine available for general use, HLRS (Stuttgart, Germany) had invited end of 2014 national scientists and researchers from various fields to apply large-scale simulation projects called XXL-projects on Hornet. The goal was to deliver evidence that all related HPC hardware and software components required to smoothly run highly complex and extreme-scale compute jobs are up and ready for top-notch challenges.

Two teams from RWTH Aachen University in Germany performed 2 big CFD LES simulations on 92 000 and 94 646 cores of Hornet related to:



The better understanding of the development of vortical flow structures and the turbulence intensity in the tip-gap of a ducted axial fan



The analysis of the impact of internal perturbations due to geometric variations on the flow field and the acoustic field of a helicopter engine jet.

Following previous reports, some of the European leading edge HPC codes have been scaled out to petascale machines with results on up to 100 000 cores like ALYA a multiphysics code developed at BSC (Barcelona Supercomputing Center). BSC collaborated with NCSA (USA) for the optimisation and the scaling of the code on the Cray BlueWaters system[5]. Specific optimisations have been performed in the numerical solver and the I/O layer and tests cases coming from various domains including incompressible flow in human respiratory system, low mach combustion in kiln furnace and coupled electro-mechanical problem in a heart have been used.

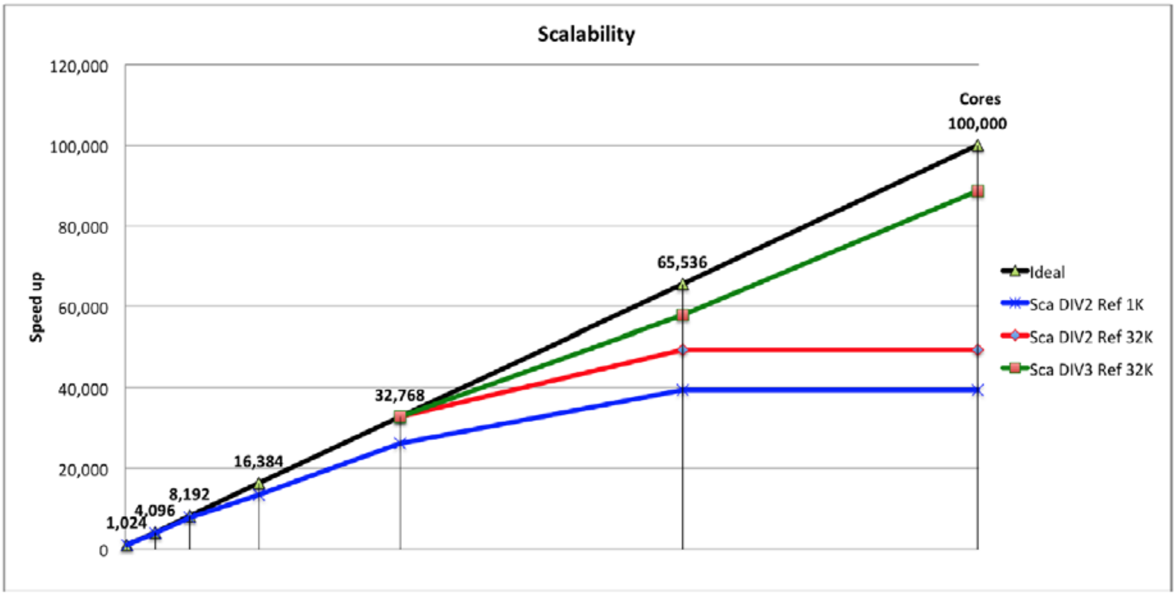


Figure 7 - Reported scalability of ALYA on up to 100 00 cores of BlueWaters

In aeronautics very promosing results have been obtained by Airbus together with Inria on the use of H-Matrices for electromagnetic simulations.

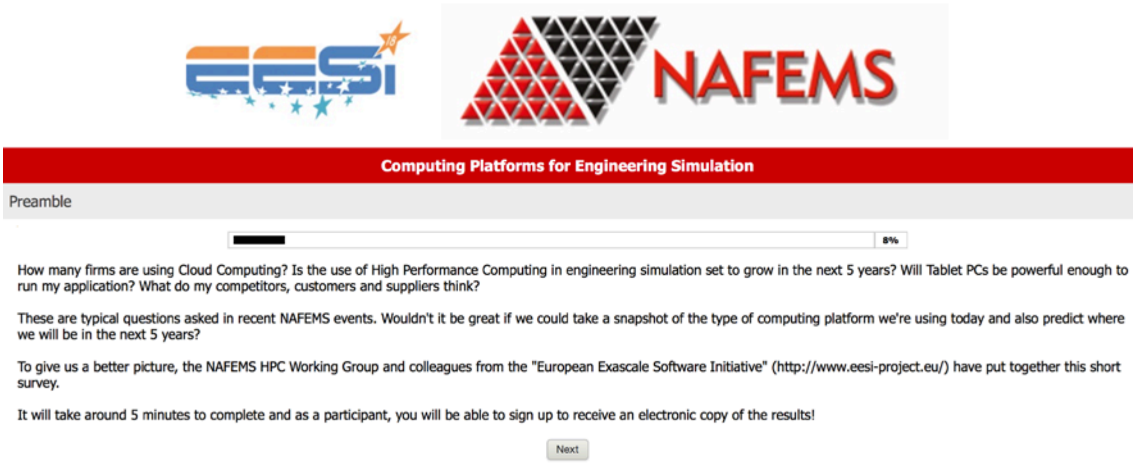
The Boundary Element Method (BEM) used for these simulations require the solving of large, ill-conditioned, dense linear systems with a large number of righ-hand sides ; the solving difficulty is the main practical obstacle to its use.

A hierarchical matrix (H_ Matrix) is a hierarchical, approximate, data-sparse storage format for

[5] http://arxiv.org/pdf/1404.4881v1.pdf

matrices that can be manipulated to produce a direct linear solver with an asymptotic space and time complexity of O(Log(N))
The work considers the relevance of the H -Matrices in terms of precision and computation time, in comparison with a classical direct solver and with an iterative solver based on the Fast Multipole Method. The first results reported, even for the moment on low number of cores (up to 32 for the moment) are quite promising in comparison with direct solver or even optimised iterative solvers with FMM methods, both in term of time to solution and also accuracy of the results.

Finally, the NAFEMS association, one of the biggest worldwide association in the field of engineering, in collaboration with EESI2 launched mid 2014 a survey among the Engineering community about the growth of simulation and high performance computing needs among companies, the use of cloud computing, how competitors or suppliers are doing, …


Computing Platforms for Engineering Simulation

Preamble

8%

How many firms are using Cloud Computing? Is the use of High Performance Computing in engineering simulation set to grow in the next 5 years? Will Tablet PCs be powerful enough to run my application? What do my competitors, customers and suppliers think?

These are typical questions asked in recent NAFEMS events. Wouldn't it be great if we could take a snapshot of the type of computing platform we're using today and also predict where we will be in the next 5 years?

To give us a better picture, the NAFEMS HPC Working Group and colleagues from the "European Exascale Software Initiative" (http://www.eesi-project.eu/) have put together this short survey.

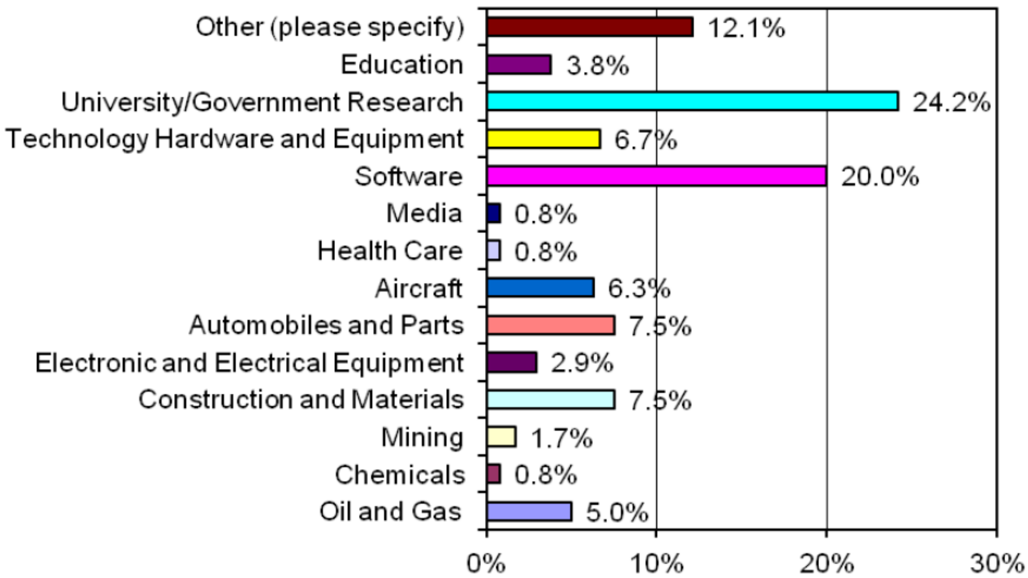
It will take around 5 minutes to complete and as a participant, you will be able to sign up to receive an electronic copy of the results!

Next

Figure 8 - Front Panel of the joint NAFEMS/EESI Survey

The results of this survey will be presented by Lee Magrets (Chairman of the HPC Working Group) at the 2015 NAFEMS World Congress in June 2015 and full information will be provided into the EESI2 WP3 final deliverable.

Nevertheless, the survey collected 231 respondents coming from Europe (58.8%), Americas (23,6%), Asia (5.4%), Middle East and Pacific (3.4%). Such respondent were working in engineering firms (30%), public sector (29%), software (20%), …



SMEs were quite well represented with 47% of the respondent working in companies with less than 250 employees.

One of the objectives of the survey was to get an idea of the size of the largest simulations carried out in the community and the size of the largest facilities used for those simulations. Respondents were asked to indicate the size of the largest simulation carried out in their organisation, measured in terms of the number of degrees of freedom, for a list of different types of engineering analysis. They were also required to indicate the maximum number of cores used in each of the categories of engineering simulations. Examples of first results when splitting between academia and industry are the followings:

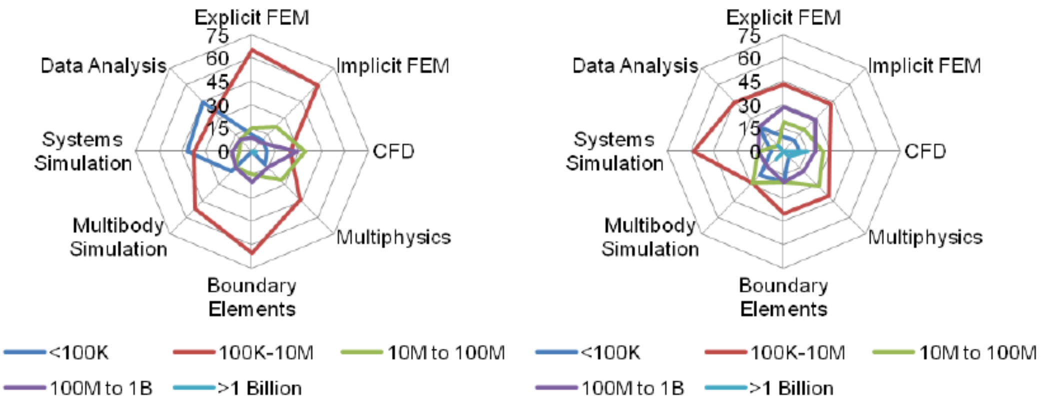


Figure 9 - Maximum number of DoF solved by type of simulation in industry (left) and academia (right)

Considering responses from industry, Figure 9 shows that the largest problem sizes tackled in finite element and boundary element analyses are in the 100,000 to 10 million degree of freedom range. There is a strong bias towards smaller problems for systems simulation and data analysis and towards larger problems for CFD.

The profile of analyses carried out by respondents working in research (universities and government laboratories) is very different. Whilst problems in the 100,000 to 10 million degree of freedom range dominate, there is a higher proportion of respondents tackling larger problems.

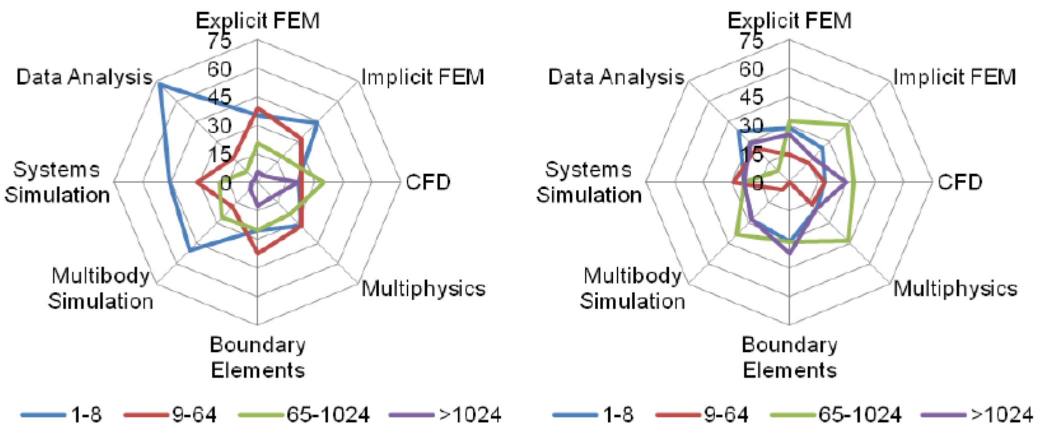


Figure 10 - Maximum corresponding number of cores uses in industry (left) and academia (right)

Figure 10 highlights strong differences in the maximum number of cores used between industry and academia. For all categories of simulation, larger core counts are used in academia compared with industry. The use of 1-8 cores in industry for data analysis, systems simulation, multibody simulation and implicit FEM is particularly striking. The differences may be due to different licensing fees for ISV software (cheaper for large core counts in academia compared with industry) or a greater tendency to use academic and/or open source software in academia.

Finally the study tries also to assess more specifically the roadmap of development toward future HPC technologies of the software editors, even ISV (Independent Software Vendors) or OSS (Open Source Software).

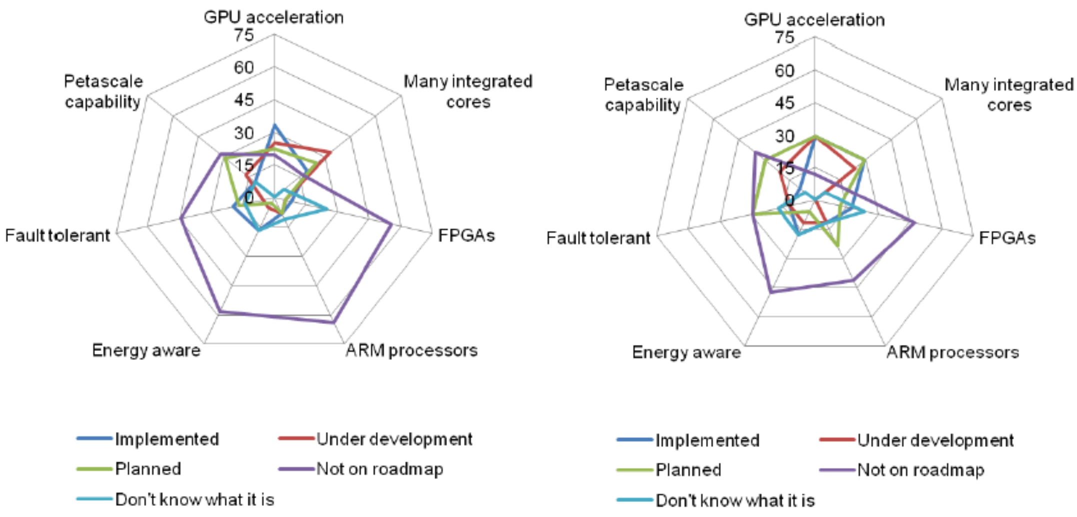


Figure 11 - Roadmap for organisation that produce software - Comparison between ISV (left) and OSS (right)

The results of these analysis are clearly showing that while GPU are starting the be adopted (because with workstation they represent an important market for software vendors), the others technologies/challenges for Exascale are not very considered/known by vendors. The situation is a bit better for Open Source Software (right) but there is clearly an issue for such communities to take the HPC waggon.

This study which need to be more analysed is already leading to the three following EESI recommendations:

1. Improve engagement with trade associations.

Engagement with NAFEMS members has provided industry insights that have not been gained elsewhere by the Exascale community.

> Increase efforts to engage with industrial trade associations across a broad range of business sectors.
>
> Disseminate EESI findings and recommendations through industrial trade association publications. The responses “Not on roadmap” and “Don’t know what it is” for Exascale technologies are significant; indicating that the Exascale message might not be getting through to organisations carrying out software development.

2. Support academic and open source software.

ISVs are motivated by quarterly sales and are unlikely to invest significantly in emerging technologies, including novel hardware or new algorithms.

> This report provides further evidence that ISVs and industry should be supported in the future by investing now in academic, open source software. Once the emerging technologies become mainstream, the ISVs will have solutions at the ready.
>
> Investment in software for systems simulation is particularly recommended as it offers a unique opportunity to make use of large HPC capability by releasing suppressed capability in

existing software components. Systems simulation joins together FEM, BEM, CFD and other technologies in a workflow to look at virtual machines, rather than virtual components for machines. It is clear from this survey that the current individual capabilities of FEM, BEM and CFD in terms of problem size and core counts, when used to look at components, are not used when these technologies are used to look at machines. Whatever the type of engineering simulation carried out, it seems to be mainly confined to a workstation or shared memory node. Thus systems appear to be over-simplified. R&D in larger-scale systems simulation could be carried out re-using existing software components, but adapted for emerging technologies.

This is a different philosophy to pushing a single CFD simulation, for example, to Exascale.

3. Attempt to disrupt the market.

Some ISVs have commented that they do not have access to large HPC systems, particularly for emerging technologies, for testing. Easier access mechanisms to academic facilities would help ISVs test and optimise their software.

The engineering simulation community does not have a suite of “typical” benchmarking problems for “performance”. It is therefore difficult for end-users to compare/evaluate the performance of different combinations of software and hardware for their particular problem. Investment in domain specific benchmarking suites may help firms indentify ISV software with the best price vs performance vs feature. Those ISVs would out compete those that are slower in adapting to the support of HPC platforms.

This first analysis elaborated by Lee Magrets is attached as annex of this deliverable.

3.3 Gap Analysis

As stated in the paragraph 3.1, the lower level of investment in the Oil & Gas industry had a little impact in the roadmap of development of new seismic models and the level of equipment of oil & gas companies.

At the beginning of the project, 18 months ago, the following roadmap for the Oil & Gas industry, derivated from the one from TOTAL but adopted by a lot of other companies was the following:

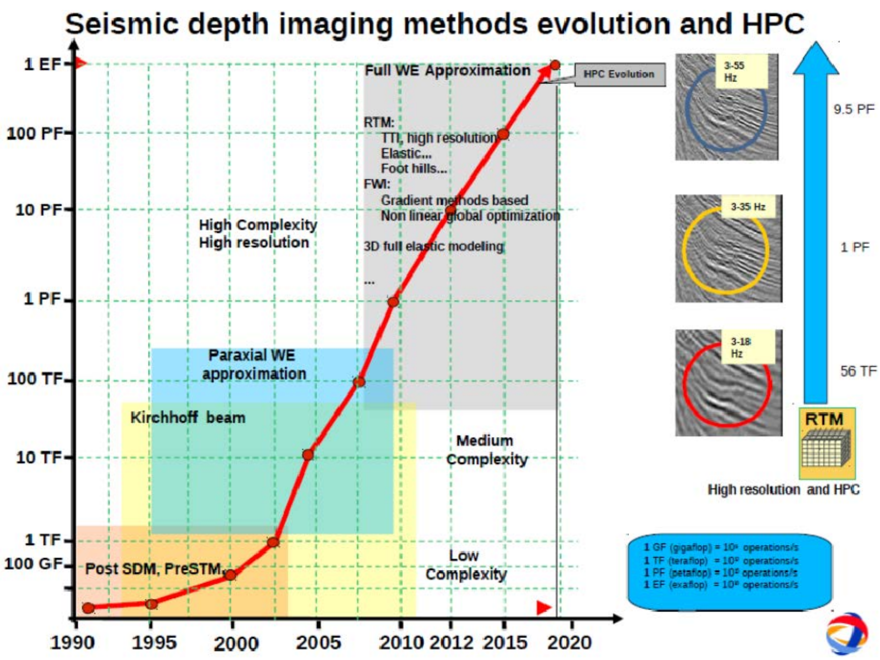


Figure 12 - Previous Oil & Gas Exascale Roadmap produced during EESI1

This roadmap was expecting to see the rise of 100 PFlops systems in the Oil & Gas domain for 2015 for addressing new models like elastic RTM, TTI (tilted transverse isotropy) or high resolution RTM.

A new roadmap provided by TOTAL (H. Calandra) end of 2014 is now the following:

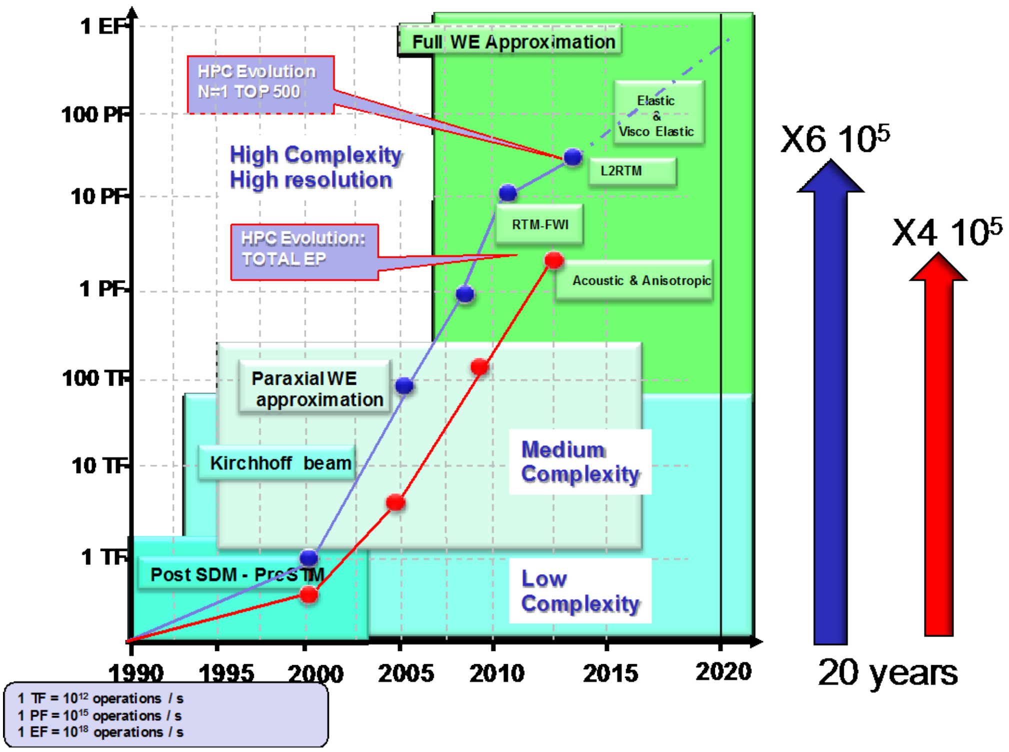


Figure 13 - Updated Oil & Gas Exascale Roadmap (from TOTAL)

This figure 13 shows the increase of computational power (up to 3 PFlops now, close to 7PFlops at the end of the year) needed for addressing the development and the use of such new seismic methods and a little shift of 3 years in the provision of capacity of HPC systems for sustaining such developments. In 20 years this curve shows also that Oil & Gas companies are following a trend very similar to the #1 of the top500, making this sector one of the most important user of HPC.

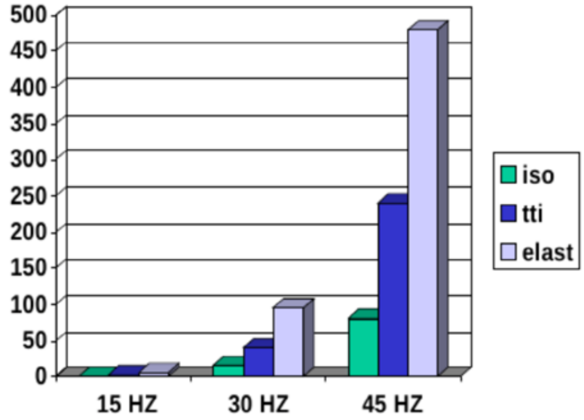


Figure 14 - increase of complexity of seismic methods wrt frequency

But it's also important to take into account the computational needs of such algorithms when increasing the frequency of acquisition. The figure on the left represents the increase in computational intensity of the different RTM methods when increasing the frequency of seismic acquisition from 15 to 45 Hz.

Finally experts and communities worked in the development of the first EESI2 recommendations produced in 2013, especially in the field of the development of mini applications for applications performance analysis, characterisation, prediction and porting on emerging architectures. It was the case for example in the field of the Exa2CT[6] European project with the development of a mini application derivated from the YALES2 CFD and combustion code developed by CORIA.

3.4 Project recommendation " Towards flexible and efficient Exascale software couplers"

Experts from WG3.1 have been involved into the recommendation toward the development of flexible and efficient software couplers. Coupling of multi physics at multi scales codes is a critical issue for aeronautics, automotive, oil & gas, energy, … almost all the fields of engineering. Like in climate and wheater forecast, Europe owns the development of many couplers used by engineering like OpenPALM for the coupling of CFD codes in turbomachines (CERFACS), the coupling components in the SALOME platform for the efficient coupling of neutronics and thermohydrolic codes in energy, the IFS coupler again in energy for FSI (fluids Structure Interactions) or the MpCCI Coupling interface developed by Fraunhofer for general coupling of engineering codes.

The recommendation "High Productivity Programming Models" is born with the motivation to optimise the couplers used and it is described in detail in the document "D7.2 EESI2 Second Annual Report 2014 Update Vision & Recommendations", and it is also attached as an annex to this deliverable.

EESI2 recommendation towards couplers went to the establishment of IP projects over a period of 4 years, with a budget of 8 to 12 M€ for addressing the following improvements and developments:

- On the coupler itself: development of a standard coupling API in order to enable interoperability, improvement of the performance of localization process and optimization of this process for geometrical or mesh changes during the simulation, optimization of the

[6] http://www.exa2ct.eu/index.html

communications, perform intelligent search, … and adapt couplers to heterogeneous manycore architectures using new programming models and systems for parallelism (PGAS, hybrid MPI/OpenMP …).
It's important to note that Europe owns a big majority of the couplers available, so developing
- On the coupled models: perform advanced comparison between single and multi executables and reduce overall communication cost
- On the software environment: development and optimization of related tools for mesh connection between model, quick verifications of conformity, evaluation of physical quantities during computations, joint exploitation of massive results …

3.5 Project recommendation " Identification of turbulent flow features into massively parallel Exascale simulations"

Experts of WG3.1 were also involved into the EESI2 recommendation related to the development of

The rationale relies on the fact that the rise of multi petascale and upcoming exascale HPC facilities will allow to turbulent simulations based on LES and DNS methods to address high fidelity complex problems in climate, combustion, astrophysics or fusion. These massive simulations performed on tens to hundreds thousands of threads will generate a huge volume of data, difficult and inefficient to post process asynchronously later after by a single researcher. The proposed approach consists on post processing this rough data on the fly by smart tools able automatically to extract pertinent turbulent flow features, store only a reduced amount of information or provide feedback to application in order to steer its behaviour.

To address such issues, it is mandatory to develop a complete toolbox of efficient parallel algorithms based on:

- Massively parallel high-order low-pass and band-pass filters
- Conservative high-order interpolation kernels for the interpolation of fine grids to coarser grids
- Massively parallel Singular Value Decomposition algorithms for Dynamic Mode Decomposition of large sets of data
- Highly efficient linear solvers for symmetric matrices as those encountered in implicit filters

This recommendation is very close to the other recommendations related to in-situ and in-transfer post processing of massive amount of data, providing here underlying tools. Its important to note that such tools and methodologies will become mandatory for providing to industry the way to address large volume of data, asses uncertainties and reduce time and cost to solution.

3.6 Project recommendation " Enhanced unified framework for model verification & validation and uncertainty quantification"

Experts of WG3.1 were also finally involved with experts from WG3.2 and WG5.2 on the definition of a recommendation related to the development of an European framework for model verification and validation and uncertainty quantification.

Computer simulation has become a major technology in daily engineers' work for understanding, forecasting and guiding decision based on modelisation of large-scale complex multi-scale and multi-physics phenomena dealing with massive amount of data generated by instruments or by simulations themselves.

The rationale of such recommendation relies on the fact that relying on extreme computing with data intensive capacities has become mandatory for addressing such scientific and industrial challenges but a quantitative uncertainty assessment of input, output data as well as models is a fundamental issue for assuring the credibility of computer model based studies, and represents a challenge too.

Here again, VVUQ (Verification, Validation and Uncertainty Quantification) is a domain where Europe has a strong card to play with existing frameworks like URANIE by CEA/DEN (France) or OpenTurns (France).

In that context its proposed that targeted IP funding tools over 4 years could host projects aiming the development of a unified European wide UUVQ framework. It should mean an approximate 15 million Euros budget

The proposed recommendation aims at:

- Preparing VVUQ and optimisation software for exascale computing by identifying and solving problems limiting the usability of these tools on many-core configurations,
- Facilitating access to the VVUQ techniques to the HPC community by providing software that is ready for deployment on supercomputers,
- Making methodological progresses on the VVUQ and optimisation methods for very large computations.

4. Weather, Climatology and Solid Earth Sciences

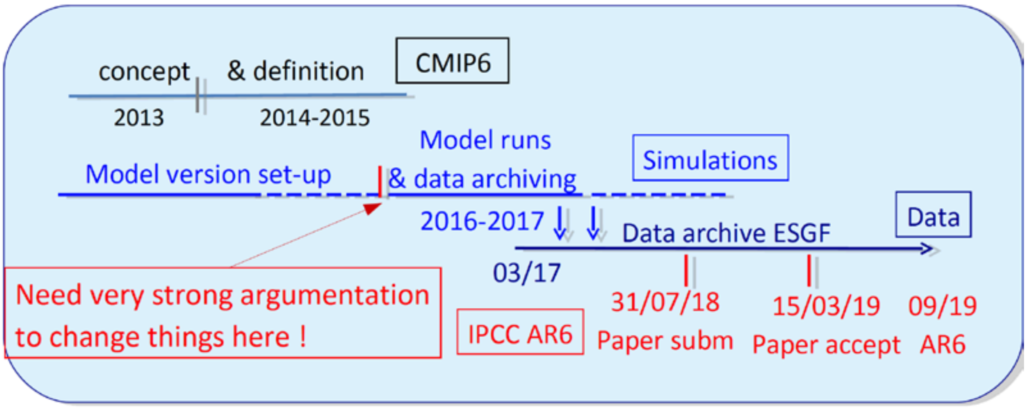
4.1 Advances, results and breakthroughs in the international context

In the climate domain, just few months after the publication of the last report of the fifth IPCC campaign (AR5), the community is now focussed on the definition of the goals of the next campaign, expected to start in 2016 until mid 2018 for a publication of the first research papers. This exercise based on coordinated climate models experiments involving multiple international modelling teams will focus on 6 grand challenges on clouds, circulation and climate sensitivity, changes in cryosphere, climate extremes, regional climate information, regional sea-level rise, water availability and biogeochemical forcing and feedbacks.

Participants will need to assess the following scenarios and simulations:

- AMIP simulation between 1979 to 2014
- Pre industrial control simulation
- Increase if 1% per year of CO_2
- an instantaneous 4xCO2 run to derive the equilibrium climate sensitivity
- Historical simulation using CMIP6 forcing (1850 to 2014)

The current agenda of CMIP6 is the following:



Regarding the previous exercise (CMIP5), CMIP6 will lead to longer simulations, bigger ensemble, more variables, higher resolutions, … and by consequence more data generated and made available through the Earth System Grid Federation (ESGF), an open source software infrastructure for the management and analysis of Earth Science data on a global scale.

In this context and before the start of the real operational phase of CMIP6, climate modelling teams are currently intensifying their efforts for developing and optimising their models (ocean, ice, atmosphere, chemistry, …) and pre and processing tools for fitting to the requirements of CMIP6. It is especially concrete in the field of the development of new dynamical cores where I/O, memory-bound and multi-physics constraints of such models need radical new approaches and paradigms which are better suited to exploit high levels of concurrency found in emerging HPC systems. Also in the field of post processing a lot of efforts are done by the community toward the development of new frameworks for I/O delegation taking into account the underlying storage hierarchy, in-situ and in-transit post processing, data compression and improvement of the existing parallel I/O libraries (NetCDF5 for example).

Another new scientific driver for the climate community is the rise of decadal climate forecasting. Besides improving the description of greenhouse-induced climate change and its course over the next decades and centuries, what its called climate predictions, it will be more and more important to forecast what detailed climate will occur in 5 years, or 10 years, of 30 years in the future, because such climate characteristics are crucial for planning most of the industrial and societal activities.

What will be required for forecasting climate over yearly-to-decadal time scales?

- As for meteorological forecasting, where the detailed description of the initial state of the atmosphere is of critical importance, forecasting the climate will require an adequate description of the initial state of the natural fluctuations, i.e. a detailed description of both the atmospheric and, more importantly, oceanic states. This implies processing a much larger amount of data as compared to what has been done up-to-now;
- Data assimilation will also be needed, as the initial state of the climate, both in the ocean and in the atmosphere, must be compatible with the coupled ocean-atmosphere model that will be used to forecast future time steps. Much larger computing resources for optimizing the initial climate state wrt the coupled ocean model will be crucially needed. It should, e.g., be recalled that in actual numerical meteorological forecasting, the computing resources devoted to defining an adequate initial state through data assimilation is significantly larger that the amount of resources needed for advancing in time from this initial state;
- Assuming that the above two requirements are correctly satisfied, the societal demand will increase for performing regularly updated climate forecasts, e.g., monthly for multi-year forecasts and at least yearly for multi-decadal forecasts.

Quite obviously, entering the climate forecasting era, which might happen quite soon, will require an increase of one order of magnitude, or more, for data processing capacities and for computing resources as compared to the actual state of climate modeling.

In the US, its notable to see that 8 Department of Energy national laboratories, including Lawrence Berkeley National Laboratory, are combining forces with the National Center for Atmospheric Research, four academic institutions and one private-sector company in the new effort. Other participating national laboratories include Argonne, Brookhaven, Lawrence Livermore, Los Alamos, Oak Ridge, Pacific Northwest and Sandia.

The project, called Accelerated Climate Modeling for Energy, or ACME[7,] is designed to accelerate the development and application of fully coupled, state-of-the-science Earth system high resolution models (12 to 25km for the atmosphere, 15 km for the ocean for 80 years simulations) for scientific and energy applications. The plan is to exploit advanced software and new High Performance Computing machines as they become available.

The initial focus will be on three climate change science drivers and corresponding questions to be answered during the project's initial phase:

- (Water Cycle) How do the hydrological cycle and water resources interact with the climate system on local to global scales? How will more realistic portrayals of features important to the water cycle (resolution, clouds, aerosols, snowpack, river routing, land use) affect river flow and associated freshwater supplies at the watershed scale?
- (Biogeochemistry) How do biogeochemical cycles interact with global climate change? How do carbon, nitrogen and phosphorus cycles regulate climate system feedbacks, and how sensitive are these feedbacks to model structural uncertainty?
- (Cryosphere Systems) How do rapid changes in cryospheric systems, or areas of the earth where water exists as ice or snow, interact with the climate system? Could a dynamical instability in the Antarctic Ice Sheet be triggered within the next 40 years?

Over a planned 10-year span, the project aim is to conduct simulations and modelling on the most sophisticated HPC machines as they become available, i.e., 100-plus petaflops machines and

[7] https://climatemodeling.science.energy.gov/sites/default/files/publications/acme-project-strategy-plan_0.pdf

eventually exascale supercomputers. The team initially will use U.S. Department of Energy (DOE) Office of Science Leadership Computing Facilities at Oak Ridge and Argonne national laboratories.

The model will also be optimized for deployment on the National Energy Research Scientific Computing Center (NERSC), which is located at Berkeley Lab.

4.2 Developments of the working group of the past year

Experts of the working group have been involved together with ENES in the organisation in Hamburg on April 2014 at DRZK of the third edition of the « Exascale Technologies & Innovation in HPC for Climate Models»[8] workshop.

This even gathered more than 70 international participants coming from climate, weather forecast, HPC and applied mathematics, PRACE and European computing centers, and allowed discussions around the evolution of the climate models in terms of future trends in climate sciences and related HPC challenges, status of the EU Exascale or the G8 projects, new parallel approaches, performance inter-comparisons of climate models, porting of climate codes on leading edge systems, …

During this workshop some of the following issues were addressed:

- P. Luigi Vidale (NCAS/UK Met, Reading UK) highlighted the results of the UPSCALE project submitted to PRACE in 2011 with 144 million core hours on Hermit (HLRS). This project completed multi-decadal simulations at a global 25km high resolution of a set of ensemble runs for assessing extreme events (hurricanes, flooding, heat waves, …). It also allowed the development of a new dynamical core (called EndGame) in order to reduce the amount of data generated (2 to 4 TB/day) and the next Global climate models at a global resolution of 12 km.
- Y. Meurdesoif (IPSL) presented the development of DYNAMICO, a new icosaedral dynamic core for atmosphere models (replacing LMDZ) with a current scalability on up to 50k cores (the aim is to go to 200k cores). This development relies on the use of MPI/OpenMP and a specific I/O layer called XIOS for delegating asynchronously I/O operations and first level of in-situ post processing to dedicated servers.
- Multiple presentations showed current state of the porting (or the refactoring) of climate models (ocean, atmosphere) on GPU (often using OpenACC) and Intel MIC architectures.

Weather and climate research communities heavily use national or European (PRACE) HPC facilities, and will need more and more of these resources with the development of high-resolution simulations and of climate prediction activities (see above). Some organisation also renewed their internal facilities like:

- UKMet announced in November 2014 an investment of £97m into a new HPC center, including a 16 PFlops CRAY XC40 and Sonexion attached storage (more than 20 PB @ 1.5 TB/s). Following the proof of concept of the UPSCALE project on PRACE resources in 2011, these advanced capabilities deployed between 2015 and 2017 will deliver an estimated £2 bn of socio-economic benefits for the UK through enhanced resilience to severe weather and related hazards.
- MétéoFrance installed a 1 PFlops BULL solution (2 clusters of 0.5 PFlops each in 2014, upgraded in 2016 to 2.2 PFlops each) for performing high resolution predictions using its 3 models (ARPEGE, ALADIN and AROME) on resolutions of up to 1.3km, and also preparing the next IPCC campaign with high resolution climate models.
- DKRZ in Germany acquired to BULL a 3 PFlops cluster with 45 PB of storage for running high-resolution climate models in the perspective of the next 6th IPCC campaign.

[8] https://verc.enes.org/ISENES2/events/ws3

From PRACE allocations or use of national resources, some major achievements in scalability of weather, climate and solid earth sciences during the period were performed like:

- A team composed by researchers from Technische Universitat Munchen, Germany, received the PRACE Award in 2014 for the a sustained petascale performance for a 3D seismic application on the SuperMUC system in LRZ (Germany)[9]. These simulations performed with up to 90% parallel efficiency and 45% floating point peak efficiency on 147k cores reached a sustained performance of 1.09 PFLOPS. In order to reach this level of performance, specific improvements on the Arbitrary high-order accurate DERivative Discontinuous Galerkin (ADER-DG) method on fully adaptive unstructured tetrahedral meshes, hardware-aware optimization of the innermost sparse/dense matrix kernels, MPI/OpenMP parallelisation and optimisations of the reading of the mesh and the parallel writing of the simulation results.
- ICON a dynamical cores model developed by The Max Planck Institute for Meteorology (MPI-M) and the German Weather Service (DWD) and which solves the fully compressible non-hydrostatic equations of motion for simulations at very high horizontal resolution (120m) has been scaled out to 458,752 cores (1,835,008 compute threads) on BlueGene/Q (JUQUEEN) and to up to 65536 cores on SuperMUC (LRZ, Munich).

- In the early installation phase of their new 3.8 PFlops CRAY XC40 and prior to making this machine available for general use, HLRS (Stuttgart, Germany) had invited end of 2014 national scientists and researchers from various fields to apply large-scale simulation projects on Hornet. The goal was to deliver evidence that all related HPC hardware and software components required to smoothly run highly complex and extreme-scale compute jobs are up and ready for top-notch challenges. One of these 6 projects was related to running a latitude belt simulation around the Earth at a resolution of few km enough to cover various extreme events of the Northern Hemisphere (such as the typhoon Soulik in the Pacific from July 10-12, 2013) and to study the model performance.
- The storage capabilities of the new Hornet system allowed the scientists to run the simulation without any interruptions for more than three days on 84 000 cores. Using the combination of MPI+OpenMPI including PNetCDF libraries, the performance turned out to be excellent. Another finding was that not the computing time but the I/O performance became the limiting factor for the duration of the model run.

Experts of WG3.2 were also involved in the BDEC meetings organised in Fukuoka (Japan, Feb. 27 2014) and Barcelona (Spain, Jan 30, 2015) with contributions in the field of the development of mini-apps in weather and climate models for performance optimization[10] (numerical solvers with a high resolution version of the ocean model NEMO, I/O and data assimilation mini-app), software infrastructure for big data analytics or issues related to the numerical convergence of climate simulations.

In the field of weather forecast a Scalability Programme has been initiated at ECMWF with the aim of coordinating developments towards a more scalable Integrated Forecasting System (IFS) and the ancillary support software. The Programme will provide accurate, efficient and scalable algorithms and code structures to cater for a variety of potential future high-performance computer (HPC) architectures. The Programme will coordinate resources from across the Centre to define the future forecasting system across all scales. Collaboration is crucial for the Programme’s success. In addition to Member States, some partnership with consortia like HIRLAM, ALADIN, COSMO, NEMO, NEMO-VAR, HPC centres, the climate modelling community and hardware companies, have been established. Vendors have an important role in the design and providing advice and access to the latest computing architectures. The collaboration with HPC centres allows to run the development codes on emerging novel computer hardware. Two different approaches have been considered:

[9] http://www5.in.tum.de/pub/isc14_seissol_supermuc.pdf

[10] http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/whitepapers/andre_0.pdf

- In the short-term some actions, such as to overlap communication/computation, to outsource (e.g. radiation on GPUs), to extend co-array structures, to employ OpenMP 4.1 once available, to use OpenACC, to migrate towards a single executable, to distribute I/O.
- In the long-term other actions, such as alternatives to global spectral model, local data structures, more flexible top-level control structures, sub-windows in sequential data assimilation.

Finally, experts of WG3.2 in collaboration with ENES have been involved into EU H2020 first calls for proposals:

- CHANCE, in the field of the FET-HPC1-2014 call, leaded by G. Aloisio (CMCC) and Mike Ashworth (STFC) which aims at exploiting new mathematical and algorithmic approaches for existing or emerging applications on extreme scale systems. This project will implement some of the EESI2 2014 recommendations in implementing into NEMO (or its successor) communication avoiding and parallelization in time methods.
- ESiWACE (Excellence in Simulation of Weather and Climate in Europe) In the field of the H2020-EINFRA-2015-1 call, which aims to setup a Center of Excellence in the field of climate research and numerical weather prediction, leaded by Joachim Biercamp (DKRZ) with specific goals in scaling out new climate models, developing tools for environment and workflows and data management.
- CliM-ERI, which aims to create a new ESFI research infrastructure, dedicated to climate research.
- PRIMAVERA (PRocess-based climate sIMulation: AdVances in high-resolution modelling and European climate Risk Assessment) leaded by PL. Vidale (Univ. of Reading, UK) which aims to follow the work done by Vidale and all. using PRACE resources in high resolution climate models (UPSCALE project using the HadGEM3 model) but going to even higher resolutions and capable of simulating and predicting regional climate with unprecedented fidelity, out to 2050.
  The list of participants and the different ocean models involved is the following :

Institution	MO/NCAS/ NOCS	KNMI/SMHI/ IC3/CNR	CERFACS	MPI	CMCC	ECMWF	AWI
Model names	UM / NEMO	ECEarth / NEMO	Arpege / NEMO	ECHAM / MPIOM	CCESM / NEMO	IFS / NEMO	ECHAM/ FESOM
Atmospheric resolution	60-25	T239-T799	T359	T255	25km	T239-T799	T255
Oceanic resolution	¼-1/12°	¼°-1/12°	¼-1/12°	¼-1/10°	¼	¼	¼ - 1/12 spatially variable

Or have been involved in the field of data management in collaborations with EUDAT (as one of the 9 scientific communities providing needs and feedback for the development of data management methodologies and tools) and the Research Data Alliance (RDA, through the participation of DKRZ in working groups related to Data Citation or PID information types).

4.3 Gap Analysis

The gap analysis described in the previous deliverable is still up to date. Its by the way important to notice that the climate community and the numerical weather prediction community together with private partners answered to the first call for proposals for setting up a European Center of Excellence. Even if the expected budget allocated to each of the 8 to 10 CoE will be limited (4.5M€) and 10 times less than the EESI1 recommendations, these 2 communities are committed to address Exascale challenges in scalability of the models, data management and tools for ensemble and workflows.

Experts and communities worked in the development of the first EESI2 recommendations produced in 2013, especially in the field of the development of mini applications for applications performance analysis, characterisation, prediction and porting on emerging architectures. Experts from the working groups presented their work during BDEC meetings[11] applied to the following codes It is especially the case for the following applications:

- The development of a mini application derivated from the solver of the NEMO high-resolution oceanic model. The mini-app built around this kernel allows to evaluate the performance portability on several computing platforms and estimate the computational performance achievable by hybrid parallel approaches, as well as to evaluate how new parallel architectures, based on accelerators (GPU, MIC), can improve the code scalability;
- The development of a mini application dealing with I/O: Flexible methods for the management of I/O are necessary for, e.g., either minimizing the call of subroutines related to I/O definition (file creation, axis and dimensions management, adding and output field...) or minimizing arguments of I/O call. Such methods should not be invasive but allow high performance targeted for massive parallel simulation (10 000 or more cores): writing data must not slow down the computation, by, e.g., using one or more “server” processes dedicated exclusively to the I/O management;
- The development of a mini application used in the context of data assimilation: The data assimilation method used in the oceanic NEMO model is treated as a separate part of the code, referred to as the NEMOVAR software. It can be used to test either 4-dimensional assimilation method, where the adjoint code is integrated backward at a resolution lower than the direct code (the so-called incremental method), or ensemble methods based upon the Kalman filter theory, or even hybrid method.

4.4 Project recommendation " Towards flexible and efficient Exascale software couplers"

Experts from WG3.2 have been involved into the recommendation toward the development of flexible and efficient software couplers. Coupling of multi physics codes for ensemble simulation is a critical issue for climate and weather forecast experts and one of the most popular software coupler in Europe called OASIS[12] is supported by ENES through the IS-ENES FP7 project.

Regarding previous evolutions of OASIS, the new version of this coupler called OASIS3-MCT, interfaced with the Model Coupling Toolkit (MCT) from the Argonne National Laboratory, offers today a fully parallel implementation of coupling field re-gridding and exchange. Low-intrusiveness, portability and flexibility are OASIS3-MCT key design concepts as for all previous OASIS versions. An important difference with respect to previous OASIS3.3 is that there is no longer a separate coupler executable: OASIS3-MCT is a coupling library that needs to be linked to the component models, with the main function of interpolating and exchanging the coupling fields between these components.

The recommendation "High Productivity Programming Models" is born with the motivation to optimise the couplers used and it is described in detail in the document "D7.2 EESI2 Second Annual Report 2014 Update Vision & Recommendations".

EESI2 recommendation towards couplers went to the establishment of IP projects over a period of 4 years, with a budget of 8 to 12 M€ for addressing the following improvements and developments:

- On the coupler itself: development of a standard coupling API in order to enable interoperability, improvement of the performance of localization process and optimization of this process for geometrical or mesh changes during the simulation, optimization of the communications, perform intelligent search, … and adapt couplers to heterogeneous manycore architectures using new programming models and systems for parallelism (PGAS, hybrid MPI/OpenMP …).

[11] http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/whitepapers/andre_0.pdf

[12] https://verc.enes.org/oasis

It's important to note that Europe owns a big majortity of the couplers available, so developing

- On the coupled models: perform advanced comparison between single and multi executables and reduce overall communication cost
- On the software environment: development and optimization of related tools for mesh connection between model, quick verifications of conformity, evaluation of physical quantities during computations, joint exploitation of massive results …

5. Fundamental Sciences

5.1 Advances, results and breakthroughs in the international context and developments of the working group of the past year

In the field of materials, the US Materials Genome Initiative (MGI) published end of 2014 its Strategic Plan[13]. MGI is an effort to double the pace of advanced-materials discovery, innovation, manufacture, and commercialization. It is part of a broader set of concrete actions launched by President Obama in 2011 to revitalize American manufacturing and the Initiative has made tremendous progress in its first three years. This Strategic Plan will serve as a framework for Federal agencies as they work to execute on the goals of MGI. The Strategic Plan highlights four sets of goals of the MGI:

- Leading a culture shift in materials-science research to encourage and facilitate an integrated team approach;
- Integrating experiment, computation, and theory and equipping the materials community with advanced tools and techniques.
- New computational methods implemented in software must be verified against known solutions and developed in concert with experiments to validate the output. As outlined in the next objective, specialized experimental tools often are required to provide the data necessary for validation. In addition, the integration of these advanced computational tools into experimental designs will drive faster and more robust experimental results from materials discovery through testing and integration of components.
- Specific technical barriers in simulation also impede substantial advancement in the field of materials. For example, the materials science and engineering community has long recognized the challenges of multiscale theory and modeling. Since a material’s performance is influenced by dynamics encountered at all length scales—from the atomic to macroscale—effective material design requires the integration of models, as well as the information derived from models, from many length scales. Equally important are the needs to quantitatively characterize and model a material’s evolution to capture phenomena over the timescales relevant to application targets for industrial use. Directed efforts within MGI can address these specific technical needs; community input is needed to define the major scientific and technical challenges for theory, modeling, and simulation for all material types.
- Making digital data accessible; and
- Creating a world-class materials science and engineering workforce that is trained for careers in academia or industry.

New advances with HPC came also from specific materials so called zeolites, which are used as both molecular sieves and catalysts to help make fuels and chemical feedstocks. To date, more than 200 types of zeolites have been synthesized and more than 330,000 potential zeolite structures have been predicted based on previous computer simulations.

With such a large pool of candidate materials, using traditional laboratory methods to identify the optimal zeolite for a particular job presents a time- and labor-intensive process that could take decades.

[13] https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/mgi_strategic_plan_-_dec_2014.pdf

That's why scientists from the University of Minnesota and Rice University are developing computational tools that can quickly screen thousands of materials to pinpoint promising candidates for further research and development.

The research team recently put their large-scale, multi-step computational screening process to the test using Mira, the ALCF's 10-petaflops IBM Blue Gene/Q supercomputer. And it passed with flying colors, as the method successfully identified new zeolites for two important applications.

Enabled by Mira's massively parallel architecture, advanced algorithms and accurate intermolecular potentials, the team's calculations expedited hundreds of thousands of virtual experiments for highly complex systems. The researchers gained access to Mira through the ALCF's Director's Discretionary program and with support for optimising the code on Mira BG/Q architecture. A co-author on the Nature Communications article, Knight assisted in porting their MCCCS-MN code to Mira, guided the developers in adding OpenMP support to permit hybrid MPI/OpenMP parallelism, and helped design an MPI-based framework to allow high-throughput calculations capable of using all of Mira's 786,432 cores.

The code performance enhancements allowed Siepmann and his team to carry out simulations for two zeolite applications of industrial relevance. For one study, they looked for zeolites that could improve the purification of ethanol from fermentation broths. The simulations pointed them to a zeolite with a pore/channel system that effectively accommodates ethanol molecules while discouraging hydrogen bonding with water molecules. With the ability to purify ethanol in a single separation step, this material displays the potential to replace an energy-intensive, multi-step distillation process currently used by industry. To validate the simulation results, University of Minnesota researchers synthesized and tested the promising zeolite, providing experimental data that was in very good agreement with the predictions.

For the second study, the team investigated potential zeolite catalysts for a de-waxing process called hydro-isomerization, in which linear long-chain alkanes are transformed into slightly branched alkanes to reduce the pour point and increase the viscosity of lubricant oils and other fuel products. Their simulations identified zeolites with up to 100 times better adsorption capability than current technology used for this process.

In the field of cosmology and astrophysics experts from the WG participated end of 2013 in a major conference in the field called "Exascale computing in Astrophysics[14]" in Ascona, Switzeland. The conference was organized by Romain Teyssier and George Lake (University of Zurich), together with Claudio Gheller and Thomas Schulthess (Swiss Supercomputing Centre).

The following topics were addressed:

- cosmological simulations: science case and code evolution;
- galaxy formation, star formation, astrophysical fluid dynamics: science case and code evolution;
- supernovae simulations: science case and code evolution;
- turbulence simulations: science case and code evolution;
- stellar, solar and disk accretion physics challenges and code evolution.

Key challenges lie into dealing efficiently in exascale codes with multi scales multi-physics problems, to optimize turbulence, MHD, N-body, gravity or radiative hydrodynamics like problems, to consider non fluid behaviour such as particle acceleration in plasmas or the presence of dust in accretion disks and low temperature atmospheres, or the efficient dynamical coupling of various codes to span the huge temporal and spatial range of scales encountered in astrophysical problems. New algorithms using SPH and Lattice Boltzmann methods and hybrid cpus/gpus solutions have been proposed or are under development as are fault tolerant solutions. AMR for turbulent plasma flow is being push forward in some astrophysical problems but can't be the solution for all of them if the structures evolve

[14] http://www.hpc-ch.org/slidecasts-of-exascale-computing-in-astrophysics-conference-in-ascona-switzerland/

too much/fast. Parallel in time algorithm for long temporal evolution in astrophysical fluid dynamics have started to be considered. A community effort towards implementing such algorithms in multi-purpose open source codes should be started/reinforced.

During its last call for proposals PRACE started to support more and more projects linked to data assimilation and design of future instruments using HPC facilities in the field of fundamental sciences for dealing with large scale instruments. For example research projects in astrophysics with Van Allen Belt Probes, ALMA or EUCLID, or in particle physics with CERN and LHC, in laser physics with the LaserLab Europe consortium or in fusion with ITER, or in materials with the Graphene Flagship have been granted by PRACE resources.

Same applies in other fields like climate/weather forecast and life sciences and medicine (with the Human Brain Project).

From PRACE allocations or use of national resources, some major achievements in scalability of weather, climate and solid earth sciences during the period were performed like:

- in the field of the modelisation of high strength steels, a German project called “EX- ASTEEL – Bridging Scales for Multiphase Steels” developed an approach called FE^2 allowing computational micro-macro scale bridging approach directly incorporating micromechanics in macroscopic simulations. The code developed is based on nonlinear FETI-DP ((Finite Element Tearing and Interconnecting) methods and has been scale out using JUQUEEN, an IBM BG/Q system located in Juelich Supercomputing Center (Germany). By performing optimizations on collectives communications (for using low level BG/Q communications routines) it has been possible to scale out the code to up to 458 752 MPI tasks with no use of simultaneous multithreading (SMT) features of the BG/Q and even 1 835 008 MPI tasks with a 4x SMT.
- In cosmology the team of Dr. Volker Springel from Heidelberg Institute for Theoretical Studies, and Heidelberg University has been able using PRACE and German national facilities (20 million core hours on CURIE in France) and their AREPO code to trace the history of the universe, starting soon after the Big Bang and continuing through to the present day, capturing 13.8 billion years of change with unprecedented fidelity. Results of Prof. Volker’s research were published on 7 May 2014 in edition #509 of Nature, the international weekly journal of science.
  The model called Illustris provides clues on the tendency of matter to redistribute in the universe, prodded by supernovas and other phenomenon. This finding could be used to fine-tune experiments performed with space-based telescopes, such as NASA’s WFIRST survey, and EUCLID, the European Space Agency’s program.
- During its last call for proposals PRACE started to support more and more project linked to data assimilation and design using HPC facilities in the field of fundamental sciences for dealing with large-scale instruments. For example research projects in astrophysics
- In materials, the KKRnano code for quantum description of nano-materials solving the Korringa-Kohn-Rostoker Green function has been scaled out to 458,752 cores (1,835,008 parallel threads) on BlueGene/Q (JUQUEEN). One of the determining factor for achieving this level of performance was the use of the SIONLib to reduce drastically the time spent for I/O.
- In fusion the PP code a particle-in-cell code for simulating relativistic and non-relativistic plasmas using either a full particle model or a hybrid ion-particle / electron-fluid model has been also scaled out to 458,752 cores (1,835,008 parallel threads) on BlueGene/Q (JUQUEEN)

Finally, experts and communities worked in the development of the first EESI2 recommendations produced in 2013, especially in the field of the development of mini applications for applications performance analysis, characterisation, prediction and porting on emerging architectures. It is especially the case for the following applications:

- DUMSES, a mini application derived from the RAMSES application in astrophysics. DUMSES uses the same algorithm as RAMSES but with simpler data structures. It performs hydrodynamics and MHD equations on a uniform grid but does not perform AMR, nor implement Poisson solver and radiative transfer. DUMSES is targeted for the development of new I/O and memory placement strategies or for hybrid parallelization on GPU or manycore devices. DUMSES has been used recently on 131 000 threads of an IBM BG/Q at IDRIS (France) for a performance study/optimization of I/O strategies.

- A reduced version like a mini application for the fusion code GYSELA 5D have been developed for working on the reduction of the memory footprint of the code on IBM BlueGene/Q systems. Based on a prediction tool coupled to the mini application it has been possible to assess the limits of the code in term of memory usage and optimize the overall consumption, leading to a reduction by 50% of the memory used by MPI process.
- In the field of the CRESTA FP7 European project, an activity of co design of the fusion code Nek5000 is conducted by using a mini application derived from Nek5000 called NEKBONE. Nekbone Kernel is a single-core code focused on the matrix-vector product at the heart of the spectral element method. The code allows for analysis and optimization of the performance of the matrix-vector product kernel, which is recast as a set of computationally intense matrix-matrix products with relatively low operation count and minimal data movement.

5.2 Gap Analysis

The gap analysis described in the previous deliverable is still up to date. Its by the way important to notice that the materials and the high energy physics worked for setting up European Center of Excellences. Even if the expected budget allocated to each of the 8 to 10 CoE will be limited (4.5M€) and 10 times less than the EESI1 recommendations, the communities are committed to address Exascale challenges in scalability of the next generation codes, convergence between HPC and data management and link with major societal challenges.

One CoE proposal is leaded by Maison de la Simulation (France) with participation of teams from Meteo France and EDF (France), Juelich (Germany), BSC (Spain), UK, Poland and Belgium.

The aim of this proposal is to establish an Energy Oriented Centre of Excellence for computing applications, (EoCoE). EoCoE (pronounce Echo²) will use the prodigious potential offered by the ever-growing computing infrastructure to foster and accelerate the European transition to a reliable and low carbon energy supply. To achieve this goal, we believe that the present revolution in hardware technology calls for a similar paradigm change in the way application codes are designed. EoCoE will assist the energy transition via targeted support to four renewable energy pillars: Weather forecast, Materials, Water and Fusion, each with a heavy reliance on numerical modelling. These four pillars will be anchored within a strong transversal multidisciplinary basis providing high-end expertise in applied mathematics and HPC. EoCoE is structured around a central Franco-German hub coordinating a pan-European network, gathering a total of 8 countries and 23 teams. Its partners are strongly engaged in both the HPC and energy fields; a prerequisite for the long-term sustainability of EoCoE and also ensuring that it is deeply integrated in the overall European strategy for HPC. The primary goal of EoCoE is to create a new, long lasting and sustainable community around computational energy science. At the same time, EoCoE is committed to deliver high-impact results within the first three years. It will resolve current bottlenecks in application codes, leading to new modelling capabilities

and scientific advances among the four user communities; it will develop cutting-edge mathematical and numerical methods, and tools to foster the usage of Exascale computing. Dedicated services for laboratories and industries will be established to leverage this expertise and to foster an ecosystem around HPC for energy. EoCoE will give birth to new collaborations and working methods and will encourage widely spread best practices.

The second initiative toward the creation of Centers of Excellence is coming from CECAM (Centre Européen de Calcul Atomistique et Moléculaire[15]) which proposed to setup a CoE called E-CAM leaded by the Irish node of CECAM with the support of 14 other members and a strong link with PRACE.

---

[15] http://www.cecam.org/index.html

E-CAM aims to develop and maintain common libraries of molecular and atomistic multiscale sotware adapted to Exascale architectures and fitting the needs of academic and industrial research communities. It also aims to develop education and training of young researchers and consulting towards industry in order to foster the use of modelisation and numerical simulation.

Finally, a third one called EMERGE is a 10 partner CoE proposal led by Dr. Michael Nolan of Tyndall National Institute University College Cork in Ireland. HPC oriented partners are ICHEC and STFC Daresbury Labs. The proposal aims to support the design of energy materials (PV, fuel cells, batteries, catalysts etc.) in academia and industry. Emphasis would be placed on growing the design community and enabling the most effective use of emerging HPC platforms for select codes.

6. Life Sciences

6.1 Advances, results and breakthroughs in the international context and developments of the working group of the past year

During the last stage of the project the panel, mainly addressed by panel chairs, reviewed the recommendations of EESI-2.

Advances in genomics

Since the sequencing of the first human genome in 2001[16,17,] DNA sequencing technology experienced a fast development and an explosive disruption in 2007, when the Next generation sequencing instruments (beating Moore's law[18]). Today computer and data management systems are the bottleneck for current grand challenges in Health (for example the sequencing of 100,000 patients in UK[19]). Managing genomic information in the public health care system (and other areas as insurances) represents one of the biggest opportunity (personalized treatments) and thread (privacy) of our society.

The International Cancer Genome Consortium (ICGC[20]) is an international initiative with the original goal to study 50 cancers, sequencing at least 500 patients for reach cancer. The Cancer Genome Atlas (TCGA[21]) is similar initiative that begun in US ten years ago to catalogue identify genetic mutations associated to cancer. In 2014 both initiatives decided to join forces and start the first pan-cancer (transversal among cancer types) genetic study (see figure 11). The amount of data available on both initiatives represents a challenge from many points of views: IT system, computing environment, data management, legal issues, etc. When the ICGC project wraps in 2018, it expects to have collected an estimated 10 to 15 petabytes of information from more than 50,000 genomes. Pan-Cancer Whole Genome Analysis (PAWG) project designed a strategy to gradually increment the volume of the data (data trains) at the time that infrastructure and analysis tools are gradually improved. The EBI and the PRACE Tier-0 Marenostrum III are currently participating to the project. Thanks to PAWG researchers are studying activity in the non-protein coding portions of the cancer genome, which makes up about 95 percent of the tumor genome. There are 130 research projects and 16 working groups involved in the PAWG in areas such as novel mutation calling, structural variations, clinical translation, evolution and heterogeneity, and more.

[16] International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome". Nature 409 (6822): 860–921

[17] Venter, JC et al. (2001). "The sequence of the human genome" (PDF). Science 291 (5507): 1304–1351.

[18] http://www.genome.gov/sequencingcosts/

[19] http://www.genomicsengland.co.uk/

[20] https://dcc.icgc.org/

[21] http://cancergenome.nih.gov/

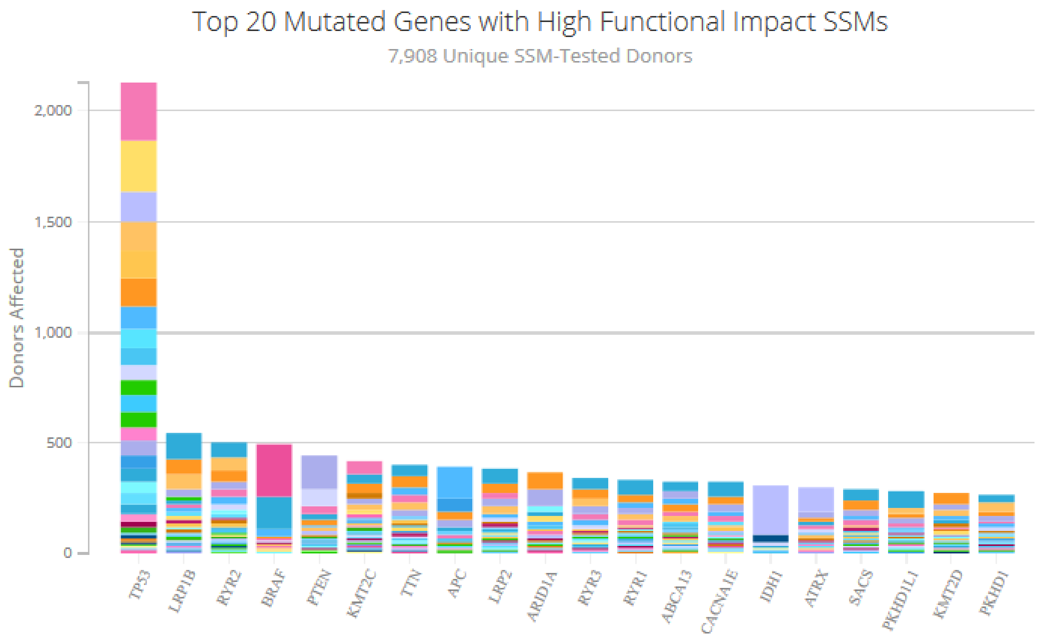


Figure 15 - top cancer-related gens from the Pan Cancer Genome analysis

Thanks to the know-how from these challenges we are making reality the idea of integrate genomics to clinical (Hospitals). In UK, the Genomics England (a company owned by the Department of Health) was set some moth ago to deliver the 100,000 Genomes Project (sequence 100,000 whole genomes from NHS patients by 2017[22]). Today, patients of Newcastle Hospitals are already receiving a diagnosis for cancer[23]. In US, Barack Obama announced a project to increase one order of magnitude these flagship initiative, proposing to invest $215M to develop a voluntary national research cohort of more than 1 million people. The idea is to obtain sequencing data for as many of those individuals as possible, identify genomic drivers of cancer, improve how next-generation sequencing-based tests are evaluated and marketed, and develop methods for managing and analyzing large patient data sets while protecting individual privacy[24].

Synthetic Biology

Synthetic Biology represents one of the major opportunities for our society in the coming years. The implications of the advances in this field have a strong impact in our society. For example, in February 2012 a group of researchers achieved some success in modifying influenza A/H5N1 viruses to be transmitted efficiently between mammals while maintaining high pathogenicity. In a decision with few precedents, the results were banned[25].

It took 15 years and US$40 million to synthesize the genome of a bacterial parasite in 2010. In 2014, a team of bioengineers redesigned and produced a fully functional chromosome of S. cerevisiae[26], and

[22] http://www.genomicsengland.co.uk

[23] http://www.ncl.ac.uk/press.office/press.release/item/health-secretary-meets-first-patients-diagnosed-through-genomics-medical-breakthrough

[24] http://www.reuters.com/article/2015/01/30/us-usa-obama-precisionmedicine idUSKBN0L313R20150130

[25] Policy: Adaptations of avian flu virus are a cause for concern. Nature 482,153–154

[26] Science 4 April 2014: Total Synthesis of a Functional Designer Eukaryotic Chromosome

eukaryote organism, more sophisticated than bacteria. Looking the massive amount efforts required to re-build life (see Figure 13), there is no doubt that computing will play a key role in synthetic biology.

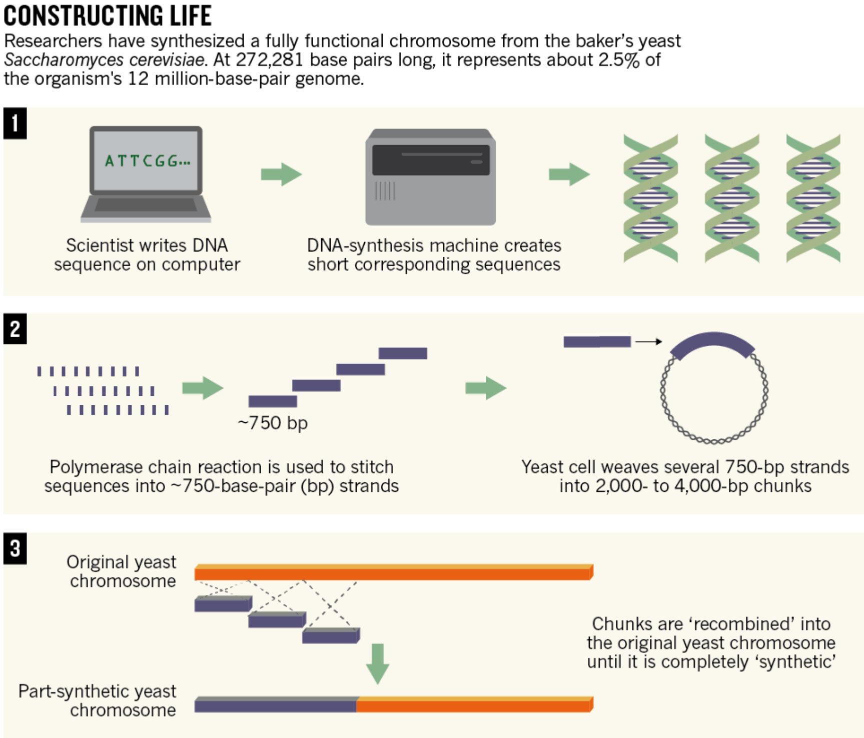


Figure 16 - Protocol to synthetize a full chromosome (source : Science 2014)

In parallel to this initiative, researchers of the Scripps Research Institute (US) expanded the genetic alphabet, integrating new symbols (beyond A, C, G, T) to a bacteria[27]. The team created in laboratory a semi-synthetic organism with several unnatural base pairs (UBPs). The UBPs did not introduce a notable growth burden to the synthetic organism although they found that DNA repair pathways is not efficiently excised. Thus, the resulting bacterium is the first organism to propagate stably an expanded genetic alphabet.

Another of tremendous interesting is the possibility to use synthetic DNA to store information. In 2013, the EBI together with Agilent Technologies, developed a method to code information using the DNA alphabet and latter generate a DNA sequence that store such information. The work published in Science[28] demonstrated that was possible to store and send over five million bits of data, consisting of text and audio files, that were latter successfully retrieved and reproduced. The accuracy of the inflation transmitted was in the range of 99.99% and 100%. Computer algorithm is one of the core technologies of the method, as it requires an error-correcting encoding scheme to ensure the extremely low data-loss rate[29]. The method is still not cost-effective, but considering the exponential decrease of DNA synthesis it may be a competitive alternative in some years. A recent study in

27 Nature 509,385–388 (15 May 2014)

[28] Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; Leproust, E. M.; Sipos, B.; Birney, E. (2013). "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA". Nature 494 (7435): 77–80.

[29] Yong, E. (2013). "Synthetic double-helix faithfully stores Shakespeare's sonnets". Nature

Angewandte Chemie [30] it is estimated that information encoded in DNA can be recovered after 1 million years if stored at -18 °C.

Molecular Simulation

One of the major advances in Molecular simulation is the assemblage of more than 1,300 identical proteins – in atomic-level detail for the simulation of HIV-1 capsid[31]. This work (that was cover in Nature) leaded by K. Schulten was a step forward the simulation of whole organisms using Molecular dynamics and supercomputing. The 64 million-atom system was simulated with Blue Waters supercomputer, at the National Center for Supercomputing Applications at the University of Illinois using the NAMD code.

“The work of matching the overall capsid, made of 64 million atoms, to the diverse experimental data can only be done through computer simulation using a methodology we have developed called molecular dynamic flexible fitting,” Schulten said. “You basically simulate the physical characteristics and behavior of large biological molecules but you also incorporate the data into the simulation so that the model actually drives itself toward agreement with the data.”

The simulations revealed that the HIV capsid contained 216 protein hexagons and 12 protein pentagons arranged just as the experimental data had indicated. The proteins that composed these pentagons and hexagons were all identical, and yet the angles of attachment between them varied from one region of the capsid to another. Possessing a chemically detailed structure of the HIV capsid will allow researchers to further investigate how it functions, with implications for pharmacological interventions to disrupt that function, Schulten said.

NAMD's parallel scalability was based on the prioritized message-driven execution capabilities of the Charm++/Converse parallel runtime system. The dynamic components of NAMD are implemented in the Charm++ parallel language. It is composed of collections of C++ objects, which communicate by remotely invoking methods on other objects. This supports the multi-partition decompositions in NAMD. Also data-driven execution adaptively overlaps communication and computation. Finally, NAMD benefits from Charm++'s load balancing framework to achieve unsurpassed parallel performance.

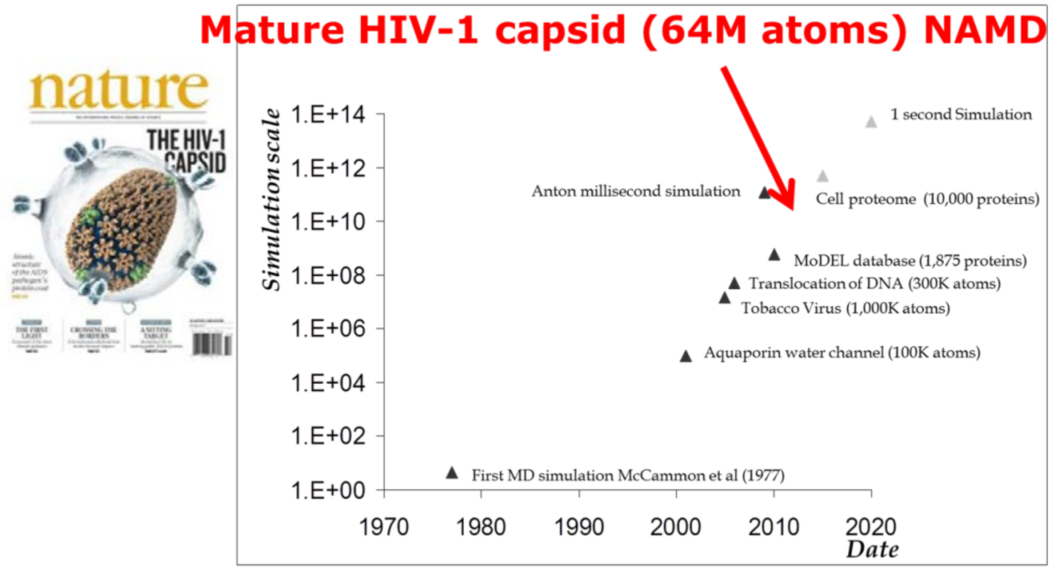
Looking back the projections made by EESI-1 panel, we see how this experiment matches our expectations and that in the close future we will require extreme computing resources, fitted to Molecular dynamics problems (see for example Anton[32]).

---

[30] Grass, R. N.; Heckel, R.; Puddu, M.; Paunescu, D.; Stark, W. J. (2015). "Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes". Angewandte Chemie International Edition 54 (8): 2552

[31] Nature 497, 643–646

[32] Shaw et al. "Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer," Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC14), Piscataway, NJ: IEEE, 2014, pp. 41–53



Simulation scale for long and large simulations is calculated multiplying simulation length (nanoseconds) with size (number of atoms).

Figure 17 - HIV-1 capsid simulation on Nature' cover. The plot adapted from EESI-1 roadmap shows that the magnitude of the project reaches the expected projection

In Europe pharma companies like Novartis are investing more and more on Cloud computing for accelerating time to solution for virtual screening techniques. Virtual screening is similar to gathering as many keys into a bucket as possible, then trying each of them individually in a keyhole to see if there's a fit. In this case, the keyhole is a chemical bond — an indication that a compound could have an effect on a particular protein. From such a point, potentially valuable compounds can be further tested with the hopes of a drug that can target these proteins.

Organ simulation

The Human Brain Project is one of the European Flagship initiatives, involving many different research fields (computer science, hardware design, neurology, bioinformatics, etc.) with the ambitious goal of model and simulate the human brain. The project recently published the first technical report (October 2013-September 2014). Although the set up of the project (ramp up stage) is complex, there are already some relevant advances (many of them related to data gathering and processing). The project is currently focusing on mouse models, building a simplified version of the virtual mouse brain (200,000 neurons). The model maps different parts of mouse body (spinal cord, eyes, skin; see Figure 16) and reacts to stimulus (touching)[33]

[33]https://www.humanbrainproject.eu/es/-/a-simulated-mouse-brain-in-a-virtual-mouse-body?redirect=https%3A%2F%2Fwww.humanbrainproject.eu%2Fes%2Fnews-events%3Fp_p_id%3D101_INSTANCE_qvWAPKvcO4xA%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-2%26p_p_col_count%3D1

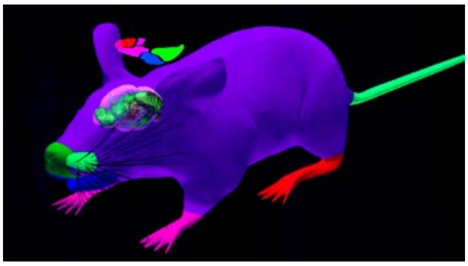


Figure 18 - Virtual mouse body (source HBP)

In the other hand, the Virtual Physiological Human (VPH; a project aims to provide digital representations of the entire human body) entered in 2015 as part of EUDAT2020 community.

VPH will be able to build on the generic data services provided by EUDAT to create rich, community-specific analysis platforms. The fact that many EUDAT partners are also large HPC centers participating in PRACE will make it easy for VPH researchers to integrate their data with high performance computing resources. This is an important step towards a more flexible framework for computing for life Sciences.

Finally, in the US a new initiative called BigNeuron, a community effort to define and advance the state of the art of single neuron reconstruction and analysis and create a common platform for analyzing 3D neuronal structure has been launched in early 2015. In an attempt to find a standard neuron reconstruction algorithm, BigNeuron will sponsor a series of international hackathons and workshops where contending algorithms will be ported onto a common software platform to analyze neuronal physical structure using the same core dataset. All ported algorithms will be bench-tested at Department of Energy supercomputing centers—including the National Energy Research Scientific Computing Center (NERSC) at the Lawrence Berkeley National Laboratory (Berkeley Lab) and the Oak Ridge Leadership Computing Facility (OLCF) at the Oak Ridge National Laboratory (ORNL)—as well as Human Brain Project supercomputing centers, allowing the community to standardize optimal protocols for labeling, visualizing and analyzing neuronal structure and key biological features.

The data generated by these benchmark runs will be used to develop a comprehensive annotated database of complex neuronal morphology, generate a searchable tool for discovering annotated and unique characteristics of neuronal morphology and lay the groundwork for potentially integrating this tool with large-scale data sets linking form to neuronal function. In addition, researchers will be able to access it via the Collaborative Research in Computational Neuroscience (CRCNS) data-sharing portal hosted by NERSC, which allows neuroscience researchers worldwide to easily share files without having to download any special software.

Systems Biology

Finally, during the last years we are observer of how Systems Biology is gradually gaining an important role in computer simulations. If in EESI-1 we debated the exponential increment of data and the maturity of models, in 2012 it was published the first whole-cell computational model for a human pathogen (Mycoplasma genitalium) using only genetic information. The experimental analysis directed by model predictions identified previously undetected kinetic parameters and biological functions. The entire organism (with 525 genes) was modeled in terms of its molecular components, that required integration of multi-format data and very fragmented data. This point us to the need of designing HPC systems “connected” to wet lab and able to have a flexible access to experimental information (not only large repository files).

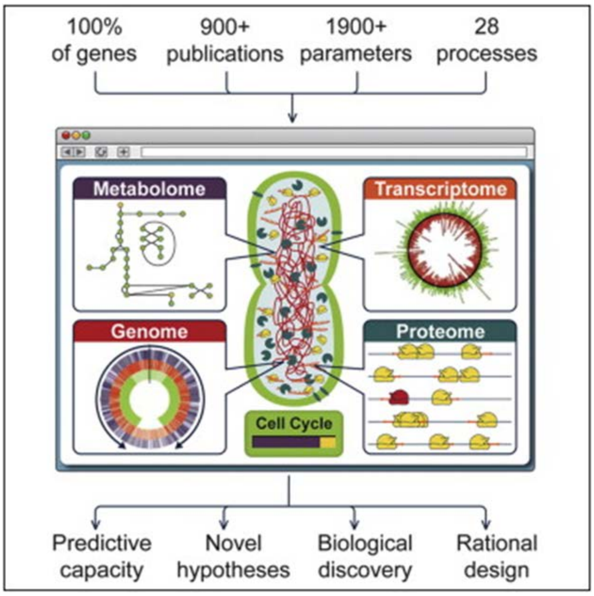


Figure 19 - Computer model for the whole-cell system

The EESI2 Life Sciences panel reformulates its 7 main conclusions:

1. *The Human Simulation is an Exascale challenge, but biology is today a bottleneck.*
2. *Experts are worried with a future scenario where million to billion core systems are seen as the only possible hardware offer, as real problems may not be able to scale to such machines*
3. *HPC computing should be designed to be easily integrated with wet-lab work*
4. *Training is important and requires sustained effort and investment*
5. *Life science problems are heterogeneous and should not run on rigid frameworks*
6. *Real-time response is critical for clinical applications*
7. *Industry needs an improved access to flexible, scalable and secure HPC resources that can deliver results in a useful time frame*

EESI-2 Recommendations

Life Sciences have a special interest to support Data-related recommendations and work towards a data centric computing. Flexible frameworks for efficient computing-data coupling, that could integrate analysis tools is a major demand of the experts.

6.2 Gap Analysis

The gap analysis described in the previous deliverable is still up to date. Its by the way important to notice that European Life Science community worked for setting up European Center of Excellences. Even if the expected budget allocated to each of the 8 to 10 CoE will be limited (4.5M€) and 10 times less than the EESI1 recommendation there were a strong activity during 2014 to setup such proposals.

During 2014 many of the members of the panel started a set of independent meetings to promote the proposal of a Center of Excellence (CoE) in Life Science. Some members participated in more than one proposal (covering different areas that do not overlap or compete among them). During the preparation process the experts exchanged information and tried to align the goals of the CoE, to the point that it has been agreed that CoE can lead the Life Science panel of EXCDI (the successor project of EESI-2). The evaluation of CoE is currently under evaluation, but its preparation has been already set a strong network among European stockholders (involving pharmaceutical and food industry and Hospitals) and defined a clear roadmap for the community.

The two projects of CoE in LifeSciences reported are the following :

- HPC4PM. Personalized medicine is increasingly computational demand by the “health community” and we need to ensure that HPC centres dedicated expertise in the needs of personalized medicine. Biology is not only computational, but also data intense and there is a critical need to ensure availability of reference data for genetic analysis and alignment of new data in the context of available knowledge. The proposal of this CoE addresses the scalability of Personalized Medicine data provision to remote HPC sites, which will give support to future massive use of HPC by companies and clinical practitioners.
- BioExcel. This center of excellence will focus on the molecular level aspects of life science. In particular, this will mean starting from modeling and prediction of biomolecular structures and higher-level complexes, using both simulations and computational biology approaches. The center has competence on docking, screening and other interactions with small molecules, and refinement and classification of experimental structural data.

Finally, experts and communities worked in the implementation of the first EESI2 recommendations produced in 2013, especially in the field of the development of mini applications for applications performance analysis, characterisation, prediction and porting on emerging architectures. It was the case for example in the field of the Exa2CT[34] European project with the development of a mini application derivated from the MUPHY[35] a multi-physics code for biomedical applications based upon the combination of microscopic Molecular Dynamics (MD) with a hydro-kinetic Lattice Boltzmann (LB) method, developed by researchers from Universisy of Roma.

34 http://www.exa2ct.eu/index.html

35 http://www.sciencedirect.com/science/article/pii/S0010465509001118

7. Disruptive applications

7.1 Advances, results and breakthroughs in the international context

In Life Sciences, The European Bioinformatics Institute (EBI) in Hinxton, UK –one location of the European Molecular Biology Laboratory (EMBL) network that is among the world's largest biology-data repositories – reported in 2013 that it held more than 20 petabytes of data and backups about genes, proteins and small molecules, according to a June, 13, 2013 article in the journal Nature. Of that data, according to the authors of the paper, 2PB were genomic data – a number that more than doubles every year.

In fact, the volume of biological data tracked by EMBL-EBI doubles every nine months, compared to the capacity of storage hardware, which doubles every 18 months, according to a 2014 update from EBI that also noted the volume of data it holds had increased to 40 petabytes during the year after publication of the Nature story.

EBI is actually addressing that problem by developing high-capacity storage technology called DNACloud that uses as little as one gram of DNA to store 2.2 petabytes of digital data[36].

That may eventually keep the world from being covered in data-storage hardware, but doesn't address the astonishing growth in the volume of data available to life-sciences companies following successful sequencing of the human genome in 2003. A single sequenced genome can equal 140GB of data, which explains how the total genetic data stored by the EBI shot up from the single-digit-terabyte range to nearly 200 terabytes by the end of 2012. Between June of 2010 and February of 2013, storage at biotech data-service company BioTeam, Inc. grew from 3TB to 21.8TB – an increase of 627 percent, or 209 percent during each year of the three-year period.

More generally in the life sciences domain the following additional disruptive applications have been identified :

- The Internet of Medicine : Wearable computers and computer accessories that exist in primitive form as fitness trackers will monitor heart rate and EKG patterns, blood sugar, blood pressure and other vital signs that can be collected periodically or broadcast in real time to caregivers monitoring a patient's health.
- Personalized medicine and integration of "omics": Advances in genomics will allow physicians to treat patients before they're sick, or select the treatments most likely to be effective based on a patient's own biochemistry and lifestyle rather than statistical norms based on a few studies measuring results in a limited numbers of patients.
- HPC has contributed to spectacular advancements in the life sciences on four levels. First, HPC is playing a crucial role in making sense of the massive amounts of data generated by modern 'omics' and genome sequencing technologies. Second, HPC is key to modeling increasingly large bimolecular systems using approaches such as quantum mechanics/molecular mechanics and molecular dynamics (see, for instance, the 2013 Noble Prize in Chemistry) to advance our understanding of bimolecular reactions and aid the discovery of new therapeutic molecules. Third, HPC is essential to modeling biological networks and simulating how network perturbations lead to adverse outcomes and disease. Finally, the simulation of organ function, such as the heart or the brain, not only depends on HPC, but also drives its development.

[36] http://arxiv.org/abs/1310.6992

Another disruptive application which could require heavily to use applied mathematics and HPC is complex manufacturing. For example, 3D printing application or additive manufacturing process, where thin layers of metal powder are molten by a high-energy electron beam that welds the powder selectively to create complex 3D metal structures with an almost arbitrary geometry by repeating the process layer by layer.

Simulation can be used for designing the electron beam gun, developing the control system, and generating the powder layer, thereby accelerating the printing process in commercial manufacturing, for example, of patient specific medical implants. Possibly the greatest simulation challenge, however, is to develop numerical models for the welding process itself—a complex 3D multi-physics problem that involves the modeling of thermal energy transfer from the electron beam to the powder bed, the melting of the powder particles, and the flow of the melt with a free surface and that is determined largely by surface tension and wetting effects. A realistic simulation with physical resolution of a few microns requires millions of mesh cells and several hundreds of thousands of time steps—computational complexity that can be tackled only with parallel supercomputers and sophisticated software.[37]

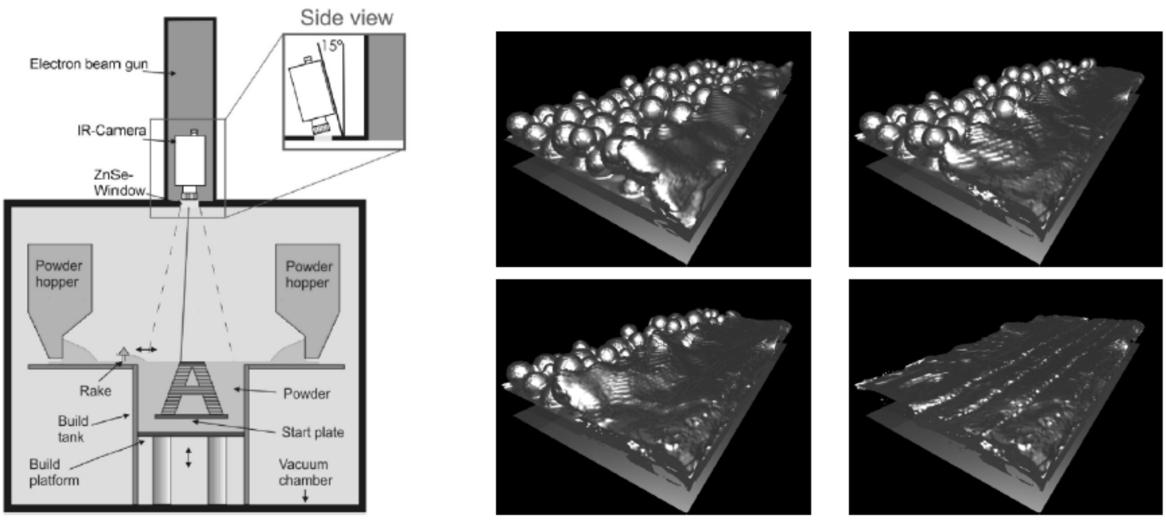


Figure 20 – (Left) Schematic illustration of a 3D printing process for metals using electron beam melting - (Right) Four snapshots of the effects of a simulated electron beam that scans over a powder bed in a sequence of parallel lines

Then, one of the last disruptive application which is now coming out with a lot of HPC requirements is machine learning. This domain which is at the crossroads of HPC and Big Data is an evolution of pattern recognition and computational learning theory by artificial intelligence.

Machine learning explores the construction and study of algorithms that can learn from and make predictions on data and is closely related to and often overlaps with computational statistics and has strong ties to mathematical optimization. When employed in industrial contexts, machine-learning methods may be referred to as predictive analytics or predictive modelling.

Such technologies were employed at the beginning by banks (Financial services is the second-largest commercial high performance computing vertical market) and insurance on historical and transactional data to identify risks and opportunities but now due to the deluge of data are used in more and more domains including web, marketing, telecommunications, retail, travel, pharmaceuticals.

---

37 M. MARKL , R. AMMER , U. RUEDE , AND C. KOERNER , Numerical investigations on hatching process strategies for powder-bed-based additive manufacturing using an electron beam , The International Journal of Advanced Manufacturing Technology, (2014), pp. 1–9.

This domain is also interesting the HPC vendors like nVIDIA or Intel which announced very recently specific offers in machine learning, as an example this field called deep learning was one of the biggest announcement of nVIDIA during their last GTC conference in Mars 2015 with target markets in using GPU enabled Bayesian models of cosmology and high-energy physics, computational models of the visual and motor cortex, deep learning systems for medical and biological image analysis, as well as machine-learning models of social behaviour and economics.

Its also useless also to present the Watson initiative from IBM, and companies like Google or Facebook are developing specific HPC solutions (Google is part of the OpenPower consortium while Facebook is also a promoter of the OpenCompute initiative) for addressing large-scale machine learning problems which involve solving huge sparse matrices.

7.2 Developments of the working group of the past year

Experts from the 3 working groups in charge of disruptive approaches in WP3, WP4 and WP5 had a face-to-face meeting in June 2014 during the EESI2 internal workshop in order to exchange about the definition and the identification of disruptive applications, algorithms and technologies. In the discussions of disruptive approaches in the context of Numerics and Applications, after identifying potential key disruptive technologies, the emphasis was on the requirements from applications that could establish disruptive breakthroughs in the application.

Weather prediction would benefit hugely if good algorithms and software were available for parallelizing the computation in the time-domain, as this is currently the bottleneck to exploiting parallelism in this area. In particular, this will be necessary if the potential of Exascale is to be realized. In that way experts from WG3.2 had been involved into the definition of the following recommendation on Parallelisation in Time and a subgroup of IS-ENES community submitted in the field of the H2020 FET HPC call a proposal called CHANCE towards the development of new ocean model involving the use of parallelisation in time.

The advent of Exascale could radically change the approach to solving many application challenges. For example, there are 10^{13} cells in the body so it is conceivable to use these as units in a computation where the computational power is at 10^{18}. Indeed breakthroughs are possible in much of the Meso scale domain and particle based methods might prove worthwhile to investigate.

The extension of data assimilation techniques used in climate modelling, to fields such as aeronautics might produce useful benefits.

In automotive applications, a great many relative small optimization computations need to be performed and techniques in both continuous and combinatorial optimization are required. The core requirement here and for many other applications is the availability of a good highly-parallel solver.

Exascale computing also opens the door to big advances in the solution of inverse problems both in the oil and gas industries but also in drug design. Computations with varying time scales (say in combining combustion with unsteady fluid flow) and in uncertainty quantification and associated stochastic approaches are also ripe for considerable advances of a truly disruptive nature.

7.3 Gap Analysis

It is difficult to perform a gap analysis since WG3.5 started later than the otters working groups and produced few outputs during the first period.

The main difference between first and second intermediate report comes from an analysis of existing distributive applications and the involvement of experts from WG3.5 in the production of one of the EESI2 recommendations related to Parallelisation in time which is a real disruptive approach.

7.4 Project recommendation " Parallel-in-Time: a major step forward in Parallel simulations"

Experts from the WG3.5 as well as 3.2 have been involved into the definition of this EESI2 disruptive recommendation towards the development of parallel in time simulations. The efficient exploitation of Exascale systems with a potential of billion threads will require massive increases in the parallelism of simulation codes, and today most time-stepping codes make little or no use of parallelism in the time domain; the time is right for a coordinated research program exploring the huge potential of Parallel-in-Time methods across a wide range of application domains.

Potential application areas include: climate research, computational fluid dynamics, life sciences, materials science, nuclear engineering, etc. These include applications of HPC with the highest return on investment in terms of economic, societal and scientific impact.

European researchers are leading Parallel-in-Time developments. There has been a series of three international workshops dedicated to these algorithms held in Europe (Lugano, 2011, Manchester, 2013, and Jülich, 2014) with 21 European speakers from six EU countries as well as invited speakers from the US, Japan and Russia. This is indicative of a diverse, thriving and world-leading European research community.

As an example a team from University of Wuppertall and LBNL scaled out to an IBM BlueGene/Q system at Juelich (Germany) a parallel multi grid code called PMG using a parallelised in time solver issued from the PFASST library. The PMG+PFASST code provides a space-time parallel solver for systems of ODEs with linear stiff terms, stemming e.g. from method of lines discretization of PDEs.

The "parallel full approximation scheme in space and time" (PFASST) joins Parareal-like iterations for time-parallel integration with multilevel spectral deferred correction sweeps in a space-time multigrid fashion. With innovative coarsening strategies in space and time, parallel efficiency can be significantly increased compared to classical Parareal.

The approach use is scalable on up to 458,752 cores on BlueGene/Q (JUQUEEN) as described in the following picture:

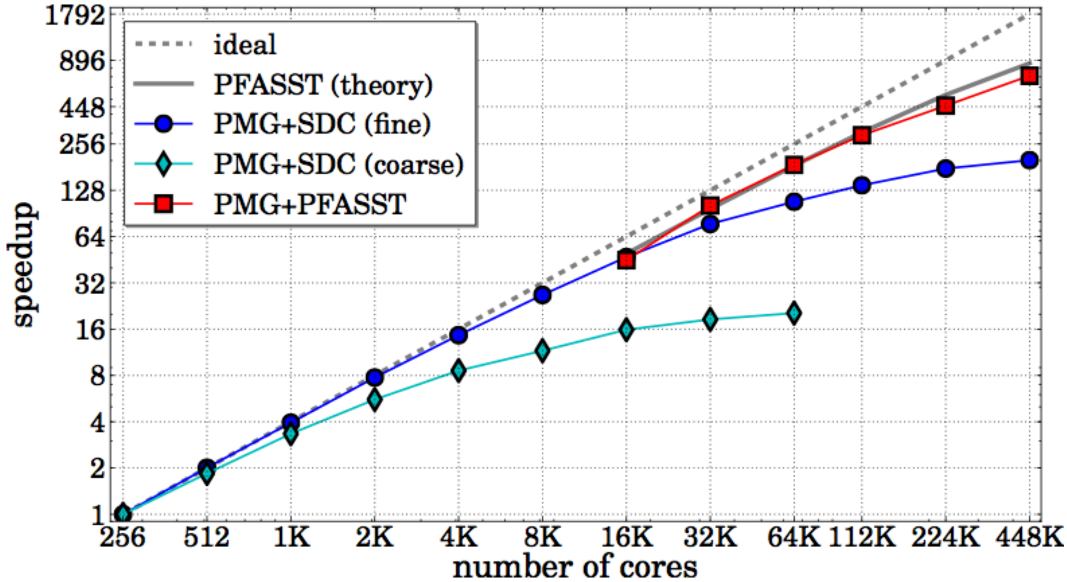


Figure 21 - scaling of PMG + PFASST on JUQUEEN

The goal of this recommendation is to fund R&D programs with the following objectives:

- Application of Parallel-in-Time methods to a wide range of application domains should now be tackled, either directly where energy conservation issues are not limiting e.g. for some life sciences problems or molecular dynamics, or after solving this issue, e.g. by projection like approaches for geophysical fluids (ocean and atmosphere), climate, seismology etc.
- The establishment of multi-disciplinary consortia to work on the deployment of Parallel-in-Time methods and applications to new fields, combining the expertise of applied mathematicians,

application scientists, computational scientists and HPC technology specialists, following a co-design approach.

- A number of different options should be looked at when preparing projects: having a rather large project, under which a number of different applications would be studied and synergistically developed, and/or a coordinated cluster of smaller projects, each of which looks at a single application, but exchanging knowledge and experience.
- A series of benchmarks and test cases should be established and maintained in order to have a clear view of the advantages, disadvantages and quality of the different Parallel-in-Time methods. The test cases should be at the same time close to the real applications and simple enough to allow both the integration and test of Parallel-in-Time methods.
- In order to maximize their exploitation across different application domains, Parallel-in-Time software should be encapsulated in reusable scalable libraries. This should be at a timely point in the development of the software and following the establishment of a stable and robust methodology born out of experience with a range of applications. Such libraries would then accelerate and facilitate the further deployment of Parallel-in-Time methods in new application areas and targeting new libraries for Exascale computing.

We envisage a number of world-leading projects utilizing the Parallel-in-Time method but focusing on different approaches and on different application areas. With 2-4 projects between €2M and €4M each, we recommend that the total amount of support should be in the range of €5 million to €10 million.

8. Conclusions

This document as second intermediate report of WP3 in charge of Applications gives an overview about latest breakthrough and developments during the year 2014.

During this period experts from WP3 worked in implementing some of the first recommendations of EESI2 (especially the ones related to the development of mini applications), participated in the elaboration of the second set of recommendations (especially the ones related to VVUQ, efficient couplers, analysis of pertinent turbulent structures, and parallelisation in time) and worked heavily in the setup of European CoE (this report give some preliminary information about some proposals elaborated by experts of the working groups) and more globally into H2020 projects proposals (including EXDCI the follow up of EESI2 with strong synergies between PRACE, ETP4HPC and future CoEs).

Major breakthroughs and results have been reported included latest developments in the Oil & Gas domain where companies are driving the HPC market with major equipment (which exceed sometimes national research facilities), the automotive where companies are (re)engaged into strong HPC roadmaps towards high fidelity combustion or system integration, climate community with the preparation of the next IPCC campaign, astrophysics with the dark matter understanding, fusion with the rise of ITER like models for the assessment of the performance and security behaviour of the first fusion reactor to be installed in 2019 in France or Life Sciences with massive molecular simulations and the development of organ modelisation especially the brain.

Some development toward extreme scalability of scientific and industrial applications using US or European (PRACE) facilities and initiatives toward the co design of next generation applications have been reported.

Finally, a survey has been conducted by the NAFEMS organisation together with EESI2 (WG3.1) and early results are showing that efforts have to be made in order to make aware ISV to the upcoming hardware and software challenges of Exascale. It's the first time that EESI gathers inputs from commercial software companies and this link now will be followed within EXDCI.